

A comparative study of MCR-based kinetic analyses for chemical reaction systems with rate constant ambiguities

Sara Mostafapour^{a,b}, Henning Schröder^{b,c}, Christoph Kubis^c, Mathias Sawall^b, Bahram Hemmateenejad^a, Klaus Neymeyr^{b,c}

^aDepartment of Chemistry, Shiraz University, Shiraz, Iran

^bUniversität Rostock, Institut für Mathematik, Ulmenstraße 69, 18057 Rostock, Germany

^cLeibniz-Institut für Katalyse, Albert-Einstein-Straße 29a, 18059 Rostock, Germany

Abstract

The results of multivariate curve resolution (MCR) methods clearly depend on the chosen MCR approach. We compare three MCR tools that support kinetic modeling in order to investigate the significant influence on the determined solutions. These methods are the Hard-Soft-Multivariate Curve Resolution, the Multivariate Curve Resolution-Net Analyte Signal and as a common roof the general approach by the computation of the set of feasible rate constants as implemented in the FACPACK kinetic hard-modeling software. The focus is on the determination of the reaction rate constants and the occurrence of corresponding solution ambiguities.

Three spectroscopic data sets are considered in this study: First, an alkaline hydrolysis (UV/Vis) with a two-step consecutive first order mechanism. Since no slow-fast-ambiguity is observed, the three MCR methods give comparable, consistent results. The two remaining data sets concern a formation of iridium catalysts (FTIR) based on a two-component reversible mechanism and the reaction of 3-chlorophenylhydrazonopropane dinitrile with 2-mercaptoethanol (UV/Vis) based on a three-component partially reversible mechanism. Even though the rate constants for these reactions can show some significant deviations, the corresponding error indicators are on an equally low level. The sets of feasible rate constants as computed by the cube enclosure algorithm enclose the particular solutions as computed by the other methods.

Keywords: factor analysis, multivariate curve resolution, ambiguity, nonnegative matrix factorization, net analyte signal, area of feasible solutions

1. Introduction

Multivariate curve resolution methods are key tools for the extraction of pure component information from spectroscopic data sets. In its simplest form, such a data set contains one time-resolved series of spectra, which can be stored row-wise in a nonnegative matrix $D \in \mathbb{R}^{m \times n}$. Then m is the number of spectra and n the number of data channels in each spectrum.

For an s -component reaction system, the matrix D we are interested in is an approximate factorization of the form

$$D = CS^T + E \quad (1)$$

with the nonnegative matrices $C \in \mathbb{R}^{m \times s}$, $S \in \mathbb{R}^{n \times s}$ and an error matrix $E \in \mathbb{R}^{m \times n}$ with matrix entries close to zero. The goal is to assign the columns of C and S to the concentration profiles and spectra of the pure components [1, 2, 3]. The matrix E comprises the residuals due to deviations from a strict bilinear decomposition model. Its absolute matrix entries should be small compared to the maximal entries of D .

The occurrence of multiple solutions of the factorization problem (1) is well known under the keyword rotational ambiguity. The rotational ambiguity is the main challenge for MCR methods [4, 5]. Usually, additional constraints are introduced for the factors based on prior knowledge of the system in order to obtain a chemically interpretable solution. The unimodality constraint and the closure constraint are very common [6, 7]. Here we focus on the implementation of kinetic models as a chemical constraint in the resolution process [8, 9]. In this manner, the concentration profiles of

the components involved in a kinetic process are shaped according to the law of mass action in terms of an initial value problem for a system of ordinary differential equations. Such a kinetic modeling can decrease the underlying rotational ambiguity drastically. In addition, the parameters of the kinetic model, namely the reaction rate constants, are obtained as well. They are of particular physicochemical and analytical interest in research problems, industrial design and process analytical chemistry [10]. It is important to remark that the application of a kinetic model (especially those of first order or pseudo-first order) does not always guarantee a unique solution. A systematic analysis of the remaining ambiguities leads to sets of feasible reaction rate constants, see [11].

In this study three different MCR methods supporting kinetic modeling are investigated: Hard-Soft-Multivariate Curve Resolution (HS-MCR) [12, 13], Multivariate Curve Resolution-Net Analyte Signal (MCR-NAS) [14, 15, 16] and FACPACK Kinetic Modeling (FKM) [11, 17]. The different methods favor slightly different solutions to the factorization problem (1) due to their theoretical basis and implementations. It is shown that the consideration of reaction rate ambiguities [17] can help to overcome the bias of these methods. Three spectroscopic data sets are studied in this context: 1. alkaline hydrolysis of dimethyl phthalate, 2. partial equilibrium of iridium catalysts and 3. reaction of 3-chlorophenylhydrazonopropane dinitrile with 2-mercaptoethanol.

1.1. Organization of the paper

Section 2 introduces the three MCR methods HS-MCR, MCR-NAS and FKM. A short overview on the theory of reaction rate ambiguities is given in section 3. The studied spectroscopic data sets are described in Section 4, followed by their analysis in Section 5. A conclusion is given in Section 6.

1.2. Notation

All variables and mathematical operators are written in italics. Matrices are denoted by capital letters. Lowercase letters are used for scalars and vectors.

2. MCR approaches with kinetic modeling

Kinetic modeling in the context of MCR problems is shortly introduced in this section. Three implementations are presented in the Section 2.1.

The general idea of kinetic modeling for MCR methods is quite similar: optimal reaction rate constants are to be determined under the constraints that the two (almost) nonnegative factors C and S satisfy Equation (1) and that the concentration profiles, as described by the columns of C , are consistent with the kinetic model that is based on the current reaction rate constants. Next the notion of *consistency* is explained. To this end, we start with a first-order kinetic model in terms of the initial value problem

$$\dot{c}(t) = M(k)c(t), \quad c(t_1) = c_0. \quad (2)$$

Therein, the matrix $M(k) \in \mathbb{R}^{s \times s}$ depends on the vector $k \in \mathbb{R}^q$ of reaction rate constants. The vectors c_0 represents the initial concentrations. The initial value problem is solved on the time grid t_0, \dots, t_m predetermined by the data set. A column-wise representation of the results in matrix form for an s -component system is given by

$$C^{\text{ode}}(k) = \begin{pmatrix} c_1(t_1) & \cdots & c_s(t_1) \\ \vdots & & \vdots \\ c_1(t_m) & \cdots & c_s(t_m) \end{pmatrix} \in \mathbb{R}^{m \times s}.$$

Then the link between the initial value problem (2) and the factorization problem (1) is established by the kinetic fit error,

$$\|C - C^{\text{ode}}(k)\|_F. \quad (3)$$

In the following, we use the relative error term $\|C - C^{\text{ode}}(k)\|_F / \|C\|_F$ due to its better mutual comparability of the results for the different data sets. If the relative error (3) equals 0 or if it is below a small positive threshold value, then the corresponding factorization CS^T is called to be consistent with the kinetic model.

2.1. MCR methods

On the basis of the introductory part of this section, three MCR methods for solving the factorization problem 1 are introduced:

2.1.1. Hard-soft multivariate curve resolution

MCR-ALS is a very popular iterative curve resolution algorithm. Here the so-called lack-of-fit (lof)

$$\text{lof} = 100 \frac{\|E\|_F}{\|D\|_F}$$

is minimized by alternating modifications of the matrices C and S .

Soft-modeling MCR-ALS uses additional constraints as unimodality or selectivity in the optimization process. Such a regularization can result in chemically meaningful concentration profiles and spectra [18].

Hard-soft multivariate curve resolution (HS-MCR) is a modification of MCR-ALS that implements kinetic models as a constraint. In each modification step of the factor C , a parameter vector k is determined such that (3) is minimal. All those columns of the factor C predicted by the model are then replaced by $C^{\text{ode}}(k)$. The remaining columns are modified according to the basic MCR-ALS approach. The optimized model parameters k^{opt} (often these are the rate constants of the kinetic model) are an additional output of this method.

2.1.2. Multivariate curve resolution-net analyte signal (MCR-NAS)

Net analyte signal (NAS) [19, 20] is defined as the part of a mixture spectrum that is directly related to the concentration of the component of interest and that is orthogonal to the spectra of the other interfering species together with the background variations. The NAS technique allows us to monitor changes in the concentration of one species (reactant or product) during the chemical process. Each row vector d_i of D which is the reaction mixture spectrum at the i th step of a reaction kinetic process is composed of the contributions of all species (namely reactants and products) plus the contribution from other sources such as inert interferences, drifts and instrumental noise. The role of NAS calculation is to extract the net contribution of one of the components involved in the reaction at different steps of the evolutionary process from the recorded spectrum. This would be feasible if the concentration of the species of interest (e.g. C_B) was known at all reaction steps. In this case the part of the mixture spectrum that is orthogonal to the spectral space spanned by interferences is used in a resolving procedure. The part of the spectrum that is not orthogonal to the spectral space of the other components can be represented as a linear combination of the spectra of the other components. Only the orthogonal part is unique to the analyte of interest.

During the MCR-NAS analysis concentration profiles are iteratively calculated using different values of rate constants. The aim of MCR-NAS approach is to find a suitable vector of reaction rate constants k so that D_{-B} is completely free from the contribution of component B . The matrix D_{-B} is defined as the rank annihilated matrix containing spectral information of all species presented in the reaction system, except B . For all values of the rate constants which are equal to true values the matrix D_{-B} contains information about all sources of variation in data except B . In this situation, a maximal correlation coefficient between the norm of the NAS vectors (at each time) and the calculated concentration profiles (C_B) by the model equations is obtained. If the process converges, then the final estimate of k is the true k [14, 15].

2.1.3. Facpack kinetic modeling (FKM)

FKM is a pure hard-modeling approach [11]. In its basic version a kinetic model is mandatory that describes the complete reaction system and thus all columns of C . In a similar way as for the previously presented methods the minimization of a cost function f (also called objective function) is used in order to determine chemically meaningful factors C and S as well as a model parameter vector k . A distinctive feature is that f solely depends on the vector k and not only implicitly on the matrices C and S . Hence the number of the degrees of freedom of f , namely the dimension of the vector k , is significantly smaller than the number of entries of C and S . Therefore the minimization of f can be done in a very efficient way.

FKM internally uses the truncated singular value decomposition $D \approx U\Sigma V^T$, wherein $U \in \mathbb{R}^{m \times s}$ and $V \in \mathbb{R}^{n \times s}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{s \times s}$ is a diagonal matrix. Then an alternative representation for a solution of Equation (1) is given by $C = U\Sigma T^+$ and $S = TV^T$ with a regular matrix $T \in \mathbb{R}^{s \times s}$ [21]. Here, T^+ denotes the pseudoinverse

of the matrix T . Thus $T = T(k) = (C^{\text{ode}}(k))^+ U \Sigma$ is a matrix that minimizes Equation (3) in a least squares sense. Summarizing, this makes it possible to calculate the factors

$$C(k) = U \Sigma (T(k))^+ \quad \text{and} \quad S = T(k) V^T \quad (4)$$

only depending on the parameters k , which are then evaluated with the typical constraints for nonnegativity and a low kinetic fit error

$$f(k) = \sum_{i=1}^m \sum_{j=1}^s \left(\min \left(\frac{C_{ij}}{\max_l(C_{lj})}, 0 \right) \right)^2 + \sum_{i=1}^n \sum_{j=1}^s \left(\min \left(\frac{S_{ij}}{\max_l(S_{lj})}, 0 \right) \right)^2 + \sum_{i=1}^m \sum_{j=1}^s \left(\frac{C_{ij} - (C^{\text{ode}}(k))_{ij}}{\max_l(C_{lj})} \right)^2.$$

3. Rate constant ambiguities

A major difficulty in the application of MCR methods is the potential occurrence of bands or continua of solutions (namely infinitely many solutions) of the factorization problem (1). These solutions may differ considerably. Their chemical importance can be evaluated and they can mutually be distinguished by adding further constraints to the problem. Consistency with a kinetic model is a possible and effective constraint. This constraint can be very restrictive. Sometimes it allows us to reduce the bands of possible solutions to a single, unique solution. However, it has been shown for first-order kinetic models that often the problem of rotational ambiguity cannot be resolved and has to be considered in the data analysis [11, 22]. In mathematical terms there are multiple combinations of factors C, S and model parameters k , such that

$$\frac{\|D - CS^T\|_F}{\|D\|_F} \rightarrow \min, \quad \frac{\|C - C^{\text{ode}}(k)\|_F}{\|C\|_F} \rightarrow \min, \quad \frac{\|\min(C, 0)\|_F}{\|C\|_F} \rightarrow \min \quad \text{and} \quad \frac{\|\min(S, 0)\|_F}{\|S\|_F} \rightarrow \min \quad (5)$$

or similar optimality criteria are fulfilled. In this paper we use the notation for the solution sets that has been introduced in [17]. For the noise-free case this includes the *set of D -consistent parameters*

$$\mathcal{K} = \left\{ k \in \mathbb{R}^q : D = C^{\text{ode}}(k) S^T \text{ for a (not necessarily nonnegative) matrix } S \in \mathbb{R}^{n \times s} \right\}$$

and the *set of feasible parameters*

$$\mathcal{K}^+ = \left\{ k \in \mathbb{R}^q : D = C^{\text{ode}}(k) S^T \text{ for a nonnegative matrix } S \in \mathbb{R}^{n \times s} \right\}.$$

In words, the set \mathcal{K} contains all those parameters, for which factors C and S exist that fulfill the first three optimality criteria in (5), namely the data reconstruction and the kinetic fit as well as the nonnegativity constraint for C . It can be seen as an intermediate step in the determination of \mathcal{K}^+ , for which the corresponding factors S also needs to be nonnegative. In summary, the set \mathcal{K}^+ is one way of representing the solution sets of MCR problems under the constraint of a kinetic model in the noise-free case. The set \mathcal{K}^+ contains the vectors of feasible reaction rate constants; the associated pure component factors C and S can be reconstructed with Equation (4).

All kinds of perturbations, for example noise or baseline errors, have to be considered for the analysis of experimental data. Generalizations of the sets \mathcal{K} and \mathcal{K}^+ are defined by introducing error tolerance values ε for the kinetic fit error and θ for the negativity of S . Then, the *set of D -approximate parameters* reads

$$\mathcal{K}_\varepsilon := \left\{ k \in \mathbb{R}^q : \text{rank}(T(k)) = s \text{ and } \frac{\|C(k) - C^{\text{ode}}(k)\|_F}{\|C(k)\|_F} \leq \varepsilon \right\} \quad (6)$$

and the *set of feasible D -approximate parameters* is given by

$$\mathcal{K}_{\varepsilon, \theta}^+ := \left\{ k \in \mathcal{K}_\varepsilon : \frac{(S(k))_{i,j}}{\max_l |(S(k))_{l,j}|} \geq -\theta \text{ for all } i, j \right\}. \quad (7)$$

Further details on the theory, recommendations on setting the tolerance values and an implementation of the cube enclosure algorithm for the numerical computation of $\mathcal{K}, \mathcal{K}^+, \mathcal{K}_\varepsilon$ and $\mathcal{K}_{\varepsilon, \theta}^+$ can be found in [17].

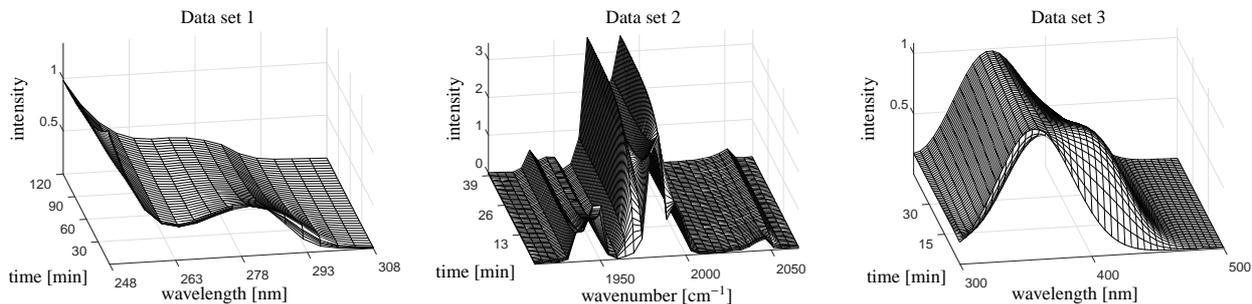


Figure 1: Visualization of the series of spectra for an alkaline hydrolysis of dimethyl phthalate (Data set 1), a formation of a partial equilibrium of iridium catalysts (Data set 2) and a reaction of 3-chlorophenylhydrazonopropane dinitrile with 2-mercaptoethanol (Data set 3).

4. Data

Three experimental data sets with different underlying reaction mechanisms are investigated here.

Data set 1 (Alkaline hydrolysis of dimethyl phthalate). The first data set contains the spectroscopic monitoring of a three-component reaction system, namely the alkaline hydrolysis of dimethyl phthalate in the presence of vanillin as an inert interference [14, 23]. The spectra were obtained in the wavelength range of 248–308 nm every 20s until 200s and then every 100s until 120min. A total of $m = 81$ spectra were recorded at $n = 31$ wavelengths each. The corresponding data matrix $D \in \mathbb{R}^{81 \times 31}$ is illustrated on the left of Figure 1.

Data set 2 (Formation of a partial equilibrium of iridium catalysts). The equilibrium between $\text{HIr}(\text{CO})_3(\text{PPh}_3)$ and $\text{HIr}(\text{CO})_2(\text{PPh}_3)_2$ in the presence of carbon monoxide and triphenylphosphine is analyzed by in-situ FTIR spectroscopy [24]. A stopped-flow unit in combination with a rapid-scan monitoring is used in this investigation. Herein, $m = 735$ spectra in the time range from 0min to 39.9min are recorded. Each spectrum contains the intensity values for $n = 325$ wavenumbers in a range from 1911cm^{-1} to 2068cm^{-1} . The corresponding data matrix $D \in \mathbb{R}^{735 \times 325}$ is illustrated in the center of Figure 1.

Data set 3 (Reaction of 3-chlorophenylhydrazonopropane dinitrile with 2-mercaptoethanol). The three-component reaction of 3-chlorophenylhydrazonopropane dinitrile with 2-mercaptoethanol is investigated by UV/Vis spectroscopy. Hereby 2-mercaptoethanol and the byproduct ethylenesulfide do not contribute to the spectra in the monitored range from 300nm to 500nm. Full details of the reaction and data acquisition conditions can be found in [25, 26, 10]. The data has been provided by the Biosystems Data Analysis Group of the University of Amsterdam [27].

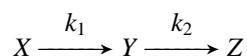
Each data set contains $m = 271$ spectra in the time range from 0.06min to 45min. The spectra are measured at $n = 201$ data channels in the wavelength range from 300nm to 500nm. The corresponding data matrix $D \in \mathbb{R}^{271 \times 201}$ is illustrated on the right of Figure 1.

5. Results and discussion

Now, we follow a kinetic study of the three data sets from Section 4 that are shown in Figure 1. For each data set this includes a comparison of the MCR methods from Section 2 regarding the optimality criteria from Equation (5) and the determination of the rate constant ambiguity (namely the set $\mathcal{K}_{\epsilon, \theta}^+$). The first data set is used in order to verify that the three MCR methods yield comparable results if no ambiguities are present. For the remaining two data sets the rate constant ambiguities are analyzed and are set in relation to the solutions of the three presented MCR methods.

5.1. Alkaline hydrolysis of dimethyl phthalate (Data set 1)

The assumed reaction model reads



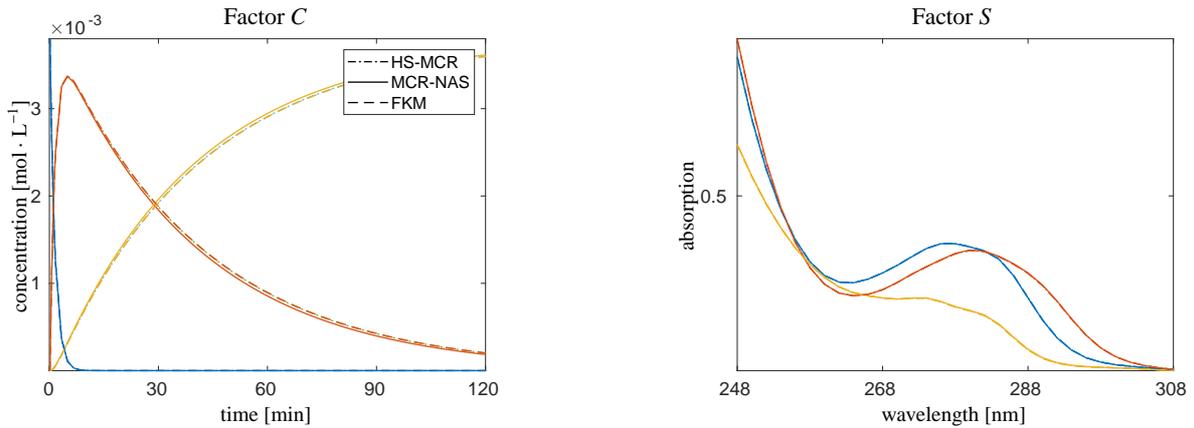


Figure 2: Resolved concentration profiles and spectra for Data set 1 by three MCR methods. The results of the three MCR methods are highly overlapping. A solution ambiguity can not be identified with respect to small tolerances.

with X = dimethyl phthalate, Y = monomethyl phthalate and Z = phthalate as well as initial concentrations $c_0 = (0.0038, 0, 0)^T$ in $\text{mol} \cdot \text{L}^{-1}$. The slow-fast-ambiguity [28] typically associated with this model is not to be found for the present data. The potential second solution can always be excluded due to its negative spectrum. Thus only a single solution is determined. The reaction rate constants, that are obtained by the three MCR methods are shown in Table 1 together with typical error indicators. Only minor differences are found in the calculated values for k_1 and k_2

Table 1: Optimized reaction rate constants and error indicators for Data set 1 broken down by chemometric methods.

| | HS-MCR | MCR-NAS | FKM |
|-----------------------------------------|---------|---------|--------|
| k_1 in min^{-1} | 0.7428 | 0.7389 | 0.7382 |
| k_2 in min^{-1} | 0.0248 | 0.0255 | 0.0247 |
| $\ D - CS^T\ _F / \ D\ _F$ | 0.0084 | 0.0008 | 0.0008 |
| $\ C - C^{\text{ode}}(k)\ _F / \ C\ _F$ | 0.0007 | 0.0092 | 0.0082 |
| $\ \min(C, 0)\ _F / \ C\ _F$ | 0.00002 | 0.0017 | 0.0013 |
| $\ \min(S, 0)\ _F / \ S\ _F$ | 0 | 0 | 0 |

among the three methods. Also the plots of the corresponding factors C and S , that are shown in Figure 2, show only minor deviations. Thus, it can be assumed that the three MCR methods result in equally well solutions.

5.2. Partial equilibrium of iridium catalysts (Data set 2)

The reversible reaction model



with $X = \text{HIr}(\text{CO})_3(\text{PPh}_3)$ and $Y = \text{HIr}(\text{CO})_2(\text{PPh}_3)_2$ as well as relative initial concentrations $c_0 = (1, 0)^T$ are used. It is known from [11] for noise-free cases that the model (8) is not sufficient in order to obtain a unique solution for the MCR problem. Analogously a similar observation can be made for the Data set 2. The reaction rate constants, that were obtained by the three MCR methods, are shown in Table 2 together with typical error indicators. Despite some significant deviations in the values of k_1 and k_{-1} , the corresponding error indicators are on an equally low level. The correlated factors C and S are shown in Figure 3 as a black dashed lines. They also highlight the major differences of the three solutions.

Table 2: Optimized reaction rate constants and error indicators for Data set 2 broken down by chemometric methods.

| | HS-MCR | MCR-NAS | FKM |
|-----------------------------------------|--------|---------|--------|
| k_1 in s^{-1} | 0.0025 | 0.0020 | 0.0030 |
| k_{-1} in s^{-1} | 0.0005 | 0.0009 | 0.0003 |
| $\ D - CS^T\ _F / \ D\ _F$ | 0.0104 | 0.0082 | 0.0082 |
| $\ C - C^{\text{ode}}(k)\ _F / \ C\ _F$ | 0.0013 | 0.0047 | 0.0046 |
| $\ \min(C, 0)\ _F / \ C\ _F$ | 0 | 0.0011 | 0.0008 |
| $\ \min(S, 0)\ _F / \ S\ _F$ | 0 | 0.0282 | 0.0046 |

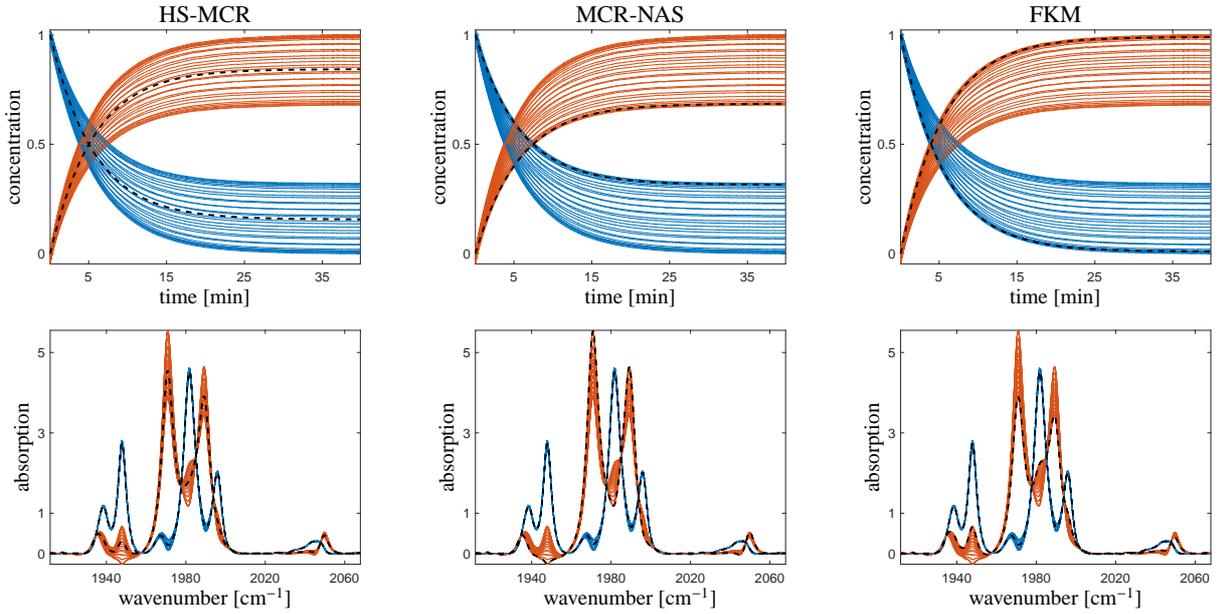


Figure 3: The factors C (top) and S (bottom) are plotted for a selection of reaction rate vectors that are contained in the set of feasible D -approximate parameters $\mathcal{K}_{0.017,0.05}^+$ for Data set 2. The black dashed lines are assigned to the resolved concentration and spectra profiles based on the optimal reaction rate vector k^{opt} for each technique.

However, the consistency of the three results can be shown by computing the set $\mathcal{K}_{\varepsilon,\theta}^+$. Here we use $\varepsilon = 0.017$ and $\theta = 0.05$. The values were determined according to the recommendations in [17] by evaluating the corresponding error indicators, see (6) as well as (7), and setting the error tolerances slightly higher. The Cube Enclosure Algorithm is applied for the approximation of $\mathcal{K}_{\varepsilon,\theta}^+$ [17]. The initial cube side length is set to $\omega = (k_1 + k_{-1})/20 \approx 0.000165$ for the reaction rate constants obtained by the FKM method. Using the reaction rate constants of the other two methods would result in the same approximation. The set $\mathcal{K}_{0.017,0.05}^+$ is presented in Figure 4 as a black dashed line. Furthermore, for each obtained vector of reaction rate constants of the three methods the corresponding set \mathcal{K} is plotted in black. Despite the fact that an analysis by the set \mathcal{K} actually presupposes perturbation-free data, it is easy to see that the three lines are very well aligned with $\mathcal{K}_{\varepsilon,\theta}^+$. The sets \mathcal{K} extend far beyond $\mathcal{K}_{\varepsilon,\theta}^+$, especially for small k_1 , because the associated rate constant vectors are assigned to factorizations whose factors S have negative entries (beyond the tolerance). The optimized rate constant vectors of HS-MCR (blue), MCR-NAS (red) and FKM (green) are all contained in the set $\mathcal{K}_{\varepsilon,\theta}^+$. This quite abstract information can also be presented in the space of concentration profiles and spectra, see Figure 3. The calculated solutions are highlighted by black dashed lines in each column. The factors that correspond to selected rate constant vectors in $\mathcal{K}_{\varepsilon,\theta}^+$ are indicated in color. In the plots of the concentration profiles as well as the plots of the spectra it can be seen that the three obtained MCR solutions differ significantly.

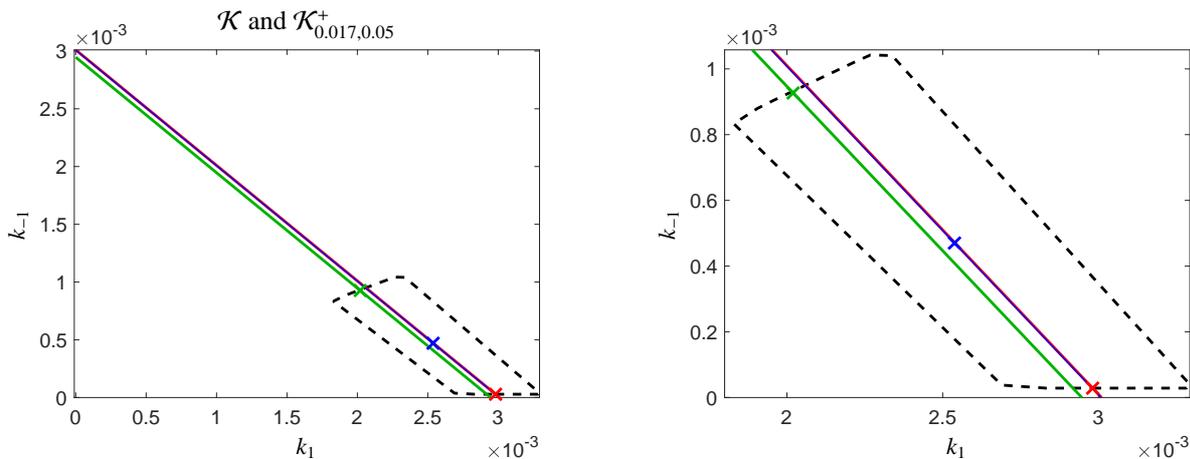


Figure 4: Reaction rate ambiguities for Data set 2. The colored lines are assigned to the sets \mathcal{K} determined by the optimal rate constant vectors of the different chemometric tools: HS-MCR (blue), MCR-NAS (green) and FKM (red). The solutions of each method are marked by crosses with the same color coding. The area marked by black dashed line illustrates the set of feasible D -approximate parameters $\mathcal{K}_{0.017,0.05}^+$. On the right a zoomed section is shown.

Table 3: Optimized reaction rate constants and error indicators for Data set 3 broken down by chemometric method.

| | HS-MCR | MCR-NAS | FKM |
|-----------------------------------------|--------|----------|----------|
| k_1 in min^{-1} | 0.2305 | 0.2182 | 0.1887 |
| k_{-1} in min^{-1} | 0.0612 | 0.0757 | 0.0971 |
| k_2 in min^{-1} | 0.0383 | 0.0388 | 0.0448 |
| $\ D - CS^T\ _F / \ D\ _F$ | 0.0067 | 0.0003 | 0.0003 |
| $\ C - C^{\text{ode}}(k)\ _F / \ C\ _F$ | 0.0010 | 0.0035 | 0.0034 |
| $\ \min(C, 0)\ _F / \ C\ _F$ | 0 | 0.00060 | 0.00056 |
| $\ \min(S, 0)\ _F / \ S\ _F$ | 0 | 0.000067 | 0.000069 |

For example the peak at 1945cm^{-1} is present for the HS-MCR and the FKM solution, but vanishes for the MCR-NAS solution. It is clear that further chemical information is needed in order to decide which solution is the one that explains the chemical reaction system in a correct way. The set $\mathcal{K}_{\epsilon,\theta}^+$ helps to give an overview of the possible solutions.

5.3. Reaction of 3-chlorophenylhydrazonopropane dinitrile with 2-mercaptoethanol (Data set 3)

The partially reversible consecutive reaction model



is used with $X = 3\text{-chlorophenylhydrazonopropane dinitrile}$, $Y = \text{intermediate}$ and $Z = 3\text{-chlorophenylhydrazono-cyanoacetamide}$ as well as initial concentrations $c_0 = (5.3 \cdot 10^{-5}, 0)^T$ in $\text{mol} \cdot \text{L}^{-1}$.

In analogy to Data set 2 the model (9) is not sufficient to obtain a unique solution of the MCR problem. This is confirmed by the following observation. The reaction rate constants, that were obtained by the three MCR methods, are shown in Table 3 together with typical error indicators. The correlated factors C and S are shown as black lines in Figure 5. Major differences can already be seen by just comparing the results for the first batch. The other nine

batch reactions result in equally low error indicators. All solutions are shown in the parameter space in Figure 6 as colored crosses. Furthermore, the three sets \mathcal{K} have been computed based on the optimal reaction rate vectors of the three methods for the first batch. They are shown as colored solid lines. The solutions are spread roughly along the sets \mathcal{K} . The different results among the methods and batches have multiple reasons. On the one hand, the three MCR methods seem to be biased in a way that the solutions shown in the parameter space form clusters, see the left plot of Figure 6. On the other hand, the solutions within such a cluster show deviations, which are most likely caused by deviations in the execution of the batch experiments.

The consistency of the calculated solutions is shown with the help of the set $\mathcal{K}_{\varepsilon,\theta}^+$. Therefore the set is approximated with the Cube Enclosure Algorithm for $\varepsilon = 0.015$ and $\theta = 0.015$. An approximation of $\mathcal{K}_{\varepsilon,\theta}^+$ is presented as a gray volume in the right of Figure 6. It contains 28 of the 30 computed solutions. The two outliers belong to the FKM method and are located slightly outside the approximation of $\mathcal{K}_{\varepsilon,\theta}^+$. The factors C and S assigned to these two solutions result in error indicators that are larger than the chosen tolerance values. Increasing these values or neglecting the corresponding batches would be a possible approach.

Again the abstract information of the set $\mathcal{K}_{\varepsilon,\theta}^+$ can be illustrated in the space of factors C and S , see the colored lines in Figure 5. While all concentration profiles are affected by the remaining ambiguity of the MCR problem, it is revealed that the characteristic peaks in the possible pure component spectra of the first and third species are nearly uniquely determined. The possible spectra of the second species shows some deviations, but the maxima of the peaks are located in a very small wavelength range around 400nm.

To conclude, the assignment of the species to the obtained spectra can be done with a great certainty. However, a quantification of the species is not possible in a reliable way without further information.

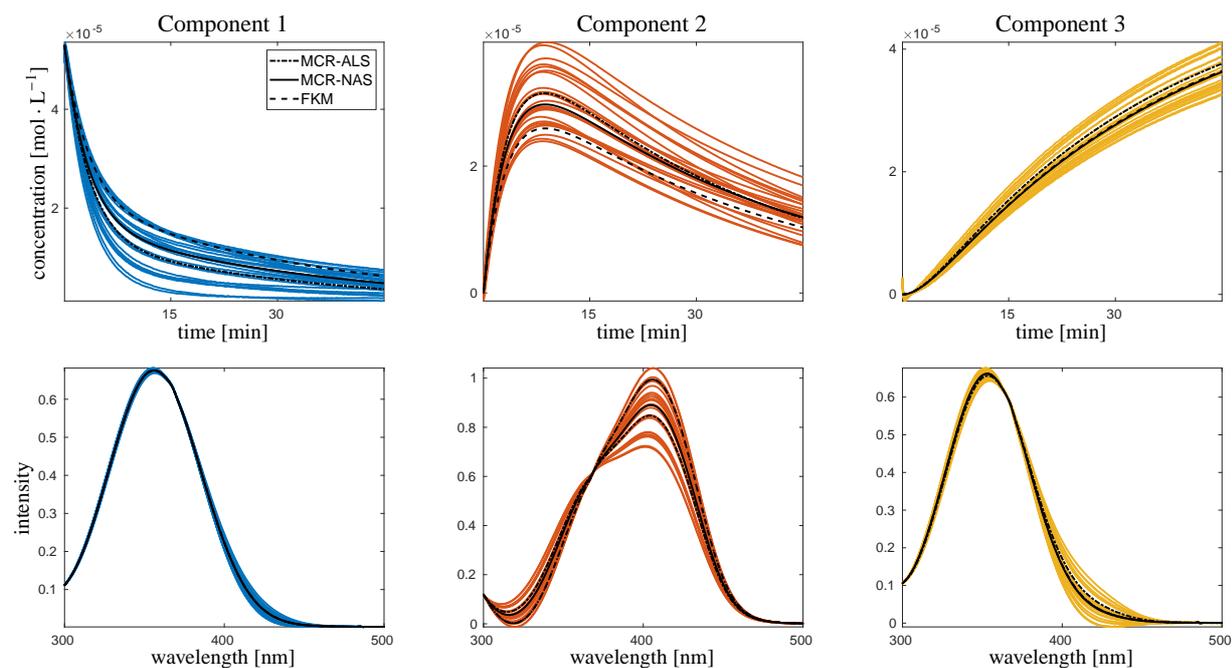


Figure 5: The factors C (top) and S (bottom) are plotted for a selection of reaction rate vectors that are contained in the set of feasible D -approximate parameters $\mathcal{K}_{0.015,0.015}^+$ for Data set 3. The black lines represent the resolved concentration and spectra profiles based on the optimal reaction rate vector k^{opt} for each technique. The detailed assignment is given in the legend.

6. Summary and conclusion

Three MCR tools for the extraction of pure component information and optimized reaction rate constants of a given kinetic model have been compared. As part of a kinetic study, they have been applied to three spectroscopic

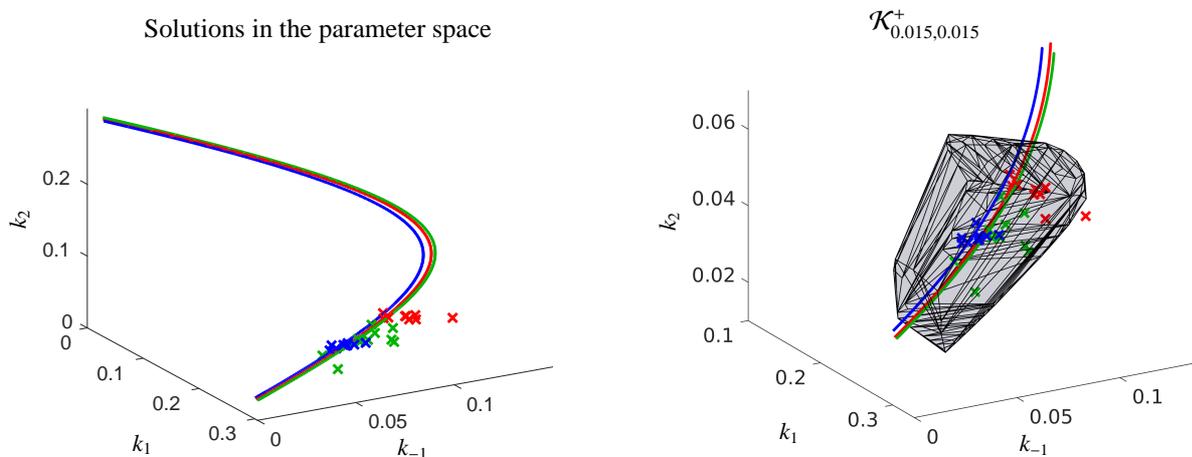


Figure 6: Reaction rate ambiguities for Data set 3. The colored lines are assigned to the sets \mathcal{K} obtained by the different chemometric tools: HS-MCR (blue), MCR-NAS (green) and FKM (red). The solutions of the 10 batches are marked by crosses with the same color coding. In the right plot the set of feasible D -approximate parameters $\mathcal{K}_{0.015,0.015}^+$ is indicated by the gray volume. It is a tube-like structure that extends around the sets \mathcal{K} . It contains all determined solutions except for two outliers.

data sets. Despite comparable low residuals of the data and model fit, significant differences of the reaction rate values occur in two of the three studied cases. This is caused by the different chemometric approaches of the tools and the intrinsic ambiguities that are correlated to the chosen kinetic models. However, the analysis by the set of feasible D -approximate parameters has shown the consistency of the results. Hence we conclude, that the analysis of parameter ambiguities should be appended to a “classical” MCR analyses in order to reveal possible uncertainties.

The comparison of the three MCR tools has yield the following results: The computation time for HS-MCR are relatively large since the method uses various constraints and all components must be considered in the optimization process. However, the resulting deviation from nonnegativity in C and S is the lowest. MCR-NAS is based on extracting the net contribution of only one component of interest from the mixture of known and unknown species. Hence it can be applied to more complex and rank deficient systems as well. FKM is a pure hard-modeling approach. The target function in FKM solely depends on the vector k and not on the matrices C and S . Thus the number of the degrees of freedom, that is the dimension of the vector k , is significantly smaller than the number of matrix entries of C and S . This makes a very efficient computational minimization of cost function possible.

Finally, the existence of solution ambiguities is a problem in MCR-based methods even under the regularization of a kinetic model. However, it is shown here that these ambiguities are computable. The results can be visualized in an abstract form as $\mathcal{K}_{\varepsilon,\theta}^+$ and in an easy-to-interpret form as band plots. Hence valuable information can be extracted even for data sets, that are affected by this solution ambiguity. Often only some of the peaks of a species spectrum are affected by such ambiguities and sometimes single species aren’t affected at all. Thus this information can be considered as reliable even if there is a solution ambiguity.

Perspectively further constraints can be applied to the set $\mathcal{K}_{\varepsilon,\theta}^+$ in the case that the interesting species are heavily affected by the ambiguity. This results in a reduction of $\mathcal{K}_{\varepsilon,\theta}^+$ to a subset which is consistent with the chosen constraints e.g smoothness or unimodality.

References

- [1] A De Juan, E Casassas, and R Tauler. Soft modeling of analytical data. *Encyclopedia of analytical chemistry: Applications, theory and instrumentation*, 2006.
- [2] Amrita Malik, Anna de Juan, and Roma Tauler. Multivariate curve resolution: a different way to examine chemical data. In *40 Years of Chemometrics—From Bruce Kowalski to the Future*, pages 95–128. ACS Publications, 2015.
- [3] C Ruckebusch and L Blanchet. Multivariate curve resolution: a review of advanced and tailored applications and challenges. *Anal. Chim. Acta*, 765:28–36, 2013.

- [4] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, and M. Maeder. On rotational ambiguity in model-free analyses of multivariate data. *J. Chemom.*, 20(6-7):302–310, 2006.
- [5] Hamid Abdollahi, Marcel Maeder, and Romà Tauler. Calculation and meaning of feasible band boundaries in multivariate curve resolution of a two-component system. *Anal. chem.*, 81(6):2115–2122, 2009.
- [6] Paul Gemperline. *Practical guide to chemometrics*. CRC press, 2006.
- [7] Hamid Abdollahi and Roma Tauler. Uniqueness and rotation ambiguities in multivariate curve resolution methods. *Chemom. Intell. Lab. Syst.*, 108(2):100–111, 2011.
- [8] Josef Diewok, Anna de Juan, Marcel Maeder, Romà Tauler, and Bernhard Lendl. Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of ph-modulated mixture samples. *Anal. chem.*, 75(3):641–647, 2003.
- [9] José Manuel Díaz-Cruz, Josep Agulló, M Silvia Díaz-Cruz, Cristina Ariño, Miquel Esteban, and Romà Tauler. Implementation of a chemical equilibrium constraint in the multivariate curve resolution of voltammograms from systems with successive metal complexes. *Analyst*, 126(3):371–377, 2001.
- [10] Sabina Bijlsma and Age K Smilde. Application of curve resolution based methods to kinetic data. *Anal. Chim. Acta*, 396(2-3):231–240, 1999.
- [11] Henning Schröder, Mathias Sawall, Christoph Kubis, Detlef Selent, Dieter Hess, Robert Franke, Armin Börner, and Klaus Neymeyr. On the ambiguity of the reaction rate constants in multivariate curve resolution for reversible first-order reaction systems. *Anal. Chim. Acta*, 927:21–34, 2016.
- [12] Anna de Juan, Marcel Maeder, Manuel Martinez, and Roma Tauler. Application of a novel resolution approach combining soft-and hard-modelling features to investigate temperature-dependent kinetic processes. *Anal. Chim. Acta*, 442(2):337–350, 2001.
- [13] Lionel Blanchet, Cyril Ruckebusch, Jean-Pierre Huvenne, and Anna de Juan. Focus on the potential of hybrid hard-and soft-mcr-als in time resolved spectroscopy. *J. Chemom.*, 22(11-12):666–673, 2008.
- [14] Mohsen Nekoeinia, Bahram Hemmateenejad, Ghodratollah Absalan, and Saeed Yousefinejad. Mcr-nas: A combined hard-soft multivariate curve resolution method based on net analyte signal concept for modeling kinetic data with inert interference and baseline drift. *Chemomet. r Intell. Lab. Syst.*, 98(1):78–87, 2009.
- [15] Bahram Hemmateenejad, Ghodratollah Absalan, Mohsen Nekoeinia, and Fatemeh Esfandiyari. Estimating the rate constant in second-order kinetics using hard-soft-net analyte signal (hs-nas) method. *Chemom. Intell. Lab. Syst.*, 102(1):35–44, 2010.
- [16] Bahram Hemmateenejad, Fatemeh Esfandiyari, and Mohsen Nekoeinia. Determination of the acidity constant of drugs using the hard-soft net analyte signal method. *J. Chem. Eng. Data*, 57(10):2802–2810, 2012.
- [17] H. Schröder, C. Ruckebusch, A. Brächer, M. Sawall, C. Kubis, D. Selent, R. Franke, A. Börner, and K. Neymeyr. Reaction rate ambiguities for perturbed spectroscopic data: theory and implementation. *To be submitted*.
- [18] R Tauler. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J. Chemom.*, 15(8):627–646, 2001.
- [19] Avraham Lorber. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.*, 58(6):1167–1172, 1986.
- [20] Avraham Lorber, Klaas Faber, and Bruce R Kowalski. Net analyte signal calculation in multivariate calibration. *Anal. Chem.*, 69(8):1620–1626, 1997.
- [21] A. Golshan, H. Abdollahi, and M. Maeder. Resolution of rotational ambiguity for three-component systems. *Anal. Chem.*, 83(3):836–841, 2011.
- [22] Henning Schröder, Cyril Ruckebusch, Olivier Devos, Rémi Métivier, Mathias Sawall, Denise Meinhardt, and Klaus Neymeyr. Analysis of the ambiguity in the determination of quantum yields from spectral data on a photoinduced isomerization. *Chemom. Intell. Lab. Syst.*, 189:88–95, 2019.
- [23] Zhong-Liang Zhu, Jun Xia, Jiang Zhang, and Tong-Hua Li. Determination of rate constants from two-way kinetic-spectral data by using rank annihilation factor analysis. *Anal. Chim. Acta*, 454(1):21–30, 2002.
- [24] Christoph Kubis, Wolfgang Baumann, Enrico Barsch, Detlef Selent, Mathias Sawall, Ralf Ludwig, Klaus Neymeyr, Dieter Hess, Robert Franke, and Armin Börner. Investigation into the equilibrium of iridium catalysts for the hydroformylation of olefins by combining in situ high-pressure fir and nmr spectroscopy. *ACS Catal.*, 4(7):2097–2108, 2014.
- [25] Sabina Bijlsma, Hans FM Boelens, and Age K Smilde. Determination of rate constants in second-order kinetics using uv-visible spectroscopy. *Appl. Spectrosc.*, 55(1):77–83, 2001.
- [26] Sabina Bijlsma and Age K Smilde. Estimating reaction rate constants from a two-step reaction: a comparison between two-way and three-way methods. *J. Chemom.*, 14(5-6):541–560, 2000.
- [27] University of Amsterdam. Biosystems data analysis. <http://www.bdagroup.nl/content/Downloads/datasets/datasets.php>.
- [28] W.G. Jackson, J.M. Harrowfield, and P.D. Vowles. Consecutive, irreversible first-order reactions. Ambiguities and practical aspects of kinetic analyses. *Int. J. Chem. Kinet.*, 9(4):535–548, 1977.