Received: 22 September 2016 | Revised: 17 November 2016 | Accepted: 20 November 2016

DOI 10.1002/cem.2875

WILEY Journal of CHEMOMETRICS

RESEARCH ARTICLE

The effect of data matrix augmentation and constraints in extended multivariate curve resolution–alternating least squares

Alejandro C. Olivieri[1] | **Romà Tauler**[2]

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

[2] IDAEA-CSIC, Barcelona, Spain

Correspondence
Alejandro C. Olivieri, 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.
Email: aolivier@fbioyf.unr.edu.ar

Funding information
MINECO Spain, Grant/Award Number: CTQ2015-66254-C2-1-P; Universidad Nacional de Rosario; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Grant/Award Number: PIP 0163; Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Grant/Award Number: PICT-2013-0136

The reliability of results obtained by multivariate curve resolution (MCR) methods is strongly dependent on the absence or presence of a small degree of rotational ambiguity associated to them. In this work, the effect of rotational ambiguities on the profiles resolved by MCR methods is examined in detail for cases of interest to analytical chemistry, where a number of calibration samples are usually prepared containing analyte standards, while test samples may contain additional uncalibrated constituents. These multiple chemical data sets having common constituents are simultaneously analyzed using matrix augmentation strategies. In these cases, conditions for better resolution and improved profiles are more easily achieved. To evaluate the extension of rotational ambiguities and to quantify their reduction after matrix augmentation, we applied the MCR-BANDS procedure. Results obtained by the application of this procedure confirmed that the simultaneous analysis of multiple data sets decreased considerably the extension of rotational ambiguities compared with those obtained when only a single data set is analyzed. Simulated and experimental data sets of interest to second-order analytical calibration are discussed.

KEYWORDS

data matrix augmentation, mixture analysis with extended MCR, multivariate curve resolution, rotational ambiguity and feasible solutions

1 | INTRODUCTION

Multivariate curve resolution (MCR) belongs to a family of chemometric methods, which are especially suitable for the investigation and analysis of multicomponent chemical mixtures.[1,2] When these multicomponent systems are measured over a set of samples using multivariate instrumental responses, every measurement in a sample can be modeled by the sum of a reduced number of contributions. Each of them can be defined as the product of 2 factors, one related to the pure component response (eg, a spectrum) and another one related to the amount of this component (concentration) in the sample. The goal of MCR methods is the estimation of the unknown contributions of the different components in their mixtures (mixture analysis). A typical example is the hyphenated chromatographic case using a multichannel detector, such as a liquid chromatograph connected to a diode array detector or gas/liquid chromatography to a mass spectrometer.[3,4]

To achieve its goal, MCR methods use the information at hand as constraints, eg, known properties of the multivariate profiles defining the resolved components.[5] The most general constraint is that chemical concentrations of mixture components (constituents) and their responses in many instrumental methods should be nonnegative. Other common constraints are unimodality (implying a single peak shape) in chromatography, closure (chemical mass balance equations), or other properties known in advance to be fulfilled by the shape of the component profiles. Of outmost relevance are constraints referring to the previously known presence or absence (zeros) of a component in certain profiles. This occurs when a chemical constituent is known to be absent in a particular mixture sample, or when it is known that this constituent does not produce a significant signal in a particular measurement (spectral) region. These constraints are generally called selectivity and local rank constraints[2] and are especially useful to limit the number of possible solutions and define

the conditions to accomplish unique solutions as postulated in the resolution theorems.[6] A relatively easy way to implement constraints in MCR methods is using an alternating least squares algorithm (MCR-ALS), where constraints are applied to the different profiles of the constituents of the analyzed mixture at every iteration cycle of the ALS optimization. This method has been shown to provide good solutions in the analysis of different data types and mixture analysis problems.[5]

When only one data set is analyzed, the use of MCR methods may be challenged by the lack of unique solutions, ie, by the presence of rotational ambiguities, even under selectivity or local rank constraints. This specific subject, ie, single data matrices, has been widely discussed in the literature.[1,2,6–9] One of the most powerful strategies to limit rotational ambiguities in MCR methods is the simultaneous analysis of multiple data sets, all arranged in a data matrix via matrix augmentation strategies, giving rise the so-called extended MCR-ALS method.[10] In these cases, using an appropriate design of the simultaneously analyzed experiments is a powerful way to improve MCR solutions, reduce or eliminate rotational ambiguities, and achieve uniqueness (or at least arrive to an acceptable solution for the purposes of the study). Multiple data matrices are available in the most usual analytical scenario, ie, when preparing a set of analyte standards for calibration, and each of them generates a second-order data array or data matrix. Joining the calibration data matrices with those for test samples, which may contain additional uncalibrated constituents, produces a complete set of matrices that can be simultaneously processed with MCR-ALS. The effect of matrix augmentation has been previously discussed.[1,2,8] However, a discussion is still lacking on the extent of rotational ambiguity that may remain in these cases of high analytical interest, as those shown in the present work, and where the full arsenal of MCR-ALS constraints can be applied.

Different powerful strategies and algorithms for the calculation of the extension of rotational ambiguities and of the bands of feasible solutions have been proposed in the recent years.[9] Among them, the MCR-BANDS method is a rather easy-to-use tool that gives a numerical estimation of the extension of the remaining rotational ambiguities for a particular MCR solution under constraints, as well as a graphical display of feasible MCR solutions at the extreme values of an optimization function defined for relative components contribution to the whole measured signal.[11]

In the present work, the effect of matrix augmentation and constraints in MCR-ALS results is investigated in detail in systems of interest to second-order analytical calibration. The goal of the paper is to show how the simultaneous analysis of multiple data sets can improve MCR results. To show this improvement, we evaluated the extension of the reduction of rotational ambiguity associated with matrix augmentation and simultaneous analysis of different simulated and experimental multiple data sets using the MCR-BANDS method.[11]

2 | DATA SIMULATIONS

Data have been simulated for systems having 3 components. Noiseless profiles at unit concentration for all sample components are shown in Figure 1A,B in both data directions and modes: elution time and spectral, as experimentally recorded when running chromatographic experiments with multivariate spectral detection (UV-visible diode array or fast-scanning spectrofluorimetric detection). Using the analyte profiles shown in Figure 1, a calibration set of samples was built having 9 samples, with concentrations of components 1 and 2 (the calibrated analytes) following a central composite design in the range 0 to 1 concentration units.

The bilinear data matrix signal for each pure sample component $\mathbf{X}_n$ is given by the product of the corresponding concentration and spectral profiles in each mode:

$$\mathbf{X}_n = \mathrm{y}_n\, \mathrm{c}_n\, \mathrm{s}^{\mathrm{T}}{}_n, \tag{1}$$

where $\mathbf{c}_n$ and $\mathbf{s}_n$ are the $(I \times 1)$ and $(J \times 1)$ unit-concentration profiles in each mode (I and J are the number of channels), y_n is the specific component concentration, and the superscript "T" indicates matrix transposition ($\mathbf{c}_n$ and $\mathbf{s}_n$ profiles are all

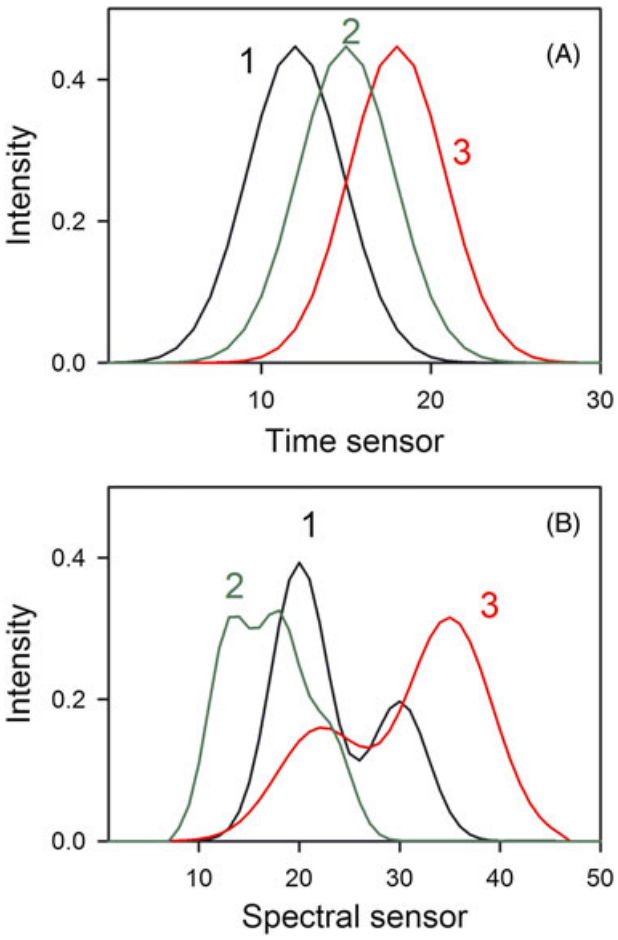

FIGURE 1 Noiseless profiles used to build the simulated data sets. A, Elution time profiles for 3 sample components. B, The corresponding spectral profiles. Components 1 (black) and 2 (green) represent the calibrated analytes, with component 3 (red) being the uncalibrated interferent

normalized to unit length). According to Figure 1, the size of the data matrices is 30 × 50 data points (temporal data points × spectral data points).

To produce the calibration data, the signal for a typical sample is given by the sum of the contributions of both analytes:

$$\mathbf{M}_{\text{cal}} = \mathbf{X}_1 + \mathbf{X}_2 + \text{Noise}, \tag{2}$$

where Noise represents a matrix of numbers from a Gaussian distribution with a standard deviation of 0.002 units, representing 2% with respect to the maximum calibration signal of each analyte at unit concentration.

For the test sample, on the other hand, both analytes were considered to be present at a concentration of 0.5 units, with the presence of a single chemical interference, at a concentration of 0.75 units (to ensure a significant amount of interference). The test matrix signal was therefore given by

$$\mathbf{M}_{\text{test}} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \text{Noise}, \tag{3}$$

where $\mathbf{X}_3$ represents the contribution of the interferent and Noise is as in Equation 2.

To compare the range of feasible solutions for a single data matrix and for an augmented data matrix, we submitted 2 different data sets to bilinear decomposition by MCR-ALS[12,13] and estimation of feasible solutions by MCR-BANDS.[11] The first one consisted of the single matrix $\mathbf{M}_{\text{test}}$ in Equation 3, applying the initialization procedure and restrictions during decomposition, which are described below.

The second data set involved a column-wise augmented data matrix $\mathbf{D}$, obtained by appending the $\mathbf{M}_{\text{test}}$ matrix with the 9 $\mathbf{M}_{\text{cal}}$ matrices along the temporal mode, in such a way that they share the spectral mode as the common mode:

$$\mathbf{D} = \begin{pmatrix} \mathbf{M}_{\text{test}} \\ \mathbf{M}_{\text{cal},1} \\ \cdots \\ \mathbf{M}_{\text{cal},9} \end{pmatrix} = \left[\mathbf{M}_{\text{test}}; \mathbf{M}_{\text{cal},1}; \ldots; \mathbf{M}_{\text{cal},9}\right]. \tag{4}$$

The size of $\mathbf{D}$ was therefore 300 × 50, and the augmented temporal data mode had 30 × 10 = 300 data points, on account of the temporal data points for each submatrix of $\mathbf{D}$ (30) and the total number of samples (10). Initialization and decomposition constraints are described below.

3 | EXPERIMENTAL DATA

The experimental data set studied in this work has already been reported and involves 2 analytes, benzo[*b*]fluoranthene (bbf) and benzo[*k*]fluoranthene (bkf), and one interferent, benzo[*j*]fluoranthene (bjf).[14] Eight calibration samples were used, consisting of duplicates of 4 different combinations of concentrations of both analytes (all in ng mL^{-1}): 0.0 and 0.0 (a blank sample), 100.0 and 2.0, 20.0 and 20.0, and 60.0 and 11.0 (the first entry corresponds to bbf and the second to bkf). These ranges were established on the basis of the analysis of the linear fluorescence-concentration range for each analyte. Test samples contained random concentrations of the studied analytes and bjf in the range 40 to 600 ng mL^{-1}.

All samples were prepared in acetonitrile/water (85:15 v/v) and submitted to high-performance liquid chromatography on a liquid chromatograph equipped with a Waters (Milford, Massachusetts) 515 HPLC pump and a Varian Cary-Eclipse luminescence spectrometer (Varian, Mulgrave, Australia) as detector. All data matrices were collected with the excitation wavelength fixed at 300 nm, using emission wavelengths from 388 to 580 nm, each at 2 nm, and times from 2.43 to 3.38 minutes, each at 2.7 seconds. The emission-time matrices were of size 121 × 22. For additional experimental details, see the work of Bortolato et al.[14]

4 | METHODS

From a mathematical point of view, the mixture analysis problem solved by MCR methods can be described by a bilinear model. In this model, experimental data are arranged in a table or matrix, $\mathbf{D}$, where a number of spectra ($I = 1, \ldots, I$, or other multivariate instrumental response) from a set of samples (eg, chemical mixtures formed by multiple constituents at different concentrations or compositions) are arranged as a row vectors of this data matrix, having wavelengths ($j = 1, \ldots, J$ instrumental channels) in the columns of this matrix. The MCR bilinear factor decomposition model can be written using linear algebra notation as (matrix sizes are given below each symbol):

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^{\mathbf{T}} + \mathrm{E}, \tag{5}$$

$$(I,J)\ (I,N)\ (N,J)\ (I,J),$$

where $\mathbf{C}$ (concentration profiles) and $\mathbf{S}^{\mathbf{T}}$ (spectra) are the factor matrices obtained by the bilinear decomposition of the experimental data matrix $\mathbf{D}$. This bilinear decomposition is performed for a number of components ($n = 1, \ldots, N$), which are contributing to the observed data variance in matrix $\mathbf{D}$. In MCR methods, this bilinear decomposition implies that the measured experimental spectra are the linear combination of the pure spectra of the constituents (components) of the analyzed mixtures, weighted by their respective concentrations.

There are different ways of solving Equation 5, one of them is using an alternating least squares (ALS) optimization algorithm (MCR-ALS method).[1,2,5,10,12,13] In this method, initial estimations of either the $\mathbf{C}$ or $\mathbf{S}^{\mathbf{T}}$ matrix are required to initiate the ALS optimization algorithm. Equation 5 is solved iteratively by linear least squares under constraints for the profiles in $\mathbf{C}$ and $\mathbf{S}^{\mathbf{T}}$ matrices.[12,13,15] The process is

repeated iteratively until the data matrix **D** is explained adequately, the residuals (**E** matrix of Equation 5) are the level of the experimental noise, and the solutions **C** and $\mathbf{S}^T$ are satisfactory, meaningful, and interpretable from a chemical point of view. Equation 5 is solved for a number of components (columns and rows of factor matrices **C** and $\mathbf{S}^T$), which are related to the number of data variance sources in **D**, ie, with the number of chemical species or constituents of the investigated mixture system having distinct concentration and spectra profiles. Constraints are the cornerstone of MCR iterative methods and can be defined as the systematic properties used to bring the iterative resolution process to optimal and chemically meaningful solutions of the concentration and response/spectra profiles. Typical constraints are nonnegativity, unimodality, closure, selectivity, local rank, and any other property known to be fulfilled by the profiles to be resolved. See previous works[12,13,15] for more details of the MCR-ALS method and possible constraints. In this work, nonnegativity constraints will be considered for the concentration and spectra profiles, and unimodality (single peak shape) for the concentration (elution mode) profiles. So as to avoid scale ambiguities in the model of Equation 5, spectra of the resolved components (rows) in $\mathbf{S}^T$ were normalized to equal length (Frobenious norm).

The MCR analysis can be significantly enhanced when multiple data sets arranged in different data matrices are simultaneously analyzed using the extension of the MCR-ALS method,[10] where the multiset data are analyzed by MCR-ALS via matrix augmentation schemes. The most commonly used is the column-wise augmented matrices, $\mathbf{D}_{aug}$, and the extension of the MCR bilinear model to it as

$$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3;\ldots;\mathbf{D}_K] = [\mathbf{C}_1;\mathbf{C}_2;\mathbf{C}_3;\ldots;\mathbf{C}_K]\,\mathbf{S}^T + [\mathbf{E}_1;\mathbf{E}_2;\mathbf{E}_3;\ldots;\mathbf{E}_K], \quad (6)$$

$$\begin{pmatrix}\mathbf{D}_1\\ \mathbf{D}_2\\ \mathbf{D}_3\\ \ldots\\ \mathbf{D}_K\end{pmatrix} = \begin{pmatrix}\mathbf{C}_1\\ \mathbf{C}_2\\ \mathbf{C}_3\\ \ldots\\ \mathbf{C}_K\end{pmatrix}\mathbf{S}^T + \begin{pmatrix}\mathbf{E}_1\\ \mathbf{E}_2\\ \mathbf{E}_3\\ \ldots\\ \mathbf{E}_K\end{pmatrix} = \mathbf{C}_{aug}\mathbf{S}^T + \mathbf{E}_{aug}, \quad (7)$$

where K is the number of matrices included in $\mathbf{D}_{aug}$. In a more compact form,

$$\mathbf{D}_{aug} = \mathbf{C}_{aug}\,\mathbf{S}^T + \mathbf{E}_{aug}. \quad (8)$$

Equation 8 is solved as Equation 5 using the MCR-ALS method under constraints to drive the algorithm to physically meaningful solutions. New constraints can be applied as a consequence of the column-wise matrix augmentation, like the possible correspondence of components in the different data sets (data matrices) simultaneously analyzed, and also because of the possibility of invariability of the concentration profiles of the same component in the different data sets as it happens in trilinear and other multilinear type of models, which are the extension of the bilinear model for 3-way and multiway type of data models (see previous works[5,10,16,17] for more details about these particular situations). In this work, however, we will only consider the possible correspondence of components among different data sets.

Solving MCR models in Equation 5 and its extension in Equation 8 by MCR-ALS under constraints provides one solution for **C** and $\mathbf{S}^T$ (or $\mathbf{C}_{aug}$ and $\mathbf{S}_{aug}^T$) factor matrices, which fits appropriately the data matrix, **D** (or $\mathbf{D}_{aug}$), and fulfills the applied constraints. However, there is no guarantee that this solution is unique, and in fact in most cases, the only application of soft constraints such as nonnegativity or unimodality does not provide unique solutions. In the absence of other stronger constraints (like selectivity or local rank), Equations 5 and 8 have an infinite number of possible solutions, because there are an infinite number of factor matrices **C** and $\mathbf{S}^T$ (when analyzing a single data matrix) providing the same result, the data matrix **D**. This indeterminacy can be described mathematically as

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^T = \left(\mathbf{C}\,\mathbf{T}^{-1}\right)\left(\mathbf{T}\,\mathbf{S}^T\right) = \mathbf{C}_{new}\,\mathbf{S}_{new}{}^T. \quad (9)$$

According to Equation 9, any invertible matrix **T** (N, N) gives a new set of equivalent solutions of the MCR model ($\mathbf{C}_{new}$ and $\mathbf{S}_{new}$ in Equation 9). Any linear combination of **C** and $\mathbf{S}^T$ solutions will produce new solutions of the bilinear model, which will be equivalent from a mathematical point of view. This type of indeterminacy in MCR methods is called rotational ambiguity, and it is the more critical and difficult type of ambiguity to be avoided in MCR solutions. Apart from rotational ambiguities, there are other 2 types of ambiguities, which are the scale and the permutation ambiguities.[9,11,18] However, the latter ambiguities are not problematic and can be easily handled by normalization and reordering of columns and rows of **C** and $\mathbf{S}^T$ matrices. In this work, only rotational ambiguities will be considered. The application of appropriate constraints in MCR methods can limit the extension of these rotational ambiguities, and in some cases eliminate them totally. Besides the natural constraints like nonnegativity or unimodality, the more powerful strategies to avoid rotational ambiguities in MCR methods involve the use of local rank and selectivity constraints,[2] the extension to simultaneous analysis of multiple data sets,[1] including the use of multilinear models,[5,10,16,17] and the use of hard (deterministic) modeling.[19,20] In this work, we will be only concerned with the use of the strategy based on matrix augmentation.

Different methods have been proposed in the literature for the evaluation of rotational ambiguities, including the calculation of the boundaries of the so-called feasible bands.[9,11,18] Feasible MCR solutions include the whole range of linear combinations of a particular MCR solution that fit the experimental data equally well and fulfill the constraints of the system, as defined by appropriate rotation matrices **T** in

Equation 9. Although different approaches have been proposed to calculate the full range of feasible solutions, most of them cannot be applied to systems with more than 4 components.[9]

Consider the possibility to define maximum and minimum values of these rotation matrices, $\mathbf{T}_{max}$ and $\mathbf{T}_{min}$, which should fulfill the following equation:

$$\begin{aligned}\mathbf{D}=\mathbf{C}_{init}\,\mathbf{S}^{T}{}_{init} &= \mathbf{C}_{init}\,\mathbf{T}_{min}\,\mathbf{T}_{min}{}^{-1}\,\mathbf{S}^{T}{}_{init} = \mathbf{C}_{min}\,\mathbf{S}^{T}{}_{min}\\ &= \mathbf{C}_{init}\,\mathbf{T}_{max}\,\mathbf{T}_{max}{}^{-1}\,\mathbf{S}^{T}{}_{init}\\ &= \mathbf{C}_{max}\,\mathbf{S}^{T}{}_{max}.\end{aligned} \tag{10}$$

In Equation 10, initial values of $\mathbf{C}$ and $\mathbf{S}^T$ matrices, $\mathbf{C}_{init}$ and $\mathbf{S}^T_{init}$, are known, while $\mathbf{C}_{min}$, $\mathbf{S}^T_{min}$ and $\mathbf{C}_{max}$, $\mathbf{S}^T_{max}$ correspond to $\mathbf{T}_{min}$ and $\mathbf{T}_{max}$ values, respectively. A possible algorithm to determine the extension of rotational ambiguities associated to a particular solution $\mathbf{C}_{init}\,\mathbf{S}^T_{init}$, obtained for instance by MCR-ALS method, is based on the definition of an objective function, which should be maximized and minimized as a function of $\mathbf{T}$ to find $\mathbf{T}_{max}$ and $\mathbf{T}_{min}$ values for every resolved component. This objective function should be a scalar function of the variables and should have well-defined boundaries (maximum and minimum). For a good performance of the optimization algorithm, this optimization function is scaled, for instance, between 0 and 1. The proposed optimization function is defined as follows[18]:

$$f_n(\mathbf{T}) = \frac{||\mathbf{c}_n(\mathbf{T})\mathbf{s}_n^T(\mathbf{T})||}{||\mathbf{C}\mathbf{S}^T||}. \tag{11}$$

This function gives the ratio between the contribution of a particular nth species (the numerator of Equation 11) with respect to the total contribution for all the components of the mixture (the denominator of Equation 11). The optimization (either maximized or minimized) of this objective function under constraints (see below) for each component $n = 1, \ldots, N$, will give an estimate of its maximum and minimum solutions ($f_n(\mathbf{T})$ max and min values respectively), from which the corresponding $\mathbf{T}_{max}$ and $\mathbf{T}_{min}$ matrices will be obtained, as well as the corresponding $\mathbf{c}_{n,max}$, $\mathbf{s}^T_{n,max}$ and $\mathbf{c}_{n,min}$ and $\mathbf{s}^T_{n,min}$ profile dyads, for each of the resolved components $n = 1, \ldots, N$. These extreme solutions should fulfill the constraints of the problem and give the relative maximum, $f_{n,max}$, and minimum, $f_{n,min}$, signal contribution of every component according to the function defined by $f_n(\mathbf{T})$ in Equation 11, ie, the ratio of the norm of $\mathbf{c}_n\mathbf{s}_n^T$ over the norm of the whole signal contribution from all components, $\mathbf{CS}^T$. In the MCR-BANDS method,[18] the optimization (maximum and minimum) of the function given by Equation 11 under constraints is performed using a nonlinear constrained nonlinear optimization problem, based on a sequential quadratic programming algorithm implemented in the MATLAB optimization toolbox *fmincon* function.[21]

The above optimization procedure produces 2 important outputs: (1) the profiles of every component in the 2 modes corresponding to the maximum, $f_{n,max}$, and minimum, $f_{n,min}$, contribution to the whole signal, which can be plotted for visual inspection; and (2) the difference between the maximum and minimum ($f_{n,max} - f_{n,min}$) of the component contribution (scaled between 0 and 1), which gives a measure of the extension of rotational ambiguity associated to this component. A value of 0 means no rotational ambiguity, while a value close to 1 corresponds to total ambiguity for this component. The effect of constraints is normally evaluated as leading to decreasing values of the difference ($f_{n,max} - f_{n,min}$). The MCR-BANDS procedure can be applied to any number of components and has already been implemented for a number of constraints, and also for the case of augmented data matrices. More details about the procedure, its implementation, and application can be found in the work of Jaumot and Tauler.[11]

In this work, the MCR-BANDS procedure is used for the evaluation of the extension of rotational ambiguities associated to a particular MCR solution. More specifically, the reduction of rotational ambiguities has been evaluated through the effect of matrix augmentation and constraints derived from this augmentation in the species correspondence between matrices.

5 | SOFTWARE

MATLAB[22] was used for producing the simulated data. The MCR-ALS was applied using the graphical interface MCR-ALS GUI 2.0 available at http://www.mcrals.info/,[12,13] which includes the MCR-BANDS[11] utility for estimating the feasible solution bands.

6 | RESULTS

6.1 | Simulation results: single data matrix

A single simulated data matrix having 3 components was first studied, built as described above from the profiles in both data modes shown in Figure 1A,B. The initial profiles for starting MCR-ALS decomposition were estimated by computing the purest variables,[23] assuming the presence of 3 sample components. During the least squares optimization to retrieve the final profiles, the applied constraints were nonnegativity in all profiles in both data modes, unimodality in the 3 elution time profiles, and normalization to unit norm in spectral profiles. Using the same restrictions, MCR-BANDS was used to compute the feasible solution bands. The results are shown in Figure 2A (elution time profiles) and Figure 2B (spectral profiles), which can be compared with the known pure profiles (Figure 2C,D). As can be seen, significant rotational ambiguity is present in both data modes, with the only exception of component 1 in the spectral mode

(Figure 2B), which was recovered with low ambiguity because of its selectivity in the chromatographic mode (region with local chemical rank equal to 1 in Figure 2C), at the beginning of its chromatographic elution. This higher selectivity in the chromatographic mode did not imply a better recovery of its elution profile, but a better recovery of its spectrum, in agreement with the application of resolution conditions described in previous works.[2,6]

Table 1 shows the differences between the maximum and minimum values of the optimization function ($f_{max} - f_{min}$), estimated by MCR-BANDS and scaled between 0 and 1. They are relatively large for the 3 components, confirming the presence of significant rotational ambiguities when only nonnegativity and unimodality constraints are applied in the analysis of the single data matrix simulated using profiles in Figure 1A,B. Moreover, from the quantitative perspective, it is the area of the elution time profile, which is proportional to the analyte concentration, and thus, significant uncertainty would exist in the calculation of the concentration from this elution profile area for both calibrated analytes 1 and 2.

TABLE 1 Summary of f_{optim} MCR-BANDS values for f_{max}, f_{init}, f_{min}, and the difference ($f_{max} - f_{min}$) (see Equation 11)

Components[a]	f_{optim}			
	f_{max}	f_{init}	f_{min}	$f_{max} - f_{min}$
Single Data Matrix (Simulated)				
1	0.603	0.461	0.409	0.194
2	0.460	0.412	0.327	0.134
3	0.576	0.449	0.372	0.204
Augmented Data Matrix (Simulated)				
1	0.581	0.579	0.579	0.002
2	0.572	0.571	0.571	<0.001
3	0.242	0.242	0.242	<0.001
Single Data Matrix (Experimental)				
bkf	0.782	0.674	0.404	0.315
bbf	0.282	0.277	0.123	0.159
bjf	0.537	0.368	0.225	0.311
Augmented Data Matrix (Experimental)				
bkf	0.452	0.443	0.419	0.033
bbf	0.664	0.644	0.595	0.069
bjf	0.038	0.037	0.037	0.001

[a]Experimental components are as follows: bbf, benzo[*b*]fluoranthene; bjf, benzo[*j*]fluoranthene; bkf, benzo[*k*]fluoranthene.

6.2 | Simulation results: augmented data matrix

In the case of the augmented data matrix, column-wise augmentation was performed along the columns or spectral mode, as is customary in MCR-ALS analysis of chromatographic-spectral matrix data sets. In this particular case, 1 test sample (sample 1 having both analytes and the interferent, as in the previous example of Figure 2) and 9 calibration samples having one or both analytes are simultaneously analyzed by MCR-ALS using the matrix column-wise augmentation strategy. The MCR-ALS initialization was also performed by estimating the profiles comprising the purest variables in the spectral mode, and optimization proceeded under the following simultaneous constraints: (1) nonnegativity in all 3 profiles in both data modes, (2) normalization to unit norm in spectral profiles, and (3) correspondence between components, which enforces the solution to have zero elements (selectivity) in the subprofiles corresponding to the absence of specific components in each submatrix. For example, the interferent (component 3) is only present in the test sample (the first subprofile from the left of Figure 3A) and absent from the calibration samples, and components 1 and 2 are

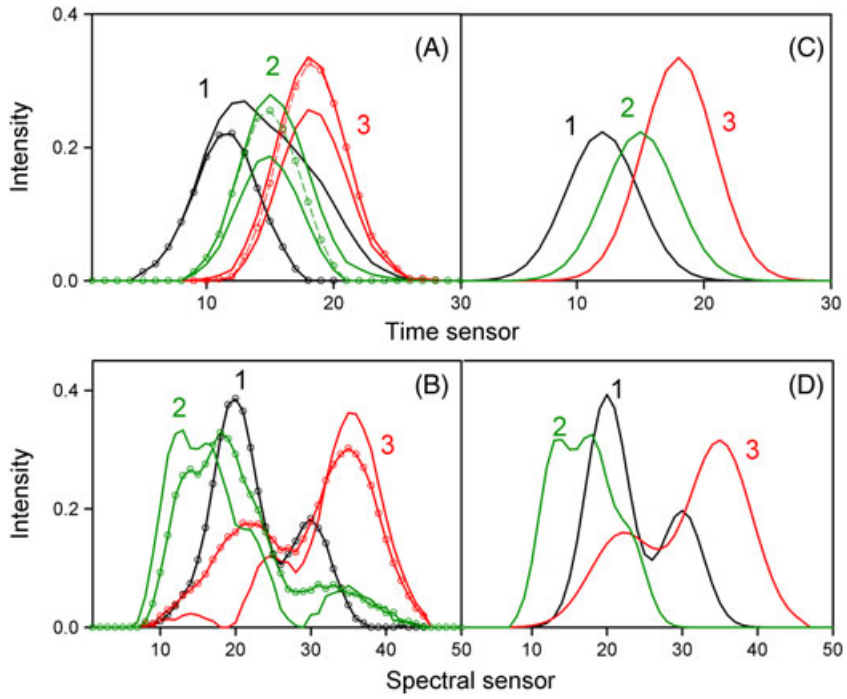

FIGURE 2 A, Profiles retrieved by multivariate curve resolution–alternating least squares study of a single simulated sample containing 3 components in the elution time mode (circles joined by dashed lines). B, Profiles in the spectral mode. In both (A) and (B), the solid lines are the feasible solution bands estimated by MCR-BANDS. Component numbers are indicated as in Figure 1. Simulated profiles from Figure 1 are repeated in plots (C) and (D) for comparison

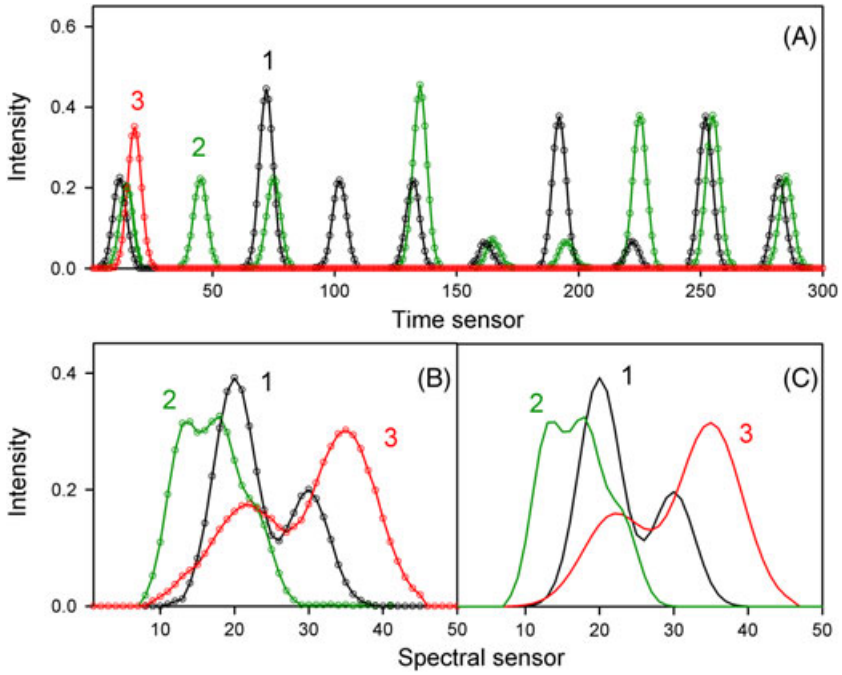

FIGURE 3 A, Profiles retrieved by multivariate curve resolution–alternating least squares study of a simulated augmented data matrix containing 3 components, built with 1 test sample and 9 calibration samples, in the augmented elution time mode (circles joined by dashed lines). B, Profiles in the nonaugmented spectral mode. The solid lines are the feasible solution bands estimated by MCR-BANDS. Component numbers are indicated as in Figure 1. Simulated spectral profiles from Figure 1 are repeated in plot (C) for comparison with (B)

absent in calibration samples 1 and 3, respectively (second and fourth subprofiles from the left of Figure 3A), because they had only one of the 2 analytes.

Results of the MCR-ALS decomposition and of the application of MCR-BANDS are shown in Figure 3A,B, including, for comparison purposes, the known pure spectra in Figure 3C. The elution profiles shown in Figure 3A are presented in greater detail in Figure 4A for the test sample (having 2 analytes and 1 interferent), and in Figure 4B for the calibration sample 2, which only has the 2 analytes. Figure 4C,D shows the corresponding time profiles expected for the latter 2 samples, built from the known pure temporal profiles for the 3 components and their specific concentrations. It is clear that as a consequence of the matrix augmentation strategy (simultaneous analysis of multiple data matrices having complementary information) and of the application of further constraints, which were not possible in a single matrix (such as the correspondence among components and samples), the solution obtained in this case is much closer to the true one, and therefore, rotational ambiguity was drastically reduced. In Figure 3B, the spectra of the 3 components (including the interferent) at the extreme values of the MCR-BANDS optimization function were practically the same, showing that there is nearly no ambiguity in their recovery. In Figure 4A, the elution profile of analyte 2 has still some small degree of ambiguity, because the 2 profiles corresponding to the maximum and minimum of the MCR-BANDS function did not coincide totally. The elution profile for analyte 2 in sample 1 is the most difficult to recover without ambiguity, because it is embedded in the other 2 elution profiles (from analyte 1 and the interferent).

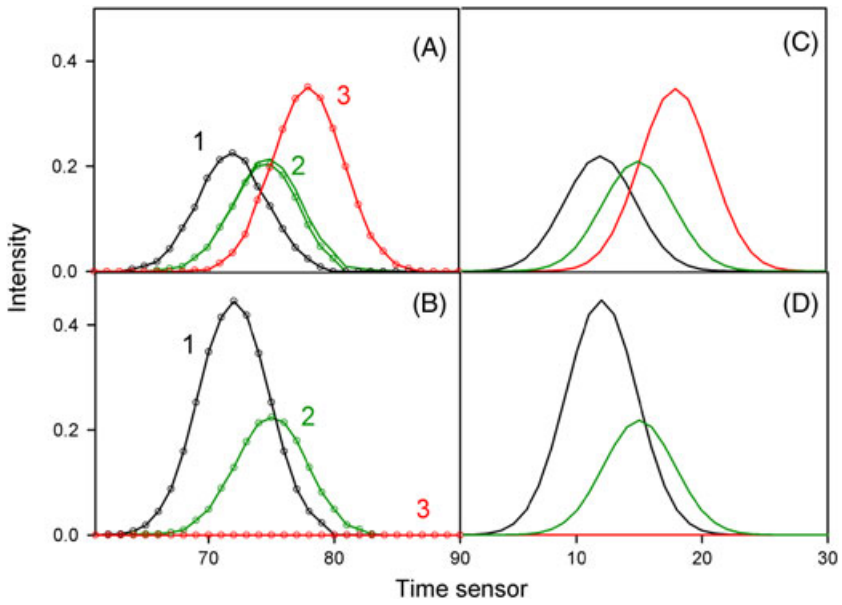

FIGURE 4 Selected subprofiles from the augmented profiles in Figure 3A. A, Test sample profiles. B, Calibration sample 2. Component numbers are indicated as in Figure 1. Simulated spectral profiles from Figure 1 are repeated in plots (C) and (D) for comparison with (A) and (C), respectively. In plots (C) and (D) the simulated profiles were scaled to the appropriate vertical intensity according to the component concentrations

These results are confirmed in the MCR-BANDS results shown in Table 1 for the analysis of the augmented data matrix. The MCR-BANDS values of $(f_{max} - f_{min})$ drastically decreased when compared with those obtained for the individual analysis of the first single data matrix.

6.3 | Experimental results: single data matrix

The experimental system consists of 3 chemical components, and the data matrices are of chromatographic-spectral type, with detection proceeding by measuring fluorescence emission spectra at a fixed excitation wavelength. Thus, conceptually, it is analogous to the simulated system described above for UV diode array detection. However, in this particular case, the experimentally recorded data had an additional signal arising from an almost constant background signal, which was modeled along with those from the 3 chemical constituents (ie, a total of 4 components were present during MCR-ALS resolution).

To perform a similar comparison to that discussed above for the simulated cases, we analyzed a single test sample data matrix by applying MCR-ALS, using the same initialization method and constraints as those used during simulations. The results of retrieved component profiles and feasible solution bands are shown in Figure 5 for the analysis of a single data matrix. Only the results for the chemical components are shown, implying a significant degree of rotational ambiguity, especially in the elution time profiles, which define the relative component concentrations, as was the case for the analogous simulated single data matrix.

Figure 5 shows the MCR-ALS resolved elution and spectral profiles of the 3 fluorescent components, together with those furnished by MCR-BANDS according to f_{min} and f_{max} results. Ambiguity is present in the elution profiles of the 3 components, and especially for the spectra of 2 of them (bkf and bbf), and much lower for the spectrum of the third one (bjf), again because its elution profile had higher selectivity at the beginning of the chromatographic elution. In this particular case, there was no need of explicitly using a selectivity constraint,[1,2,18] in this chromatographic region, because only using the nonnegativity constraint already favored the presence of unique resolution conditions, as has been also shown in other situations in previous works.[24,25] In Table 1, these results are again confirmed: MCR-BANDS values of

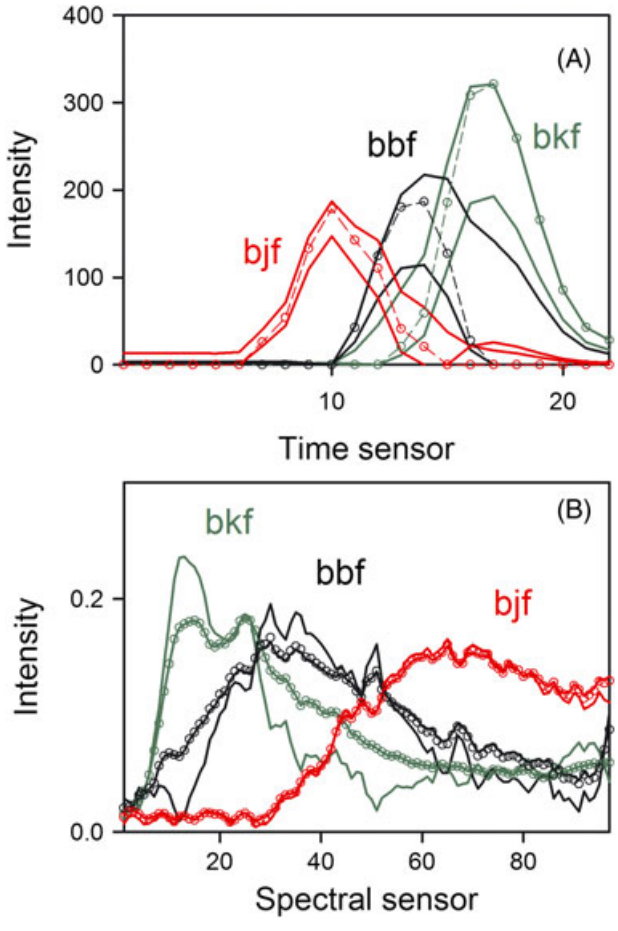

FIGURE 5 Circles joined by dashed lines, profiles retrieved by multivariate curve resolution–alternating least squares study of a single experimental sample containing 3 components: A, elution time mode; B, spectral mode. Solid lines, feasible solution bands estimated by MCR-BANDS. Component labels are indicated

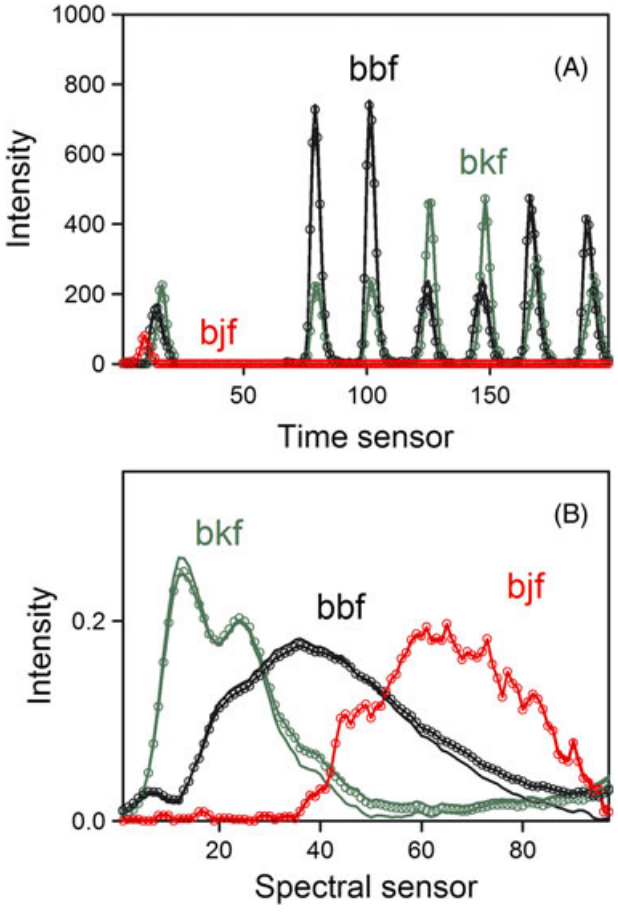

FIGURE 6 Circles joined by dashed lines, profiles retrieved by multivariate curve resolution–alternating least squares study of an experimental augmented data matrix containing 3 components, built with 1 test sample and 8 calibration samples: A, augmented elution time mode; B, nonaugmented spectral mode. Solid lines, feasible solution bands estimated by MCR-BANDS. Component labels are indicated as in Figure 5

($f_{max} - f_{min}$) for the 3 components are rather high, as a consequence of the remaining ambiguity, especially in the elution profiles.

6.4 | Experimental results: augmented data matrix

In this case, the augmented data matrix includes the experimental chromatographic determinations of the test sample (bbf and bkf as analytes and bjf as interferent) and of 8 calibration samples (2 blank samples and 6 samples having bbf and bkf as analyte standards). Figures 6 and 7 show respectively the MCR-ALS and MCR-BANDS results obtained for the whole set of elution and spectra profiles obtained in the simultaneous analysis. Specifically, Figure 7 includes the elution profiles for the test and for one of the calibration samples, which were included in the analysis of the augmented data matrix. Recovery of the elution and spectra profiles improved significantly, compared with those given previously for Figure 5. The MCR-BANDS results obtained and shown in Figures 6 and 7 corroborate the fulfillment of our previous results,[1,2] and of Manne resolution theorems.[6] Because of the experimental design used to perform the different chromatographic experiments, which included the presence and absence of interferent species, rotation ambiguities were drastically reduced.

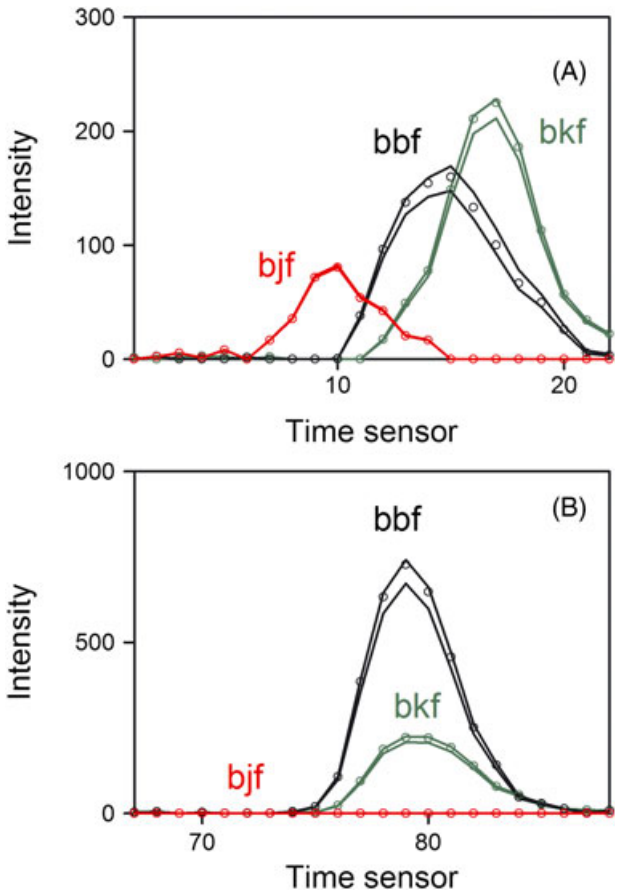

FIGURE 7 Selected subprofiles from the augmented profiles in Figure 6A. A, Test sample profiles. B, Calibration sample 3. Component labels are indicated as in Figure 5. Multivariate curve resolution–alternating least squares solutions are represented by circles joined by solid lines

In particular, the elution profiles in Figure 7 are shown in detail for the test sample (with bbf, bkf, and bjf elution profiles) and for one of the calibration samples (only with bbf and bkf, but not bjf as interferent). In Table 1, MCR-BANDS ($f_{max} - f_{min}$) values did drastically decrease compared with those obtained for the previous single data matrix analysis. This again confirms the improvement of the MCR-ALS results when the simultaneous analysis of multiple data sets is compared with MCR-ALS results of the single data matrix analysis, because of the reduction of rotational ambiguities in the latter case.

7 | CONCLUSIONS

Application of the MCR-BANDS method to MCR profiles (concentration and spectra) obtained by simultaneous analysis of multiple data sets having common constituents showed that rotational ambiguities are drastically reduced compared with those obtained by individual analysis of each data set separately. Conditions for better resolution results with improved profiles and reduced ambiguities are more easily achieved when the matrix augmentation strategy of joint MCR analysis of multiple data sets is applied. This demonstration opens the possibility to improve MCR solutions using a proper design of the experimental conditions that give better local rank and selectivity resolution constraints.

ACKNOWLEDGEMENTS

Romà Tauler acknowledges research funding from MINECO Spain grant number CTQ2015-66254-C2-1-P. Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) project number PIP 0163, and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica) project number PICT-2013-0136 are gratefully acknowledged for the financial support.

REFERENCES

1. Tauler R. Multivariate curve resolution applied to second order data. *Chemom Intel Lab Syst.* 1995;30(1):133–146.
2. Tauler R, Smilde A, Kowalski B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J Chemometr.* 1995;9:31–58.
3. Arancibia JA, Damiani PC, Escandar GM, Ibañez GA, Olivieri AC. A review on second- and third-order multivariate calibration applied to chromatographic data. *J Chromatogr B.* 2012;910:22–30.
4. Xu L, Tang LJ, Cai CB, et al. Chemometric methods for evaluation of chromatographic separation quality from two-way data—a review. *Anal Chim Acta.* 2008;613(2):121–134.
5. Tauler R, de Juan A. Multivariate curve resolution for quantitative analysis. In: Muñoz de la Peña A, Goicoechea HC, Escandar GM, Olivieri AC, eds. *Data Handling in Science and Technology, Vol. 29, Fundamentals and Analytical Applications of Multiway Calibration*. Amsterdam: Elsevier; 2015:247–292.
6. Manne R. On the resolution problem in hyphenated chromatography. *Chemom Intel Lab Syst.* 1995;27(1):89–94.

7. Vosough M, Mason C, Tauler R, Jalali-Heravi M, Maeder M. On rotational ambiguity in model-free analyses of multivariate data. *J Chemometr.* 2006;20(6-7):302–310.
8. Alinaghi M, Rajkó R, Abdollahi H. A systematic study on the effects of multi-set data analysis on the range of feasible solutions. *Chemom Intel Lab Syst.* 2016;153:22–32.
9. Golshan A, Abdollahi H, Beyramysoltan S, et al. A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data. *Anal Chim Acta.* 2016;911:1–13.
10. Tauler R, Maeder M, de Juan A. Multiset data analysis: extended multivariate curve resolution. In: Brown S, Tauler R, Walczak B, eds. *Comprehensive Chemometrics*. Vol. 2. Oxford: Elsevier; 2009:473–505.
11. Jaumot J, Tauler R. MCR-BANDS: a user friendly MATLAB program for the evaluation of rotation ambiguities in multivariate curve resolution. *Chemom Intel Lab Syst.* 2010;103(2):96–107.
12. Jaumot J, Gargallo R, de Juan A, Tauler R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom Intel Lab Syst.* 2005;76(1):101–110.
13. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: new features and applications. *Chemom Intel Lab Syst.* 2015;140:1–12.
14. Bortolato SA, Arancibia JA, Escandar GM, Olivieri AC. Time-alignment of bidimensional chromatograms in the presence of uncalibrated interferences using parallel factor analysis. Application to multi-component determinations using liquid-chromatography with spectrofluorimetric detection. *Chemom Intel Lab Syst.* 2010;101(1):30–37.
15. de Juan A, Rutan SC, Tauler R. Two-way data analysis: multivariate curve resolution—iterative resolution methods. In *Comprehensive Chemometrics*, S Brown, R Tauler, B Walczak (Eds.), Elsevier: Oxford, Vol. 2, 325-344.
16. de Juan A, Tauler R. Comparison of three-way resolution methods for non-trilinear chemical data sets. *J Chemometr.* 2001;15(10):749–771.
17. Malik A, Tauler R. Extension and application of multivariate curve resolution-alternating least squares to four-way quadrilinear data-obtained in the investigation of pollution patterns on Yamuna River, India—a case study. *Anal Chim Acta.* 2013;794:20–28.
18. Tauler R. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J Chemometr.* 2001;15(8):627–646.
19. de Juan A, Maeder M, Martínez M, Tauler R. Combining hard- and soft-modelling to solve kinetic problems. *Chemom Intel Lab Syst.* 2000;54:123-141
20. Diewok J, de Juan A, Maeder M, Tauler R, Lendl B. Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of pH-modulated mixture samples. *Anal Chem.* 2003;75(3):641–647.
21. Gill PE, Murray W, Wright MH. *Practical Optimization*. London: Academic Press; 1981.
22. *MATLAB Version R2012a*. Natick, Massachusstes, USA: The Mathworks; 2012.
23. Windig W, Guilment J. Interactive self-modeling mixture analysis. *Anal Chem.* 1991;63(14):1425–1432.
24. Paatero P, Hopke PK, Song X-H, Ramadan Z. Understanding and controlling rotations in factor analytic models. *Chemom Intel Lab Syst.* 2002; 60(1-2):253–264.
25. Paatero P, Hopke PK, Begum BAB, Biswas SK. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. *Atmos Environ.* 2005;39(1):193–201.

How to cite this article: Olivieri AC, Tauler R. The effect of data matrix augmentation and constraints in extended multivariate curve resolution–alternating least squares. *Journal of Chemometrics*. 2017;e2875. https://doi.org/10.1002/cem.2875