A multi-method chemometric analysis in spectroelectrochemistry: case study on molybdenum mono-dithiolene complexes.

Mathias Sawall[a], Christian Fischer[b], Benedict J. Elvers[b], Sebastian Pätsch[b], Klaus Neymeyr[a,c]

[a] *Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18057 Rostock, Germany*
[b] *Universität Greifswald, Bioanorganische Chemie, Felix-Hausdorff-Straße 4, 17487 Greifswald, Germany*
[c] *Leibniz-Institut für Katalyse, Albert-Einstein-Straße 29a, 18059 Rostock*

Abstract

Spectroelectrochemical (SEC) analyses combine spectroscopic measurements with electrochemical techniques and can provide deep insight into complex multi-component chemical reaction systems. SEC experiments typically produce large amounts of spectroscopic data. Chemometric techniques are required for the data analysis and aim at extracting the underlying pure component information. Here we analyze spectroelectrochemically gained UV-vis data from five molybdenum mono-dithiolene complexes with changing redox states. SEC enables an electrochemical control of the mixture composition which supports the application of chemometric curve resolution techniques. The factor ambiguity problem is addressed by a multi-method approach combining chemometric tools from the evolving factor analysis (EFA) and from the area of feasible solutions (AFS) methodology in combination with factor duality arguments. EFA enables a subsystem analysis. Two subsystems with three species each are identified, which belong to a reductive and to an oxidative region. A joint species is contained in both regions. A complete pure component decomposition becomes possible in a final step.

Key words: Spectroelectrochemistry, cyclic voltammetry, multivariate curve resolution, area of feasible solutions, duality principle, FACPACK.

1. Introduction

Spectroelectrochemistry (SEC) combines the fields of spectroscopy and electrochemistry in highly versatile ways. SEC applies spectroscopic techniques as UV-vis, IR, Raman, photoluminescence and others [27] to chemical reaction systems in which the oxidation or reduction of the reagents due to electrode processes is controlled by a potentiostat. SEC is a powerful analytical technique. For instance, it can be used for the study of catalytic processes [21, 22] including transition metal based complexes of changing redox states. Such sophisticated chemical processes let expect complex superimposed spectroscopic data. The challenge consists in recovering the pure component information from the spectral mixture data. Multivariate curve resolution (MCR) methods can help to unravel the hidden information by an arsenal of different approaches and techniques.

Let the mixed spectral data be given on a time×frequency grid and stored in a $k \times n$ matrix D. The number of spectra is k and each spectrum has n channels. The bilinear Lambert-Beer law [26, 19, 24] states an approximate equality of D with the product of two matrices containing column-wise the concentration profiles in $C \in \mathbb{R}^{k\times s}$ and the spectra in $S \in \mathbb{R}^{n\times s}$ of the s pure components

$$D = CS^{\mathrm{T}} + E. \tag{1}$$

The $k \times n$ matrix E is assumed to be close to the null matrix and collects small error terms due to deviations from ideal bilinearity or noise. Spectral recovery aims at determining C and S from given D on the assumption of nonnegative factors. This is a blind source separation problem. Nonnegativity of C and S is often not enough in order to get a unique factorization. Besides the trivial ambiguities (by mutual scaling or column permutation of the factors) there is nearly always a transformational ambiguity which traditionally is called the rotational ambiguity, see [20, 4, 1, 11] and others. In order to compute chemically meaningful factorizations one can apply additional constraints on the factors in the form of soft constraints like monotonicity, unimodality, smoothness, closure and others [6, 47, 29, 2, 30] or hard constraints that, e.g., impose the consistency of the concentration profiles to a kinetic model [13, 5, 16, 24, 34, 45].

Typically the number of anticipated chemical species s is much smaller than the number of spectra and the number of channels. A truncated singular value decomposition (SVD, [12]) of D of the form $U\Sigma V^{\mathrm{T}}$ yields mathematical bases for

the two factors. The s left singular vectors form the columns of U and the s right singular vectors appear columnwise in V. An invertible $s \times s$-matrix T serves to form the factors

$$C = U\Sigma T^{-1}, \qquad S^{\mathrm{T}} = TV^{\mathrm{T}},$$

see [20, 26, 24, 29]. This approach reduces the degrees of freedom underlying C and S to the number of s^2 matrix elements of T. The set of possible row vectors of T forms the Area of Feasible Solutions (AFS) for the spectral factor, see [4, 32, 11, 39, 40, 10, 36, 37]. This determines the set of possible columns of S, namely the set of possible spectra. The row vectors of T can even be truncated by their first component if a proper scaling is assumed. Analogously, T^{-1} is associated with a second AFS representing the possible profiles of the factor C. An AFS of a chemical s-component system is a bounded subset of the $(s-1)$-dimensional space. Its volume correlates to some extent with the ambiguity of the possible factors. For chemical systems with two ($s = 2$) or three ($s = 3$) species the AFS sets are bounded intervals ($s - 1 = 1$) or bounded planar sets ($s - 1 = 2$). For four species ($s = 4$) the AFS computation is time consuming and its visualization in the 3D space is complex to handle. For $s > 4$ all this is even more complex. For $s = 2, 3, 4$ various AFS-computation methods are available [4, 32, 11, 39, 40, 17, 37]. Here the polygon inflation methods [39, 40] are used as implemented in the FACPACK software. The software provides adaptive numerical methods that are robust for noisy data. Sometimes additional information on the system is available, e.g., in the form of single, initially known pure component spectra or known concentration windows. This information can be used in order to lock such known profiles in the AFS. The duality theory [14, 31, 35] makes it possible to translate known parts in one AFS into restrictions for the dual AFS of the other factor [3, 41].

Here we present an approach for breaking up the limitation of AFS techniques on their applicability to systems with not more than three or in ideal situations four chemical species. The data originate from a cyclovoltammetric spectroelectrochemical (SEC) experiment in combination with UV-vis measurements. The system contains five absorbing species. AFS techniques can successfully determine a unique pure component decomposition of the experimentally gained data. We show that the SEC conditions allow a partitioning of the spectral data into subsets that each comprise a maximum of three absorbing species. These subsets are analyzed separately. The partitioning is done in time direction starting with a concentration window that represents all five components. The data partitioning strategy benefits from certain properties of SEC data in which steps of the external electric potential during the cyclovoltammetric experiment are coupled to a recording of UV-vis spectra. See Figure 1 (left) for a typical cyclovoltammogram and the same figure (mid and right) for the associated series of spectra. The evolving factor analysis (EFA), see [8, 23, 25, 18, 24], helps to identify the emergence of new species and to define the disassembly points for the data partition. To this end, EFA monitors the evolving character of the significant singular values (those which are well separated from zero) depending on series of data matrices containing row-wise the increasing number of measured spectra. The occurrence of a new significant singular value indicates that a new component emerges for the first time. Usually EFA is applied in forward and backward direction, but it can also be applied to data subsets. Such applications of EFA to certain data windows can also reveal important information on the presence and absence of certain species.

The three building blocks (EFA, AFS and duality arguments) of our approach are well-established chemometric methods. The novelty in this work consists in their combination in order to break the hurdle of a factor ambiguity analysis by AFS techniques for a five-species chemical system. Our method includes factor ambiguity analyses on subsystems and merging steps which are supported by duality arguments. Sole purity-based methods are not sufficient to solve the factor ambiguity problem. EFA helps to identify proper subsystems for the analysis. The focus of this paper is on describing the methodological aspects underlying the combination of chemometric techniques. We describe general principles whose fields of application are not restricted to SEC experiments and which can be understood as a blueprint. Further, we note that SEC-plots typically use the electric potential on the abscissa. In contrast to this, we prefer here to represent the concentration profiles and left singular vectors with respect to the spectra index of the underlying data matrices. We confirm that the axis labeling is interchangeable to the electric potential or the time coordinate. However, the strong connection of the EFA plots to certain other quantities make it simple to relate qualitative changes if consistently the same axis labeling is used. All this helps to determine the disassembly points for the data partition. However, we also present the concentration-voltage dependency in Sec. 3.3.

The paper is organized as follows. Section 2 introduces the spectroscopic data set. Series of SVDs of data subsets are computed and EFA is applied to detect proper supports for the data partitioning. On this basis we analyze the appearance respectively the absence of certain species in the subsystems. An AFS analysis for two decisive subsystems is presented in Section 3. Pure component factorizations are computed for the subsystems. Finally, the pure components as identified for the subsystems are merged in order to determine a pure component decomposition for the complete spectral data set.

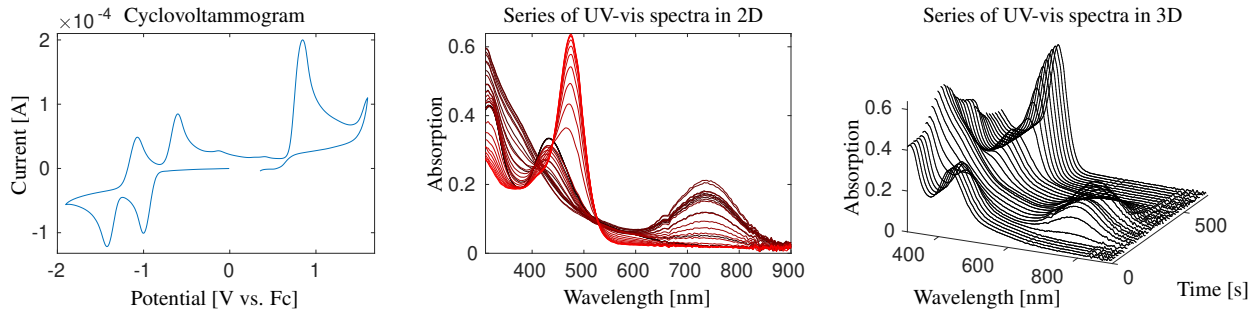

Figure 1: Plot of the applied potential versus the resulting current (left). Plot of the series of spectra in 2D form (center) and 3D-presentation (right). The color shift goes from black for the first spectrum to red for the last one. Only every 30th spectrum is plotted.

2. Spectral data partitioning

2.1. SEC analyses of molybdenum complexes

The molybdenum complex $[Mo^{II}(CO)_2(dt)(PP)]$ with dt = cyclohex1ene1, 2dithiol and PP = 1, 2-Bis(diphenylphosphino)ethane) is studied by cyclovoltammetry (CV) in acetonitrile with 0.1 M NBu_4PF_6. A platinum wire is used as a pseudo-reference electrode. The initial concentration of the molybdenum complex is $8.06 \cdot 10^{-4}$mol/L.

The changes of the electric potential during the CV experiments lead to a depletion or formation of certain species related to the potential applied at the working electrode. Consequently, after passing a potential at which a species of a certain redox state is either reduced or oxidized, it does no longer exist in the diffusion layer built in front of the working electrode. Thus, the concentration values in the diffusion layer do not behave according to global kinetic equations but rather fluctuate in dependence of the potential applied to the working electrode. A scan rate of 10mV s^{-1} is used and a number of $k = 948$ spectra are taken. Originally a number of 998 spectra were taken, but the last 50 suffer from interfering influences and are therefore discarded. We arrange the series of spectra according to the passed time coordinate. This results in an equidistant time grid with $t \in [0,\ 665.4]$ s. Each spectrum contains $n = 591$ absorption values for an equidistant grid with $\lambda \in [310,\ 900]$ nm. The applied current and the spectral data are presented in 2D and 3D form in Fig. 1. For the sake of an easier discussion the voltammograms were divided in first a reductive region $[-2.3, -1.0]$V vs. Fc/Fc^+ and second in an oxidative region for $[-1.0,\ 0.7]$V vs. Fc/Fc^+. In the reductive region the species have the formal metal oxidative states 0, I and II and in the oxidation region the species have the complex all over oxidation states II, III and IV. See [7] for more details on the data and the chemical interpretation.

2.2. Experimental setup

Spectroscopic data were obtained with a photodiode array MMS/100-1 manufactured by J&M. Cyclic voltammograms independent of the electrode setup were referenced with $Fc/Fc^+ = 0$ V as the internal standard. Tetrabutylammonium hexafluorophosphate (0.1 M) in CH_3CN was employed as supporting electrolyte in all cases. The SEC-UV-vis studies (Spectro-Electrochemistry in the UV-vis range) were conducted in a TSC Spectro manufactured by rhd-instruments and purchased from Metrohm. The counter electrode was made of platinum and a silver wire served as pseudo reference electrode. As working electrode a fine mesh of platinum was used which eventually allows the orthogonal light beam to pass the surface layer during the electrochemical experiments. Every experiment was started by registering a new baseline of the electrolyte. The cell length was determined beforehand spectroscopically to be $l = 0.12$ cm. A closer look on the geometry of the applied SEC-UV-vis cell reveals some special characteristics: 1. The large platinum mesh (0.7×0.8 cm) can be approximated as a planar electrode located in between both ends of the cell. 2. The white light beam emitted by the diode array orthogonally passes the center of the electrode with an estimated radius of only $\sim$ 0.1 mm. In consequence, both light beam and diffusion heading towards the same spatial direction. 3. Caused by diffusion towards or from the working electrode dependent on the applied potential species in solution are inhomogeneously distributed only along that spatial direction. The Lambert-Beer law is applicable since he absorption is proportional to the integral of the concentration profile along the light beam path.

2.3. Detection of absorbing species by SVDs

In order to determine the number of absorbing species we first compute an SVD of the complete data set D. The so-called abstract profiles by the first 5 left and also by the first 5 right singular vectors are shown in Fig. 2 together

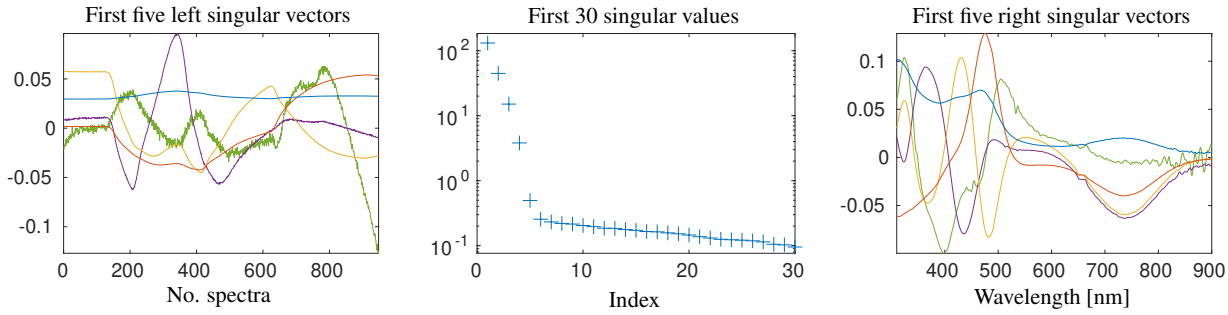

Figure 2: Characteristic parts of the SVD which contribute to the pure component recovery. The color code for the (i)th singular vector ($i = 1, \ldots, 5$) is blue (1), red (2), ochre (3), purple (4) and green (5). The singular values indicate a number of $s = 5$ absorbing chemical species.

with the 30 largest singular values plotted semi-logarithmically. In Fig. 2 the five largest singular values appear to be well-separated from the remaining smaller ones. Hence five species show a significant absorbance and should be reconstructed. The very good signal-to-noise-ratio of UV-vis data gives rise to the assumption that such a pure component recovery can be successful.

The underlying chemical process can be summarized as follows: As shown by Elvers et al. in [7] the initial uncharged complex $[Mo^{II}(CO)_2(cydt)(dppe)]$ (shortly named in bold face letters **Mo**) with molybdenum in the oxidation state +II can be reduced by two one-electron steps. First the the monoanion $\mathbf{Mo^{1-}}$ is formed and then the dianion $\mathbf{Mo^{2-}}$. Subsequent oxidation leads back to the uncharged species **Mo** and finally to the doubly oxidized dication $\mathbf{Mo^{2+}}$. The following sequence summarizes the redox events within one CV scan

$$\mathbf{Mo} \xrightarrow{+e^-} \mathbf{Mo^{1-}} \xrightarrow{+e^-} \mathbf{Mo^{2-}} \xrightarrow{-e^-} \mathbf{Mo^{1-}} \xrightarrow{-e^-} \mathbf{Mo} \xrightarrow{-2e^-} \mathbf{Mo^{2+}}.$$

While no change in the ligand sphere has been observed for the species introduced yet, the doubly oxidized complex $\mathbf{Mo^{2+}}$ is further substituting the carbonyl ligands with solvent (acetonitrile) molecules to eventually form a seven-coordinated solvent complex $[Mo(CH_3CN)_3(cydt)(dppe)]^{2+}$ ($\mathbf{MoACN_3}$). This non-redox substitution reaction is indirectly coupled to the potential at which $\mathbf{Mo^{2+}}$ is generated at the working electrode. Thus, it does not exist when passing the former potentials.

2.4. Evolving factor analysis and SVD-based analyses of subsystems

Due to the inherent structure of the data SEC measurements provide interesting opportunities and challenges for the computation of pure component decompositions. The key observation is that the reductive and oxidative regions contain different chemical species. Such additional information on the presence and absence of certain species is extremely helpful for the extraction of pure component information. For instance, we show below that the first spectra are solely determined by the initial species **Mo**. Further, after the first two reduction steps we expect that $\mathbf{Mo^{2-}}$ is the main species (although it cannot be chemically prepared due to the short time scale of the experiment).

EFA plots are the key tool for stating the appearance or absence of single chemical species or for identifying chemical subsystems. For the given data the results are presented in Fig. 3. If in an EFA plot a certain number of singular value curves appear to be separated from the bundle of superimposed curves (in the lower part of each EFA plot), then this number of independent chemical species can be expected in the reaction system at the respective time. We have computed forward and backward EFA plots for the complete data set as well as a forward EFA plot for the last 249 spectra. In all plots the logarithms of the dominant singular values are plotted versus the series of matrices with increasing or decreasing dimensions. The application of EFA to a series of matrices formed by the last 249 spectra helps to decide how many chemical species can be expected around the spectra indexes 700. EFA indicates that only two chemical species are of major importance within this window. Further, our aim is to detect concentration windows in which subsystems of the five ($s = 5$) chemical components are active. In particular **Mo**, $\mathbf{Mo^{1-}}$ and $\mathbf{Mo^{2-}}$ are checked concerning their contribution to the last 249 spectra. The EFA analysis justifies the following statements:

1. **Mo** is the only existing component within the first 125 spectra (see the blue curve in forward EFA). In order to answer the question if **Mo** is also present at the end of the CV cycle (which however is not a complete cycle, see Fig. 1), we compare the singular values of the matrix formed by the last 59 spectra with the singular values of the matrix formed by the last 59 spectra and extended by a column containing the first spectrum. The first spectrum represents only the initial species. The plot of these singular values is presented on the left in Fig. 4.

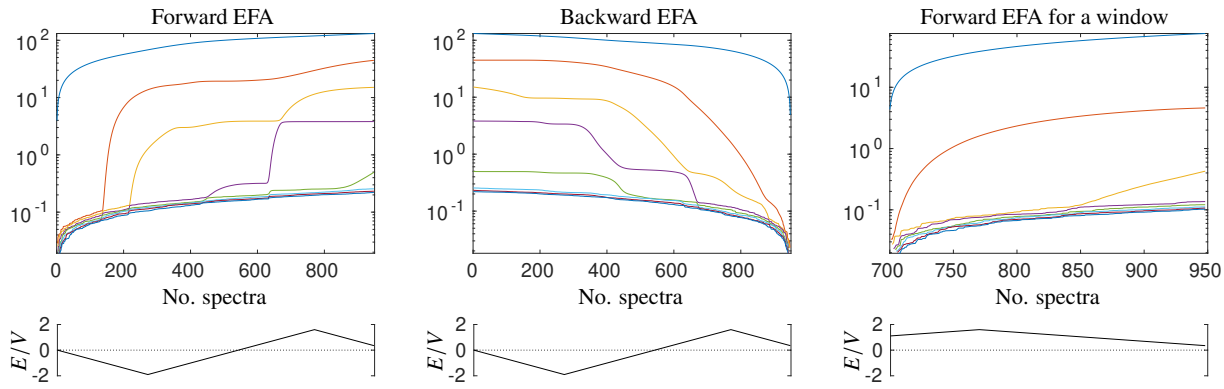

Figure 3: Forward (left) and backward (center) EFA plots for the given data set. The right plot is on an application of EFA only to the spectra 700 – 948. EFA allows to state the emergence of a new chemical species if a further isolated singular value curve develops from the bundle of superposed curves. The form of the bundle of superimposed curves is mainly caused by noise and measurement errors. Areas in which only one curve has come off the bundle indicate the prevalence of only one chemical species. The associated electrostatic potentials are drawn below each plot.

The last 59 spectra clearly indicate the presence of two absorbing species, but its extension by the first spectrum contains one further dominant singular value. We conclude that the initial species **Mo** is absent in the last 59 spectra.

2. Further, a similar comparison for a larger number of the last 249 spectra is presented in the centered plot of Fig. 4. The augmented matrix which includes the first spectrum does not show a further dominant singular value. This clearly indicates that **Mo** is present in the series of spectra with numbers between 700 and 948.

3. Similar comparisons of the dominant singular values are presented in the right plot of Fig. 4. The singular values indicate that the species $\mathbf{Mo}^{\mathbf{1}-}$ and $\mathbf{Mo}^{\mathbf{2}-}$ do not appear in the oxidative region of the last 249 spectra.

2.5. Presence and absence windows & data segmentation

Next the results on the presence or absence of chemical species are summarized. Based on the EFA curves as shown in Fig. 3 and the distribution of the largest singular values by computing SVDs of the subsystem matrices, see Fig. 4, it can be concluded that:

1. **Mo** is the only and pure component within the first 125 spectra, it is also present somewhere between spectra 700-948 (early oxidative region) and is absent in the spectra 890-948 (late oxidative region).

2. $\mathbf{Mo}^{\mathbf{1}-}$ arises approximately at the spectrum 125 and is absent in the spectra 700-948 (oxidative region).

3. $\mathbf{Mo}^{\mathbf{2}-}$ arise approximately at spectrum 200 and is absent in the spectra 700-948 (oxidative region).

4. $\mathbf{Mo}^{\mathbf{2}+}$ and $\mathbf{MoACN}_3$ are absent in the first 400 spectra (reductive region) and present somewhere in the spectra 700-948.

5. The solvent complex $\mathbf{MoACN}_3$ mainly absorbs in the spectra 850-948.

These findings allows us to separate the following two submatrices of the spectral data matrix D, namely

$$D^{(1)} = D(1:400,:), \qquad D^{(2)} = D(700:948,:).$$

The first submatrix $D^{(1)}$ relates to a subsystem in the reductive region and contains the $s = 3$ species **Mo**, $\mathbf{Mo}^{\mathbf{1}-}$ and $\mathbf{Mo}^{\mathbf{2}-}$. The second submatrix $D^{(2)}$ describes a ($s = 3$)-component subsystem in the oxidative region and contains the species **Mo** and $\mathbf{Mo}^{\mathbf{2}+}$ as well as $\mathbf{MoACN}_3$. Next these subsystems are separately analyzed. We compute the associated AFS sets and apply duality arguments. The analysis of the two subsystems has formal similarities to multiset MCR analyses, for details see 3.7. But here the subsystem definition is steered by information extracted from the EFA plots and in the subsystems bias-free AFS and duality techniques are applied.

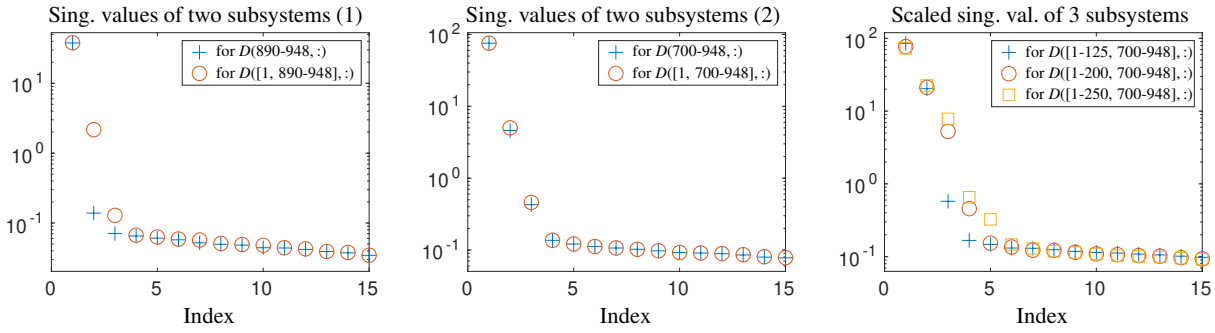

Figure 4: The fifteen largest singular values are drawn in a semilogarithmic plot with respect to either two subsystems (study of two cases) or three subsystems of the spectral data. Matching singular values help to identify the same chemical species whereas different singular values indicate different species and also the absence of certain species. Left: The singular values of the submatrix containing only the spectra of the last 59 spectra with indexes 890 to 948 are plotted by blue plus symbols. Only two chemical species can be assumed if singular values below 10^{-1} are attributed to noise. If the matrix of these spectra is augmented by a further column vector, namely the first spectrum which represents only **Mo** as a pure component, then the singular values (red circles) indicate that a new linear independent component has been added as three singular values are larger than 10^{-1}. This shows that **Mo** is absent at the end of the CV cycle. Center: A similar analysis is applied to the last 249 spectra. If again the the first spectrum is added as a new column to the subsystem matrix, then the largest singular values of the two matrices nearly match each other. We conclude that **Mo** is present in some of the last 249 spectra. Right: The subsystem consisting of the first 125 and the last 249 spectra includes 3 components (by blue plus symbols). Further the extended subsystem of the first 200 and the last 249 spectra includes 4 components (by red squares). Since $\mathbf{Mo^{1-}}$ arises around spectrum 125, it must be absent in the last 249 spectra. Finally, since the subsystem consisting of the first 250 and the last 249 spectra includes 5 components (ochre squares) and $\mathbf{Mo^{2-}}$ arises around spectrum 200, $\mathbf{Mo^{2-}}$ must be absent in the last 249 spectra. In the right plot the values are scaled in order to counter the upscaling effect of growing singular values with growing matrix sizes. For the left and the centered plot the upscaling effect can be neglected.

3. Subsystem analyses and pure component decompositions by AFS techniques

The MatLab software package FACPACK is used to determine the pure profiles underlying the subsystems $D^{(1)}$ and $D^{(2)}$. All five components of the complete data set can already be detected in the submatrices $D^{(1)}$ and $D^{(2)}$. The middle part (formed by the spectra 401-699) does not require an additional subsystem analysis since due to missing presence or absence windows no additional information gain can be expected. Conversely, this justifies to assume that no relevant chemical information is missed by ignoring the middle part of the spectra as the assembled matrix $\left[D^{(1)}, D^{(2)}\right]$ has the full rank of D. In a final step we compute the concentration profiles of the complete data D by merging the results of the two subsystems. Details are given below.

These computations succeed with the module *Duality/complementarity & AFS (3 comp. systems)* of FACPACK. Therein the AFS-sets are computed by the (standard and inverse) polygon inflation algorithm. Then the factorization is constructed interactively in the low-dimensional AFS representation by using the graphical user interface of the software. During this interactive construction, by moving the mouse pointer through the AFS and simultaneously watching the associated profiles, we determine those profiles which comply with the known conditions on the presence and absence of respective species. For the given experimental and noisy data the two control parameters ε_C and ε_S serve to accept small negative entries in the profiles of C and S. We use $\varepsilon_C = \varepsilon_S = 0.01$ so that negative entries are accepted which have an amplitude of 1% related to the maximal amplitude of each profile. See [39, 40, 36, 43] for more details on AFS-computations by polygon inflation.

3.1. Pure component decomposition for the first subsystem in the reductive region

The two AFS-sets, namely for the concentration and the spectra factors, for the first subsystem $D^{(1)}$ in the reductive region are presented in Fig. 5. The additional information on the windows in which **Mo**, $\mathbf{Mo^{1-}}$ and $\mathbf{Mo^{2-}}$ are present helps to determine a feasible solution by defining a proper simplex (triangle). For this step while working with the graphical user interface of FACPACK one can move the mouse pointer through the AFS. This yields only one triangle whose vertices present the concentration profiles of the three species and which comply with the known conditions on the presence and absence windows.The AFS sets are each connected sets with a hole around the origin. (The inner white sets are not the so-called inner polygons from the AFS constructions. These inner polygons are not drawn here.) The two triangles in the AFS plots of Fig. 5 are feasible as they are included in the respective outer polygons and are enclosing the respective inner polygons. The information from EFA that $\mathbf{Mo^{1-}}$ respectively $\mathbf{Mo^{2-}}$ arise after or near the spectra with indexes 125 respectively 200 supports the solution construction. This yields a unique pure component decomposition. Hence the pure component spectra of **Mo**, $\mathbf{Mo^{1-}}$ and $\mathbf{Mo^{2-}}$ and their concentration profiles with respect to the first 400 spectra can be locked in the given windows for the following computations. See Fig. 5 for the complete pure component decomposition of the subsystem $D^{(1)}$.

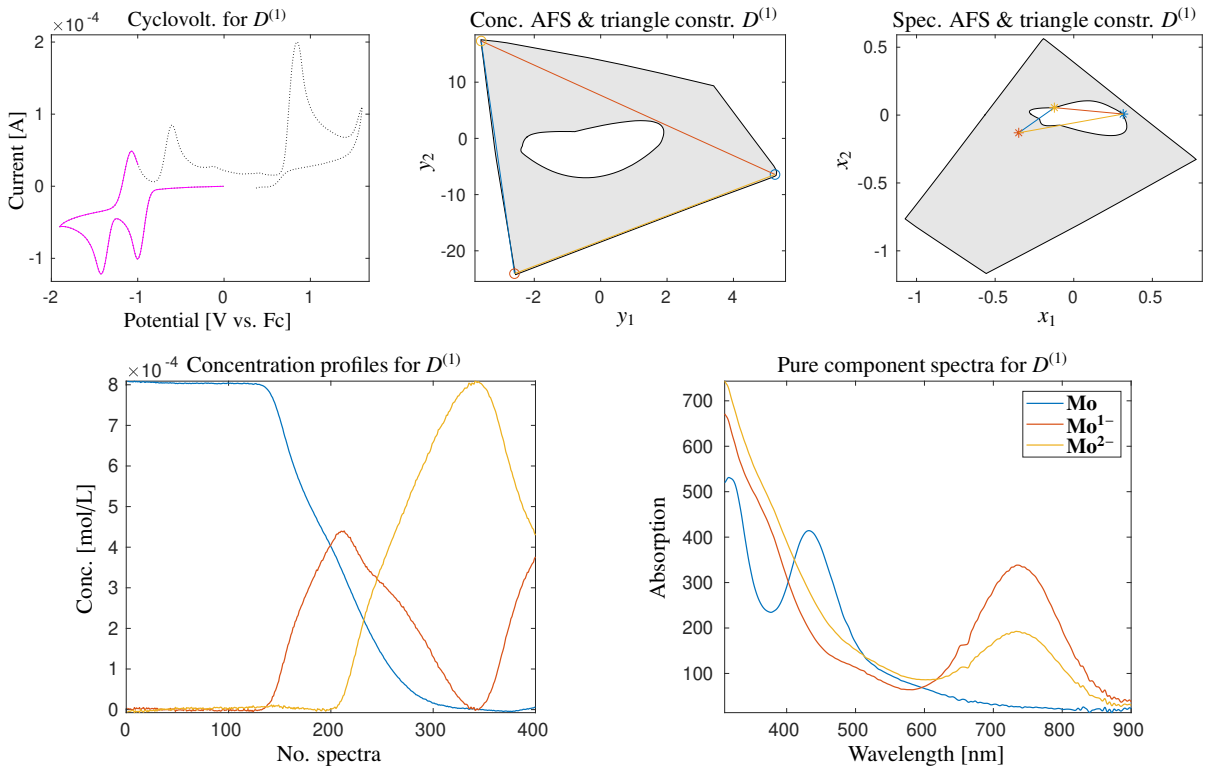

Figure 5: Reductive region: Construction of a partial pure component decomposition for the subsystem $D^{(1)}$ by means of the two associated AFS-sets (gray) and additional information on the absence windows. Top row: The cyclovoltammogram for the selected subsystem (left) shows the used spectra in magenta. The two AFS-sets include the feasible triangles whose vertices determine the associated profiles (center and right). Bottom row: The final concentration profiles and the spectra. The concentration profiles are normalized by closure constraints. The AFS-sets are computed for $\varepsilon_C = \varepsilon_S = 0.01$ so that a maximum of 1% of negative entries are acceptable.

3.2. Pure component decomposition for the second subsystem in the oxidative region

Next the subsystem $D^{(2)}$ in the oxidative region is analyzed. This analysis is much easier since some information is already available from the first subsystem analysis. Hence the AFS sets for the concentration profiles and the spectra (see top row of Fig. 6) show only for one species an associated area of uncertainty (plotted as gray areas). If this additional information were not available, then the AFS sets would be much larger. The boundaries of these larger sets are drawn by broken lines. Since the pure component spectrum of **Mo** is already determined by the first subsystem analysis it can be marked in the spectral AFS (blue star). We call this a locked profile. Additionally, the pure component spectrum of the solvent complex $\mathbf{MoACN_3}$ is available from a separate measurement (black dash-dotted line with a slightly different wavelength interval) and is also locked in the spectral AFS (green star). The remaining ambiguity for the spectrum of $\mathbf{Mo^{2+}}$ in terms of the AFS representation is the gray area.

Based on the presence and absence windows, the concentration profiles of **Mo** and $\mathbf{MoACN_3}$ are uniquely determined aside from scaling and are marked by blue and green circles in the concentration factor AFS. Once again, the remaining ambiguity for $\mathbf{Mo^{2+}}$ is plotted by a gray area. Duality arguments can even resolve this ambiguity. The point is that the pure component spectra of **Mo** and $\mathbf{MoACN_3}$ have already been determined so that, by duality, the concentration profile of $\mathbf{Mo^{2+}}$ is uniquely determined. A purple circle marks the position of this profile in the concentration factor AFS. Thus all three concentration profiles are uniquely determined and this implies the uniqueness of the full spectral factor. This finalizes the successful decomposition of the subsystem $D^{(2)}$.

3.3. Merging the results and computation of the final solution

The results of the two subsystem analyses in the reductive and the oxidative regions are now used in order to compute a pure component factorization of the original, complete CV cycle data set. Therefore the pure component spectra of all five species are first merged (**Mo** appears in both windows). Their best fits are computed with respect to the basis of the first 5 right singular vectors of the initial matrix D. The final step is solving a least-squares problem for computing the factor C so that

$$\left\| D - CS^{\mathrm{T}} \right\|_F^2 \to \min .$$

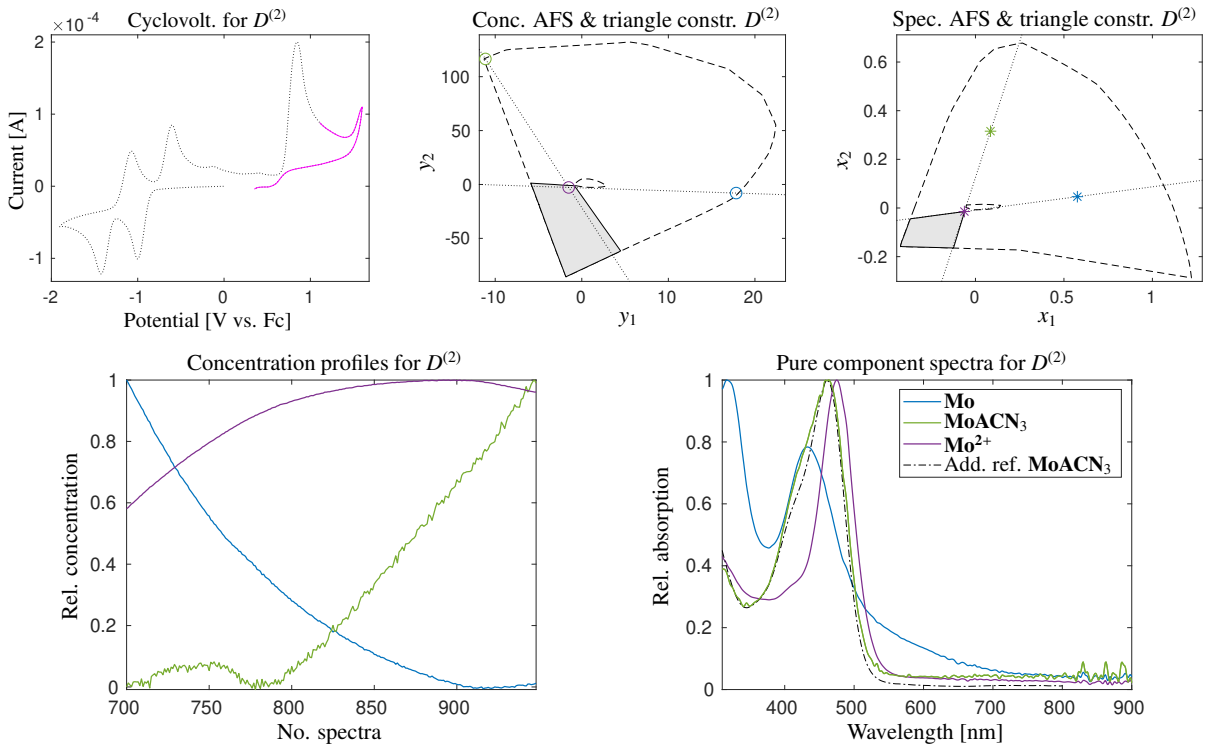

Figure 6: Oxidative region: Construction of a partial pure component decomposition for the subsystem $D^{(2)}$ by means of the two AFS-sets (gray) and additional information on the absence windows. Top row: The cyclovoltammogram for the selected subsystem (left) shows the used spectra in magenta. The two AFS-sets are again connected sets with a hole around the origin and are plotted by broken lines. However, some parts of the factors are already known; the respective known concentration profiles and spectra are marked by circles and stars. Only the remaining ambiguity for one profile in each AFS is marked by a gray area (center and right). Top row and right plot: Two pure component spectra are known (**Mo** from the decomposition of $D^{(1)}$ and **MoACN**$_3$ from a separate reference measurement, see the black dashed-dotted line in the lower right plot). Top row and centered plot: Due to duality, the single, remaining concentration profile is uniquely determined and marked by a purple circle. Since the other two concentration profiles are already known from the presence- and absence analysis above (blue and green circles in the AFS) all profiles can be computed. Bottom row: Resulting normalized concentration profiles and the spectra.

This amounts to computing the pseudo-inverse S^+ of S since $C = D(S^{\mathrm{T}})^+$. Alternatively, classical least squares (CLS) can be applied in order to compute C. The difference is that in our approach the least-squares problem is solved with respect to the basis of the five dominant singular values and in the alternative case with respect to the full basis underlying the data matrix D. However, the differences of the respective concentration factors are negligible. Then a scaling step is applied so that the initial concentration value $c_0 = 8.06 \cdot 10^{-4}$ mol/L is reproduced and closure constraints for C are fulfilled. The results are plotted in Fig. 7.

3.4. Result representation with respect to the electric potential

Up to here we have plotted the concentration profiles and the left singular vectors versus the spectra index. The benefit of this representation is that it allows a direct comparison of all these plots with the associated EFA results. EFA naturally uses the spectra index on the abscissa. In contrast to this, SEC plots use a representation versus the applied (cyclic) electric potential. Next we present the results plotted against the electric potential. Fig. 8 shows the EFA-information, namely the largest singular values, and Fig. 9 shows the concentration profiles for the five chemical species. The right sub-plot of Fig. 8 includes a certain scaling of the largest five singular values by $1/\sqrt{m}$ with m being the number of current spectra. This approximately allows us to split off effects of the growth of the singular values by the matrix dimensions, see [46, 28] for a discussion of this effect.

3.5. Discussion of the resulting pure component decomposition

The subsystem analyses for the given SEC experiment with five chemical species are the basis for a successful MCR analysis of the complete CV cycle. EFA contributes the key information on the presence and absence windows of the species. These windows clearly correlate with the reductive and oxidative regions of the CV cycle. The final merging of the subsystems with a joint species confirms that the pure component profiles (as determined for the subsystems) are still the correct profiles for the full data set. The experimental nature of the data D with the inevitable perturbations

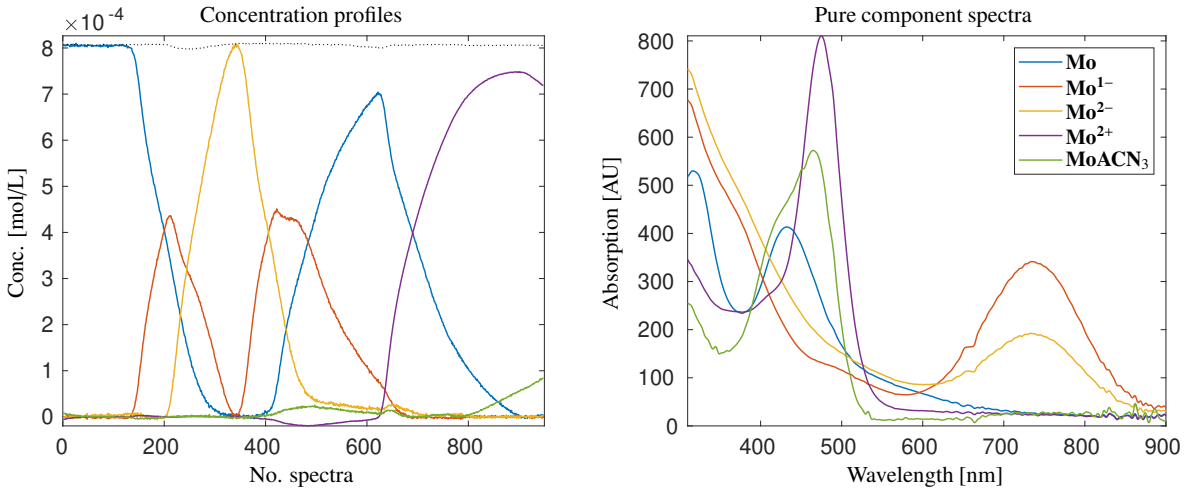

Figure 7: Final concentration profiles and spectra of all five species, namely the four oxidation states **Mo**, $\mathbf{Mo}^{1-}$, $\mathbf{Mo}^{2-}$, $\mathbf{Mo}^{2+}$ and the solvent complex $\mathbf{MoACN}_3$. The profiles are scaled by mass balance (closure) and the initial concentration $c_0 = 8.06 \cdot 10^{-4}$ mol/L. The sum of the concentration values of all species is plotted by a black dotted line.

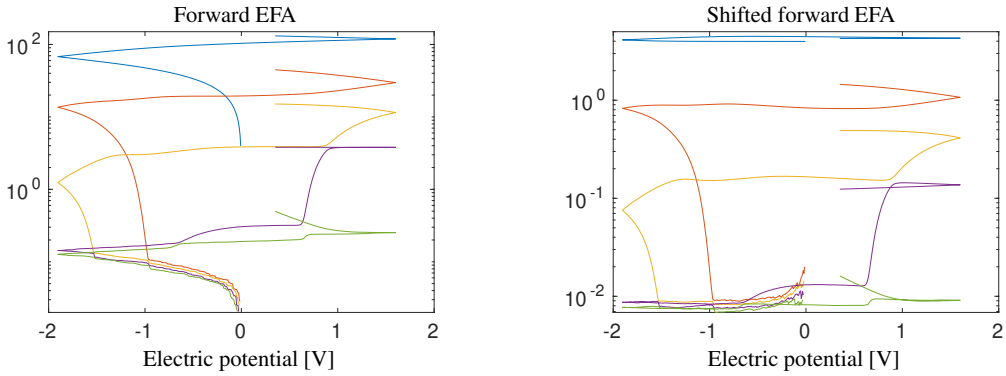

Figure 8: The forward EFA plots (for the dominant five singular values) against the electric potential. The right plot includes a certain scaling of the singular values by $1/\sqrt{m}$ with m being the spectra index. This approximately splits off the effect of the matrix dimensions on the size of the singular values.

and noise entail some uncertainty on the found windows of presence and absence. For example, the concentration of the solvent complex is not exactly equal to zero for all of the first 800 spectra. Furthermore, **Mo** is not completely absent for all of the last 59 spectra. However, the deviations are sufficiently small and do not prevent the subsequent steps of the decomposition process.

3.6. Final technical remarks on the maximal concentration combination within the first subsystem

In the first subsystem $D^{(1)}$ in the reductive region the two concentration profiles of **Mo** and $\mathbf{Mo}^{2-}$ show extremal pairwise ratios of concentration values with respect to the present species. For such a situation the analysis in [38] (on how to design chemical experiments in a way so that the factorization ambiguity of the associated factorization problem is as small as possible) shows that the ambiguity for **Mo** and $\mathbf{Mo}^{2-}$ is as small as possible for the given situation. The underlying idea is that the pure component spectra of a given chemical experiment are predefined and fixed; they may cause a considerable amount of factor ambiguity in the case of strongly overlapping spectra. In such a situation extremal pairwise ratios of concentration values are to be aimed at by means of a proper setup of the chemical experiment. This can help to reduce the inherent factor ambiguity. For details on the effect of extremal pairwise ratios of concentration values on extremal simplices in the AFS construction see [38]. Therein, the keywords in the context of the design of experiments are the so-called relevant spectra and minimal AFS-sets. See also [9, 33] on essential information respectively spectral pixels as well as [44] for locked points, the size of the remaining AFS-sets and the variability of the associated profiles. The data subset $D^{(1)}$ is an interesting example in this sense. Nevertheless, the remaining level of ambiguity is still relatively large in terms of the large AFS-sets which are typical for the given UV-vis data with their broad and few peaks. In this paper the key tools for reducing the ambiguity up to a unique

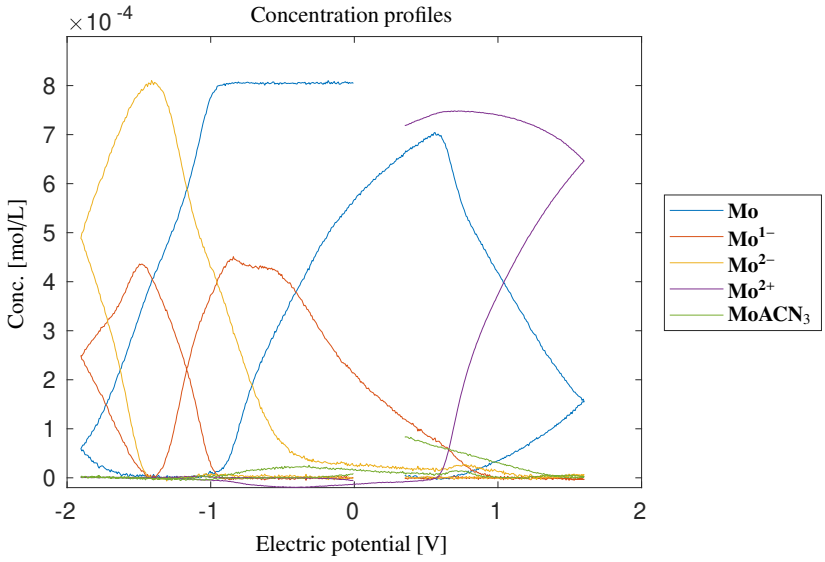

Figure 9: Plot of the final concentration profiles against the electric potential.

pure component factorization are the two-step subsystem analyses, duality arguments and the final merging of the subsystem analyses.

3.7. Limits on the applicability of the approach

Multiset analyses in chemometrics are powerful and versatile tools for solving MCR problems. They merge system information from multiple experiments, different spectroscopic techniques or repeated experiments. This includes techniques as data fusion or multi-way data analyses, see, e.g., [48] or the MCR-ALS software [15]. To some extent, our sub-window analyses can be considered as a multiset data analysis.

However, the present approach combines the submatrix analyses with a bias-free factor ambiguity analysis by means of AFS techniques. The focus is on reducing factor ambiguities. The EFA analysis supports the data partitioning. The approach has an adaptive nature. At its starting point the data partitioning scheme is unknown and also the number of chemical species is not given. Instead, EFA information steers the factor ambiguity analysis of the subsystems and indicates the emergence of new species. Even unexpected species can be detected. For the suggested method the subsystems should not have more than three chemical species since the AFS techniques work effectively only for systems with not more than three components. For systems with four or more components the factor ambiguity analysis by AFS techniques is complicated and costly. In the given SEC experiment the electrical potential serves as a control parameter which makes accessible subsystems with three components. Hence SEC experiments turn out as an ideal field of application. However, other applications are possible if the experimental conditions (like pH value changes or temperature changes) serve to move the reaction system to new equilibria including at least one new or at least one less species.

Rank deficiency can also complicate the data analysis especially for SVD-based techniques. The AFS for rank-deficient problems was recently introduced in [42], but further investigations are pending. Closely related to this are linearly dependent spectral profiles for which EFA cannot indicate the emergence of new species. However, such rank deficiencies are sometimes known beforehand or can be guessed at.

4. Summary and conclusion

The narrative of the Baron Munchhausen (the novel *Baron Munchausen's Narrative of his Marvellous Travels and Campaigns in Russia* was published in 1785) reports on Munchhausen's ability to pull himself out of the swamp by his own hair. Allegorically, the MCR analysis for the given SEC data shows a comparable problem solution with own means: A relatively complex chemical reaction system with five chemical species is decomposed into the pure component contributions and simultaneously all factor ambiguity is suppressed. The process steps are a proper EFA-based splitting of the data, subsystem analyses in the reductive and oxidative regions and a final merging procedure. We think that SEC experiments on the basis of the cyclic voltammetry in combination with recent MCR techniques

have a considerable potential for further developments. The electric potential is the decisive control parameter for equilibrium settings and for the definition of chemical subsystems containing chemical species of nearby redox levels. Depending on the electric potential some species have vanishing concentration values. Chemometric techniques such as EFA can be applied along with AFS techniques, duality arguments and experimental control methods. Experimental plans can be constructed by means of purest variables which finally support successful pure component decompositions.

References

[1] H. Abdollahi and R. Tauler. Uniqueness and rotation ambiguities in multivariate curve resolution methods. *Chemom. Intell. Lab. Syst.*, 108(2):100–111, 2011.

[2] S. Beyramysoltan, H. Abdollahi, and R. Rajkó. Newer developments on self-modeling curve resolution implementing equality and unimodality constraints. *Anal. Chim. Acta*, 827(0):1–14, 2014.

[3] S. Beyramysoltan, R. Rajkó, and H. Abdollahi. Investigation of the equality constraint effect on the reduction of the rotational ambiguity in three-component system using a novel grid search method. *Anal. Chim. Acta*, 791(0):25–35, 2013.

[4] O.S. Borgen and B.R. Kowalski. An extension of the multivariate component-resolution method to three components. *Anal. Chim. Acta*, 174:1–26, 1985.

[5] A. de Juan, M. Maeder, M. Martínez, and R. Tauler. Combining hard and soft-modelling to solve kinetic problems. *Chemom. Intell. Lab. Syst.*, 54:123–141, 2000.

[6] A. de Juan, Y. Vander Heyden, R. Tauler, and D. L . Massart. Assessment of new constraints applied to the alternating least squares method. *Anal. Chim. Acta*, 346(3):307–318, 1997.

[7] B. J. Elvers, M. Sawall, E. Oberem, K. Heckenberger, R. Ludwig, K. Neymeyr, C. Schulzke, V. Krewald, and C. Fischer. Towards operando IR- and UV-vis-Spectro-Electrochemistry: A Comprehensive Matrix Factorisation Study on Sensitive and Transient Molybdenum and Tungsten Mono-Dithiolene Complexes. *Chemistry-Methods*, 1(1):22–35, 2021.

[8] H. Gampp, M. Maeder, C. J. Meyer, and A. D. Zuberbühler. Calculation of equilibrium constants from multiwavelength spectroscopic data IV: Model-free least-squares refinement by use of evolving factor analysis. *Talanta*, 33(12):943–951, 1986.

[9] M. Ghaffari, N. Omidikia, and C. Ruckebusch. Essential spectral pixels for multivariate curve resolution of chemical images. *Anal. Chem.*, 91(17):10943–10948, 2019.

[10] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, and R. Tauler. A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data. *Anal. Chim. Acta*, 911:1–13, 2016.

[11] A. Golshan, H. Abdollahi, and M. Maeder. Resolution of rotational ambiguity for three-component systems. *Anal. Chem.*, 83(3):836–841, 2011.

[12] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2012.

[13] H. Haario and V.M. Taavitsainen. Combining soft and hard modelling in chemical kinetics. *Chemom. Intell. Lab. Syst.*, 44:77–98, 1998.

[14] R.C. Henry. Duality in multivariate receptor models. *Chemom. Intell. Lab. Syst.*, 77(1-2):59–63, 2005.

[15] J. Jaumot, A. de Juan, and R. Tauler. MCR-ALS GUI 2.0: new features and applications. *Chemom. Intell. Lab. Syst.*, 140:1–12, 2015.

[16] J. Jaumot, P. J. Gemperline, and A. Stang. Non-negativity constraints for elimination of multiple solutions in fitting of multivariate kinetic models to spectroscopic data. *J. Chemom.*, 19(2):97–106, 2005.

[17] A. Jürß, M. Sawall, and K. Neymeyr. On generalized Borgen plots. I: From convex to affine combinations and applications to spectral data. *J. Chemom.*, 29(7):420–433, 2015.

[18] H.R. Keller and D.L. Massart. Evolving factor analysis. *Chemom. Intell. Lab. Syst.*, 12(3):209–224, 1991.

[19] R. Kellner, J.-M. Mermet, M. Otto, M. Valcárcel, and H. M. Widmer, editors. *Anal. Chem.* Wiley-VCH, Weinheim, 2004.

[20] W.H. Lawton and E.A. Sylvestre. Self modelling curve resolution. *Technometrics*, 13:617–633, 1971.

[21] K. J. Lee, N: Elgrishi, B. Kandemir, and J. L. Dempsey. Electrochemical and spectroscopic methods for evaluating molecular electrocatalysts. *Nat. Rev. Chem.*, 1(5), 2017.

[22] C. W. Machan. Recent advances in spectroelectrochemistry related to molecular catalytic processes. *Curr. Opin. Electrochem.*, 15:42–49, 2019.

[23] M. Maeder. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.*, 59(3):527–530, 1987.

[24] M. Maeder and Y.M. Neuhold. *Practical data analysis in chemistry*. Elsevier, Amsterdam, 2007.

[25] M. Maeder and A. Zilian. Evolving factor analysis, a new multivariate technique in chromatography. *Chemom. Intell. Lab. Syst.*, 3(3):205–213, 1988.

[26] E. Malinowski. *Factor analysis in chemistry*. Wiley, New York, 2002.

[27] A. Neudeck, F. Marken, and R. G. Compton. UV/Vis/NIR Spectroelectrochemistry. In *Electroanalytical Methods*, pages 179–200. Springer Berlin Heidelberg, 2009.

[28] K. Neymeyr, M. Beese, and M. Sawall. On properties of EFA polts. Report University of Rostock, Submitted, 2021.

[29] K. Neymeyr, M. Sawall, and D. Hess. Pure component spectral recovery and constrained matrix factorizations: Concepts and applications. *J. Chemom.*, 24:67–74, 2010.

[30] N. Rahimdoust, M. Sawall, K. Neymeyr, and H. Abdollahi. Investigating the effect of flexible constraints on the accuracy of self-modeling curve resolution methods in the presence of perturbations. *J. Chemom.*, 30(5):252–267, 2016.

[31] R. Rajkó. Natural duality in minimal constrained self modeling curve resolution. *J. Chemom.*, 20(3-4):164–169, 2006.

[32] R. Rajkó and K. István. Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry. *J. Chemom.*, 19(8):448–463, 2005.

[33] C. Ruckebusch, R. Vitale, M. Ghaffari, S. Hugelier, and N. Omidikia. Perspective on essential information in multivariate curve resolution. *Trend Anal. Chem.*, 132:116044, 2020.

[34] M. Sawall, A. Börner, C. Kubis, D. Selent, R. Ludwig, and K. Neymeyr. Model-free multivariate curve resolution combined with model-based kinetics: Algorithm and applications. *J. Chemom.*, 26:538–548, 2012.

[35] M. Sawall, C. Fischer, D. Heller, and K. Neymeyr. Reduction of the rotational ambiguity of curve resolution techniques under partial knowledge of the factors. Complementarity and coupling theorems. *J. Chemom.*, 26:526–537, 2012.

[36] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. *On the analysis and computation of the area of feasible solutions for two-, three- and four-component systems*, volume 30 of Data Handling in Science and Technology, "Resolving Spectral Mixtures", Ed. C. Ruckebusch, chapter 5, pages 135–184. Elsevier, Cambridge, 2016.

[37] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. Simultaneous construction of dual Borgen plots. I: The case of noise-free data. *J. Chemom.*, 31:e2954, 2017.

[38] M. Sawall, C. Kubis, H. Schröder, D. Meinhardt, D. Selent, R. Franke, A. Brächer, A. Börner, and K. Neymeyr. Multivariate curve resolutions methods and the design of experiments. *J. Chemom.*, 32(6):e3012, 2019.

[39] M. Sawall, C. Kubis, D. Selent, A. Börner, and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: Concepts and applications. *J. Chemom.*, 27:106–116, 2013.

[40] M. Sawall and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. II: Theoretical foundation, inverse polygon inflation, and FAC-PACK implementation. *J. Chemom.*, 28:633–644, 2014.

[41] M. Sawall and K. Neymeyr. On the area of feasible solutions and its reduction by the complementarity theorem. *Anal. Chim. Acta*, 828:17–26, 2014.

[42] M. Sawall and K. Neymeyr. On the area of feasible solutions for rank-deficient problems: I. Introduction of a generalized concept. *J. Chemom.*, n/a(n/a):e3316, 2020. e3316 cem.3316.

[43] M. Sawall, H. Schröder, D. Meinhardt, and K. Neymeyr. On the ambiguity underlying multivariate curve resolution methods. In S. Brown, R. Tauler, and B. Walczak, editors, *In Comprehensive Chemometrics: Chemcial and Biochemical Data Analysis*, pages 199–231. Elsevier, 2020.

[44] M. Sawall, S. Vali Zade, C. Kubis, H. Schröder, D. Meinhardt, A. Brächer, R. Franke, A. Börner, H. Abdollahi, and K. Neymeyr. On the restrictiveness of equality constraints in multivariate curve resolution. *Chemom. Intell. Lab. Syst.*, 199:103942, 2020.

[45] H. Schröder, M. Sawall, C. Kubis, D. Selent, D. Hess, R. Franke, A. Börner, and K. 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.

[46] G. W. Stewart. Perturbation theory for the singular value decomposition. In: R. J. Vaccaro, editor. *In: R. J. Vaccaro, Editor, SVD and signal processing II: Algorithms, analysis and applications*, pages 99–109, 1991.

[47] 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.

[48] R. Tauler, M. Maeder, and A. de Juan. Multiset data analysis: Extended multivariate curve resolution. In S.D. Brown, R. Tauler, and B. Walczak, editors, *Comprehensive Chemometrics, Chemical and biochemical data analysis*, volume 2, pages 473–505. Elsevier, 2009.