Microchemical Journal 142 (2018) 219–228

Contents lists available at ScienceDirect

Microchemical Journal

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

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 interferents 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/

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 (ITTFA) [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 interferents, 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

of two potential interferents in real samples, using high-performance liquid chromatography coupled to fast scanning fluorescence spectrometry under isocratic conditions, which notably reduces the analysis time.

2. Theory

2.1. FDA background and basic concepts

Functional data analysis proposes a new action field in chemometrics. The statistical and mathematical methods make this new approach have a unique ability to deal with functional data. FDA methods can unravel complex data structures which may be characterized as continuous functions, such as, data arising from chromatographic or spectroscopic analyses, where they are represented as a function of time or wavelength, respectively.

The fundamental assumption of FDA is that there are functions which explain the behavior of the recorded data, and proposes the replacement of the raw data by these functions, or approximations thereof, in order to perform the respective analysis. While it is true that FDA has been used successfully in a wide variety of research areas (medicine [52], economics [53], etc.), it is still an incipient methodology in chemistry. In chemometrics, for instance, the reported works are yet very scarce, despite the fact that the idea of representing spectra with functions was already introduced in 1993 [54].

This section briefly sets forth the basic concepts used to expand a discrete and finite set of data on a complete basis of functions. When working with 'soft' data sampled in very short time intervals, it is logical and almost intuitive to think that the information obtained is the product of the discrete manifestation of an underlying function. Once the functional structure of the information has been identified, it is convenient to analyze the data in a functional space that is able to describe its behavior more completely.

FDA begins by acquiring a sequence of observations $y = (y_1, y_2, \ldots, y_n)$, measured in $(t_1, t_2, \ldots, t_n)$ times. The functional data are typically observed as a discrete number of n-tuples (t_j, y_j) where y_j is the function value at time t_j (it should be noted that t_j is not always referred as time). There are several continua over which data can be measured, including wavelength, frequency, chemical shift, among others. In this paper, t_j is the chromatographic elution time. FDA approach supposes that there exists some function $x(t)$ that underlies the recorded data, and it assumes that this function satisfies the requirements of continuity and derivability, at least in the first derivative (i.e., two adjacent data points, y_j and y_{j+1} should have similar values). Nevertheless, this fact may not be so clear in raw data due to observational error and noise. Consequently, the following general model for each y_j observation could be proposed:

$$y_j = x_{(t_j)} + \varepsilon_j;\ \varepsilon_j \sim N(0, \sigma^2) \tag{1}$$

where term ε_j represents the error of the fit. ε_j values must be independent with mean zero and a finite variance ($\varepsilon_j \sim (0, \sigma^2)$).

The precision with which the behavior of the data is estimated at a point t_j, depends on certain factors that must be considered especially when interpolating, among them are the density of the values of t_j with respect to the curvature of the function $\frac{\partial^2 x_{(t)}}{\partial t^2}$ estimated at that point, and the intensity and type of error in the measurements.

If the observations are arranged in a vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, Eq. (1) is transformed into

$$\mathbf{y} = x(\mathbf{t}) + \mathbf{e} \tag{2}$$

where $\mathbf{y}$, $x(\mathbf{t})$, and $\mathbf{e}$ are column vectors of length n. In general, the function $x(\mathbf{t})$ cannot be represented analytically, but it admits an approximation through an expansion in the basis functions $\Psi_k = \{\psi_{k(t)}\}$:

$$x(\mathbf{t}) = \sum_{k=1}^{K} c_k \psi_{k(t)} \tag{3}$$

A basis function system consists of a set of independent functions ψ_k, $k = 1, 2, \ldots, K$, so that a linear combination of these approximates the function $x(t)$ for an appropriate K, being K = number of basis functions. The ratio between the number of basis functions and the number of observation defines the detail degree with which you want to describe the discrete phenomenon. It is advisable to find a balance between useful information and overfitting.

There are many types of basis function, and an adequate choice should help to highlight those aspects of the data that are useful for the analysis. In the case of chemometric applications, Fourier and spline basis function can be mentioned [43,55], while in the case of applications in other scientific areas, it has been shown that function systems including wavelets, exponential or power bases are very useful as well [40]. Fourier basis function are especially suitable for data with some periodicity. For strictly non-periodic data, other basis function may be considered in order to obtain a better representation of the underlying functional form of the signal [56–59]. In our case, splines and Gaussian basis have been used to represent the chromatographic data. Since the implementation of the basis function follows a common structure, only the elementary details of generic basis functions will be discussed in this section. For more details, see Ref. [44].

Regardless of the chosen basis function, solving the functionalization problem implies estimating the coefficients c_k of Eq. (3). The coefficients c_k can be calculated using least squares criterion, in which the mean square error (MSE) is meant to be minimized:

$$MSE = \frac{1}{n} \sum_{j=1}^{n} \left[y_j - \sum_{k=1}^{K} c_k \psi_{k(t_j)} \right]^2 \tag{4}$$

Eqs. (3–4) can be written in matrix terms to implement the least squares criterion, according to:

$$\hat{x}(\mathbf{t}) = \mathbf{c}'\boldsymbol{\psi} \tag{5}$$

$$MSE = (\mathbf{y} - \boldsymbol{\Psi}\mathbf{c})'(\mathbf{y} - \boldsymbol{\Psi}\mathbf{c}) = \|\mathbf{y} - \boldsymbol{\Psi}\mathbf{c}\|^2 \tag{6}$$

where $\mathbf{c}$ is a vector containing the coefficients for the basis functions, $\boldsymbol{\psi}$ is a vector containing the values of the basis functions at time t_j, and $\boldsymbol{\Psi}$ is a $n \times K$ matrix that contains the values of $\boldsymbol{\psi}_k(t_j)$, $k = 1, \ldots, K$; $j = 1, \ldots, n$.

Solving Eq. (6) according to the minimization criterion, we obtain the estimates for the coefficients $\mathbf{c}$ of the function:

$$\mathbf{c} = (\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}(\boldsymbol{\Psi}'\mathbf{y}) \tag{7}$$

Finally, Eq. (5) is calculated with the estimated coefficients in order to obtain an effective functional representation $x(t_j)$ of our original data vector $\mathbf{y}$.

2.2. Functional Alignment of Pure Vectors (FAPV)

FAPV algorithm allows restoring trilinearity in three-way data. Although it is not discussed here, it is easily extendable to multi-way data. The only condition for its application is that data hold one mode with reasonable selectivity both for each analyte and interferent, i.e., the pure analyte signal must be different enough so that the processing algorithm chosen retrieves them adequately [60]. In our case, the required condition is met because the analytes have reasonably different fluorescence emission spectra (see Ref. [12]).

In a first phase, the hyper matrix $\boldsymbol{D}_{cal}$ is built from the calibration samples experimental data and can be expressed as

$$\boldsymbol{D}_{cal} = \{\mathbf{D}_i\}_{N_{cal}};\ \mathbf{D}_i \in \mathbb{R}^{L \times T}$$

while a hyper matrix data $\boldsymbol{D}_{test}$ is also defined for the test samples as

$$\boldsymbol{D}_{test} = \{\mathbf{D}_i\}_{N_{test}};\ \mathbf{D}_i \in \mathbb{R}^{L \times T}$$

where $\mathbf{D}_i$ indicates the ith HPLC-FSFD matrix of each data set, L represents the number of sensors of the selective mode (spectral), T is the number of sensors of the non-aligned mode (elution-time), N_{cal} is the number of the total samples of the calibrated data set, and N_{test} is the number of total samples of test data set.

Once the work variables have been defines, the algorithm runs through the following sequence:

(1) *Estimating the spectra of selective mode using the SIMPLISMA* [61] auxiliary routine. Once the initial estimates have been resolved, a new matrix **E** with the pure analytes and interferents spectra obtained from $\boldsymbol{D}_{cal}$ and $\boldsymbol{D}_{test}$ is built.

The algorithm uses SIMPLISMA to estimate the initial spectra. Although it has been reported that SIMPLISMA could promote the algorithm to end at local minimums, in our case, provided adequate initial estimates for the spectra in the analyzed systems [62]. When SIMPLISMA is not useful, it is convenient to use another initial estimator of the spectra, such as an initialization based on random orthogonalized values, or known experimental profiles, or singular value decomposition vectors or direct trilinear decomposition [63].

(2) *Estimating the misaligned elution-time profiles of non-selective mode* from **E**, $\boldsymbol{D}_{cal}$, and $\boldsymbol{D}_{test}$.
(3) *Aligning the pure vectors of elution-time mode based on FDA approach.*

In this step, two FDA-based alternatives can be selected:

a. The preservation of the shape of the chromatographic band of each analyte. The functionalization provides results visually more similar to the experimental data but the multilinearity is not restored completely, thus the results of a PARAFAC analysis are not adequate.
b. The preservation of the area of the chromatographic band of each analyte. In this way, the multilinearity of the data is completely restored.

In our case, we selected the second option. The conservation of the area can be evaluated under a "classic" paradigm by using a type 1 norm. However, a better estimation of the area may be achieved by using an FDA approach.

(4) *Reshaping each sample data* by matrix product between the pure vectors (spectra and elution-time profiles).
(5) *Second-order calibration.*

Now we will explain the algorithm in terms of the proposed transformations and the basis functions used. Let $\boldsymbol{D}_{cal}$ and $\boldsymbol{D}_{test}$ be second-order tensors which contain all the sample calibration data and test data, respectively,

$$\boldsymbol{D}_{cal} = (d^{cal}_{t,\lambda,i_1})_{T\times L\times N_{cal}} \tag{1}$$

$$\boldsymbol{D}_{test} = (d^{test}_{t,\lambda,i_2})_{T\times L\times N_{test}} \tag{2}$$

where T indicates the number of sensors of the non-aligned mode (i.e., the chromatographic way), $t = 1, \ldots, T$, L is the number of sensors of the selective mode (spectral way) where $\lambda = 1, \ldots, T$, i_1 is the i_1-th calibrated sample with $i_1 = 1, \ldots, N_{cal}$, and i_2 is the i_2-th test sample, with $i_2 = 1, \ldots, N_{test}$.

(1) *Estimating the spectra of selective mode using the SIMPLISMA algorithm.*

$$\mathscr{F}_1\colon (\boldsymbol{D}_{cal}, N_a, \varepsilon) \rightarrow \mathbf{E}_{\text{cal}} = (e^{cal}_{\lambda,j_1})_{L\times N_a} \tag{3}$$

$$\mathscr{F}_2\colon (\boldsymbol{D}_{test}, N_a + N_i, \varepsilon) \rightarrow \mathbf{E}_{\text{test}} = (e^{test}_{\lambda,j_2})_{L\times N_a+N_i} \tag{4}$$

where j_1 is the j_1-th analyte with $j_1 = 1, \ldots, N_a$, N_a is the total number of analytes, j_2 is the j_2-th compound with $j_2 = 1, \ldots, N_a + N_i$, N_i is the total number of interferents, and ε indicates the noise signal. The columns of $\mathbf{E}_{\text{cal}}$ are vectors that represent the pure spectra estimation of all analytes, while the columns of $\mathbf{E}_{\text{test}}$ contain the pure spectra estimation of all analytes together whit the interferents.

Once the initial estimates have been resolved, the following transformation is proposed

$$\mathscr{F}_3\colon (\mathbf{E}_{\text{cal}}, \mathbf{E}_{\text{test}}) \rightarrow \mathbf{E} = (e^{test}_{\lambda,j_2})_{L\times N_a+N_i} \tag{5}$$

where **E** is a new matrix built with the analytes spectra obtained from $\mathbf{E}_{\text{cal}}$, and the interferents spectra obtained from $\mathbf{E}_{\text{test}}$.

(2) *Estimating the misaligned elution-time profiles of non-selective mode.*

Two new tensors with analytical sense are defined:

$$\boldsymbol{C}_{cal} = (c^{cal}_{t,i_1,j_1})_{T\times N_{cal}\times N_a} \tag{6}$$

$$\boldsymbol{C}_{test} = (c^{test}_{t,i_1,j_2})_{T\times N_{test}\times N_a+N_i} \tag{7}$$

where $\boldsymbol{C}_{cal}$ contains the elution-time profile of each analyte in each calibration sample while $\boldsymbol{C}_{test}$ contains the elution-time profile of each analyte and interferent of each sample of the test set. Since the chromatographic profile of each component may not be aligned, different chromatographic profiles for each component from every sample can be obtained.

From Eqs. (6–7), each element of the tensors on Eqs. (1–2) can be estimated according to:

$$\hat{d}^{cal}_{\lambda,t,i_1} = \sum_{i_1}^{N_{cal}} e^{cal}_{\lambda,j_1} c^{cal}_{t,i_1,j_1} + \varepsilon^{cal}_{\lambda,t,i_1} \tag{8}$$

$$\hat{d}^{test}_{\lambda,t,i_2} = \sum_{i_2}^{N_{test}} e^{test}_{\lambda,j_2} c^{test}_{t,i_2,j_2} + \varepsilon^{test}_{\lambda,t,i_2} \tag{9}$$

To solve Eqs. (8–9), first $\boldsymbol{C}_{cal}$ and $\boldsymbol{C}_{test}$ must be estimated. In order to do this, the generalized inverse of the vector space **E** is calculated for each data set ($\mathbf{V}_{\text{cal}}$ for calibration set, and $\mathbf{V}_{\text{test}}$ for test set),

$$\mathbf{V}_{\text{cal}} = \mathbf{E}^{-1}_{\text{cal}} = (v^{cal}_{i_1,\lambda})_{N_a\times L} \tag{10}$$

$$\mathbf{V}_{\text{test}} = \mathbf{E}^{-1}_{\text{test}} = (v^{test}_{i_2,\lambda})_{N_a+N_i\times L} \tag{11}$$

Finally, individual elution-time vectors are estimated as an orthogonal projection of $\mathbf{V}_{\mathbf{cal}}$ and $\mathbf{V}_{\mathbf{test}}$ on each matrix of each data set:

$$\hat{c}^{cal}_{t,i_1,j_1} = \sum_{\lambda}^{L} \hat{d}^{cal}_{\lambda,t,i_1} v^{cal}_{j_1,\lambda} \tag{12}$$

$$\hat{c}^{test}_{t,i_1,j_2} = \sum_{\lambda}^{L} \hat{d}^{test}_{\lambda,t,i_2} v^{test}_{j_2,\lambda} \tag{13}$$

which are sorted in the estimated tensors,

$$\hat{\boldsymbol{C}}_{cal} = (\hat{c}^{cal}_{t,i_1,j_1})_{T\times N_{cal}\times N_a} \tag{14}$$

$$\hat{\boldsymbol{C}}_{test} = (\hat{c}^{test}_{t,i_1,j_2})_{T\times N_{test}\times N_a+N_i} \tag{15}$$

(3) *Aligning the pure vectors of elution-time mode based on FDA approach.*

Gaussian basis functions which will be used in the reshaping of the aligned data, are defined,

$$\varphi = \{\varphi^{\mu,\sigma}_{(x)}\} = \frac{1}{(2\pi\sigma^2)^{0.5}} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2} \tag{16}$$

$$x_t \coloneqq (x_t)_T$$

$$x_t \subset x$$

To evaluate the projection coefficients

$$\left\langle \widehat{C}_{cal_{(:,i_1,j_1)}} \mid \varphi_{\mu,\sigma(x_k)} \right\rangle = \sum_{t}^{T} \varphi_{(x_t)}^{\mu,\sigma} \hat{c}_{t,i_1,j_1}^{cal} \tag{17}$$

it is required to define a discrete and finite partition in $x_t := (x_t)_T$ so that its dimension is equal to T. A uniform and equidistant partition is convenient.

$$\varphi = \{\varphi_{(x_t)}^{\mu,\sigma}\} = \frac{1}{(2\pi\sigma^2)^{0.5}} e^{-\left(\frac{x_t-\mu}{2\sigma}\right)^2} \tag{18}$$

$$\left(\mu_{i_1,j_1}^{cal-max}, \sigma_{i_1,j_1}^{cal-max}\right) = arg_{\mu,\sigma} max\left(\sum_{t}^{T} \varphi_{(x_t)}^{\mu,\sigma} \hat{c}_{t,i_1,j_1}^{cal}\right) \tag{19}$$

$$\left(\mu_{i_2,j_2}^{test-max}, \sigma_{i_2,j_2}^{test-max}\right) = arg_{\mu,\sigma} max\left(\sum_{t}^{T} \varphi_{(t)}^{\mu,\sigma} \hat{c}_{t,i_2,j_2}^{test}\right) \tag{20}$$

The components of the vector $(\mu_{i_*}, j_*^{\,cal\ or\ test-max}, \sigma_{i_*}, j_*^{\,cal\ or\ test-max})$ contain the values of μ and σ for which the projection coefficient $(\sum_t^T \varphi_{(x_t)}^{\mu,\sigma} \hat{c}_{t,i_*,j_*}^{cal\ or\ test})$ is maximized for the same i_1, j_1 parameter set (for calibration samples) or i_2, j_2 parameter set (for samples test).

Now we build the matrices $\mathbf{A}_{\text{cal}}$ and $\mathbf{A}_{\text{test}}$ so that they contain the estimation of the area of each j_1 analyte from i_1 sample of the calibrated set, and for each j_2 analyte or interferent from i_2 sample of the test set, respectively:

$$\mathbf{A}_{\text{cal}} = (a_{i_1,j_1})_{N_{cal},N_a} = \sum_{t}^{T} \hat{c}_{t,i_1,j_1}^{cal} \tag{21}$$

$$\mathbf{A}_{\text{test}} = (a_{i_2,j_2})_{N_{test},N_a+N_i} = \sum_{t}^{T} \hat{c}_{t,i_2,j_2}^{cal} \tag{22}$$

From Eqs. (19–22) the following matrices are proposed as the replacement of non-trilinear experimental ones, both for the calibration and test data sets ($d_{t,\,\lambda,\,i_2}^{\,cal}$ and $d_{t,\,\lambda,\,i_2}^{\,test}$ in Eqs. (1–2), respectively)

$$\widehat{d}_{\lambda,t,i_1}^{acal} = \sum_{j_1}^{N_a} e_{\lambda,j_1}^{cal} \varphi_{(x_t)}^{\mu_{i_1,j_1}^{cal-max},\sigma_{i_1,j_1}^{cal-max}} a_{i_1,j_1} \tag{23}$$

$$\widehat{d}_{\lambda,t,i_2}^{atest} = \sum_{j_2}^{N_a+N_i} e_{\lambda,j_2}^{atest} \varphi_{(x_t)}^{\mu_{i_1,j_2}^{test-max},\sigma_{i_1,j_2}^{test-max}} a_{i_2,j_2} \tag{24}$$

(4) *Reshaping each sample data*

Finally, from the individual matrices of Eqs. (23–24), the following estimated sensors are built for the second order calibration

$$\widehat{\boldsymbol{D}}_{acal} = (\widehat{d}_{\lambda,t,i_1}^{acal})_{L\times T\times N_{cal}} \tag{25}$$

$$\widehat{\boldsymbol{D}}_{atest} = (\widehat{d}_{\lambda,t,i_2}^{atest})_{L\times T\times N_{test}} \tag{26}$$

where $\widehat{\boldsymbol{D}}_{acal}$ and $\widehat{\boldsymbol{D}}_{atest}$ are tensors whose trilinear structure has been completely restored, and which contain all the information of the raw data tensors (see Eqs. (1) and (2)), but aligned on each elution-time sensor.

(5) *Second-order calibration.*

In the second order calibration step, both in PARAFAC and MCR-ALS, the data used are the tensors $\widehat{\boldsymbol{D}}_{acal}$ and $\widehat{\boldsymbol{D}}_{atest}$ which are rearranged or reshaped according to the algorithm requirements.

Fig. 1 shows a schematic representation of the FAPV algorithm used to analyze three-way data corresponding to a typical calibration samples set (three calibrated analytes).

2.3. Simulations

The data were simulated to represent numerous scenarios mimicking elution time-fluorescence emission data, similar to those experimentally recorded when running chromatographic measurements with spectrofluorimetric detection. In all cases, the systems generated had three calibrated analytes and two potential interferents in the test samples along with the analytes. A detailed description of the simulated data is shown in Supporting Information.

2.4. Second-order calibration

To evaluate the performance of the FAPV algorithm, PARAFAC [63] and MCR-ALS [64] second-order calibration models were used, which were implemented following the general guidelines discussed in the aforementioned references. In both cases, the second-order calibrations were performed by imposing non-negativity constraint on both dimensions of the data, and unimodality on the temporal dimension. On the other hand, in the MCR-ALS model an augmented data matrix was created from each test data matrix and the calibration data matrices. If all the matrices are of size $T \times L$, where T is the number of data points in the dimension of the retention times and L the number of fluorescence wavelengths, the direction of rows is considered the time direction, and the direction of columns the spectral direction. Augmentation can be performed in either direction, depending on the mode that breaks the trilinearity. In our case, augmentation was implemented in the time direction (i.e. column-wise augmentation) because of the presence of warping effects and retention time shifts between different chromatographic runs. For further information, see Ref. [20].

2.5. Calculations

FAPV algorithm and simulations were made using in-house Python 3.0 [65] routines, which are available from the authors on request. The PARAFAC and MCR-ALS algorithms were written for Python 3.0 according to the MATLAB codes from [66,67], respectively. All programs were run on an IBM-compatible microcomputer with an Intel Core i5, 2.90 GHz microprocessor and 16.00 GB of RAM.

3. Experimental

The BbF, BaP, DbA, and BkF analytes were determined in water samples in the presence of the potential interferents BjF and BeP, using the chromatographic method developed in Ref. [12], i.e., HPLC with fluorescence spectral detection. The experimental procedure and sample composition were the same as those described in Ref. [12]; therefore they are not repeated here. However, a new data treatment was carried out: from the raw data matrices (collected with the excitation wavelength fixed at 300 nm, using emission wavelengths from 340 to 580 nm each 2 nm, and times from 0 to 7.20 min each 2.7 s), the temporal mode was restricted to 2.43–4.50 min (matrices were of size 46 × 121), where there occurs coelution of the four analytes together with the two interferents mentioned above.

The calibration set included 18 samples: 16 corresponded to the concentrations provided by a fractional factorial design at two levels, and the remaining two correspond to a blank and to a solution containing all the studied PAHs at an average concentration. The tested concentrations were in the ranges 0.0–100 ng mL^{-1} for BbF, 0.0–50.0 ng mL^{-1} for BaP and DbA, and 0.0–20.0 ng mL^{-1} for BkF. The test set contained twenty samples of the studied analytes at random concentrations, with BjF and BeP as interferences (the concentrations of the latter were in the range 0–600 ng mL^{-1} and 0–1000 ng mL^{-1}, respectively). HPLC-fluorescence data were collected using a liquid chromatograph equipped with a Waters 515 pump connected to a Varian Cary-Eclipse luminescence spectrometer as a detector. For additional instrumental details see [12].

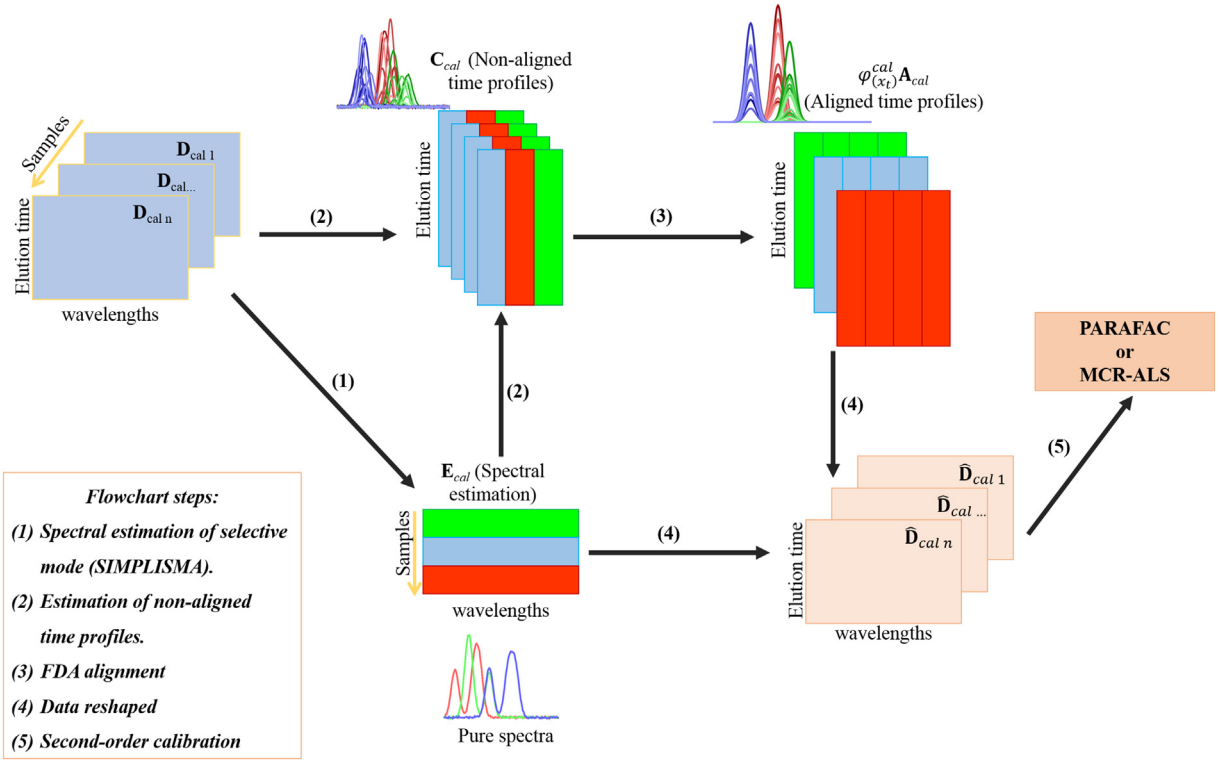

Fig. 1. Flowchart of the FAPV algorithm for the treatment of three-way chromatographic-spectral calibration data (three calibrated analytes, indicated in blue, red and green solid lines). The symbols used are explained in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Results and discussion

4.1. Simulated data

In Section 2.3 we describe the generation of the simulated data sets. Before the second-order multivariate calibration steps, the most relevant details of the huge variety of simulated chromatographic scenarios analyzed are discussed.

In order to evaluate the effectiveness of the algorithm used to restore trilinearity in the data set, a total of 10^5 possibility sets were generated, considering different noise intensities, number of sensors, number of interferents, and misalignment degrees (artifact parameters), among others (for further details, see Supporting Information). Fig. 2b SI illustrates eight chromatographic possibilities given by the modification of the artifact parameters for a single analyte in two-run superposed, from a fully aligned data set (Fig. 2b SI-I) to a totally misaligned one (Fig. 2b SI-VIII). Once FAPV algorithm processed the data as described in Section 2.2, the second-order calibrations were performed with PARAFAC and MCR-ALS. The same calibrations had been previously made with raw data for comparative purposes.

As regards the study of the performance of FAPV algorithm, a global mean square error of predictions (MSEP) was defined for the three calibrated analytes, according to:

$$MSEP_m = \frac{1}{3n}\sum_{i=1}^{n}\sum_{j=1}^{3}(y_{ji} - \bar{y}_{ji,jm})^2;\ m = \{\text{PARAFAC without FPVA, MCR with FPVA, PARAFAC with FPVA, MCR with FPVA}\}$$

where m indicates the evaluated model, n is the total number of test samples, y is the nominal concentration of the j analyte, and $\bar{y}$ is the predicted concentration of the j analyte. In this way, each set of analytical possibilities was transformed into an answer vector.

Fig. 2a shows the MSEP results of PARAFAC and MCR-ALS models applied to the raw data as a boxplot. The two models were applied following the general guidelines of references [63, 64], i.e., imposing the non-negativity constraints for the spectral and elution-time mode, both for PARAFAC and MCR-ALS. For the latter, the data matrix for each test sample was augmented with the calibration data matrices in the elution-time mode by joining the elution time-spectral data matrices alongside each other (i.e., by column-wise augmentation), creating the so-called augmented data matrix. Evidently, the performance of PARAFAC is always worse than that of MCR-ALS, except in the case when the original data are perfectly trilinear. In contrast, when the misalignment of the data is remarkable, PARAFAC is overshadowed by MCR-ALS. This is mainly due to the fact that PARAFAC decomposes the data into pure vectors assuming a linear model in each of the analysis modes; this assumption brings advantages when the data are trilinear, but produces large errors when they are not. On the contrary, MCR only requires that data be linear in only one of their modes. These facts place the analyst in a compromise situation in which he must opt for an adequate algorithm. Luckily, the decision is usually clear if he knows the nature of the data he works with.

Finally, the results obtained when performing the second-order calibrations with the aligned data were compared (Fig. 2b). It is observed that the performance of FAPV/PARAFAC is marginally better than FAPV/MCR-ALS, since the data under analysis turned perfectly trilinear in all situations and PARAFAC was able to take advantage of this property to make better predictions. In the series analyzed, a similar tendency is observed in the errors of the FAPV/PARAFAC and FAPV/MCR-ALS models, but always the boxes that correspond to the first ones are slightly smaller, as well as the whiskers and the medians (Fig. 2b).

In Supporting Information, a comparison among PARAFAC and MCR-ALS results for raw data and aligned data with FAPV is performed (Fig. 2 SI). Also, the simulated data were processed with PARAFAC2 and MCR/COW-PARAFAC models (Fig. 4 SI). The obtained results show that these models have a lower performance than FAPV/PARAFAC. For further details, see Supporting Information.

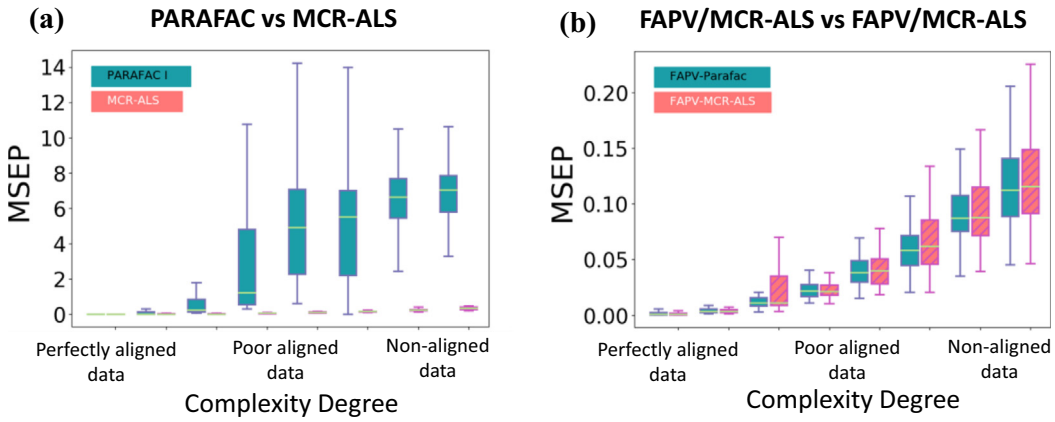

Fig. 2. Boxplot corresponding to the MSEP results for a) PARAFAC and MCR-ALS without FAPV pre-processing, and b) PARAFAC and MCR-ALS with FAPV pre-processing. Abscissa axis indicates the complexity degree of the system, from zero complexity to a highly complex scenario. Note that the ordinates axis does not have the same scale. For further details, see Theory section and Fig. 1 in Supporting Information.

4.2. Experimental data

The experimental data correspond to the analytical determination of BbF, BkF, BaP and DbA in a set of samples that contain BjF and BeP as potential interferents together with the analytes. In the course of the chromatographic analysis of these samples using fast scanning fluorescence emission detection, remarkable overlapping in both data modes takes place, as described in Ref. [12]. Likewise, the presence of interferents in the test samples intensifies the above mentioned drawbacks, making the proposed experimental system a relevant analytical challenge for the developed algorithm, in line with the examples discussed in the previous section.

Fig. 3 shows the chromatograms of DbA at the calibration and test samples retrieved by PARAFAC from the raw data (Fig. 3a) and estimated by the FAPV algorithm (Fig. 3b). As regards the raw data, a deformation and/or duplication of the chromatographic bands of DbA were observed. On the contrary, the data processed with the FAPV algorithm are perfectly trilinear, thus, PARAFAC becomes a good method to make the second-order calibration.

The general procedure applied to this experimental system was analogous to that discussed above in connection with the simulated data systems. The data are first aligned and then processed with MCR-ALS and PARAFAC. As it is was discussed above, in MCR-ALS analysis, the matrix data for each test sample was augmented by column-wise augmentation. The decomposition was performed according to a bilinear model proposed in Ref. [64], by imposing the constraint of non-negativity in both dimensions and unimodality in the temporal dimension. The MCR-ALS algorithm requires the exact number of components responsible for the analytical signal to be known, and is preferable to initialize the algorithm with the profiles of the components as close as possible to the final result. The number of components can be estimated using principal component analysis on the basis of singular value decomposition of the augmented matrix and the pure spectra was provided by SIMPLISMA algorithm. On the other hand, the PARAFAC implementation was followed from Ref. [63], also imposing the constraint of non-negativity in both dimensions and unimodality in the temporal dimension for the six components modeled. Finally, for comparative purposes, the raw data were also processed by the two algorithms.

Fig. 4a shows the prediction results corresponding to the application of MCR-ALS to raw data from the twenty test samples. The predicted values for the PAHs fit reasonably well with their respective nominal values. This was expected since MCR-ALS can get reliable information from non-trilinear data. Fig. 4b shows the poor performance of PARAFAC when it processes the raw data, which is explained by the fact that the model cannot deal with non-trilinear data, as observed in the simulated data set and as has been reported in the literature [20]. Although it is true that for this type of case, the alternative proposed by R. Bro and his collaborators is PARAFAC2 [68,69], it has already been demonstrated that the problems of the experimental system analyzed in

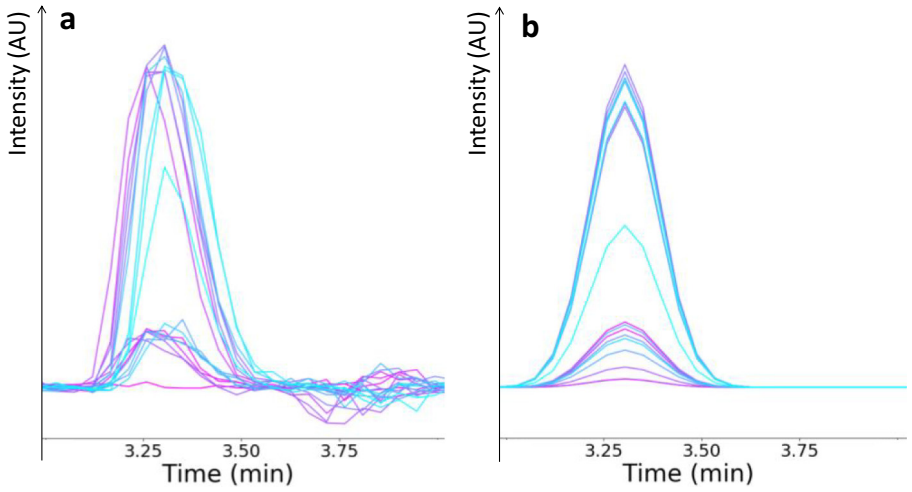

Fig. 3. Elution-time profiles retrieved by PARAFAC for DbA in test and calibration samples: (a) raw data, (b) data pre-processed with FAPV.

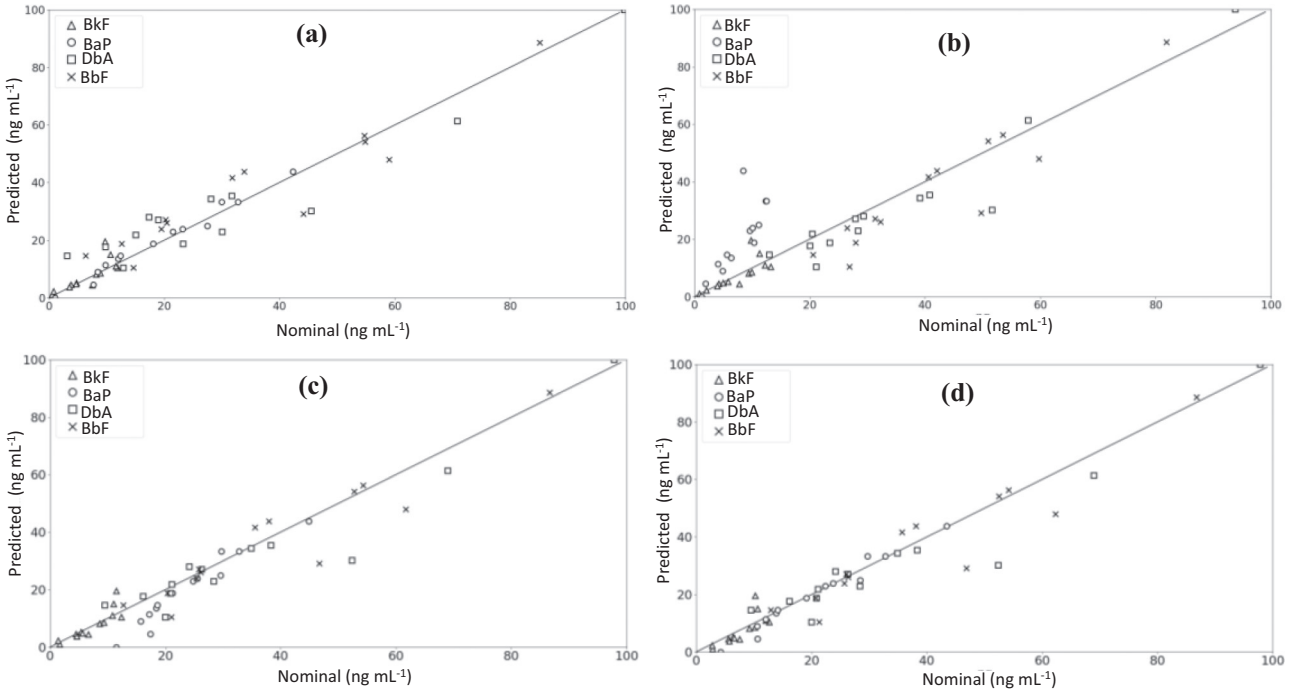

Fig. 4. Plots for predicted concentrations of the four studied PAHs as a function of the nominal values in test samples with interferents for a) raw data/MCR-ALS, b) raw data/PARAFAC, c) FAPV/MCR-ALS, and d) FAPV/PARAFAC.

this work cannot completely solve by the PARAFAC2 model [27]. In addition, the experimental data were processed with PARAFAC2 (Fig. 5 SI, Table 1 SI). As suspected, the obtained results show that this model has a lower performance than FAPV/PARAFAC (for further details, see Supporting Information).

In a second step, the data were aligned with the proposed algorithm and the transformed data were re-processed with the models. Fig. 4d shows the prediction results obtained with PARAFAC for the data aligned with the FAPV algorithm. In this case, the predictions for the four PAHs are in good agreement with the corresponding nominal values, consistently with the results that MCR-ALS finds when it processes the raw data (Fig. 4a), but with a marginal improvement. Interestingly, the MCR-ALS results obtained from data aligned with FAPV algorithm show an enhancement over the results of the same model on raw data (Fig. 4d). While it is true that the presence of interferents should not prevent the accurate quantification of analytes by means of MCR-ALS algorithm (i.e., the exploitation of the second-order advantage) [16,70,71], it has been shown that the presence of several chromatographic artifacts apart from to spectral collinearity (due to the own analytes or interferents) makes the results worse [29,72]. Consequently, the observed improvement is reasonable.

Additionally, the experimental data were also processed with the alternative MCR/COW-PARAFAC model (Fig. 5 SI, Table 1 SI), that had been mentioned in the Introduction section. The results obtained are less satisfactory than those obtained with our algorithm (for further details, see Supporting Information).

For a detailed comparison, the relative error prediction (REP) was calculated for each analyte. Table 1 shows the REP results for the analytes from the test set obtained with the models used.

The following fact emerges from Table 1: the best REPs were achieved with the PARAFAC model with the aligned data, indicating that trilinearity was restored by the FAPV algorithm. Moreover, the REPs of FAPV/PARAFAC are even a slightly better than the results proposed by MCR-ALS for the data with the same pre-processing. Finally, the comparison could turn even more towards PARAFAC if the analytical figures of merit (AFOMs) were analyzed since it has been shown that for MCR-ALS, better REPs do not correlate with better AFOMs, while REPs and AFOMs improve in accordance for PARAFAC [29,60]. Nevertheless, AFOMs in the models built from functional data are not defined, hence changes to the current formulas should be done in order to make a fair comparison with the established models.

Table 1
PARAFAC and MCR-ALS prediction results (REPs).

	PARAFAC[a]	FAPV/PARAFAC[a]	MCR-ALS[a]	FAPV/MCR-ALS[a]
BbF	45.5	7.69	10.5	9.23
BkF	32.5	1.60	4.80	4.70
BaP	14.6	3.52	9.84	8.52
DbA	15.5	3.51	9.63	8.48

[a] Six components modeled both for PARAFAC and MCR-ALS, corresponding to the four analytes (BbF, BkF, BaP, and DbA) and the two interferents (BeP and BjF). The REPs values are indicated in %.

5. Conclusions

A new algorithm (FAPV) based on functional data analysis was developed to align three-way chromatographic-fluorescence data. Unlike most of the available algorithms, the method allows performing matrix alignment when the test data matrix contains unexpected chemical interferences.

Simulated and experimental three-way chromatographic systems were analyzed to show the capability of the FAPV algorithm to restore the trilinearity in several analytical conditions. MCR-ALS and PARAFAC second-order calibration models were applied to the simulated and experimental data, both for raw data and for data aligned with the FAPV algorithm. In all cases, the PARAFAC model applied to the aligned data provided the better predictions for the analytes in samples with potential interferents. Furthermore, the low relative prediction errors obtained in the determinations of benzo [k]fluoranthene, benzo [*a*]pyrene, dibenzo[*a*,*h*]anthracene and benzo[*b*]fluoranthene (1.6 to 7.7%) demonstrate that the method is accurate, and that it can be a reliable alternative to solve problems of similar nature. It is important to remark that while PARAFAC rendered good results only when the

data were previously aligned, MCR-ALS was able to provide successful predictions of the concentration of the four studied PAHs in all real samples without a pre-process the data.

Acknowledgments

The following institutions are gratefully acknowledged for financial support: Universidad Nacional de Rosario (RCS Nº 141/18), CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2016-4423). We would like to thank the staff from the English Department (Facultad de Ciencias Bioquímicas y Farmacéuticas, UNR) for the language correction of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2018.06.041.

References

[1] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, Fast, comprehensive two-dimensional liquid chromatography, J. Chromatogr. A 1168 (2007) 3–43.
[2] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, Recent advancements in comprehensive two-dimensional separations with chemometrics, J. Chromatogr. A 1184 (2008) 341–352.
[3] M. Daszykowski, B. Walczak, Methods for the exploratory analysis of two-dimensional chromatographic signals, Talanta 83 (2011) 1088–1097.
[4] M. Vosough, M. Rashvand, H.M. Esfahani, K. Kargosha, A. Salemi, Direct analysis of six antibiotics in wastewater samples using rapid high-performance liquid chromatography coupled with diode array detector: a chemometric study towards green analytical chemistry, Talanta 135 (2015) 7–17.
[5] A. Gałuszka, Z. Migaszewski, J. Namieśnik, The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices, Trends Anal. Chem. 50 (2013) 78–84.
[6] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J. Poppi, Second and third-order multivariate calibration: data, algorithms and applications, Trends Anal. Chem. 26 (2007) 752–765.
[7] A.C. Olivieri, G.M. Escandar, A. Muñoz de la Peña, Second-order and higher-order multivariate calibration methods applied to non-multilinear data using different algorithms, Trends Anal. Chem. 30 (2011) 607–617.
[8] G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, Second- and higher order data generation and calibration: a tutorial, Anal. Chim. Acta 806 (2014) 8–26.
[9] C.J. Appellof, E.R. Davidson, Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents, Anal. Chem. 53 (1981) 2053–2056.
[10] R. Ferrer, J. Guiteras, J.L. Beltran, Development of fast-scanning fluorescence spectra as a detection system for high-performance liquid chromatography determination of polycyclic aromatic hydrocarbons in water samples, J. Chromatogr. A 779 (1997) 123–130.
[11] J.L. Beltran, J. Guiteras, R. Ferrer, Three-way multivariate calibration procedures applied to high-performance liquid chromatography coupled with fast-scanning fluorescence spectrometry detection. Determination of polycyclic aromatic hydrocarbons in water samples, Anal. Chem. 70 (1998) 1949–1955.
[12] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, Non-trilinear chromatographic time retention – fluorescence emission data coupled to chemometric algorithms for the simultaneous determination of 10 polycyclic aromatic hydrocarbons in the presence of interferences, Anal. Chem. 81 (2009) 8074–8084.
[13] R.A. Gimeno, J.L. Beltran, R.M. Marce, F. Borrull, Determination of naphthalenesulfonates in water by on-line ion-pair solid-phase extraction and ion-pair liquid chromatography with fast-scanning fluorescence detection, J. Chromatogr. A 890 (2000) 289–294.
[14] F. Cañada Cañada, J.A. Arancibia, G.M. Escandar, G.A. Ibañez, A. Espinosa Mansilla, A. Muñoz de la Peña, A.C. Olivieri, Second-order multivariate calibration procedures applied to high-performance liquid chromatography coupled to fast-scanning fluorescence detection for the determination of fluoroquinolones, J. Chromatogr. A 1216 (2009) 4868–4876.
[15] A. Mancha de Llanos, M.M. de Zan, M.J. Culzoni, A. Espinosa Mansilla, F. Cañada Cañada, A. Muñoz de la Peña, H.C. Goicoechea, Determination of marker pteridines in urine by HPLC with fluorimetric detection and second-order multivariate calibration using MCR-ALS, Anal. Bioanal. Chem. 399 (2011) 2123–2135.
[16] M.J. Culzoni, A. Mancha de Llanos, M.M. de Zan, A. Espinosa Mansilla, F. Cañada Cañada, A. Muñoz de la Peña, H.C. Goicoechea, Enhanced MCR-ALS modeling of HPLC with fast scan fluorimetric detection second-order data for quantitation of metabolic disorder marker pteridines in urine, Talanta 85 (2011) 2368–2374.
[17] M. Daszykowsky, B.T. Walczak, Use and abuse of chemometrics in chromatography, Trends Anal. Chem. 25 (2006) 1081–1096.
[18] A. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution, Anal. Chim. Acta 500 (2003) 195–210.
[19] H. Parastar, R. Tauler, Multivariate curve resolution and hyphenated and multi-dimensional chromatographic measurements: a new insight to address current chromatographic challenges, Anal. Chem. 86 (2014) 286–297.
[20] A.C. Olivieri, G.M. Escandar, A road map for multi-way calibration models, Analyst 142 (2017) 2862–2873.
[21] A.R. Tôrres, S.G. Junior, W.D. Fragoso, Multivariate control charts for monitoring captopril stability, Microchem. J. 118 (2015) 259–265.
[22] S. De Luca, M. De Filippis, R. Bucci, A.D. Magrì, A.L. Magrì, F. Marini, Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics, Microchem. J. 129 (2016) 348–361.
[23] C.M. Monzón, C.M. Teglia, M.R. Delfino, H.C. Goicoechea, Multiway calibration strategy with chromatographic data exploiting the second-order advantage for quantitation of three antidiabetic and three antihypertensive drugs in serum samples, Microchem. J. 136 (2018) 185–192.
[24] A.C. Olivieri, G.M. Escandar, Practical Three-Way Calibration, Ed. Elsevier, MA, USA, 2014.
[25] J.M. Amigo, T. Skov, J. Coello, S. Maspoch, R. Bro, Solving GC-MS problems with PARAFAC2, Trends Anal. Chem. 27 (8) (2008) 714–725.
[26] J.M. Amigo, M.J. Popielarz, R.M. Callejón, M.L. Morales, A.M. Troncoso, M.A. Petersen, T.B. Toldam-Andersen, Comprehensive analysis of chromatographic data by using PARAFAC2 and principal components analysis, J. Chromatogr. A 1217 (2010) 4422–4429.
[27] S.A. Bortolato, A.C. Olivieri, Chemometric processing of second-order liquid chromatographic data with UV-vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2, Anal. Chim. Acta 842 (2014) 11–19.
[28] J.M. Amigo, T. Skov, R. Bro, ChroMATHography: solving chromatographic issues with mathematical models and intuitive graphics, Chem. Rev. 110 (2010) 4582–4605.
[29] R.B. Pellegrino Vidal, F. Allegrini, A.C. Olivieri, The effect of constraints on the analytical figures of merit achieved by extended multivariate curve resolution-alternating least-squares, Anal. Chim. Acta 1003 (2018) 10–15.
[30] X. Liu, S.D. Sidiropoulos, Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays, IEEE Trans. Signal Process. 49 (2001) 2074–2086.
[31] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, 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. Intell. Lab. Syst. 101 (2010) 30–37.
[32] N.P. Vest Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning on single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J. Chromatogr. A 805 (1998) 17–35.
[33] T. Skov, F. van den Berg, G. Tomasi, R. Bro, Automated alignment of chromatographic data, J. Chemom. 20 (2006) 484–497.
[34] T.G. Bloemberg, J. Gerretzen, A. Lunshof, R. Wehrens, L.M. Buydens, Warping methods for spectroscopic and chromatographic signal alignment: a tutorial, Anal. Chim. Acta 781 (2013) 14–32.
[35] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Standardization of second-order chromatographic: spectroscopic data for optimum chemical analysis, Anal. Chem. 70 (1998) 218–225.
[36] E. Comas, R.A. Gimeno, J. Ferré, R.M. Marcé, F. Borrull, F.X. Rius, Quantification from highly drifted and overlapped chromatographic peaks using second-order calibration methods, J. Chromatogr. A 1035 (2004) 195–202.
[37] W. Jiang, Z.-M. Zhang, Y.H. Yun, D.-J. Zhan, Y.-B. Zheng, Y.-Z. Liang, Z.Y. Yang, L. Yu, Comparisons of five algorithms for chromatogram alignment, Chromatographia 76 (2013) 1067–1078.
[38] Y.-B. Zheng, Z.-M. Zhang, Y.-Z. Liang, D.-J. Zhan, J.-H. Huang, Y.-H. Yun, H.-L. Xie, Application of fast Fourier transform cross-correlation and mass spectrometry data for accurate alignment of chromatograms, J. Chromatogr. A 1286 (2013) 175–182.
[39] H. Parastar, N. Akvan, Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements, Anal. Chim. Acta 816 (2014) 18–27.
[40] J. Ramsay, B. Silverman, Functional Data Analysis, Springer, New York, USA, 1997.
[41] J. Ramsay, B. Silverman, Applied Functional Data Analysis, Springer, New York, USA, 2002.
[42] W. Saeys, B. De Ketelaere, P. Darius, Potential applications of functional data analysis in chemometrics, J. Chemom. 22 (2008) 335–344.
[43] A.M. Aguilera, M. Escabias, M.J. Valderrama, M.C. Aguilera-Morillo, Functional analysis of chemometric data, Open J. Stat. 03 (05) (2013) 334–343.
[44] R. Burfield, C. Neumann, C.P. Saunders, Chemom. Intell. Lab. Syst. 149 (2015) 97–106.
[45] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Peak alignment using wavelet pattern matching and differential evolution, Talanta 83 (2011) 1108–1117.
[46] W.E. Wallace, A. Srivastava, K.H. Telu, Y. Simón-Manso, Pairwise alignment of chromatograms using an extended fisher–Rao metric, Anal. Chim. Acta 841 (2014) 10–16.
[47] T. Wenzl, R. Simon, J. Kleiner, E. Anklam, Analytical methods for polycyclic aromatic hydrocarbons (PAHs) in food and the environment needed for new food legislation in the European Union, Trends Anal. Chem. 25 (2006) 716–725.
[48] SCF, Scientific Committee on Food, Opinion of the Scientific Committee on Food on the Risks to Human Health of Polycyclic Aromatic Hydrocarbons in Food, 4 European Commission, Brussels, 2002 December 2002.
[49] European Commission Regulation (EC) No 1881/2006, Off. J. Eur. Communities L364 (2006) 5.

[50] European Food Safety Authority (EFSA), Polycyclic aromatic hydrocarbons in food, scientific opinion of the panel on contaminants in the food chain (adopted on 9 June 2008), EFSA J. 724 (2008) 1–114 http://www.efsa.europa.eu/EFSA/efsa locale-1178620753812 1211902034842.htm.
[51] European Commission Regulation (EC) 835/2011, Off. J. Eur. Communities 215 (2011) 4.
[52] S. Houlgrave, G.M. Laporte, J.C. Stephens, J.L. Wilson, The classification of inkjet inks using AccuTOFTM DARTTM (direct analysis in real time) mass spectrometry—a preliminary study, J. Forensic Sci. 58 (2013) 813–821.
[53] S.K. Guharay, G.S. Thakur, F.J. Goodman, S.L. Rosen, D. Houser, Analysis of non-stationary dynamics in the financial system, Econ. Lett. 121 (3) (2013) 454–457.
[54] B.K. Alsberg, Representation of spectra by continuous functions, J. Chemom. 7 (1993) 177–193.
[55] W. Saeys, B. De Ketelaere, P. Darius, Potential applications of functional data analysis in chemometrics, J. Chemom. 22 (2008) 335–344.
[56] G. Whaba, Spline Models for Observational Data, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.
[57] C. De Boor, A Practical Guide to Splines (Applied Mathematical Sciences 27), Revised Edition, Springer, New York, 2001.
[58] G. Nürnberger, J.W. Schmidt, G. Walz (Eds.), Multivariate Approximation and Splines (ISNM International Series of Numerical Mathematics 125), Springer, Basel, 1997.
[59] G. Micula, S. Micula, Handbook of Spline (Mathematics and its Applications 462), Springer, Netherlands, 1999.
[60] Data Handling in Science and Technology, in: A. Muñoz de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivier (Eds.), Fundamentals and Analytical Applications of Multiway Calibration, Vol. 29 Elsevier, Amsterdam, 2015Chapter 13.
[61] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425–1432.
[62] L. Valderrama, R.P. Goncalves, P.H. Marc, D.N. Rutledge, P. Valderrama, Independent components analysis as a means to have initial estimates for multivariate curve resolution-alternating least squares, J. Adv. Res. 7 (2016) 795–802.
[63] R. Bro, PARAFAC, Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[64] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemom. Intell. Lab. Syst. 76 (2005) 101–110.
[65] https://www.python.org/download/releases/3.0/.
[66] http://www.models.life.ku.dk/nwaytoolbox.
[67] https://mcrals.wordpress.com/download/mcr-als-2-0-toolbox/.
[68] H.A.L. Kiers, J.M.F. ten Berge, R. Bro, PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemom. 13 (1999) 275–294.
[69] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2—Part II. Modeling chromatographic data with retention time shifts, J. Chemom. 13 (1999) 295–309.
[70] R.Q. Aucelio, G.M. Escandar, High-performance liquid chromatography with fast-scanning fluorescence detection and multivariate curve resolution for the efficient determination of galantamine and its main metabolites in serum, Anal. Chim. Acta 740 (2012) 27–35.
[71] M. Vosough, H.M. Esfahani, Fast HPLC-DAD quantification procedure for selected sulfonamids, metronidazole and chloramphenicol in wastewaters using second-order calibration based on MCR-ALS, Talanta 113 (2013) 68–75.
[72] H. Abdollahi, R. Tauler, Uniqueness and rotation ambiguities in Multivariate Curve Resolution methods, Chemom. Intell. Lab. Syst. 108 (2011) 100–111.