Chemometrics and Intelligent Laboratory Systems 141 (2015) 1–11

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Novel augmented parallel factor model for four-way calibration of high-performance liquid chromatography–fluorescence excitation–emission data

Santiago A. Bortolato [a], Valeria A. Lozano [a], Arsenio Muñoz de la Peña [b], Alejandro C. Olivieri [a,*]

[a] Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, 2000, Rosario, Argentina
[b] Department of Analytical Chemistry, Faculty of Sciences, University of Extremadura, 06006 Badajoz, Spain

ARTICLE INFO

Article history:
Received 11 September 2014
Received in revised form 24 November 2014
Accepted 27 November 2014
Available online 5 December 2014

Keywords:
Liquid chromatography–excitation–emission fluorescence data
Multi-way calibration
Augmented parallel factor analysis

ABSTRACT

A new augmented parallel factor analysis model (Augmented PARAFAC) is presented, inspired by the useful augmentation concept employed in multivariate curve resolution-alternating least-squares (MCR-ALS), applicable to calibration based on non-quadrilinear four-way data, such as those produced by high-performance liquid chromatography with matrix excitation–emission fluorescence detection. The new model involves creating an augmented three-way array in the elution time direction, containing data for the calibration sample set and for each of the test samples, subsequently analyzed with an Augmented PARAFAC version. To test the properties of this approach, chromatographic data were simulated with different degrees of overlapping and misalignment among the chromatographic peaks. Additionally, experimental data from olive oil samples were tested with the new model, aimed at the quantitation of the level of chlorophylls and pheophytins. The results were compared with those obtained by data processing with MCR-ALS. Relative prediction errors (%) were: Augmented PARAFAC, 9.7, 21.0, 14.7 and 9.3, and MCR-ALS, 5.9, 14.5, 20.0 and 14.7 for Chl *a*, Chl *b*, Phe *a* Phe *b*, respectively, for concentrations in the range 0.00–1.00 μg mL^{-1}. Both MCR-ALS and Augmented PARAFAC allow one to obtain a detailed and realistic description of the analyzed samples, in terms of pure elution time, excitation and emission spectral profiles, which can be independently retrieved for every component.

1. Introduction

Chemical multi-way calibration has become an important frontier in chemometric research. The multivariate analysis of excitation–emission fluorescence matrix (EEM) data, constituting three-way arrays when joining data for a group of samples, has been summarized in two recent reviews and a tutorial, all reporting up-to-date applications in the biomedical, environmental and food analysis fields [1–3].

High-performance liquid chromatography (HPLC), when combined with spectroscopic techniques, such as UV–visible diode-array detection (DAD) or fast-scanning fluorescence detection (FSFD) is also able to yield spectral-elution time matrix data. The response can be arranged as a data matrix, where each column corresponds to a different wavelength and each row to a different elution time. When full selectivity in the chromatographic separation is not achieved, and even in the presence of unexpected components, multivariate calibration can be successfully applied to the three-way arrays obtained by joining data for a group of samples. Additional benefits are decreasing cost and times of analysis. Several recent reports deal with the advantages and drawbacks associated with the combination of multivariate calibration and chromatography, and pertinent references on the successful processing of spectroscopic–chromatographic data can be found [4–14].

Four-way data can be obtained by joining third-order data for a set of samples into a four-dimensional mathematical object. The latter data not only retain the 'second-order advantage' inherent to three-way/second-order calibration [15], but can also have additional advantages [16]. They would display the obvious advantage of providing richer analytical information, implying more stable methods towards interferences and matrix effects, and less sensitive to minor changes in reaction conditions, which should allow for an improvement in predictive ability. In addition, improvements in sensitivity and the resolution of collinearity problems have been reported [17]. However, it is interesting to note that only few experimental four-way data have been recorded and used to develop analytical methodologies to date, which can be attributed to the fact that the practical acquisition of these data arrays is still difficult to implement. In addition, a thorough understanding of their analytical advantages is still needed.

Four-way data can be collected with a single instrument; the most common approach is the recording of luminescence EEMs as a function of some factors such as reaction time, decay time or any additional variable affecting the analytical signal (pH, sample volume, quenching effects, etc.). These factors are introduced as independent analytical modes, to construct multi-way data and to exploit the additional

* Corresponding author. Tel./fax: +54 341 4672704.
E-mail address: olivieri@iquir-conicet.gov.ar (A.C. Olivieri).

http://dx.doi.org/10.1016/j.chemolab.2014.11.013
0169-7439/

information they provide. Several different experimental approaches, describing practical analytical applications, can be found in the literature [18–41].

In the case of chromatographic data, four-way arrays can be generated with comprehensive gas or liquid two-dimensional chromatographic systems equipped with detection based on time of flight mass spectrometry (TOFMS) or DAD, leading to GC × GC–TOFMS or LC × LC–DAD hyphenated data. This is a currently growing scenario, and examples of different analytical problems in samples of high complexity have been reported [42–47].

Unidimensional chromatography can also provide four-way data, and Bro was the first to explore this possibility by recording full fluorescence EEMs during detection [48]. An alternative are LC–DAD-kinetic data, collected while following the hydrolysis of the Aly pesticide [49]. Recently, two different approaches have been reported by recording four-way LC–EEM data, which were employed for the analysis of: chlorophylls and pheophytins in olive oil samples [50], and fluoroquinolones in water samples [51].

When component profiles change from sample to sample, as it usually happens during the collection of chromatographic data because of elution time shifts or peak shape changes, the four-way LC–EEM data are not quadrilinear (strictly speaking, low-rank quadrilinear) [52]. In this case, the elution time mode is considered to be the 'quadrilinearity breaking' mode, or the mode suspected of breaking the quadrilinearity of the data [51,53]. Only a few multivariate techniques have been duly tested for data processing of these data: multivariate curve resolution with alternating least-squares (MCR-ALS) is the most common one [51], although the variant parallel factor analysis-2 (PARAFAC2) has also been applied [48]. However, it should be noted that the use of MCR-ALS implies unfolding the four-way data set into a super-augmented matrix. The statistical efficiency of decomposing multiway arrays is higher (and consequently the sensitivity is larger) when the original data structure is maintained, in comparison with unfolding into arrays of lower dimensions [54]. PARAFAC2, on the other hand, has been shown to be less efficient than MCR-ALS when processing multiway data in the presence of unexpected interferents not included in the calibration phase [55].

To avoid the above mentioned potential disadvantages, we propose a new model based on three-way PARAFAC, taking advantage of the augmentation philosophy applied in MCR-ALS studies. In the present work, both simulated and experimental four-way liquid chromatographic data with EEM detection are analyzed using the new Augmented PARAFAC model which allows to model inter-sample chromatographic profile variations, with an exhaustive comparison with the MCR-ALS algorithm. In the case of experimental data, samples already analyzed in Ref. [50] were now processed by Augmented PARAFAC, MCR-ALS and PARAFAC2.

2. Theory

2.1. Simulations

Data have been synthesized for a system having two calibrated analytes and a single potential interferent in the test samples along with the analytes. All data arrays were built mimicking four-way chromatographic data with EEM detection (elution time-excitation wavelength-emission wavelength), similar to those recorded for the experimental systems. The simulated signal for component n at unit concentration ($\mathbf{m}_n$) is governed by the following equation:

$$\mathbf{m}_n = \mathbf{t}_n \otimes \mathbf{w}_n \otimes \mathbf{z}_n \quad (1)$$

where $\mathbf{m}_n$ is the $JKL \times 1$ vectorized signal at unit concentration (J, K and L are the number of channels in each mode – elution time, excitation and emission wavelength, respectively – and are equal to 30, 20 and 50), $\mathbf{t}_n$, $\mathbf{w}_n$ and $\mathbf{z}_n$ are the individual profiles in each data mode (of size $J \times 1$, $K \times 1$, and $L \times 1$, respectively), and the symbol '⊗' indicates the Kronecker product.

Representative Gaussian elution time profiles $\mathbf{t}_n$ ($n = 1, 2$ and 3), partially overlapped in the time mode, are shown in Fig. 1A, although they change randomly from sample to sample during the simulations. Various types of chromatographic shifts were introduced into these time profiles, in order to generate a comprehensive set of

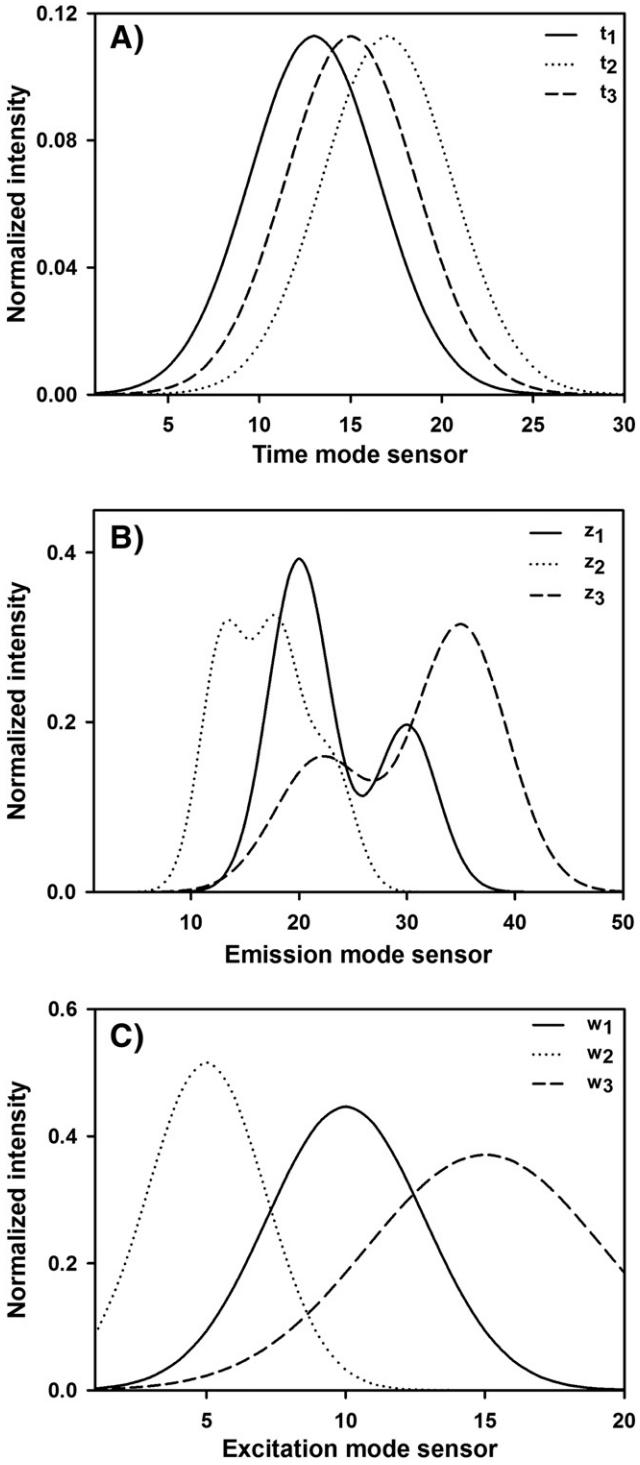

Fig. 1. Noiseless profiles employed for the simulations, in the elution time mode (A), in the emission mode (B) and in the excitation mode (C), for sample components at unit concentration. Solid line, analyte 1, dotted line, analyte 2, dashed line, potential interferent. The time profiles in (A) are scaled to unit area under each profile, while in (B) and (C) they are normalized to unit length.