Analytica Chimica Acta 1003 (2018) 10–15

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

The effect of constraints on the analytical figures of merit achieved by extended multivariate curve resolution-alternating least-squares

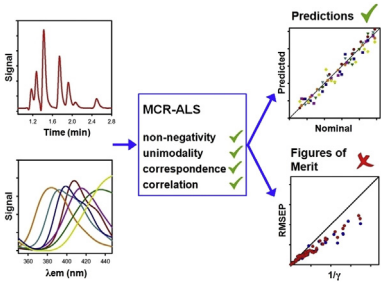
Rocío B. Pellegrino Vidal, Franco Allegrini, Alejandro C. Olivieri*

Instituto de Química Rosario (CONICET-UNR), Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, 2000, Rosario, Argentina

HIGHLIGHTS

- Multivariate curve resolution figures of merit are analyzed.
- Simulations and experimental data have been studied.
- Prediction errors do not follow the trend expected from analytical sensitivities.
- The effect is due to lack of consideration of constraints in the figures of merit.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 10 August 2017
Received in revised form
30 October 2017
Accepted 11 December 2017
Available online 28 December 2017

Keywords:
Multivariate curve resolution
Analytical figures of merit
Constraints
Sensitivity

ABSTRACT

Multivariate curve resolution-alternating least-squares (MCR-ALS) is the model of choice when dealing with some non-trilinear arrays, specifically when the data are of chromatographic origin. To drive the iterative procedure to chemically interpretable solutions, the use of constraints becomes essential. In this work, both simulated and experimental data have been analyzed by MCR-ALS, applying chemically reasonable constraints, and investigating the relationship between selectivity, analytical sensitivity (γ) and root mean square error of prediction (RMSEP). As the selectivity in the instrumental modes decreases, the estimated values for γ did not fully represent the predictive model capabilities, judged from the obtained RMSEP values. Since the available sensitivity expressions have been developed by error propagation theory in unconstrained systems, there is a need of developing new expressions or analytical indicators. They should not only consider the specific profiles retrieved by MCR-ALS, but also the constraints under which the latter ones have been obtained.

1. Introduction

Multivariate curve resolution-alternating least-squares (MCR-ALS) is a versatile tool to extract meaningful information from bilinear data, i.e., data that can be described in terms of a small number of pure bilinear contributions [1]. Many analytical instruments or combination of instruments generate bilinear data, with MCR-ALS being the model of choice to deal with second-order calibration. This is especially so when extended MCR-ALS is applied to adequately augmented data matrices, allowing to achieve the second-order advantage [2]. In particular, the resolution of chromatographic data with multivariate detection into pure constituent profiles is possible, even in the presence of changes in peak positions and/or shapes from sample to sample [3].

One of the main limitations of MCR-ALS in reaching unique

* Corresponding author.
E-mail address: olivieri@iquir-conicet.gov.ar (A.C. Olivieri).

https://doi.org/10.1016/j.aca.2017.12.008
0003-2670/

solutions is the phenomenon of rotational ambiguity, due to the existence of different combinations of concentration and response profiles satisfying the bilinear model [4,5]. The application of constraints has the object of minimizing the degree of rotational ambiguity, reducing the number of possible solutions, and decreasing the uncertainty in predicted concentrations [6]. The most widely used constraints are: (1) non-negativity, because chemical concentrations of mixture constituents and their responses in many instrumental methods should be nonnegative, (2) unimodality, implying that a single peak is observed in chromatography, (3) closure, which is related to chemical mass balance equations, (4) correspondence between constituents and samples, (5) concentration correlation, which builds a linear regression between the area of resolved profiles and reference values during the alternating least-squares optimization [7], and (6) selectivity, if it is known that a constituent does not respond in a certain spectral region [1].

Recently, the reduction of rotational ambiguity resulting from matrix augmentation has been rationalized based on the data matrix augmentation strategy and the application of suitable constraints [8]. This helps to get a better insight into MCR-ALS resolving power under different conditions. However, information about the quality of future analytical predictions is also required, as would be measured by the root mean square error of prediction (RMSEP) for a group of validation samples. The latter parameter should in principle be inversely related to the sensitivity, a figure of merit that may anticipate the prediction quality of a model. However, the current expression for the MCR-ALS sensitivity does not consider the influence of the applied constraints, since it was derived by error propagation theory, estimating the prediction error brought about by a small perturbing signal noise [9,10]. During this noise propagation, no constraints are involved.

In this work, simulations have been carried out with the objective of studying the effect of constraints in the relationship between RMSEP values and figures of merit, for systems generated under different conditions of profile overlapping and noise level. The same analysis was extended to an experimental second-order data set, aimed at the determination of polycyclic aromatic hydrocarbons (PAHs) *via* high performance liquid chromatography (HPLC) with multivariate fluorescence detection (FLD). It is shown that sensitivities estimated for both simulated and experimental data do not present the expected correlation with RMSEPs, and are therefore unrepresentative of the predictive capabilities of the model.

2. Simulations

Second-order data sets were simulated for a system containing two constituents. Simulations were performed to mimic chromatographic experiments with spectral detection. Profiles in the elution time mode were kept at a fixed overlapping level, while profiles in the spectral mode were digitally moved to obtain eleven different degrees of selectivity. The matrix of signals **X** for a typical sample was generated by the following expression:

$$\mathbf{X} = y_1\mathbf{c}_1\mathbf{s}_1^{\mathrm{T}} + y_2\mathbf{c}_2\mathbf{s}_2^{\mathrm{T}} + Noise \quad (1)$$

where y_1 and y_2 are the concentrations of analytes 1 and 2, respectively, $\mathbf{c}_n$ and $\mathbf{s}_n$ ($n = 1, 2$) are the ($J \times 1$) and ($K \times 1$) constituent profiles in the temporal and spectral modes (J and K are the number of channels in each mode, in this case $J = 100$ and $K = 110$), and the superscript 'T' indicates matrix transposition. Seven noise levels were applied, in the range 0.1%–3% with respect to the maximum calibration signal for each analyte at unit concentration. This led to 77 different data sets, each with a specific combination of noise level and spectral resolution. Fig. 1 shows the noiseless temporal profiles used to build all data sets, as well as the noiseless spectra for both the highest and lowest level of spectral overlapping.

All data sets comprised 21 calibration samples, with concentrations of both compounds in the range 0–1 following a central composite design (duplicates of the factorial and star samples and five center samples), and 50 validation samples containing random concentrations of both compounds in the range 0.4–0.6. Analytical calibration using MCR-ALS is usually accomplished through matrix augmentation. In the present simulations, augmentation was performed in the temporal direction (column-wise), by appending each validation data matrix with all 21 calibration data matrices. Initial spectral profiles employed to start the MCR-ALS fitting were obtained from the so-called purest variables in the spectral domain [11]. The following constraints were imposed during the ALS fit: non-negativity in both spectral and temporal modes, unimodality in the temporal mode, and correspondence between constituents and samples. After convergence of the ALS optimization (the

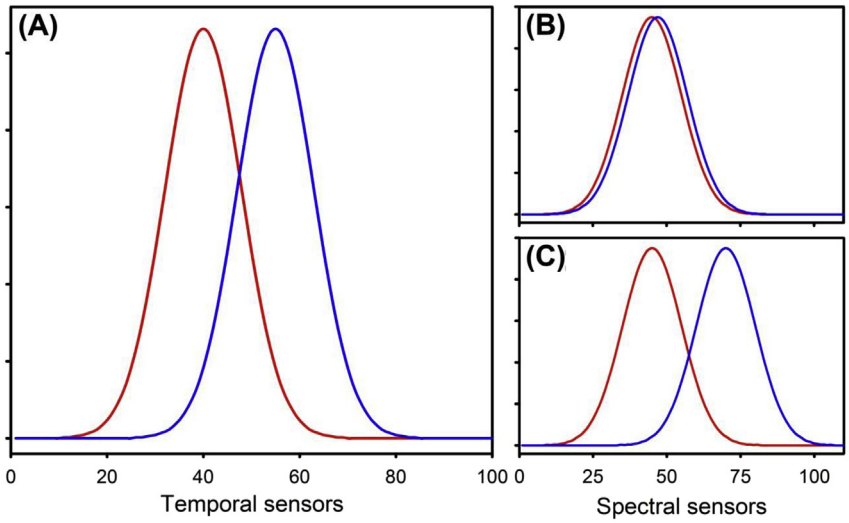


Fig. 1. Noiseless simulated profiles for compound 1 (blue) and compound 2 (red). (A) Temporal mode profiles, common to all data sets.(B) Spectral mode profiles for the highest spectral overlapping. (C) Spectral profiles for the lowest spectral overlapping. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

tolerance for convergence was set at a 0.01% relative change in fit for successive iterations), analytes were identified by their spectral profiles and their quantitation was performed through the corresponding pseudo-univariate calibration curves. Figures of merit were estimated according to ref. 10.

3. Experimental data

The experimental data set studied in this work has already been reported, and involves seven PAHs: fluoranthene (FLT), pyrene (PYR), benzo[*a*]anthracene (BaA), benzo[*b*]fluoranthene (BbF), benzo[*a*]pyrene (BaP), dibenz[*a,h*]anthracene (DBA), and benzo[*g,h,i*]perilene (BghiP) [12]. Eighteen calibration samples were prepared in the range 0–50 ng mL^{-1}, sixteen with concentrations provided by a fractional factorial design, a blank, and a sample containing all analytes at the average concentration. A validation set of ten samples was also prepared with random concentrations in the corresponding calibration ranges. All samples were prepared in acetonitrile/water (85:15 v/v) and analyzed by HPLC on an Agilent 1200 liquid chromatograph (Agilent Technologies, Waldbronn, Germany), equipped with a quaternary pump, a thermostated column compartment set at 35 °C, and a fluorescence detector irradiating at 222 nm and collecting emission spectra from 295 to 450 nm, each 2 nm. A Rheodyne valve with a 20.0 μL loop was employed to inject the sample on to a Poroshell 120 EC C18 column (4.6 mm × 50 mm, 2.7 μm particle size). The mobile phase was a mixture of acetonitrile/water (85:15 v/v) and flow was set at 1.25 mL min^{-1}, leading to a run time of 3.2 min. The data were collected using the software HP ChemStation for LC Rev HP 1990–1997.

Elution time-fluorescence emission wavelength matrices were subjected to asymmetric least-squares baseline correction [13] before data processing by MCR-ALS. The number of components was estimated by principal component analysis (PCA). MCR-ALS was initialized by estimating the so-called purest variables in the spectral domain. The modeling stage was carried under the same constraints that were applied in the simulated systems, including in this case the concentration correlation constraint. Finally, analytes were identified by their spectral profiles and their quantitation was performed through the corresponding pseudo-univariate calibration curves. Figures of merit were estimated according to ref. 10.

4. Theory

4.1. MCR-ALS

A detailed discussion of the MCR-ALS algorithm can be found in the literature [1]. Owing to the lack of reproducibility in elution time profiles, second-order data of chromatographic origin are best processed by applying this model to the augmented matrix **D** generated by appending each validation data matrix with the calibration data matrices. Augmentation is performed in the temporal direction to preserve the bilinearity, while decomposition of **D** follows the bilinear expression:

$$D = C S^T + E \quad (2)$$

where the columns of **D** contain the chromatograms acquired at J times for $(I_{cal} + 1)$ different samples at K wavelengths, the columns of **C** contain the augmented elution time profiles of the intervening species, the columns of **S** their corresponding spectra, and **E** is a matrix of residuals not fitted by the model. The sizes of these matrices are **D**, $J(I_{cal}+1) \times K$, **C**, $J(I_{cal}+1) \times N$, **S**, $K \times N$, **E**, $J(I_{cal}+1) \times K$ (N is the number of responsive constituents). The augmented **D** matrix contains data for the calibration samples (I_{cal}) and for a given validation sample.

MCR-ALS requires estimations of either the spectral or temporal profiles to initiate the ALS optimization algorithm. While different options are available, such as supplying spectra obtained from pure analyte standards, or providing estimated elution time profiles, as obtained from procedures such as evolving factor analysis (EFA) [14], in this work MCR-ALS was initialized by estimating the so-called purest variables in the spectral domain.

Decomposition of **D** is achieved by constrained iterative least-squares minimization of the residuals contained in **E** [1], as already explained. Once **D** is decomposed, analyte scores are defined as the

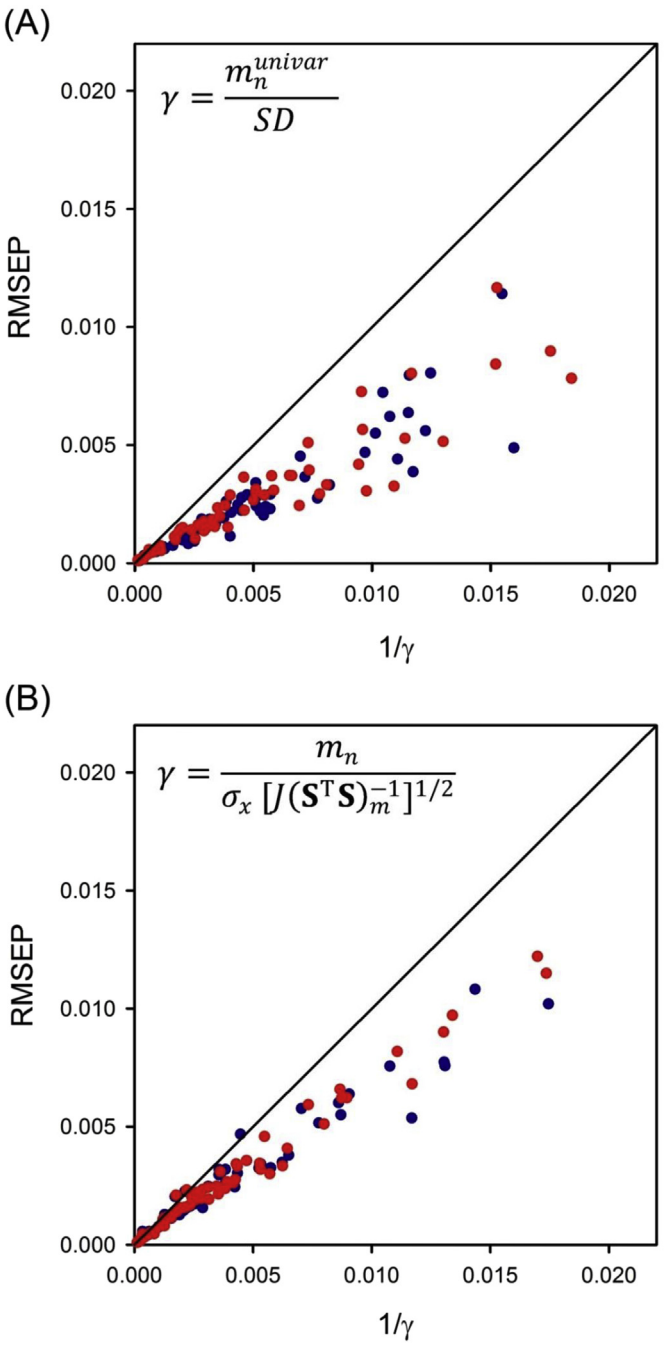


Fig. 2. Root mean square error of prediction (RMSEP) as a function of the inverse of γ for all the simulated data sets modeled by MCR-ALS. Compound 1, blue, compound 2, red. (A) Values of γ calculated following the pseudo-univariate approach, see eq. (4). (B) Values of γ estimated with eq. (5). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

area under the time elution profile for the *i*th sample, as:

$$a(i,n) = \sum_{j-1+(i-1)J}^{ij} c(j,n) \quad (3)$$

where $a(i,n)$ is the score for analyte n in sample i, and $c(j,n)$ is the element of the analyte profile in the augmented mode. Finally, the analyte scores in the calibration samples are employed to build a pseudo-univariate calibration curve against the nominal analyte concentrations, and the concentration in the validation sample is predicted by interpolation of the validation sample score.

4.2. Software

The routines employed were written in MATLAB 7.0. (Mathworks, MA, USA) [15]. MCR-ALS was implemented both by an in-house MATLAB routine and using the graphical interface of the MVC2 toolbox [16], freely available on the Internet [17].

5. Results and discussion

5.1. Simulations

As was previously discussed, all simulated data sets were processed by MCR-ALS, applying non-negativity in both spectral and temporal modes, unimodality in the temporal mode, and correspondence between components and samples. Afterwards, with the aim of evaluating the effect of the applied constraints, the RMSEPs and analytical sensitivities (γ) were calculated for each data set. RMSEP is widely regarded as a measure of a model predictive power. On the other hand, the analytical sensitivity can be expressed as the inverse of the uncertainty in the predicted concentration for a given model. The latter was first estimated from the corresponding pseudo-univariate calibration curves, according to the classical univariate expression [18]:

$$\gamma_{univar} = \frac{m_n^{univar}}{SD} \quad (4)$$

where m_n^{univar} and SD are, respectively, the slope and the standard deviation of the residuals obtained for the pseudo-univariate curve corresponding to analyte n.

This figure of merit was also estimated using eq. (5), as proposed in Ref. [9] by error propagation analysis:

$$\gamma_{MCR} = \frac{m_n^{univar}}{\sigma_x \left[J (S^T S)^{-1}_{nn} \right]^{1/2}} \quad (5)$$

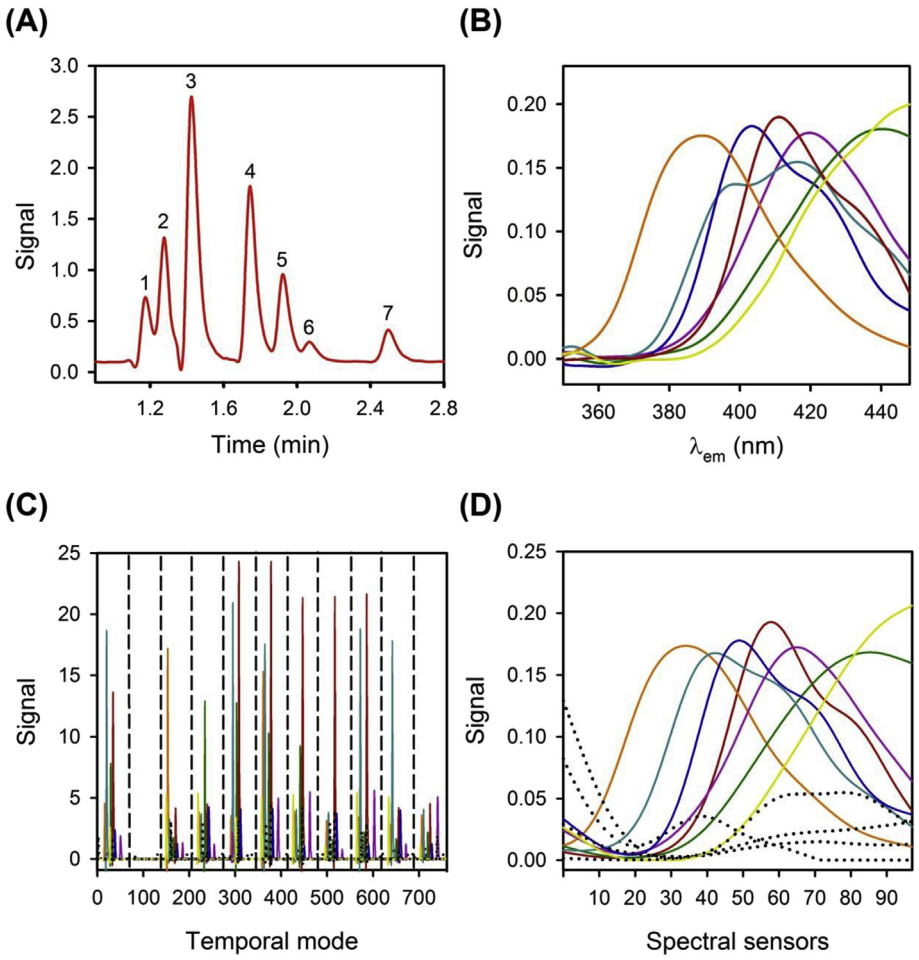


Fig. 3. (A) Chromatogram of a selected calibration sample (fluorescence detection at λexc = 220 nm and λem = 420 nm). Peak numbers refer to (1) FLT, (2) PYR, (3) BaA, (4) BbF, (5) BaP, (6) DBA, and (7) BghiP. (B) Normalized fluorescence emission spectra for the assayed analytes. (C) Augmented temporal mode as retrieved by MCR-ALS. (D) Spectral profiles retrieved by MCR-ALS. BaP in burgundy, PYR in orange, BbF in green, BaA in dark cyan, BghiP in purple, DBA in dark blue, FLT in yellow, and blank (black dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

where all symbols have already been defined, except σ_x which is the experimental signal noise. In eq. (5), the subscript *nn* indicates the (*n*,*n*) element of the matrix product ($\mathbf{S}^T\mathbf{S}$).

Fig. 2 shows the changes in RMSEP as a function of the inverse analytical sensitivity for the simulated data sets described in Section 2., calculated according to eq. (4) (Fig. 2A) and to eq. (5) (Fig. 2B). In both cases there is a significant deviation from the ideal linear relationship at high values of RMSEP, i.e. the estimated analytical sensitivities are smaller than expected. This deviation is apparent when analyzing the systems with high noise and high spectral overlapping. We propose that the deviation in the plot of RMSEP vs. $(1/\gamma)$ is a consequence of the analytical sensitivity expressions not taking into account the applied constraints.

It has already been suggested that the pseudo-univariate approach to MCR-ALS sensitivity does not reflect the degree of overlapping among spectral component profiles [9]. On the other hand, eq. (5) does consider the overlapping, and thus the plot of Fig. 2B is less disperse than the one in Fig. 2A. However, the present results suggest that in cases where the application of constraints is essential to successfully retrieve the profiles, analytical sensitivities are still underestimated. This shortcoming in the calculation of γ_{MCR} by eq. (5) can be attributed to the fact that it fails to acknowledge the way in which data were processed to obtain the constituent profiles.

An attempt has also been made to correlate RMSEP values vs. inverse analytical sensitivities estimated for three-way parallel factor analysis (PARAFAC) [19] and also with an average of the MCR-ALS and the PARAFAC sensitivity. However, this approach proved to be unsatisfactory, mainly due to the lack of correlation between the MCR-ALS and PARAFAC sensitivities, as already reported [10].

5.2. Experimental data

The presently discussed experimental system involves the analysis of seven PAHs by HPLC coupled to fluorescent detection and second-order calibration. Due to unavoidable shifts in retention times between successive chromatographic runs, the data matrices were modeled by MCR-ALS. Fig. 3 shows a typical chromatogram for a calibration sample, as well as the analyte fluorescence emission spectra. Even in the presence of a high degree of overlapping among spectral profiles, applying the previously mentioned constraints allowed the algorithm to successfully retrieve all profiles (Fig. 3). Hence, the importance of applying adequate constraints when modeling complex systems is demonstrated, in agreement with previous reports [20,21].

Fig. 4 displays the plot of predicted *vs.* nominal concentrations for the validation samples, where good correlation is observed for all analytes. The elliptical joint confidence region (EJCR) statistical test was performed to verify the accuracy of the predictions [22]. Fig. 4 shows that the theoretically expected point (slope = 1, intercept = 0) lies inside the elliptical region, indicating the accuracy of the proposed model. The quality of the predictions is also verified by the low RMSEP values (Table 1), which imply relative errors of prediction (REPs) in the range 4–22%. Analytical sensitivities, selectivities (SEN) and limits of detection (LOD) and quantification (LOQ) are shown in Table 1. We note that the values for the analytical sensitivity γ, calculated according to eq. (5), were significantly lower than the expectations based on the achieved RMSEP values. Furthermore, while concentrations as low as 10 ng mL^{-1} were successfully predicted by the proposed model, the estimated LODs and LOQs were unreasonably large, in the range 8–508 ng mL^{-1} and 24–1523 ng mL^{-1} respectively. These results suggest that the currently accepted expressions for the calculation of figures of merit, particularly the sensitivity and analytical sensitivity, do not adequately anticipate the predictive capabilities of the proposed model.

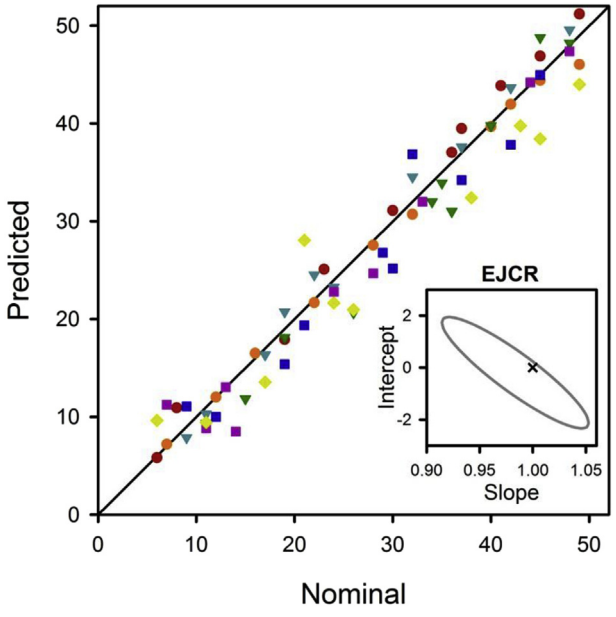


Fig. 4. Plots of BaP (burgundy circles), PYR (orange circles), BbF (green triangles), BaA (dark cyan triangles), BghiP (purple squares), DBA (dark blue squares), and FLT (yellow diamonds) predicted concentrations as a function of the nominal values in validation samples. Inset: elliptical joint region (at 95% confidence level) for the slope and intercept of the predicted *vs.* nominal concentrations plot. The cross marks the theoretical (intercept = 0, slope = 1) point. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

6. Conclusions

The study of both simulated and experimental data demonstrated that in challenging analytical systems, where the application of MCR-ALS constraints is vital to retrieve adequate solutions, the accepted expression for analytical sensitivity does not fully represent the predictive capability of the model. While this expression considers the overlapping among constituents' profiles in the non-augmented mode, it does not reflect the way in which these profiles are obtained, e.g. the constraints that allowed MCR-ALS to arrive to such results. As a consequence, additional figures of merit derived from the sensitivity, such as limit of detection and limit of quantification, can also be highly overestimated. This precludes their use in assessing the predictive quality of the multiway model, as well as the comparison with other methodologies. Thus, there is an apparent need to devise new sensitivity expressions, which should include the effect of the constraints applied during MCR-ALS resolution.

Table 1
Figures of merit for the studied analytes by HPLC-FLD in validation samples.[a]

	BaA	PYR	DBA	BaP	FLT	BbF	BghiP
Calibration range	0–50	0–50	0–50	0–50	0–50	0–50	0–50
RMSEP	1.5	1.1	4.2	2.0	5.6	2.9	2.6
γ_{MCR}	0.2	0.5	0.04	0.3	0.007	0.01	0.03
SEL	0.1	0.2	0.06	0.1	0.02	0.03	0.05
LOD	16	8	101	13	508	290	107
LOQ	48	24	306	40	1523	879	324

[a] Calibration range and RMSEP values in ng mL^{-1}, γ_{MCR}, analytical sensitivity calculated according to ref. [9] (ng^{-1} mL), SEL, selectivity, LOD and LOQ, limits of detection and quantitation calculated according to ref. [10] (ng mL^{-1}).

Acknowledgements

The authors wish to thank Universidad Nacional de Rosario (Project 19/B487), CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP 0163) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2016-1122) for financial support. RPV and FA thank CONICET for their fellowships.

References

[1] R. Tauler, M. Maeder, A. de Juan, Multiset data analysis: extended multivariate curve resolution, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Elsevier, Oxford, UK, 2009, pp. 473–505.
[2] A.C. Olivieri, Analytical advantages of multivariate data processing. One, two, three, infinity? Anal. Chem. 80 (2008) 5713–5720.
[3] A.C. Olivieri, G.M. Escandar, Practical Three-way Calibration, Elsevier, Waltham, US, 2014.
[4] R. Tauler, Multivariate curve resolution applied to second order data, Chemometr. Intell. Lab. Syst. 30 (1995) 133–146.
[5] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, 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 (2016) 1–13.
[6] G. Ahmadi, H. Abdollahi, A systematic study on the accuracy of chemical quantitative analysis using soft modeling methods, Chemometr. Intell. Lab. Syst. 120 (2013) 59–70.
[7] A.C. de, O. Neves, R. Tauler, K.M.G. de Lima, Area correlation constraint for the MCR–ALS quantification of cholesterol using EEM fluorescence data: a new approach, Anal. Chim. Acta 937 (2016) 21–28.
[8] A.C. Olivieri, R. Tauler, The effect of data matrix augmentation and constraints in extended multivariate curve resolution-alternating least squares, J. Chemometr.. DOI: 10.1002/cem.2875.
[9] M.C. Bauza, G.A. Ibañez, R. Tauler, A.C. Olivieri, Sensitivity equation for quantitative analysis with multivariate curve resolution-alternating least-squares: theoretical and experimental approach, Anal. Chem. 84 (2012) 8697–8706.
[10] A.C. Olivieri, Analytical figures of merit: from univariate to multiway calibration, Chem. Rev. 114 (2014) 5358–5378.
[11] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425–1432.
[12] R.B. Pellegrino Vidal, G.A. Ibañez, G.M. Escandar, Advantages of data fusion: first Multivariate Curve Resolution analysis of fused liquid chromatographic second-order data with dual diode array-fluorescent detection, Anal. Chem. 89 (2017) 3029–3035.
[13] P.H.C. Eilers, I.D. Currie, M. Durbán, Fast and compact smoothing on large multidimensional grids, Comput. Stat. Data Anal. 50 (2006) 61–76.
[14] M. Maeder, A. Zilian, Evolving factor analysis, a new multivariate technique in chromatography, Chemometr. Intell. Lab. Syst. 3 (1988) 205–213.
[15] MATLAB, TheMathworks, (Natick, Massachusetts, USA).
[16] A.C. Olivieri, H.L. Wu, R.Q. Yu, MVC2: a MATLAB graphical interface toolbox for second-order multivariate calibration, Chemometr. Intell. Lab. Syst. 96 (2009) 246–251.
[17] http://www.iquir-conicet.gov.ar/descargas/mvc2.rar (Accessed August 2017).
[18] R.K. Skogerboe, C.L. Grant, Comments on the definitions of the terms sensitivity and detection limit, Spectrosc. Lett. 3 (1970) 215–220.
[19] A.C. Olivieri, K. Faber, New developments for the sensitivity estimation in four-way calibration with the quadri-linear parallel factor model, Anal. Chem. 84 (2012) 186–193.
[20] M.D. Carabajal, J.A. Arancibia, G.M. Escandar, A green-analytical chemistry method for agrochemical-residue analysis in vegetables, Microchem. J. 128 (2016) 34–41.
[21] M. Vosough, H. Mashhadiabbas Esfahani, Fast HPLC-DAD quantification procedure for selected sulfonamides, metronidazole and chloramphenicol in wastewaters using second-order calibration based on MCR-ALS, Talanta 113 (2013) 68–75.
[22] A.G. González, M.A. Herrador, A.G. Asuero, Intra-laboratory testing of method accuracy from recovery assays, Talanta 48 (1999) 729–736.