analytical chemistry

Article

Cite This: Anal. Chem. 2018, 90, 7040−7047

pubs.acs.org/ac

Quantifying the Prediction Error in Analytical Multivariate Curve Resolution Studies of Multicomponent Systems

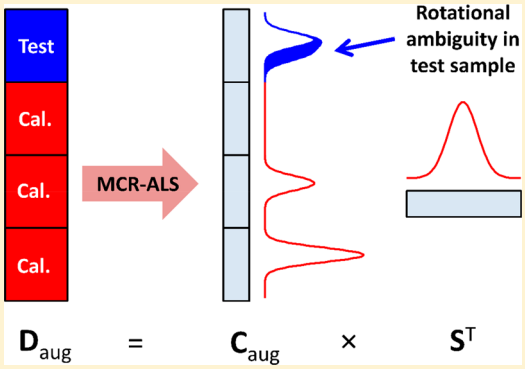
Rocío B. Pellegrino Vidal,[†] Alejandro C. Olivieri,[†,*] and Romà Tauler[‡]

[†]Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química de Rosario (IQUIR-CONICET), Suipacha 531, Rosario S2002LRK, Argentina

[‡]IDAEA-CSIC, Jordi Girona 18-24, Barcelona 08034, Spain

Supporting Information

ABSTRACT: In multivariate curve resolution (MCR) analysis, a range of feasible solutions is often encountered, because of the rotational ambiguities associated with the bilinear decomposition of data matrices. For quantitative purposes, the analysis is usually applied to a carefully designed set of calibration and test samples having uncalibrated interferents. Under the usual minimal constraints (non-negativity, unimodality, species correspondence, etc.), concentration and spectral profiles of the analyte in the test samples are not univocally recovered, unlike those in the calibration samples, especially when profile overlapping with the interferents is significant and selective regions do not exist for the analyte. In this report, a quantitative measure of the prediction errors due to rotational ambiguities is discussed, based on the calculation of the differences between the maximum and minimum area under the analyte concentration profiles calculated by the MCR-BANDS procedure. This methodology can be applied in different analytical scenarios with any number of analytes and interferents. Both absolute and relative quantitative errors due to rotation ambiguities are estimated and discussed in both simulated and experimental examples derived from liquid chromatography with diode array detection. The proposed procedure can be generalized to most of the analytical situations where every instrumentally measured sample produces a data table or data matrix.



Multivariate calibration based on the simultaneous analysis of multiple datasets from calibration and test samples is a well-established and powerful analytical methodology.[1] When these datasets are of chromatographic origin, i.e., one of the data modes is the elution time, one possible chemometric model to use is Multivariate Curve Resolution−Alternating Least-Squares (MCR-ALS) in its extended format,[2,3] i.e., MCR-ALS applied simultaneously to multiple datasets. The MCR-ALS model is based on a bilinear data matrix decomposition subjected to several chemically driven constraints. Even under naturally occurring constraints (non-negativity, unimodality or closure), this bilinear decomposition may not be unique, producing a range or area of feasible solutions (AFS), which is a phenomenon usually called "rotational ambiguity".[4]

Several approaches have been described in the literature for estimating the MCR feasible solutions. In a recent review, the methods have been classified in three groups.[4] One of them involves their explicit calculation based in geometrical principles,[5] which is of limited application to real systems with more than three components in the presence of noise. The second group includes computer-intensive grid searches or other similar comprehensive search tools.[6,7] For multicomponent analytical systems, as is often the case in real-world applications, these comprehensive searches become prohibitive.

Finally, a third approach was proposed by Gemperline[8] and Tauler,[9] based on a nonlinear optimization with nonlinear constraints, which seeks to maximize or minimize the signal contribution of the different components of the system. This is an attractive alternative, which can be applied to systems having any number of components. In the MCR-BANDS method, for every component of the system, a function defined by the ratio of its contribution to the overall signal contribution by all components is maximized and minimized under a variety of sensible constraints.[10]

For a two-component system, the maximum and minimum of the signal contribution function defined by MCR-BANDS coincide with the boundaries of the set of solutions provided by the comprehensive grid search methods in their two-dimensional (plane) plots.[11] However, for more than two components, the solutions corresponding only to the two extremes of the relative signal contribution function from each sample component may not adequately describe the entire multidimensional space of feasible solutions.[4] This may be due to the problem of defining a multidimensional volume or

Received: March 29, 2018
Accepted: May 11, 2018
Published: May 11, 2018

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

hyper-volume of feasible solutions using only two (maximum and minimum) profiles in a two-dimensional space.

In the experimental systems of analytical interest, where multiple samples are analyzed simultaneously by the extended MCR-ALS procedure,[2] the individual data matrices are arranged in an augmented data matrix, typically including a set of calibration samples with analytes at known nominal concentrations, and one or more test samples having the analytes at unknown concentrations together with uncalibrated interferents. Data matrix augmentation is performed along one of the experimental data modes (along the rows or the columns of the data matrix), in such a way that the bilinearity of the system is preserved. In chromatographic-spectral datasets, the natural augmentation mode is in the direction of the elution time (the matrix columns), preserving the spectral direction common for all data matrices simultaneously analyzed. This is because (1) the spectral measurements are reproducible, (2) the component spectra (the matrix rows) are usually invariant from sample to sample, and (3) the chromatographic time profiles are seldom constant, because of run-to-run retention time shifts. Once the column-wise augmented matrix is decomposed into bilinear components defined by their elution/concentration and spectral profiles, the area of the concentration subprofiles for each sample can be integrated to yield a series of score values: one for each analyte and for each sample. A univariate pseudo-calibration procedure can be then performed using the calibration scores and the known nominal concentrations. Prediction of the analyte concentration proceeds by interpolation of the analyte scores for the unknown samples.[1]

If the elution/concentration profiles are subjected to rotational ambiguity, this will be reflected in an uncertainty in the quantitative estimations, as has already been suggested based on grid search studies.[12] However, the latter approach is only applicable for samples with a small number of components. In the general case of complex multicomponent samples, the degree of uncertainty brought about by rotational ambiguity in MCR-ALS studies can be still estimated by the proposed MCR-BANDS procedure,[10] adapted so that the function to be minimized or maximized is the relative area under the analyte profiles, instead of the relative signal contribution function for each system component. This new procedure is still a component-wise optimization, since, for each component, the relative area of its concentration profile is minimized or maximized, relative to the total area of all components. Moreover, the spectral profiles of the different components are all normalized to unit length, and thus the different quantitative contribution of the components to the entire signal is only measured by the different contribution of the concentration profiles.

When the simultaneous analysis of multiple datasets from carefully designed calibration sets includes samples with pure analytes, unique solutions for their spectra and concentration subprofiles in the calibration samples are obtained.[13−15] However, in the presence of unknown interferents with strong profile overlapping with the analyte, both in the spectral and temporal modes, the recovery of the concentration profile for the analyte in the test samples may not be unique, although the number of feasible solutions can be significantly reduced, because of the imposed constraints.[16]

In this work, both simulated and experimental data will be investigated to illustrate the impact of the applied constraints and of remaining rotational ambiguities in predicted concentrations using extended MCR-ALS for quantitative purposes. Different simulated systems with two and three components are employed to compare the results by the comprehensive grid search and MCR-BANDS methods. For more components, the grid search method becomes prohibitive, from a computational point of view. However, conclusions can be generalized to systems with a larger number of components. The experimental system includes the determination of the herbicides carbaryl, norfluorazon, and 1-naphthol, and the xenoestrogen bisphenol A, in the presence of test sample interferents, using datasets from liquid chromatography with diode array detection (LC-DAD). MCR-BANDS was applied to assess the AFS under various constraints, to estimate the prediction uncertainty that can be expected for the analytes. No other approach would be useful for this system, because the samples contain up to six different components, which is a prohibitive number for comprehensive search approaches.

In a previous MCR-ALS analytical study, prediction uncertainties due to experimental noise were estimated using error propagation calculations.[17] However, rotational ambiguity generates a different type of uncertainty in analyte predictions. Although their differentiation can be difficult in practical situations, in this work, a measure of the degree of uncertainty (δ_{RA}) introduced by rotation ambiguities is calculated by the modified MCR-BANDS approach and the obtained results are discussed in detail.

THEORY

General. When MCR is used for quantitative purposes where a data matrix is obtained for every analyzed sample, a column-wise augmented data matrix $\mathbf{D}_{aug}$ is built up from the individual matrices obtained for calibration and test samples. Decomposition of $\mathbf{D}_{aug}$ is accomplished by means of a bilinear model:[2,13,14,16]

$$\mathbf{D}_{aug} = \mathbf{C}_{aug}\mathbf{S}^{T} + \mathbf{E}_{aug} \quad (1)$$

where $\mathbf{C}_{aug}$ is the column-wise augmented matrix of concentration profiles of the various components in the different samples (individual data matrices), $\mathbf{S}^{T}$ the matrix of associated spectra, and $\mathbf{E}_{aug}$ a matrix of model residuals. The individual spectra can be scaled in different manners; here, their 2-norms (the square root of the sum of squared vector elements) are set to 1. The concentration profiles, on the other hand, reflect the relative component concentrations in each sample; the area under each subprofile (its 1-norm) is proportional to the component concentration in each of the samples.

When the MCR-ALS procedure is used,[2,13,14] the bilinear decomposition of $\mathbf{D}_{aug}$ is subjected to a variety of chemically meaningful constraints, such as (1) non-negativity, because concentrations of mixture constituents and their instrumental (spectral) responses should be non-negative; (2) unimodality, implying that a single peak is observed in the concentration profiles (similar to that observed in chromatography); (3) closure, which is related to chemical mass balance; (4) correspondence between constituents in the different simultaneously analyzed samples (which forces to zero all elements of the concentration profile of the components known to be absent in a particular sample, e.g., the interferents in the calibration samples); (5) concentration correlation between the area of resolved profiles and reference values; and (6) local rank and selectivity for those profiles that are known to be zero in a

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

certain spectral or concentration region (which forces only some of the elements of a concentration or spectral profile of a particular component to zero, because it is known that the pure component does not appear in those elements). In the case of the simultaneous analysis of multiple datasets, it is also possible to implement multilinear (trilinear or quadrilinear) constraints.[13,14]

Generally, except for the cases of multilinear and selectivity/local rank constraints, the application of the above constraints during the ALS optimization may produce a nonunique pair of solutions for matrices $\mathbf{C}_{aug}$ and $\mathbf{S}^T$. This implies that a set of feasible solutions may exist, fitting equally well the data and satisfying the same constraints, which can be described by a rotation matrix $\mathbf{T}$:

$$\mathbf{C}_{RA}\mathbf{S}_{RA}^T = \mathbf{C}_{aug}\mathbf{T}\mathbf{T}^{-1}\mathbf{S}^T \quad (2)$$

where the set of $\mathbf{C}_{RA}$ and $\mathbf{S}_{RA}$ profiles form the AFS in the concentration and spectral modes, respectively. The AFS can be estimated by different methods,[4] such as the grid search[6,7] or the MCR-BANDS[10] methods, as described in more detail in the Supporting Information.

In this work, the MCR-BANDS method was modified to maximize and minimize the relative area values under the concentration profiles of the maximum and minimum feasible solutions for the analyte in the test sample, i.e., $\max(a_{test})$ and $\min(a_{test})$ (see more details in the Supporting Information). They can be used to estimate the prediction error due to rotational ambiguity (δ_{RA}) in concentration units as

$$\delta_{RA} = \frac{\max(a_{test}) - \min(a_{test})}{s} = \frac{\Delta a_{test}}{s} \quad (3)$$

where s is the slope of the univariate MCR-ALS calibration graph of scores versus nominal analyte concentrations. If $\delta_{RA} = 0$, it implies that the concentration profiles have been uniquely recovered in the test sample. The corresponding relative error (expressed in units of %) can be computed as

$$RE_{RA} = 100\frac{\Delta a_{test}}{a_{test}} \quad (4)$$

where a_{test} is the area of the analyte concentration profile in the test sample.

Simulations. Simulated datasets are intended to mimic experimental LC-DAD data. Three datasets were obtained, covering different scenarios, with the following number of analytes and interferents: dataset 1, a single analyte and a single interferent; dataset 2, two analytes and a single interferent; and dataset 3, a single analyte and two interferents. Table 1 lists the nominal concentrations of the analytes and interferents in the three sets.

The individual sample matrices for these three simulated LC-DAD cases, $\mathbf{D}$, were built as the sum of the analyte and interferent contributions defined by the bilinear model as

$$\mathbf{D} = \sum_{n=1}^{N} \mathbf{c}_n\mathbf{s}_n^T \quad (5)$$

where $\mathbf{c}_n$ and $\mathbf{s}_n$ are the concentration (elution) and spectral profiles for the nth component (analyte or interferent). The size of the $\mathbf{D}$ matrices is $(J;K)$, where J and K are the number of data points in the concentration (elution time) and spectral directions respectively: $J = 80$ and $K = 100$ in the presently simulated datasets.

Table 1. Composition of the Samples in the Simulated Datasets

sample	analyte 1	analyte 2	interferent 1	interferent 2
		Dataset 1		
Dtest	0.5		0.5	
Dcal1	0		0	
Dcal2	0.5		0	
Dcal3	1		0	
		Dataset 2		
Dtest	0.5	0.5	0.5	
Dcal1	0	1	0	
Dcal2	0.5	0.5	0	
Dcal3	1	0	0	
		Dataset 3		
Dtest	0.5		0.5	0.5
Dcal1	0		0	0
Dcal2	0.5		0	0
Dcal3	1		0	0

In all cases, the augmented data matrix $\mathbf{D}_{aug}$ was then built by column-wise matrix augmentation, i.e., by appending the individual data matrices $\mathbf{D}$ from the different samples analyzed along the concentration (elution time) mode. In MATLAB notation:[18]

$$\mathbf{D}_{aug} = [\mathbf{D}_{test}; \mathbf{D}_{cal1}; \mathbf{D}_{cal2}; \mathbf{D}_{cal3}] \quad (6)$$

meaning that the matrices are placed on top of each other, and that the size of $\mathbf{D}_{aug}$ is $(4 \times J;K)$.

For the three simulated systems, the grid search method was applied to estimate the presence of rotational ambiguity under different constraints. The relevant elements of the rotation matrices $\mathbf{T}$ describing the AFS according to eq 2 were computed as detailed in the Supporting Information. In addition, MCR-BANDS was independently applied to calculate the set of profiles giving minimum and maximum area under the analyte concentration profiles in the test samples for each of the simulated systems and under the same constraints (see the Supporting Information). The results were used to estimate the values of δ_{RA} and RE_{RA} in each case.

Software. MATLAB[18] version R2012a was used for all simulations, for grid search studies, and for the application of MCR-BANDS.[10] The latter requires access to the MATLAB optimization toolbox.

EXPERIMENTAL SECTION

The experimental system involves the agrochemicals 1-naphthol (NAP), norfluorazon (NOR) and carbaryl (CBL), and the xenoestrogen bisphenol A (BPA).[19] Eight calibration samples were prepared with concentrations following a factorial design in the range 0−50 ng mL^{-1} for NAP, BPA and CBL, and 0−100 ng mL^{-1} for NOR, with two additional samples having only pairs of analytes at intermediate concentrations. Four spiked water samples were also prepared by adding aliquots of standard analyte solutions. The compositions of calibration and test samples are collected in Table 2. Chromatographic analysis was performed using an Agilent 1200 liquid chromatograph (Agilent Technologies, Waldbronn, Germany), equipped with a DAD. Absorbance spectra were collected every 0.4 s, from 200 to 320 nm, each 2 nm. Further information can be found in ref 19.

The number of components for MCR-ALS was estimated by principal component analysis: five components were required

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

Table 2. Composition of the Calibration and Test Samples in the Experimental Dataset

	Composition (ng mL^{-1})			
sample	NAP	NOR	BPA	CBL
	Calibration			
1	50.0	20.0	50.0	10.0
2	50.0	20.0	10.0	50.0
3	10.0	100.0	10.0	10.0
4	50.0	100.0	10.0	10.0
5	10.0	100.0	50.0	50.0
6	50.0	100.0	50.0	50.0
7	10.0	20.0	10.0	50.0
8	10.0	20.0	50.0	10.0
9	30.0	0.0	0.0	30.0
10	0.0	50.0	30.0	0.0
	Test			
1	30.0	82.0	25.0	12.0
2	8.0	26.0	42.0	16.0
3	10.0	34.0	38.0	9.0
4	7.0	91.0	28.0	8.0

for samples 1−3 and six components were required for sample 4, indicating that a single interferent occurred in the former ones and two in the latter. MCR-ALS was initialized by estimating the so-called purest spectra.[20] Analytes were identified by their spectral profiles, and their quantitation was performed through the corresponding pseudo-univariate calibration curves.[1]

RESULTS AND DISCUSSION

General. Details about the rotational ambiguity expected to be present in the three simulated datasets are given in the Supporting Information. All these systems meet the following requirements: (1) a calibration procedure is designed with pure analyte samples, and (2) test samples contain uncalibrated interferents that are not present in the calibration set. Based on the knowledge of the sample compositions, a correspondence constraint can be applied both to the interferent and to individual analytes to fix their concentration to zero where they are known to be absent.

The main result from the Supporting Information is that, in a general multicomponent case, the AFS of the augmented concentration profiles of the analyte ($\mathbf{c}_{RA}$) under the species correspondence constraint is defined by

$$\mathbf{c}_{RA} = \mathbf{c} + \sum p_{int}\mathbf{c}_{int} \tag{7}$$

where $\mathbf{c}$ and $\mathbf{c}_{int}$ are the MCR-ALS augmented concentration profiles for the analyte and for each of the interferents respectively, and p_{int} represents a rotation ambiguity parameter measuring the degree of "mixing" of the pure analyte profile with those of the interferents (see the Supporting Information). The specific values of p_{int} are dependent on the applied constraints and on the degree of profile overlapping in both data modes (spectral and concentration). In the event that p_{int} is practically zero, because of all of the imposed constraints, the solution will be unique. Notice that, in eq 7, the profile for a given analyte is not combined with those for other analytes, because of the use of calibration samples, where every analyte is present in its pure form.

The relationship between the signal contribution function (f) and the relative component area (α) is

$$f = \frac{\|\mathbf{c}_{RA}\mathbf{s}^T\|_2}{\|\mathbf{C}\mathbf{S}^T\|_2} = \frac{\left\|\left(\mathbf{c} + \sum p_{int}\mathbf{c}_{int}\right)\mathbf{s}^T\right\|_2}{\|\mathbf{C}\mathbf{S}^T\|_2} \tag{8}$$

$$\alpha = \frac{\left\|\left(\mathbf{c} + \sum p_{int}\mathbf{c}_{int}\right)\right\|_1}{\|\mathbf{C}\|_1} \tag{9}$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ indicate the 2-norm and 1-norm, respectively. Because of the different meaning of the two norms, and, consequently, the different behavior of f and α, as a function of p_{int}, the conditions for maximizing or minimizing f for the analytes may differ, generally, from those maximizing or minimizing α. Therefore, the separate maximization/minimization of the relative areas when employing MCR-BANDS is required to better estimate the effects of rotational ambiguity in the quantitative predictions of the analyte concentrations, which are better measured from the difference between maximum and minimum analyte areas. This difference can be directly ascribed to the test sample, because the concentration profiles in the calibration samples can be uniquely recovered.

In the case of the interferents, neither their spectral nor their concentration profiles may be recovered uniquely. This is because only limited information on the interferents is available, since they are only present in the test sample, and extensive overlapping may occur among profiles in both data modes (i.e., with practically no selective regions). A complete discussion on the interferent results can be found in the Supporting Information. Notice that the ambiguity in the concentration interferent profiles for datasets where a single interferent occurs is only due to different vertical scales; if these AFS profiles are normalized to unit 2-norm, they will show a unique profile.

Simulated Data. Figure 1 shows the concentration and spectral profiles for the analyte(s) and interferent(s) for the

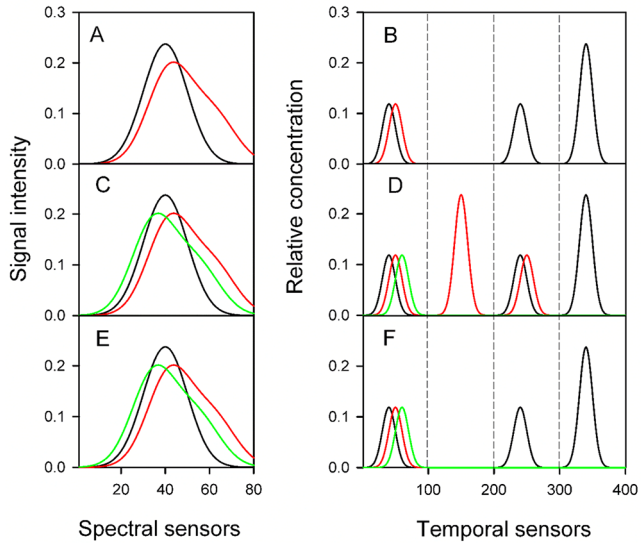


Figure 1. Spectral (left) and augmented concentration (right) pure component profiles for the simulated data sets. (A and B) dataset 1, analyte, black line, interferent, red line; (C and D) dataset 2, analyte 1, black line, analyte 2, red line, interferent, green line; (E and F) dataset 3, analyte, black line, interferent 1, red line, interferent 2, green line. In panels (B), (D), and (F), concentration profiles are on the left for the test sample and on the right for the remaining three calibration samples. Different samples are separated by vertical dashed lines.

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

pure components in the three simulated systems. A rather high degree of spectral overlapping was introduced to highlight the rotational ambiguity effects. In a first stage, the AFS was computed by imposing the following constraints: non-negativity in the profiles in both modes, and species correspondence for analyte and interferent. In the case of the simulated dataset 1 (Table 1), Figure 2 shows the AFS

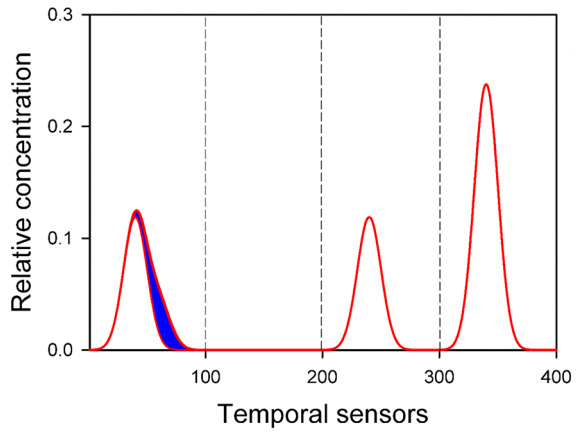


Figure 2. Area of feasible solutions for the augmented concentration profile for the analyte 1 (see Figure 1B), calculated by the grid search method. Red lines are the profiles giving the maximum and minimum area and signal contribution values below them. Non-negativity and correspondence constraints were applied. The left-hand side shows data for the test sample, and the right-hand side shows data for the three calibration samples. Different samples are separated by vertical dashed lines.

computed using the grid search method for the analyte concentration profile: the first left panel corresponds to the test sample, and the subsequent three correspond to the three calibration samples. The spectrum of the analyte is not shown, because it is uniquely recovered. As can be seen in Figure 2, the concentration profiles of the analyte in the calibration samples are unique, while rotational ambiguity is concentrated in the concentration profile of the analyte in the test sample (left panel). The profiles corresponding to maximum values of f and α are coincident, and shown as red lines in Figure 2.

The AFS of the interferent in its concentration and spectral profiles are given in Figure S1 in the Supporting Information. Grid search computations agree with the theoretical results, i.e., there is a significant rotation ambiguity for the interferent profiles in both data modes under the applied constraints. However, the concentration interferent profile becomes unique after normalization to unit 2-norm (Figure S1A in the Supporting Information).

It is important to note that the rotation ambiguities and the AFS could be further reduced and even eliminated if additional constraints are imposed to the system, e.g., local rank and selectivity constraints. In Figures 1A and 1B, the presence of concentration and spectral regions where only the interferent responds is reflected. If this information is employed as a selectivity constraint during the grid search, the profiles are uniquely recovered for both sample components and in both data modes. There are some limitations for the application of local rank constraints, because of possible rank deficiencies in some particular cases.[21] However, in the present case local rank constraints are applied in regions where it is known that the analyte is not present. Therefore, generally, if proper local rank conditions exist[13,15] and they are applied as additional restrictions, the solution will be unique. On the other hand, when these local rank conditions do not exist or they exist but are not used as a constraint (as in Figure 2), rotation ambiguities will most probably be present and the AFS might be significant.

Table 3 presents the results of the minimum and maximum areas for the analyte in the test sample a_{test}, as estimated by

Table 3. Prediction Errors Due to Rotation Ambiguities in the Investigation of Simulated Data Systems 1, 2, and 3[a]

	Dataset 1	Dataset 2		Dataset 3
parameter	analyte	analyte 1	analyte 2	analyte
Non-negativity Constraints				
$\max(a_{test})$	4.0	5.3	3.7	5.8
$\min(a_{test})$	3.0	3.0	3.0	1.9
Δa_{test}	1.0	2.3	0.7	3.9
a_{test}	3.0	3.0	3.0	3.0
slope, s	6.0	6.0	6.0	6.0
δ_{RA}	0.17	0.39	0.11	0.57
RE_{RA} (%)	34	78	22	>100
Non-negativity, Correspondence, and Local Rank for the Analyte in Test Sample Constraints				
$\max(a_{test})$	3.0	3.0	3.0	3.0
$\min(a_{test})$	3.0	3.0	3.0	3.0
Δa_{test}	0	0	0	0
a_{test}	3.0	3.0	3.0	3.0
slope, s	6.0	6.0	6.0	6.0
δ_{RA}	0	0	0	0
RE_{RA} (%)	0	0	0	0

[a]For the definition of Δa_{test}, a_{test}, δ_{RA}, and RE_{RA}, see eqs 3 and 4. Rotational ambiguity error (δ_{RA}) is expressed in arbitrary concentration units. Maximum and minimum areas were estimated by MCR-BANDS.[10]

MCR-BANDS under different imposed constraints. For dataset 1, and under non-negativity and correspondence constraints, the quantitative relative error due to rotation ambiguity for the single analyte under study is estimated as 34%. On the other hand, the prediction uncertainty due to rotation ambiguity is reduced to zero when local rank constraint is applied in addition to the former restrictions.

In the case of dataset 2, Figures 3A and 3B show the concentration profiles for analytes 1 and 2, respectively, because of rotational ambiguity, when only non-negativity and species correspondence are applied. As for dataset 1, the profiles for maximum and minimum values of f and α are coincident (red lines). The analyte concentration profiles in the calibration samples, as well as their spectral profiles, are recovered uniquely; however, rotational ambiguity significantly affects the concentration profile of the analytes in the test sample. The expected relative errors for the two analytes derived from rotational ambiguity are reported by MCR-BANDS as 78% and 22%, respectively (see Table 3). In the case of the interferent, the AFS shows a similar behavior to dataset 1, i.e., a significant rotation ambiguity in the AFS for its spectral and concentration profiles (if not normalized, see Figure S2 in the Supporting Information). As in the previous dataset, when additional restrictions are applied, e.g., local rank information for the analyte elution profiles in the test sample (Figures 1C and 1D), the solutions can be uniquely recovered (Table 3).

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

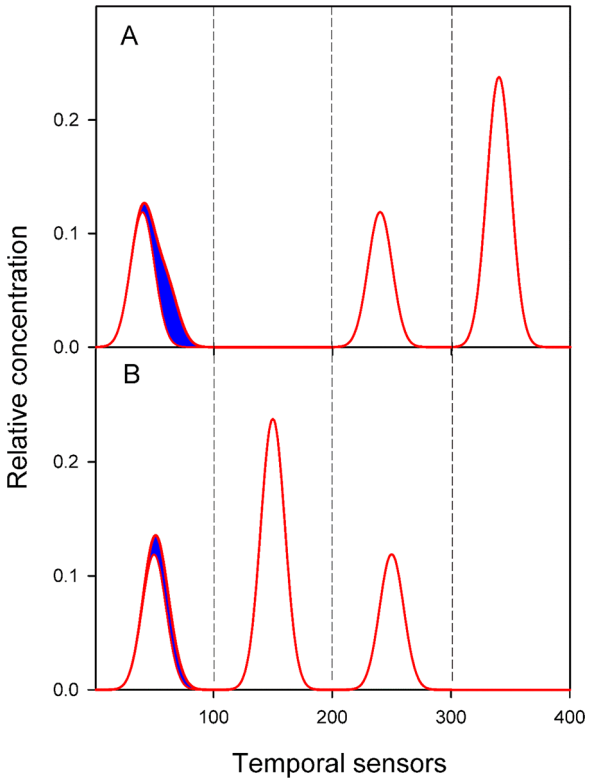


Figure 3. Area of feasible solutions for the augmented concentration profile for (A) analyte 1 and (B) analyte 2 in dataset 2 (see Figure 1D), calculated by the grid search method. Red lines are the profiles giving the maximum and minimum area and signal contribution values below them. Non-negativity and correspondence constraints were applied. The left-hand side shows data for the test sample, and the right-hand side shows data for the three calibration samples. Different samples are separated by vertical dashed lines.

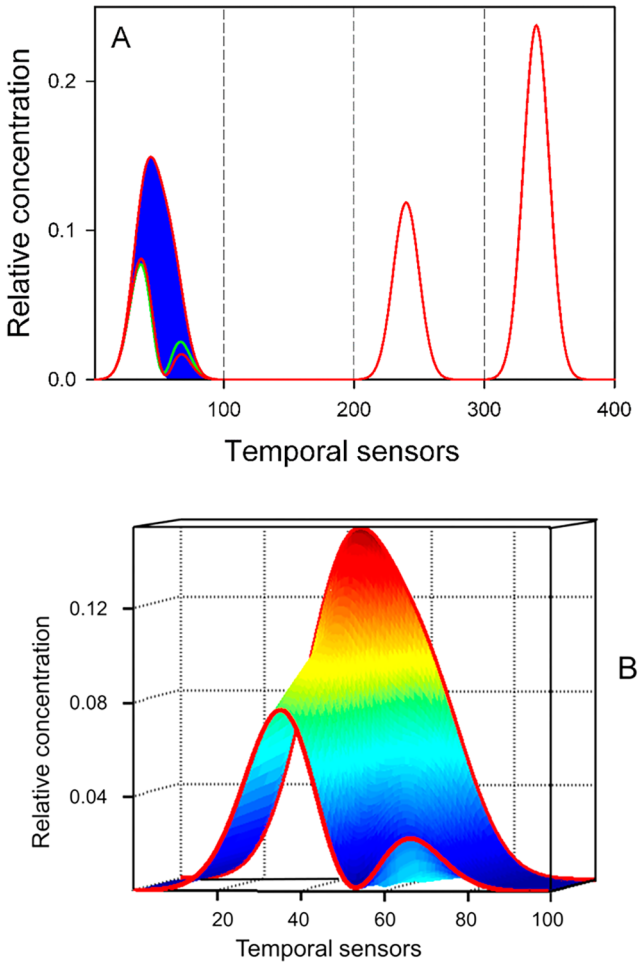


Figure 4. (A) Area of feasible solutions for the augmented concentration profile for the analyte in dataset 3 (see Figure 2F), calculated by the grid search method. Red lines define the profiles giving the maximum and minimum area values below them. Green line is the profile giving the minimum signal contribution function *f*, as calculated by MCR-BANDS (the maximum one is coincident with the upper one for maximum area in red). Non-negativity and correspondence constraints were applied. The left-hand side shows data for the test sample, and the right-hand side shows data for the three calibration samples. Different samples are separated by vertical dashed lines. (B) Three-dimensional plot of the AFS profiles for the test sample only. The red lines show the profiles for minimum and maximum signal contribution, according to MCR-BANDS.

Dataset 3 differs from the previous sets (two interferents occur in the test sample). Figure 4A shows the AFS (blue area) found by grid search for the augmented data matrix and for the test data matrix, under non-negativity in both data modes and species correspondence as constraints. The concentration profiles for the analyte in the calibration samples are unique (Figure 4A). In this dataset, the profile for the maximum of the analyte signal contribution function *f* is coincident with the one for the maximum relative area α. In contrast, the profile for the minimum of the analyte signal contribution function *f* (green line in Figure 4A) does not match with the one for minimum relative area α of the analyte. Although these findings correspond to the particular case under discussion, they may appear for systems with more than one interferent. For the interferents, AFS regions are shown in Figures S3 and S4 in the Supporting Information. They follow the same trend as in the previous two systems. If additional local rank constraints are applied, as described above, the analyte profiles can be uniquely recovered and those for the interferents are significantly reduced (see Figures S5 and S6 in the Supporting Information).

Notice also in Figure 4A that the profiles for minimum and maximum signal contribution, as provided by MCR-BANDS (red lines), do not completely describe the borders of the volume defined by the AFS profiles. This is better appreciated in the expanded three-dimensional plot of Figure 4B.

The prediction errors computed with MCR-BANDS under different constraints are reported in Table 3. As expected, when only non-negativity and species correspondence are employed as restrictions, quantitative analysis in dataset 3 is naturally hindered by the large extent of rotational ambiguity in the analyte score, leading to a very large error. However, under application of the additional local rank constraint, the rotation ambiguity is reduced to zero.

Experimental Data. Each of the four experimental test samples was separately processed, joining each test data matrix with the calibration samples in turn to build the augmented data matrix before MCR-ALS processing. For these systems having up to six components, comprehensive searches are prohibitive to find the AFS, which can only be assessed by MCR-BANDS.

When applying only non-negativity constraints, large prediction errors are computed by the MCR-BANDS method,

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

Table 4. Prediction Errors Due to Rotation Ambiguities in the Investigation of the Experimental System[a]

test sample[b]	NAP Δa_{test}	NAP a_{test}	NAP δ_{RA}	NAP RE_{RA} (%)	NOR Δa_{test}	NOR a_{test}	NOR δ_{RA}	NOR RE_{RA} (%)
Only Non-negativity Constraints								
1	47.1	67.8	15.9	69	31.2	83.6	24.9	37
2	20.1	33.7	6.3	60	13.4	34.3	10.6	39
3	19.5	29.4	6.7	66	13.1	34.2	10.5	38
4	413.9	79.7	142.5	>100	376.2	94.6	329.4	>100
Non-negativity, Unimodality, and Correspondence Constraints								
1	0	64.2	0	0	0	80.7	0	0
2	0	14.5	0	0	0	27.9	0	0
3	0	19.9	0	0	0	33.4	0	0
4	0	7.5	0	0	0	105.1	0	0
	BPA Δa_{test}	BPA a_{test}	BPA δ_{RA}	BPA RE_{RA} (%)	CBL Δa_{test}	CBL a_{test}	CBL δ_{RA}	CBL RE_{RA} (%)
Only Non-negativity Constraints								
1	28.8	24.8	27.5	>100	35.4	41.8	9.5	85
2	16.2	20.7	21.2	78	20.1	56.2	5.5	36
3	13.1	30.7	12.2	43	13.7	26.9	3.6	51
4	266.2	26	279.6	>100	304.6	38.8	83.5	>100
Non-negativity, Unimodality, and Correspondence Constraints								
1	0	27.9	0	0	0	37.1	0	0
2	0	30.5	0	0	0	53.4	0	0
3	0	34.0	0	0	0	11.0	0	0
4	0	10.8	0	0	0	13.6	0	0

[a]For the definition of Δa_{test}, a_{test}, δ_{RA}, and RE_{RA}, see eqs 3 and 4. The rotational ambiguity error δ_{RA} is expressed in ng mL^{-1}. [b]Test samples 1−3 contain a single interferent; test sample 4 contains two interferents.

ranging from 37% to more than 100% (see Table 4). Notice that when analyzing test sample 4, which has all four analytes and two interferents, the rotation ambiguity errors are considerably larger than for test samples 1−3, which only have a single interferent. As expected, increasing the profile overlapping with additional sample interferents increases the analyte prediction errors.

The application of further constraints, such as unimodality (for the analytes), as well as the correspondence of analyte species between samples and constituents, leads, in this experimental case, to unique decompositions. No analyte local rank constraints were needed in this case. In this experimental case, the overlapping among the elution/concentration profiles of the analyte and interferents was not so severe, in comparison with the simulated cases, and the remaining rotation ambiguities could be easily drastically reduced. This is proved in Table 4 by the fact that all prediction errors due to rotational ambiguities are negligible for the analytes in all samples. MCR-ALS results obtained in the analysis of the experimental dataset using the non-negativity, unimodality, and species correspondence constraints are shown in Figure 5 in the analysis of test sample 4 (see spectral profiles in Figure 5A and augmented concentration profiles in Figure 5B).

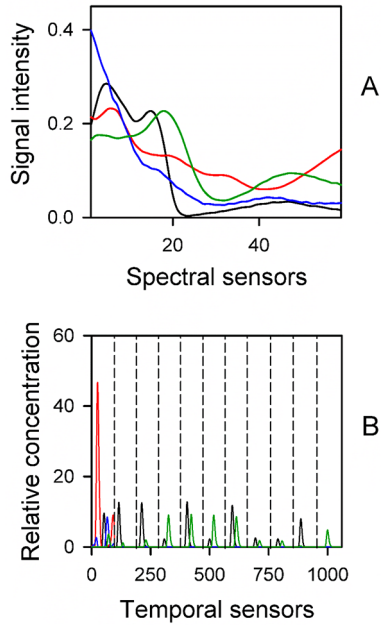


Figure 5. (A) MCR-ALS resolved spectra of the four analytes and two interferents. (B) MCR-ALS resolved elution (time) profiles. The left-hand side shows data for test sample 4 with the four analytes and two interferents, and the right-hand side shows data for the 10 calibration samples. [Legend: green lines, analyte NOR; blue lines, analyte NAP; purple lines, analyte BPA; red lines, analyte CBL; and gray lines, interferents in the real water sample. Vertical dashed lines separate the different samples.] Applied MCR-ALS constraints show non-negativity in both data modes, unimodality for the analytes in all subprofiles, and species correspondence.

CONCLUSIONS

Quantitative determinations using multivariate curve resolution methods should be accompanied by the corresponding estimation of prediction errors stemming from propagation of experimental uncertainties and from rotation ambiguities. The latter can be estimated, for general multicomponent systems, by the MCR-BANDS method based on a nonlinear optimization with nonlinear constraints, which maximizes and minimizes the

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047

signal contribution of the different species, adapted in this work to the maximization and minimization of the relative area under the concentration profiles of the different components or species. The difference between the extreme area values, converted to analyte concentrations, provides a way to estimate the prediction errors due to rotation ambiguities. In this work the MCR-BANDS method has been adapted to accomplish this goal with satisfactory results. Critical application of constraints apart from the basic non-negativity are needed to decrease and eventually eliminate rotation ambiguities and have good quantitative results, in particular, local rank and selectivity constraints and the possibility to have pure analyte standard samples. Results obtained in the present work can be generalized to most of the situations where MCR-ALS method is applied to get quantitative information from instrumentally measured signals producing a data table or data matrix in the analysis of single samples.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b01431.

(1) Detailed analysis of the mathematical expressions for the rotationally ambiguous bilinear profiles in the simulated datasets, under the species correspondence constraint, and a general conclusion regarding systems with multiple analytes and interferents; (2) description of the grid search methodology, employed to estimate the AFS for the bilinear decompositions in the simulated datasets; and (3) brief discussion of the original MCR-BANDS algorithm and the presently modified procedure based on the minimization and maximization of the relative component area under the concentration profiles (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel./fax: 54-341-4372704. E-mail: olivieri@iquir-conicet.gov.ar.

ORCID

Alejandro C. Olivieri: 0000-0003-4276-0369

Romà Tauler: 0000-0001-8559-9670

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.C.O. and R.B.P.V. thank Universidad Nacional de Rosario (Project No. 19B/487), CONICET (Project No. PIP 0163), and ANPCyT (Project No. PICT-2016-1122) for financial support. R.B.P.V. thanks CONICET for a doctoral fellowship. R.T. acknowledges the Ministerio de Economía y Competividad, Spain (Grant No. CTQ2015-66254-C2-1-P).

■ REFERENCES

(1) Olivieri, A. C.; Escandar, G. M. *Practical Three-Way Calibration*; Elsevier: Waltham, MA, USA, 2014.

(2) Tauler, R.; Maeder, M.; de Juan, A. Multiset data analysis: Extended multivariate curve resolution. In *Comprehensive Chemometrics*, Vol. 2; Brown, S., Tauler, R., Walczak, B., Eds.; Elsevier: Oxford, U.K.; 2009, pp 473−505.

(3) Escandar, G. M.; Olivieri, A. C. *Analyst* **2017**, *142*, 2862−2873.

(4) Golshan, A.; Abdollahi, H.; Beyramysoltan, S.; Maeder, M.; Neymeyr, K.; Rajkó, R.; Sawall, M.; Tauler, R. *Anal. Chim. Acta* **2016**, *911*, 1−13.

(5) Rajkó, R. *Anal. Chim. Acta* **2009**, *645*, 18−24.

(6) Golshan, A.; Maeder, M.; Abdollahi, H. *Anal. Chim. Acta* **2013**, *796*, 20−26.

(7) Sawall, M.; Kubis, C.; Selent, D.; Börner, A.; Neymeyr, K. *J. Chemom.* **2013**, *27*, 106−116.

(8) Gemperline, P. J. *Anal. Chem.* **1999**, *71*, 5398−5404.

(9) Tauler, R. *J. Chemom.* **2001**, *15*, 627−646.

(10) Jaumot, J.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2010**, *103*, 96−107.

(11) Abdollahi, H.; Maeder, M.; Tauler, R. *Anal. Chem.* **2009**, *81*, 2115−2122.

(12) Ahmadi, G.; Abdollahi, H. *Chemom. Intell. Lab. Syst.* **2013**, *120*, 59−70.

(13) Tauler, R.; Smilde, A. K.; Kowalski, B. R. *J. Chemom.* **1995**, *9*, 31−58.

(14) Tauler, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133−146.

(15) Manne, R. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 89−94.

(16) Olivieri, A. C.; Tauler, R. *J. Chemometr.* In press, DOI: 10.1002/cem.2875.

(17) Bauza, C. M.; Ibañez, G. A.; Tauler, R.; Olivieri, A. C. *Anal. Chem.* **2012**, *84*, 8697−8706.

(18) *MATLAB*; The Mathworks, Inc.: Natick, MA, USA.

(19) Pellegrino Vidal, R. B.; Ibañez, G. A.; Escandar, G. M. *Anal. Chem.* **2017**, *89*, 3029−3035.

(20) Windig, W.; Stephenson, D. A. *Anal. Chem.* **1992**, *64*, 2735−2742.

(21) Akbari Lakeh, M.; Rajkó, R.; Abdollahi, H. *Anal. Chem.* **2017**, *89*, 2259−2266.

DOI: 10.1021/acs.analchem.8b01431
Anal. Chem. 2018, 90, 7040−7047