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

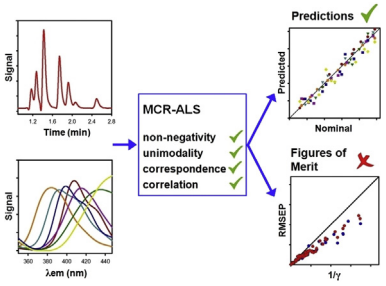
Rocío B. Pellegrino Vidal, Franco Allegrini, Alejandro C. Olivieri*

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

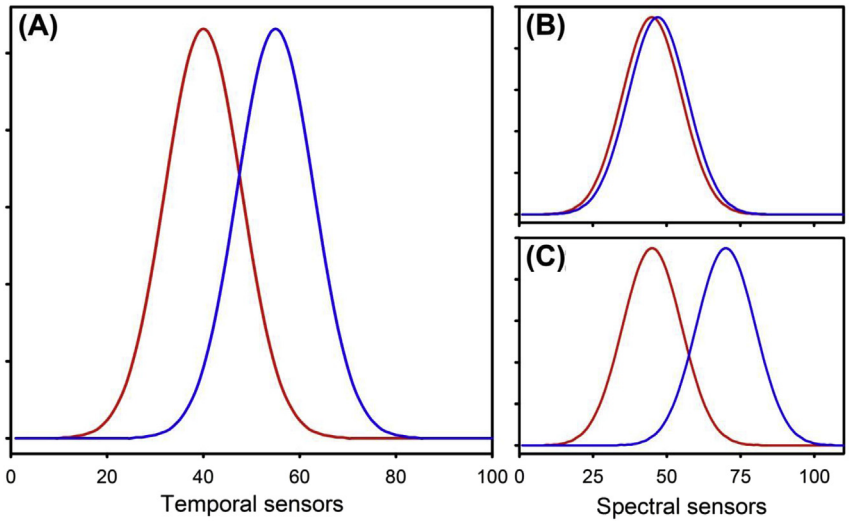


Fig. 1. Noiseless simulated profiles for compound 1 (blue) and compound 2 (red). (A) Temporal mode profiles, common to all data sets.(B) Spectral mode profiles for the highest spectral overlapping. (C) Spectral profiles for the lowest spectral overlapping. (For interpretation of the references to colour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