



Functional data analysis, a new approach to aligning three-way liquid chromatographic with fluorescence detection data

Juan M. Lombardi^a, Santiago A. Bortolato^{a,b,*}

^a Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, Rosario S2002LRK, Argentina

^b Departamento de Matemática, Estadística y Procesamiento de Datos, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario S2002LRK, Argentina



ARTICLE INFO

Keywords:

FDA
Chromatography
Three-way data
Trilinearity
Second-order advantage

ABSTRACT

Functional data analysis (FDA) arises as a promissory auxiliary methodology designed to help the analytical chemists, especially chemometricians. However, although the innovative ideas of this approach have barely spread into the chemical research field. In this work, a novel approach for aligning three-way chromatographic-spectral data based on FDA methodology is proposed. Unlike most of the available algorithms, this novel method allows performing data alignment when the test data matrix contains unexpected chemical interferences. Simulated and experimental analytical systems composed of calibrated analytes and potential interferents in the test samples are studied. The experimental system corresponds to the determination of the four polycyclic aromatic hydrocarbons (benzo[*a*]pyrene, dibenzo[*a,h*]anthracene, benzo[*b*]fluoranthene and benzo[*k*]fluoranthene) in the presence of benzo[*j*]fluoranthene and benzo[*e*]pyrene as potential interferences. These compounds are priority pollutants; hence, its reliable quantification in environmental samples is an analytical challenge. The method consists in decomposing the three-way data in each sensor mode and in making a functional alignment of the pure vectors obtained from the individual chromatograms of each analyte and interferent. The functional preprocessing step enables the analysis of the data set with second-order multivariate calibration algorithms, such as PARAFAC. The results illustrate that the proposed method restores the trilinearity of the three-way data, thus being able to successfully quantify the analytes in the presence of the interferences, that is, exploiting the second-order advantage. MCR-ALS is also applied to both simulated and experimental data to evaluate the performance of the PARAFAC second-order calibration model proposed. The performance of the PARAFAC and MCR-ALS models are compared and discussed.

1. Introduction

Three-way chromatographic data are increasingly used in analytical chemistry research since, by appropriate processing with chemometric algorithms, they yield quantitative information of the analytes in the presence of potential interferences by using simple chromatographic procedures that meet most of the Green Analytical Chemistry principles [1–5]. Among a large repertoire of options, high-performance liquid chromatography (HPLC), combined with spectroscopic techniques, such as UV–visible diode-array detection (DAD) or fast-scanning fluorescence detection (FSFD), is suitable to obtain three-way data, which are usually called spectral-elution time matrices. These methodological alternatives were profusely discussed in recent reviews showing their latest applications in the biomedical, environmental and food analysis fields [6–8]. Nevertheless, most of the examples use UV–visible DAD detection and, surprisingly, very few reported works

concerning the processing of three-way HPLC -FSFD. Previous works used a videofluorimeter as a detector [9]. However FSFD is usually preferable, for instance for determining of polycyclic aromatic hydrocarbons (PAHs) [10–12], naphthalenesulphonate derivatives [13], fluoroquinolones [14] and metabolic disorder marker pteridines [15, 16].

The experimental response is structured as a data matrix, where each column corresponds to a wavelength and each row corresponds to a different elution time. The three-way arrays are processed by second-order multivariate calibration methodologies, allowing a successful simultaneous quantification of the analytes of interest, even when full selectivity in the chromatographic separation is not achieved. These approaches give access to reliable analytical information even in the presence of unexpected compounds (i.e., exploiting the second-order advantage). In the last few years, several works discussing the benefits and drawbacks of the combination of multivariate calibration and

* Corresponding author at: Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, Rosario S2002LRK, Argentina.
E-mail address: bortolato@iquir-conicet.gov.ar (S.A. Bortolato).

<https://doi.org/10.1016/j.microc.2018.06.041>

Received 14 March 2018; Received in revised form 21 June 2018; Accepted 29 June 2018
Available online 06 July 2018

0026-265X/ © 2018 Elsevier B.V. All rights reserved.

spectroscopic-chromatographic techniques have been published. In each case, a large number of successful applications of the analyzed strategies are shown [17–23].

From the above mentioned works, it is clear that liquid chromatography presents certain experimental facts that define the chemometric model chosen. One of these facts, temporal misalignment in the data, which means that a given constituent peak in different chromatographic runs appears at different positions and/or with different shapes along the elution time axis, is one of the most difficult problems to model [20]. The retention time shifts critically compromise the chemometric resolutions when complete chromatographic data are used. Specifically, this situation is commonly known as “trilinearity-breaking” [24] and is described as leading to a loss of the property of trilinearity in the data, which basically requires that each chemical component should present a unique profile (both in the spectral and elution time mode) in all samples [24].

It is worth mentioning that the problems mentioned in the discussion are not only typical for liquid chromatography. In fact, the aforementioned phenomena can also affect gas chromatography [25,26]. Although in this report the discussion is circumscribed to liquid chromatography, it would be interesting in the future to analyze gas chromatography experimental data with the method presented in this work.

In general, in the chromatographic data, retention time shifts are closely linked to the elution-time mode, and then it is said that the latter is the mode which is suspected of breaking the trilinearity. These types of cases coincide with the one presented in this work, since elution-time is the only non-reproducible mode. This condition, which is usually referred to “alignment problem”, is ubiquitous in the liquid chromatographic topic.

Although there exists a wide variety of chemometric methods that can achieve the second-order advantage, only some can deal with data including shifts and/or deformations in the chromatographic bands of the different sample components [20]. In theory, Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) and Parallel Factor Analysis 2 (PARAFAC2) models process non-trilinear data properly. Nevertheless, in a recent paper, it is shown that for systems with remarkable chromatographic artifacts, the results for PARAFAC2 may not be entirely appropriate [27]. However, it should be noted that PARAFAC2 is also a proper option to solve LC data with alignment problems in many experimental systems of varying complexity [28]. On the other hand, in a recent work it has been shown that, for extremely complex analytical systems, the accepted expressions for analytical figures of merit do not fully represent the predictive capability of the MCR-ALS model since they are often overestimated [29]. Therefore, whenever possible, it is preferable to restore the trilinearity of the data, given that linear models tend to be more robust or, at least, to have more reliable performance parameters [29–31].

Restoring trilinearity often involves aligning the chromatographic profiles of the analytes in the different samples before the construction of the second order calibration model. Several methods have been developed to restore trilinearity, which can be classified into two large groups: (1) methods that seek the maximum correlation between chromatograms for analyte peak alignment, and (2) methods that try to take full advantage of the matrix data structure. In the former group, the multi-wavelength correlation optimized warping (COW) [32,33] and their variants [34] stand out. In the second group the rank alignment [35], and the iterative target transformation factor analysis (ITTTFA) [36], are found among others. It is also worth mentioning that the methods in the first group have some implementation problems: they generally need to optimize a series of input parameters, also change the shape of the peak (especially in COW [37]) and do not fully use the spectral information from the datasets with multi-way detectors [38]. Recently, a third variant has been published, where a new combination of MCR-ALS and COW with PARAFAC is developed to exploit the second-order advantage in complex chromatographic measurements [39]. In MCR-COW, the complexity of the chromatographic data

is reduced by arranging the data in a column-wise augmented matrix, using MCR bilinear model and aligning the resolved elution profiles by means of COW. Then, the aligned chromatographic data is then decomposed using a PARAFAC trilinear model of in order to obtain the information required.

A few years ago, Ramsay and Silverman [40,41] proposed a whole new approach to deal with problems of a similar nature, which was called functional data analysis (FDA). The methodology is based on thinking the experimental data as the manifestation of an underlying functional dependency between the observed variables. Generally speaking, in spectroscopic or chromatographic data, the variable of interest is often measured as a function of wavelength, frequency, time, or some other continuum. Despite this fact, the predominant view involves processing the data with chemometric algorithms that treat them as if they were discrete sets. The principal component analysis, partial least squares and related techniques, even Parallel Factor Analysis (PARAFAC), were not specifically proposed for the analysis of spectral or chromatographic data: they consider the multidimensional data block as a set of n different variables that can be ordered in any way to give equivalent results [42]. On the contrary, the framework proposed by Ramsay and Silverman advocate the use of mathematical tools that take advantage of the singular characteristics of chemistry data. Recently, a review reports giving an overview on applications of FDA to chemistry data have been published [43,44]. These reports highlight that FDA is not yet widely used to analyze chemistry data, and both papers contain very few examples of chemometric applications of FDA. However, as regards the functional analysis of chromatographic data, there is a growing interest in FDA-based methods as they can provide better solutions than usual approaches, especially in relation to the alignment problem [38,45,46]. Nevertheless, these methods align only first-order chromatographic data, while, to the best of our knowledge, FDA treatment of the three-way chromatographic-fluorescence data with second-order advantage has not been reported yet.

In order to exploit the second-order advantage, an alignment algorithm based on FDA with trilinear decomposition methodologies is proposed. Briefly, the new algorithm decomposes the three-way data in their two modes (spectral and elution-time vectors). It uses an adequate basis function to process the nonlinear mode, aligns the functionalized pure vectors and reshapes the transformed vectors in matrices, thus restoring the trilinearity of the data. The algorithm is referred to as FAPV (Functional Alignment of Pure Vectors).

In summary, in this work, both simulated and experimental three-way HPLC-FSFD systems with trilinearity breaking elution-time mode are processed through a new alignment algorithm based on functional data analysis (FAPV) and analyzed using a PARAFAC model. Additionally, an exhaustive comparison of this model with the MCR-ALS methods is performed.

The simulations provide a series of chromatographic scenarios with a huge variety of artifacts (changes in peak position and band shapes) in order to examine the capacity of the new algorithm to restore the trilinearity of the data.

The experimental data system corresponds to the analysis of four polycyclic aromatic hydrocarbons (PAHs), benzo[*b*]fluoranthene (BbF), benzo[*k*]fluoranthene (BkF), dibenzo[*a,h*]anthracene (DbA) and benzo[*a*]pyrene (BaP), in water samples that also contain two potential interferences, benzo[*j*]fluoranthene (BjF) and benzo[*e*]pyrene (BeP). The PAHs are ubiquitous and toxic compounds formed by two or more aromatics rings originated from the incomplete combustion of organic matter [47]. For the target analytes, genotoxic and mutagenic activities have also been proved [48]. The regulations on the identification and/or determination of PAHs in environmental samples are continually revised, demanding ever lower maximum allowed values [49–51]. It is crucial to promote improved analytical techniques for determining of these compounds in different environmental matrices. The present report indicates that our method is an adequate analytical alternative for simultaneous quantification of BaP, DbA, BkF, and BbF, in the presence

Download English Version:

<https://daneshyari.com/en/article/7640189>

Download Persian Version:

<https://daneshyari.com/article/7640189>

[Daneshyari.com](https://daneshyari.com)