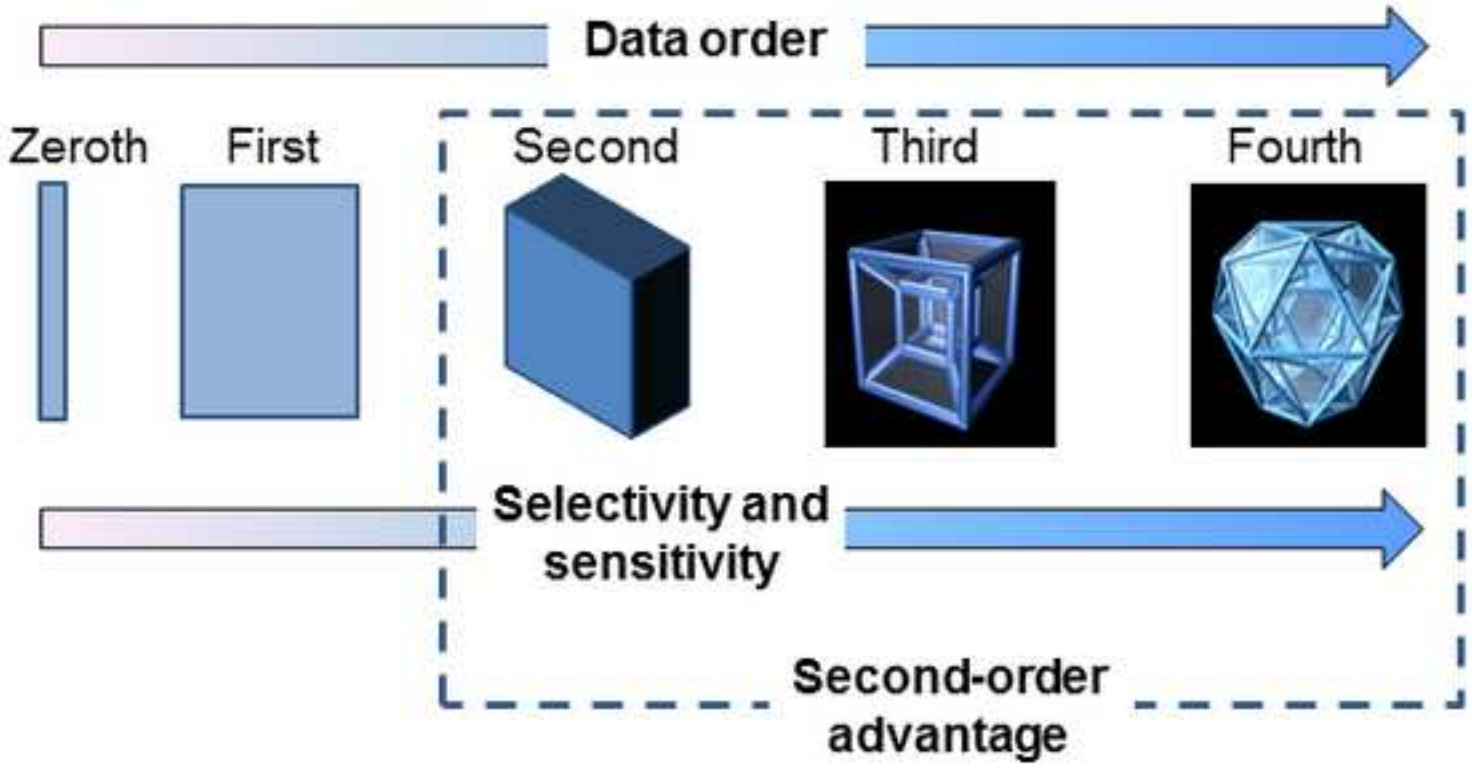
*Graphical Abstract



Highlights

Title: "Second- and higher-order data generation and calibration: A tutorial"
Authors: Graciela M. Escandar, Héctor C. Goicoechea, Arsenio Muñoz de la Peña, Alejandro C. Olivieri

► Second- and third-order (multi-way) data and algorithms are reviewed. ► Suitable examples of different complexity are provided. ► The advantages of multi-way calibration are illustrated. ► Multi-way analytical figures of merit are discussed.

Second- and higher-order data generation and calibration: A tutorial

Graciela M. Escandar,[a] Héctor C. Goicoechea,[b] Arsenio Muñoz de la Peña,[c] Alejandro C. Olivieri[a,*]

[a] *Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química de Rosario (IQUIR-CONICET), Suipacha 531, Rosario, S2002LRK, Argentina*

[b] *Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional de Litoral, Santa Fe, S3000ZAA, Argentina*

[c] *Department of Analytical Chemistry, University of Extremadura, 06006, Badajoz, Spain*

[*] To whom correspondence should be addressed. E-mail: olivieri@iquir-conicet.gov.ar. TE/Fax: +54-341-4372704

Abstract

An introduction to multi-way calibration based on second- and higher-order data generation and processing is provided, with emphasis on practical experimental aspects. After a discussion concerning a proper nomenclature scheme, a suitable classification of the obtainable data, and the general features of the available algorithms and their underlying models, a series of examples is discussed in detail, with the purpose of illustrating the great potentiality of the field for the analytical community. Emphasis is directed towards the most popular multi-way data, i.e., second-order or matrix data, which can be conveniently measured in a variety of instruments. Third-order data are being increasingly studied and are also discussed, along with the less explored field of fourth-order data. The estimation of figures of merit, which analysts need to report during method development, is now sufficiently mature to be provided for the general audience.

Keywords: Multi-way calibration; Parallel factor analysis; Multivariate curve resolution; Partial least-squares

41 # Contents

63
64
65

1. Introduction

The variety of second- and higher-order instrumental data which are being produced by modern instruments, and the subsequent enhancement in analytical properties which is obtained by processing this kind of data, have made multi-way analysis a subject of high interest for the analytical community. Multi-way analysis has produced a significant impact on the development of analytical methods, especially for the quantitation of analytes of interest in complex matrices, such as those found in environmental, biological and food samples, among others [1-3]. The subject started with the pioneering work of Kowalski [4], among others, and specifically with the first experimental realization of a multi-way calibration of complex samples in the presence of interferents in 1978 [5]. A literature search reveals that more than 360 papers have been published in the last five years on the subject, including interesting comprehensive reviews [6-9].

The emphasis directed to processing complex data has been accompanied by the development of a diversity of mathematical algorithms based on various data models, which are available to analytical chemists for the convenient study of this body of information. A casual reader may be confused by the variety of experimental and theoretical developments which have been flourishing in this area in recent years. This provides the main motivation for the present tutorial, i.e., to consistently classify and compare the various participants of the scene: different types of second- and higher-order instrumental data and the various available models and algorithms, as well as the selection of the most convenient ones for a given experimental application. In this context, several examples extracted from the literature are discussed, showing details regarding the data generation and the advantages and disadvantages of the application of

some specific models and algorithms. In addition, a discussion about recent developments concerning analytical figures of merit is included.

An important message is to be left from the present tutorial to the analytical community, which is of paramount importance and is perhaps the major legacy of multi-way calibration. It may be better understood using an example: suppose one wish to quantitate an analyte in a sample from absorbance measurements at a single wavelength, but there are other constituents whose responses overlap with that of the analyte. It is well-known that this is not possible in this univariate context. If spectra were measured at many wavelengths (i.e., vectorial absorbance data), it would be possible to accomplish the goal, but only on the condition that a calibration model is produced from a large set of samples, which contain all varying concentrations of possible future sample constituents. However, if matrix data (or higher-order arrays) are recorded, and sufficient selectivity is present in the various data modes, it is in principle possible to predict analyte concentrations in any future sample, no matter how many signal-overlapping constituents this sample contains, and having calibrated with *pure analyte standards*. This is the experimental realization of the main advantage offered by second- and third-order calibration, i.e., the so-called 'second-order advantage'.

The tutorial is organized as follows: an introduction to multi-way nomenclature is first provided, along with a discussion on data properties, models and algorithms, followed by a series of examples concerning the most explored data for multi-way calibration, such as luminescence matrix spectroscopy and chromatography with multivariate detection. A final section is devoted to the latest developments in the estimation of multi-way analytical figures of merit.

2. Nomenclature

In this section the usual nomenclature concerning sample constituents, algorithms, methods and data is presented. It should help the general reader to follow the remaining of the tutorial.

It is first important to classify the different constituents which may be present in the various sample types which are normally studied during multi-way calibration. Constituents present in the samples employed for calibration and validation are regularly called 'expected', i.e. they are included in these sets because they are expected to be present in future samples. On the other hand, constituents which are only present in the unknown samples are called 'unexpected', and also 'potential interferents'. The expected constituents can be further sub-classified into 'calibrated' and 'uncalibrated': the former are those for which calibration concentrations are known, whereas the latter are constituents producing measurable signals but whose concentrations are not known [1,10].

The so-called potential interferents, however, will not always generate an interference, leading to a systematic error in the analyte quantitation [11]. Depending on the measured instrumental signals and calibration methodology, the interference may only remain as potential. In first-order calibration, for example, unexpected constituents are usually true interferences. However, in higher-order calibration achieving the second-order advantage, the signal from the unexpected constituents can be modelled and mathematically removed, in such a way that their effect is negligible [12].

As a general note, we prefer to refer to chemical sample *constituents* rather than to *components*, because the latter word may imply abstract linear combinations of real constituents for some methodologies.

It is also important to distinguish among *technique*, *method*, *model*, and *algorithm*, terms which are sometimes interchanged. In the context of calibration, an analytical technique is a procedure used to determine an analyte concentration, whereas an analytical method is more specific concerning the sample and measuring conditions. A model is a description of the data properties, and an algorithm is a detailed set of instructions for accomplishing a computational task. Therefore, the specific mathematical procedures for processing second- and higher-order data are all algorithms. They allow for data analysis based on a certain model, i.e., on certain assumptions concerning the properties of the data. Different algorithms may apply to the same model. The term *method* is employed in a more general sense and sometimes replaces the term *algorithm* [1].

Table 1 shows the natural progression from the simplest zeroth-order to multivariate data. Zeroth-order corresponds to instruments producing a single response per sample (a zeroth-order tensor), such as the absorbance at a single wavelength or the reading of an ion-selective electrode. First-order data for a given sample are arranged as a vector or first-order tensor, such as spectra: UV-visible spectrophotometry, spectrofluorimetry, infrared, near infrared (NIR), nuclear magnetic resonance, electrochemical scans (voltammograms, chrono-amperograms), among others. On the other hand, matrix data for a single sample are considered to be second-order. They can be recorded in two ways: (1) using a single instrument, such as a spectrofluorimeter registering excitation-emission matrices (EEMs) or a diode-array spectrophotometer following the kinetics of a chemical reaction, or (2) coupling two 'hyphenated' first-order instruments, as in tandem gas chromatography-mass spectrometry (GC-MS), GC-GC, MS-MS, etc. When second-order data for a group of samples are joined into a single, three-dimensional array, the resulting object is known as three-way array and these data are usually known

as three-way data. Finally, introducing an extra mode in the data leads to higher-order data, in which case the mathematical object obtained by grouping third-order data for several samples into the fourth direction, for example, is known as a four-way array. Examples of four-way arrays are those obtained by following the kinetics of an excitation-emission (EEM) fluorescence data matrix, or by hyphenating three instruments, e.g. two-dimensional liquid chromatography with diode array detection (LC-LC-DAD).

Notice that two alternative nomenclatures are employed for describing data and the associated calibrations. One could refer to either second-order calibration or three-way calibration; the former expression focuses on the number of modes of a single sample (two modes, second-order data) whereas the latter on the number of modes of a sample set (three modes, three-way data). The analytical community is used to the order-based nomenclature, probably because of the expression 'second-order advantage'. However, the chemometric community prefers the way-based nomenclature. In any case, it should be clear from the context which is the number of working modes and the type of multi-way calibration being discussed in each of the tutorial sections.

3. Models and algorithms

There are many algorithms available for processing multi-way data, but a few underlying models on which they are based. The choice of a particular model and algorithm should be primarily guided by the properties of the data, in the sense that the model should match the data properties. Below we summarize the most popular multi-way algorithms and their underlying models, with emphasis on which type of data can be analyzed by them.

3.1. Second-order data

An important property of second-order data is the trilinearity. A group of measured data matrices for a set of samples can in principle be arranged into a three-way data array, as shown in Fig. 1A. The array is considered to be trilinear if its elements can be reasonably fit to the following expression:

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + e_{ijk} \quad (1)$$

where a_{in} represents the relative concentration (also called *score*) of a given constituent n in the i-th sample, b_{jn} and c_{kn} are the intensities in both of the instrumental modes j and k, respectively (also called *loadings*) and e_{ijk} collects the fitting errors. Summation in equation (1) implies that the individual constituent signals are additive. Usually the loadings are normalized to unit length, and collected into the loading matrices **B** and **C**, of size $J \times N$ and $K \times N$ respectively (Fig. 2), and the scores are collected into the score matrix **A** (size $I \times N$). Strictly speaking, an array is trilinear when the number of trilinear components (N) required to describe the data through equation (1) is small, and is ideally equal to the number of chemical constituents producing measurable signals. In the remainder of this work we will simply refer to such arrays as trilinear. In principle, a non-trilinear array can also be described using equation (1), but will require a value of N significantly larger than the actual number of responsive constituents.

Equation (1) represents the trilinear model which is the basis of trilinear decomposition algorithms. The fitting of a trilinear three-way array to this model often provides unique solutions, meaning that there is a single set of **A**, **B** and **C** matrices whose elements satisfy equation (1) to a reasonable degree. The uniqueness of the decomposition is the natural basis of the second-order advantage, because the profiles

211 contained in matrices **B** and **C** are proportional to the true instrumental profiles of each
212 pure sample constituent in each data mode (except for a scaling factor), and the scores
213 contained in **A** are proportional to the pure constituent concentrations, as if the
214 constituents were physically separated from the sample. Usually the data for the
215 calibration samples and each test sample are joined into a single three-way array. The
216 information provided by the constituent scores in the calibration samples can be used in
217 the context of calibration to build a plot of analyte scores vs. nominal analyte
218 concentrations. The analyte concentration in the unknown sample is then predicted by
219 interpolating its score in the fitted line. Similar calibration plots are also employed in
220 the context of first-order multivariate calibration [13,14].

221 ** Insert Fig. 1 **

222 Algorithms based on the trilinear model are thus useful for multi-way calibration
223 from trilinear three-way data. One of the most employed trilinear algorithms is parallel
224 factor analysis (PARAFAC) [15-18]. Although there are various versions of PARAFAC
225 for fitting equation (1) [15,19], we herein refer to the one which uses an alternating
226 least-squares (ALS) procedure. PARAFAC has become the algorithm of choice for
227 calibration based on trilinear three-way data analysis, due to its efficiency, robustness,
228 ability to process multiple samples, and availability of a variety of constraints to be
229 applied during the fit, which ensures reaching physically interpretable results.
230 Additional algorithms based on trilinear modelling of three-way arrays are self-
231 weighted alternating trilinear decomposition (SWATLD) [20] and penalized alternating
232 trilinear decomposition (APTLD) [21]. Other less employed algorithms in this context
233 are generalized rank annihilation (GRAM) [22], direct trilinear decomposition (DTLD)
234 [23] and bilinear least-squares (BLLS) [24], either because the use single calibration
235 standards (GRAM. DTLD) or do not achieve the second-order advantage (BLLS).

One common cause by which a three-way array deviates from trilinearity is the presence of profiles of the constituents in one of the data modes which change from sample to sample. This is typical of chromatographic-spectral data, because the spectra can be reasonably expected to be constant for a given constituent in different samples, but the elution time profiles are likely to change due to uncontrolled irreproducibility in chromatographic runs, particularly in liquid chromatography. This may also happen for other instrumental modes: pH gradients may show differences from run to run, kinetic profiles may differ because of uncontrolled temperature changes, and even spectra may change if interactions occur with the sample background. The mode with low reproducibility across samples is said to be the trilinearity-breaking mode, because if the profiles in this mode for all constituents were equal in all samples, the data would be trilinear [25]. In the latter case PARAFAC could be conveniently applied; however if lack of reproducibility indeed occurs, calibration using trilinear models would not be recommended.

** Insert Fig. 2 **

A three-way array of non-trilinear data of this type can be unfolded into an augmented matrix, as shown in Fig. 1B. If the augmentation mode is chosen to be the elution time mode (i.e., the trilinearity breaking mode in this particular case), then the resulting augmented matrix $\mathbf{X}_{\text{aug}}$ is of size $IJ \times K$, if J is the number of data points in the elution time mode for each of the I data matrices involved in $\mathbf{X}_{\text{aug}}$. This latter matrix is bilinear, and thus its elements can be fitted to a bilinear model:

$$x_{\text{aug},mk} = \sum_{n=1}^{N} b_{\text{aug},mn} c_{kn} + e_{mk} \qquad (2)$$

where $x_{\text{aug},mk}$ is a generic element of the augmented matrix $\mathbf{X}_{\text{aug}}$ (the index m runs from 1 to IJ), $b_{\text{aug},mn}$ is an element of the augmented profile for constituent n in the augmented direction, c_{kn} an element of the profile in the direction for constituent n (this profile is

261 common to all samples), and e_{mk} is an element of the matrix of model residuals. The
262 elements $b_{\mathrm{aug},mn}$ and c_{kn} are usually collected into a matrix $\mathbf{B}_{\mathrm{aug}}$ of augmented profiles
263 and a matrix **C** of normalized profiles.

264 In contrast to the trilinear decomposition of a trilinear three-way array into the three
265 matrices **A**, **B** and **C** (see above), the bilinear decomposition of a bilinear augmented
266 matrix into $\mathbf{B}_{\mathrm{aug}}$ and **C** is not unique. However, equation (2) can often be solved in terms
267 of true constituent profiles if a proper set of initial values for $\mathbf{B}_{\mathrm{aug}}$ or **C** are found in
268 order to start the bilinear decomposition, and suitable constraints are applied during the
269 fit, to limit the number of possible solutions, and to ensure they are chemically
270 reasonable. A popular algorithm accomplishing this goal is multivariate curve resolution
271 coupled to ALS (MCR-ALS) [26-28]. MCR-ALS is usually employed for quantitative
272 analytical purposes in the so-called extended mode [29], which decomposes an
273 augmented data matrix created from calibration and unknown matrices. Many different
274 constraints are available in MCR-ALS, while initial values can be efficiently estimated
275 by a variety of methods [30,31].

276 Once a reasonable solution has been found by MCR-ALS by proper initialization and
277 natural constraints, pure constituent concentration information is contained in the areas
278 under each of the $\mathbf{B}_{\mathrm{aug}}$ profiles for each sub-matrix corresponding to each of the
279 participating samples (Fig. 3). This allows one to achieve the second-order advantage,
280 because information on potential interferents is efficiently separated from those for the
281 analytes. A univariate plot can be built by regressing the areas under the profiles for a
282 specific analyte vs. its nominal concentration in the calibration samples. The analyte can
283 then be predicted in unknown samples by interpolating, in the univariate plot, the area
284 of the $\mathbf{B}_{\mathrm{aug}}$ profile corresponding to that particular analyte in the unknown sample.

285 ** Insert Fig. 3 **

PARAFAC2 is a variant of PARAFAC which allows for some profile changes in one of the data modes [32,33]. It includes a mathematical constraint on the profiles in the latter mode (namely that the cross-products of component profiles should be constant across samples), but lesser chemically natural constraints in comparison with MCR-ALS (non-negativity, unimodality). It is thus limited to similar changes in peak positions for various constituents in different samples [25,32,34,35].

Finally, other non-trilinear data occur which are intrinsically more complex than the two types described above. For example, mass spectrometric (MS-MS) matrices are non-bilinear, because the mass spectrum of a fragment depends on its position on the mother spectrum. Thus these data cannot yield trilinear three-way arrays, as in other second-order data where the phenomena described by the instrumental data modes are mutually dependent.

More flexible models are required to cope with these data, such as non-bilinear rank annihilation (NBRA) [36], and unfolded and multi-way partial least-squares (U-PLS and N-PLS) [37,38]. In the case of U-PLS, the calibration data matrices are unfolded, and PLS is applied using a suitable number of latent variables (Fig. 1C). This provides greater flexibility. The second-order advantage is achieved after calibration, submitting the test sample in matrix form (Fig. 1C) to a procedure called residual bilinearization (RBL) [39,40], which separates the portion of the signal which can be explained by calibration from the contribution of the potential interferents (see Fig. 4). This gives rise to the useful U-PLS/RBL and N-PLS/RBL methodologies.

** Insert Fig. 4 **

It should be noticed that MCR-ALS and U-PLS/RBL, being more flexible than PARAFAC, in addition to non-linear data, can also be applied to trilinear data. In fact, trilinearity is an additional constraint that can be applied during MCR-ALS execution.

3.2. Third-order data and beyond

The models followed by third-order data are derived by analogy with second-order data, namely: (1) quadri-linear, which is the logical sequel after trilinear, (2) a bilinear augmented matrix for non-quadrilinear data with quadri-linearity breaking modes, and (3) latent variable models for other, more complex, non-quadrilinear data.

With third-order data for a group of samples, a four-way data array can be constructed, in which case the simplest model is the quadri-linear one, or multi-linear in general. Multi-linearity can be defined by extension of equation (1), and a four-way data array is quadri-linear if its elements comply with:

$$x_{ijkl} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} d_{ln} + e_{ijkl} \qquad (3)$$

where all symbols are as in equation (1), with d_{ln} describing the profile in the fourth data mode. As for the decomposition of a trilinear three-way array, uniqueness is often achieved in the decomposition of a quadri-linear four-way array.

If the data are quadri-linear then multi-way PARAFAC would be adequate, because it can be applied to data with any number of ways, employs multiple calibration samples, and includes useful constraints during the fitting phase. Complementary algorithms such as alternating penalty quadri-linear decomposition (APQLD) [41] and alternating weighted residual constraint quadri-linear decomposition (AWRCQLD) [42] are also available for processing quadri-linear data. Excitation-emission luminescence matrices measured as a function of reaction or decay time are prime examples of quadri-linear data to which these algorithms can be confidently applied (see below).

As with second-order data, a common cause of multi-linearity loss is the variation of constituent profiles in one particular mode from sample to sample, a phenomenon

usually observed in multi-way data of chromatographic origin. To be able to apply MCR-ALS to these data, it is necessary to first unfold the original third-order data into matrices, so that they could then be arranged into a bilinear augmented matrix. A typical case is two-dimensional chromatography with spectral detection, in which the data modes are two elution times and spectra. Since the former two are potentially quadri-linearity breaking, it is necessary to unfold the two chromatographic elution time modes into a single one, transforming the original three-way arrays into matrices (Fig. 5). This approach has been taken when MCR-ALS is applied to third-order chromatographic data of this kind [43,44].

** Insert Fig. 5 **

More complex data deviating from quadri-linearity by other causes can be analyzed using latent structures, coupled to residual trilinearization (RTL) to achieve the second-order advantage, as the natural extension of RBL to three data modes. This gives rise to unfolded- and multi-way PLS combined with RTL (U-PLS/RTL, N-PLS/RTL) [45,46]. The hybrid technique trilinear least-squares/RTL (TLLS/RTL) is also available but less flexible than the former ones [46]. For one way further, the combination of U-PLS with residual quadri-linearization (U-PLS/RQL) has been recently reported, and is available for processing five-way data arrays [47].

3.3. Software

Software for multi-way analysis is freely available on the Internet, in the form of MATLAB codes [48], including several useful graphical user interfaces (GUI) [49-52]. Table 2 shows a variety of free MATLAB programs for multi-way data processing.

3.4. Summary

To summarize the above discussion on models and algorithms, multi-way data should be first classified by the analyst according to their properties. There are three main data types in this regard: (1) multi-linear data, (2) non multi-linear data but unfoldable into a bilinear data matrix, and (3) other non multi-linear data which cannot be unfolded into a bilinear data matrix. This classification gives rise to the three main data models on which data processing tools should resort: (1) the multi-linear model, (2) the bilinear augmented matrix, and (3) latent variable models.

Once this classification is made, the subsequent task is to find a suitable data processing algorithm. The recommendation in this regard is to resort to an algorithm fulfilling the following requirements: (1) be based on the model the data are assumed to follow, and (2) allow one to perform calibration using multiple standard samples. For the three data types mentioned above, then, three algorithm types can be recommended: (1) PARAFAC and its TLD/QLD variants for multi-linear data, (2) MCR-ALS for non-multilinear data which are unfoldable into a bilinear augmented matrix, and (3) PLS/RML for the remaining non-multilinear data. All these algorithms allow one to calibrate using as many standards as desired, and are thus preferable over single-calibration-sample algorithms of any kind.

4. Second-order data generation and examples

Luminescence (fluorescence, phosphorescence, chemiluminescence) and other spectroscopic data, such as UV-visible spectrophotometric data, are prone to be affected by spectral interferences and, therefore, the coupling with multivariate calibration is an excellent alternative to gain selectivity in this area by mathematical means.

To obtain second-order data from UV-visible spectrophotometric measurements, they are usually combined with kinetic or pH-gradient experiments: the absorbance spectra constitute one data mode and the reaction time (or pH) the other one. Some application examples can be found in the literature regarding these data types [53,54]. A more general way is by multi-wavelength (diode array) detection in chromatography, which is discussed in a separate section.

Among luminescence signals, fluorescence is by far the most frequently employed, although examples on the use of time-resolved phosphorescence at room temperature have been also reported [55]. Due to the important advantages of methods based on fluorescence emission, they are profusely used with analytical purposes for the determination of many analytes which display either native or induced fluorescence emission. Among these advantages, the following can be mentioned: high sensitivity (which is intrinsic to the fluorescence signals) and low or null consumption of organic solvents (allowing one to work under green-chemistry principles). On the other hand, if selectivity is an issue, it can be improved by coupling to second-order calibration.

Examples of luminescence second-order data are excitation-emission fluorescence matrices (EEFM) [56], reaction time (kinetic)-excitation or emission wavelength matrices [57], and chromatography with spectral luminescence detection. In what follows, we discuss literature examples where second-order calibration is applied to luminescence data, with emphasis on different data sets and the specific algorithms that can be applied in each case. The overall idea of the following sections is to discuss examples showing a variety of experimental possibilities which can be found in practice. Although the general rule (see above) is that luminescence data are trilinear and hence PARAFAC (or its TLD variants) is the algorithm of choice for processing them, there are some instances in which this is not the case. Hence the discussed

examples should warn analytical chemists to be cautious in what concerns algorithm selection.

4.1. Trilinear luminescence data

As discussed before, an individual EEFM is an example of bilinear data, and in many cases, a set of EEFMs can be arranged into a trilinear three-way array, in which case trilinear decomposition methods can be applied to retrieve the excitation and the emission spectra for the participating fluorescent constituents. However, RBL and MCR-ALS can also be conveniently employed to process these data, because they are more general and include trilinearity as a specific case.

The most popular trilinear algorithm is probably PARAFAC (see above), which in addition to being friendly for analysts, has the advantage of providing important physical interpretation, in the form of the fluorescence profiles of the constituents under study. Many literature examples using PARAFAC to resolve systems based on EEFMs can be cited [56]. Only as a recent example, we mention the spectrofluorimetric determination, in environmental water samples, of the herbicide bentazone (BTZ), one of the most employed herbicides in countries of profuse agriculture [58].

The EEFMs were measured in the presence of methyl-β-cyclodextrin, which produces a significant increasing of the bentazone fluorescence intensity. The experiment involved the following steps, which are the usual ones in the development and validation of multi-way analytical procedures: (1) measurement of EEFMs of standard solutions of BTZ of known concentrations (calibration step), (2) measurement of EEFMs of solutions of BTZ different from those for calibration, and verification of the predictive ability in the absence of interferents (validation step), (3) measurement of EEFMs of artificial samples prepared by adding to BTZ solutions different amounts of

seven common agrochemicals (fuberidazole, thiabendazole, dichlorophene, carbendazim, carbaryl, carbofuran and 1-naphthaleneacetic acid) acting as potential spectral interferences (these samples are used for testing the predictive ability in the presence of unexpected constituents), and (4) measurement of EEFMs of real samples containing the investigated analyte. All the studied samples could be satisfactorily resolved by PARAFAC, because the systems retained their trilinearity even in the presence of interfering agents in both synthetic and real samples.

The wavelength ranges for calibration, validation and real samples (390-470 nm and 240-372 nm for emission and excitation, respectively) were selected in order to cover the regions of fluorescence emission and excitation of BTZ. On the other hand, because the constituents selected as interferents display intense fluorescence signals and were added at high concentrations in order to check for potentially unfavorable situations, a saturation of the fluorescence signal was observed in a wide spectral region of the EEFMs for the test samples. Therefore, a restricted region, where the instrumental signal could be correctly measured, was selected for data processing (390-470 nm and 340-372 nm for emission and excitation, respectively).

The number of responsive components was selected applying three procedures: (1) taking into account the results of the core consistency analysis, (2) through the analysis of PARAFAC residuals, and (3) considering that the addition of subsequent components did not generate repeated profiles (we refer here to 'components' because the algorithm is blind to whether they are real chemical constituents or not). It is important to note that the core consistency analysis is a tool based on some data structural assumptions and may fail in certain circumstances [59]. In other words, in addition to the core consistency analysis, other tools to provide evidence of the appropriate number of components should be considered [56].

The results obtained by the three procedures for determining the number of components required by PARAFAC were consistent and established that this number in the validation and test samples was two. In the validation samples, the number was ascribed to the analyte and to a background signal. In the test samples, the number was assigned to BTZ and to a combined signal corresponding to the interferents (taking into account that in this restricted wavelength zone the background signal is negligible and does not significantly contribute to the total signal). In this latter case, PARAFAC was not able to discern between the profiles of each foreign constituent, and retrieved the interference profiles as a single unexpected constituent. However, this fact did not preclude the obtainment of good analytical results in these samples. In spiked natural water samples, the number of PARAFAC factors was two or three, depending on the analyzed matrix.

As an example of the results furnished by PARAFAC, Fig. 6 shows the scores (relative concentrations) and loadings (profiles of constituents in both modes) retrieved by PARAFAC for a real assayed sample containing BTZ.

** Insert Fig. 6 **

While the loadings allowed to identify each chemical constituent of the sample, the scores corresponding to the calibrated constituent are used for building a univariate calibration curve. In the studied case, constituent 1 (blue line) was assigned to BTZ. In fact, the agreement between the normalized spectra of a standard BTZ solution and the corresponding PARAFAC loadings of constituent 1 can be appreciated in Fig. 7. Constituent 2 on the other hand (green line) constitutes an interference.

** Insert Fig. 7 **

4.2. Challenges of second-order fluorescence data

Three important sources of problems exist in second-order fluorescent data analysis: (1) deviations from the ideal trilinearity, which may occur due to the presence of inner filter effects, (2) significant spectral overlapping among several sample constituents, and (3) spectral profiles in one mode which are identical in all samples, due to a certain physical law, reducing the selectivity in the affected mode to zero [1]. The latter phenomenon may occur either between calibrated analytes or between the analyte and interferences; in the latter case some algorithms cannot successfully model the data.

4.2.1. Inner filter effects

Inner filter causes deviations from trilinearity because spectra are deformed in a specific manner for each chemical sample. This phenomenon can be handled by MCR-ALS or PARAFAC2 only when the effect occurs in the excitation or in the emission mode, but not when it occurs in both excitation and emission modes simultaneously. In this latter case, only U-PLS/RBL can correctly solve this analytical problem, because of its flexible structure, which allows it to account for these effects by including additional latent variables in the calibration phase. In addition, if unexpected sample constituents are present, the U-PLS calibration can be combined with RBL, modeling the interferent contribution and achieving the second-order advantage.

It should be noticed that inner filter effects have been traditionally corrected by sample dilution or mathematical transformations. However, this may not be possible in the case of fluorescence measurements conducted in the solid surface which is employed for pre-concentration purposes. It is likely that future developments of solid-state luminescence measurements, where constituents are concentrated on a small area of a solid membrane, will show the presently described phenomenon.

In 2006, the power of U-PLS/RBL was experimentally demonstrated for the first time, by correctly predicting the concentration of two selected analytes (the fungicides carbendazim and thiabendazole) through EEFMs read in a C18 membrane, in which the analytes were retained, in the presence of unexpected species and overcoming the inner filter effect [60]. Significant changes are produced by thiabendazole in both the excitation and emission spectra of carbendazim. While U-PLS/RBL allowed the determination of carbendazim in samples prepared with both artificial and real matrices, PARAFAC showed a considerable lack of precision. Apparently, the inner-filter phenomenon produced by thiabendazole cannot be modelled by PARAFAC.

In another 2006 report, EEFMs for samples containing mixtures of chrysene (the analyte of interest), benzopyrene (which produced strong inner filter effect across the useful wavelength range) and pyrene (the unexpected constituent) were successfully resolved by U-PLS/RBL as the only viable alternative [61].

4.2.2. Significant spectral overlapping among analytes and interferents

Four examples, arranged in increasing order of complexity, are discussed below. In the first one, one analyte and four interferents occur: galantamine (GAL), an acetylcholinesterase inhibitor, was spectrofluorimetrically determined in a micellar medium of sodium dodecyl sulfate through EEFMs in the presence of interferents [62]. Test samples constituted by GAL and the spectral interferents ibuprofen, acetyl salicylic acid, phenylephrine and atropine were evaluated with PARAFAC, U-PLS/RBL and N-PLS/RBL algorithms. While both PARAFAC and N-PLS rendered poor results, the U-PLS/RBL predictions were in good agreement with the corresponding nominal values, demonstrating the ability of this latter algorithm to successfully overcome the strong spectral overlapping between the analyte and interferences.

The second example involves two analytes and three interferents: the fungicides thiabendazole (TBZ) and fuberidazole (FBZ) were spectrofluorimetrically determined in the presence of high concentrations of carbaryl, carbendazim and 1-naphthylacetic acid, using an optosensor coupled to a flow-injection system [63]. The sensor was based on the simultaneous retention of the analytes on C18-bonded phase placed inside a flow-cell. The EEFM was read after the arrival of the analytes to the sensing zone.

Figure 8 shows the fluorescence excitation and emission spectra for TBZ, FBZ and other agrochemicals adsorbed on the C18 solid surface, where the strong overlapping among analytes and interferents is clear. U-PLS/RBL provides good predictions for both TBZ and FBZ, allowing to reach selectivity using a commercial but non-selective sensing support. The PARAFAC recoveries were comparably worse, especially for the analyte TBZ at low concentrations. This may be ascribed to significant overlapping profiles of both analytes (TBZ and FBZ) in the emission mode (Fig. 8).

** Insert Fig. 8 **

In the third example, two analytes and fourteen interferents were studied: benzo[*a*]pyrene and dibenzo[*a*,*h*]anthracene, the two most carcinogenic polycyclic aromatic hydrocarbons (PAHs), were quantitated in a very interfering environment through EEFMs measured on a nylon-membrane surface [64]. The matrices were processed by applying PARAFAC and U-PLS/RBL. The superiority of U-PLS/ RBL to quantify the selected PAHs in the presence of the remaining 14 US EPA (United States Environmental Protection Agency) PAHs was demonstrated.

Finally, six PAHs, namely benzo[*a*]pyrene (BaP), dibenz[*a*,*h*]anthracene (DBA), chrysene (CHR), benzo[*b*]fluoranthene (BbF), benzo[*k*]fluoranthene (BkF) and benz[*a*]anthracene (BaA) were simultaneously quantified in the presence of other 10 interfering PAHs, applying second-order multivariate calibration to the EEFMs

555 obtained with a flow-through optosensor interfaced to a fast-scanning
556 spectrofluorimeter [65]. The interferences were analyzed at high concentrations, in
557 order to maximize the problem they may cause in the determination.

558 PARAFAC rendered poor results for BaP, DBA, CHR, and BaA, and this fact was
559 ascribed to a lack of selectivity for these analytes. Certainly, the significant spectral
560 overlapping among analytes and interferences appears to preclude the successful
561 decomposition of the second-order data. The ability of U-PLS/RBL to resolve highly
562 overlapped analytes was demostrated, even in a very interfering medium. Among the
563 studied PAHs, the calculated values for BbF showed certain dispersion with respect to
564 the nominal ones, and this fact was ascribed to the presence of FLT as interference,
565 whose spectra seriously overlap with those for BbF. On the other hand, N-PLS was only
566 able to successfully predict the concentrations of BaP, BbF, and BkF. This demonstrates
567 a weaker capability of this algorithm to resolve this type of complex systems.
568

569 ### 4.2.3. Identical profiles for two constituents (analyte or interferent)

570 As explained above, the U-PLS/RBL algorithm is able to resolve cases with strong
571 spectral overlapping. However, in the cases of identical profiles for analyte and
572 interferent in one mode this algorithm fails, because the RBL technique is unable to
573 distinguish the constituents. Trilinear decomposition also fails in this case. The
574 phenomenon is a special case of *linear dependency*, because a linear mathematical
575 relationship exists between the profiles of two constituents along one of the data modes.

576 These problems can be appropriately resolved by MCR-ALS. As already stated,
577 MCR-ALS decomposes an augmented data matrix, built by placing matrices for
578 different samples adjacent to each other, in such a way that the augmentation mode is

579 the one affected by the profile overlapping. As a result, the null selectivity in the
580 affected direction is recovered in the augmented mode.

581 A sample with either two responsive constituents with identical profiles in one mode
582 or the presence of an interferent with the same profile as a calibrated constituent will
583 produce a data matrix with rank one; that is, the matrix will be rank-deficient, a
584 situation also known as *rank overlap*. Analytical problems involving rank overlap
585 appear, for example, (1) when the kinetics of a reaction is followed and the reaction
586 product (unique for all sample constituents) is the responsive constituent; (2) when the
587 common mode is the luminescence time decay, corresponding to a lanthanide ion whose
588 excitation spectrum varies with the constituent that complexes the ion, (3) when the
589 emission spectrum of a species is common to all constituents, but the time evolution of
590 the signal differs and (4) when the interferent has the same profile in one of the data
591 modes as the analyte.

592 As an example of this case, we can mention the determination of three
593 fluoroquinolone antibiotics (ciprofloxacin, norfloxacin and danofloxacin) in serum in
594 the presence of the interferent salicylate [66]. The method was based on processing
595 lanthanide-sensitized excitation-time decay matrix data for their terbium (III)
596 complexes. As can be appreciated in Fig. 9, while the excitation mode shows good
597 selectivity between analyte and salicylate, in the time mode the selectivity is very poor.

598 The calculated lifetimes for the studied fluoroquinolones and salicylate are very
599 similar confirming the similarity of the corresponding decay curves.

600 MCR-ALS permitted the achievement of the second-order advantage in the presence
601 of a high degree of overlapping between the time decay profiles for the analyte and the
602 interferent complexes. Due to the presence of interactions between analytes and the
603 investigated matrix (serum), it was necessary to employ the standard addition method

for a successful determination. The test data matrix was subtracted from the standard addition matrices, and quantitation was performed using classical external calibration procedure.

** Insert Fig. 9 **

Figure 10 shows the MCR-ALS decomposition results for a typical test sample containing ciprofloxacin and salicylate. Figure 10A displays the retrieved spectral profiles, which are seen to resemble those for the terbium complexes of the analyte and salicylate. Besides, Fig. 10B shows the progression of time decay profiles in the corresponding standard addition study. The left sub-plot corresponds to the test sample itself, while the three remaining matrices are those obtained after subtracting the test sample matrix from each of the standard addition matrices (in these three sub-plots the profile for the salicylate is absent).

** Insert Fig. 10 **

This example illustrates the success of MCR-ALS in decomposing the contributions from the analyte and from the interferent, even when the time decay profile for the salicylate complex is almost identical to that of the ciprofloxacin complex.

4.3. Chromatography/multivariate detection matrices

In chromatography, the retention factor (k), which is the degree of retention of the sample constituent in the column, plays an important role in most analysis: in practice analytes elute with retention factors between 1 and 20, with a peak with k equal to 0 indicating a constituent which does not interact with the stationary phase [67]. Chromatographic separations can become a difficult task when complex samples have to be analyzed, because of the presence of constituents with similar retention factors. Nevertheless, the use of multi-way calibration may provide a useful resource for

628 accurate analyte quantitation when complete separation is not accomplished, or new
629 constituents are present in the sample being analyzed [68].

630 The hyphenation of liquid (LC) or gas (GC) chromatography (or capillary
631 electrophoresis, CE) with spectroscopic techniques can yield second-order data which
632 combine instrumental signals built from both spectral and time domains. Popular
633 techniques are LC-DAD (UV-Visible diode array detection), LC-FSFD (fast scanning
634 fluorescence detection) or LC-MS (mass spectrometric detection). The responses are
635 thus arranged as a data matrix, where each column corresponds to a wavelength (or *m*/*z*
636 ratio) and each row corresponds to a different elution time. Figure 11 shows a typical
637 example of a second-order data matrix generated by HPLC-DAD, corresponding to the
638 fast determination of dyes in beverages [69]. As can be appreciated, the incomplete
639 separation of three dyes (plus interferents) was accomplished in ca. 1.9 min.
640 Implementation of a second-order calibration by using pure standard samples matrices
641 furnished highly accurate results when analyzing real non-alcoholic beverage samples.

642 The use of second-order multivariate algorithms has been shown to play a critical
643 role in several analytical fields, as can be gathered from a literature survey in relevant
644 analytical, chemometrics and applied journals. Specifically, an important number of
645 reports have been presented focusing on the resolution of really complex samples by
646 using liquid chromatography and exploiting the second-order advantage [69,70]. In this
647 context, extremely important issues such as reduction in analysis time and consequently
648 in costs and amount of contaminant solvents should be considered.

649 ** Insert Fig. 11 **

650 Several algorithms can be used to model this kind of data, but MCR-ALS has
651 become the choice in most of the published works. This may be due to the fact that
652 differential migration of the constituents originates dissimilarities in both retention

653 times and peak shapes (this is even more significant in CE data), leading to data without
654 the property of trilinearity (see above). MCR-ALS can efficienlty solve this type of
655 problems by resorting to the mathematical resource of matrix augmentation.
656 PARAFAC2 is a variant of PARAFAC attempting to model such changes, but
657 apparently is less powerful than MCR-ALS in this regard, especially when time shifts or
658 band shapes are significant, and when potential interferents are present in test samples.
659 However, when data are conveniently pre-treated in order to alleviate the above-
660 mentioned problems, using adequate alignment strategies of elution profiles, good
661 results can be obtained by using PARAFAC or RBL-based algorithms [68,70]. For a
662 detailed description of the different pre-treatment approaches see Refs. [68,70] and a
663 recent tutorial [71].

664 As an example, we consider the development of an HPLC-FSFD method for the
665 simultaneous determination of five marker pteridines in urine samples: neopterin
666 (NEO), biopterin (BIO), pterin (PT), xanthopterin (XAN) and isoxanthopterin (ISO)
667 [72]. Figure 12 shows a chromatogram registered at λ_{exc} = 272 nm and λ_{em} = 445 nm
668 and elution times between 2.0 and 6.5 min for a urine sample, after spiking it with the
669 five analytes. The elution order was NEO, XAN, BIO, ISO, and PT, as indicated by an
670 arrow in Fig. 12. As can be appreciated in this figure, the elution profiles of XAN, BIO
671 and ISO are highly overlapped. On the other hand, an unknown peak appears between
672 the NEO and XAN peaks. In addition, several unexpected substances appear in the urine
673 matrix, making necessary to model the data with a second-order algorithm capable of
674 exploiting the second-order advantage.

675 Owing to the fact that three of the constituents present the same spectrum (linear
676 dependency, see above), these chromatographic data can be modeled following two
677 different strategies: (1) dividing the data in regions in which the spectra of target

678 analytes and interferents are different, and (2) modeling the whole data, but augmenting
679 the matrix in the direction of the linear dependency, i.e. in the direction of the spectra. It
680 should be noted that this latter strategy can be implemented because there are no elution
681 profile changes in the chromatographic direction. In addition, independently of the
682 methodology followed to process the data, a baseline correction was applied to subtract
683 the background present in the chromatograms (for more details see Ref. 72).

684 The strategy 1, which is the most usually implemented one when analyzing these
685 kind of second-order data, involves as a first step the selection of appropriate regions to
686 quantitate one analyte at a time; thus different MCR models were built to carry out the
687 analysis (shown as boxes in Fig. 12). As can be seen in this figure, the regions in which
688 the total chromatographic data were divided, in order to simplify the analysis, were the
689 following: region 1, between 2.75 and 3.12 min, in which only NEO is eluting, was
690 selected for NEO quantitation, region 2, between 3.35 and 3.58 min, in which XAN and
691 BIO are coeluting, was selected for XAN quantitation, region 3, between 3.45 and 3.74
692 min, in which XAN, BIO, and ISO are coeluting, was selected for BIO quantitation,
693 region 4, between 3.58 and 3.87 min, in which BIO and ISO are coeluting, was selected
694 for ISO quantitation, and region 5, between 3.84 and 4.20 min, in which ISO and PT are
695 coeluting, was selected for PT quantitation.

696 ** Insert Fig. 12 **

697 The usual way to process these data involves the construction of a suitable
698 augmented data matrix from which the MCR-ALS analysis retrieves one matrix
699 containing the spectral profiles for all the species present in the samples, and an
700 additional one including information which is useful to perform analyte quantitation, i.e.
701 the area under each chromatographic peak (Fig. 3). The satisfactory implementation of
702 this strategy to determine the analytes building a single MCR-ALS model (including the

703 complete experimental elution time range) requires that the spectra of the constituents
704 must be different, since the resolution is based on this fact. This strategy is not viable in
705 this example, due to the fact that NEO, BIO and PT present almost identical emission
706 spectra, leading to near zero selectivity in the spectral mode. On the other hand, as can
707 be appreciated in Fig. 12, although the time profiles of XAN, BIO and ISO are highly
708 overlapped, small differences exist in their elution time profiles, enough to allow MCR-
709 ALS resolution in the different regions. For this reason, the data were divided into time
710 regions which were processed separately, as commented above.
711 First, Fig. 13A shows the MCR-ALS spectral profiles retrieved for region number 3.
712 These spectra correspond to the three analytes present in this region (ISO, BIO and
713 XAN) and an additional one ascribed to an interfering agent. An indicative measure of
714 the quality of the MCR-ALS modeling can be attained by visual inspection of the
715 retrieved spectral profiles for the standard solution and its comparison with real analyte
716 spectra.
717 On the other hand, Fig. 13B shows the time profiles retrieved by MCR-ALS when
718 strategy 1 was applied to region 3 [matrix $\mathbf{B}_{aug}$ in equation (2)], for two urine samples
719 spiked with the five analytes and two analyte standard samples. As mentioned above,
720 this region contains ISO, BIO and XAN, and was employed for BIO quantitation. The
721 successive boxes in Fig. 13B show: (1) a constant signal in the first two urine samples,
722 absent in the standards (this corresponds to a urine background interferent), and (2)
723 signals corresponding to ISO, BIO and XAN, all increased in the last two samples with
724 respect to the first two due to the addition of analyte standards to the urine samples. The
725 isolation of the signal which in this case was ascribed to BIO in each studied test sample
726 was used for the quantitation of this analyte. As can be observed, the interference profile
727 remains constant.

** Insert Fig. 13 **

On the other hand, the procedure of dividing into regions was avoided in strategy 2, in which the complete data were used to build the augmented data matrix that allows simultaneously determining all analytes. The complexity of the processing is extremely reduced if the complete data are conveniently modeled following the strategy 2. In other words, the resolution of each analyte by strategy 1 requires the building of one particular MCR-ALS model with data belonging to the region in which it is included, i.e. five MCR-ALS models are needed to quantify the five analytes present in one sample (for details see Ref. 73). Strategy 2, however, involves non-standard augmentation in the elution time direction, which in turns demands that elution time profiles do not vary from sample to sample. This situation is however scarcely found in the literature.

5. Third-order and beyond

5.1. Four-way spectral-kinetic data

Four-way data can be obtained by joining third-order data for a set of samples into a four-dimensional mathematical object. It is interesting to note that only in few reports four-way data have been recorded and used to develop analytical methodologies. This may be attributed to our lacking of a thorough understanding of their analytical advantages, or to the fact that the practical acquisition of these data arrays is still difficult to implement. Hence, although in theory a large number of possible forms of obtaining four-way data exist, those commonly used are the following: (1) with a single instrument, EEMs (either fluorescence or phosphorescence) as a function of reaction time or decay time [74] and (2) with hyphenated instruments, two-dimensional

752 chromatography with time of flight mass spectrometry (TOFMS) or diode array
753 detection (DAD), such as GC-GC-TOFMS or LC-LC-DAD, and LC-DAD as a function
754 of reaction time [75].

755 In this way, for obtaining third-order data using luminescence spectroscopy, several
756 different approaches can be found in the literature. One is based on EEM-fluorescence
757 life-time data: the first publication dates back to as early as 1990 [76], by incorporating
758 fluorescence lifetime as a third mode of information to the excitation-emission matrix
759 (EEM) using the phase modulation technique. The data were combined with the GRAM
760 algorithm, and applied to the resolution of a two constituent mixture composed of
761 benzo[*b*]fluoranthene and benzo[*k*]fluoranthene.

762 A second approach involves laser-excited time-resolved EEM fluorescence data, as
763 in the collection of fluorescence time-resolved excitation-emission data at 4.2 K *via*
764 laser-excited time-resolved Shpol´skii spectroscopy, which has been recently proposed
765 for the resolution of mixtures of 15 polycyclic aromatic hydrocarbons [77]. These data
766 could be successfully processed using PARAFAC and U-PLS/RTL. Figure 14 shows an
767 schematic representation of how the EEMs were collected following the decay time.

768 ** Insert Fig. 14 **

769 A third alternative is based on laser-excited time-resolved EEM phosphorescence
770 data, using instrumental data generated from Shpol´skii matrices at liquid helium
771 temperature. In this case, the third-order data arrays, consisting of excitation modulated
772 wavelength-time matrices, were collected with the aid of a cryogenic fiber-optic probe,
773 a tunable dye laser, and a multichannel system for phosphorescence detection, and
774 applied to the analysis of 2,3,7,8-tetrachloro-dibenzo-para-dioxin [78].

775 Another possibility is to measure EEM-reaction time data. An example is presented
776 in which four-way fluorescence data were recorded by following the kinetic evolution

of EEMs, analyzed by PARAFAC and TLLS/RTL. They were applied to the simultaneous determination of the constituents of the anticancer combination of methotrexate and leucovorin in human urine [74]. Both analytes were converted into highly fluorescent products by oxidation with potassium permanganate, and the kinetic of the reaction was continuously monitored by recording full EEMs of the samples at different reaction times. A fast scanning spectrofluorimeter was used for the acquisition of a complete EEM in 12 seconds at a wavelength scanning speed of 24,000 nm/min. The emission spectra were recorded from 335 nm to 490 nm at 5 nm intervals, exciting from 255 nm to 315 nm at 6 nm intervals. Ten successive EEMs were measured at 72 seconds intervals, in order to follow the fluorescence kinetic evolution of the mixture constituents (Fig. 15). The excitation, emission and kinetic time profiles recovered by both chemometric techniques are in good agreement with experimental observations.

Similarly, the evolution of on-line photochemically induced excitation-emission matrices with the irradiation time, allowed the determination of folic acid (FA), and its two main metabolites, tetrahydrofolic acid (THF) and 5-methyltetrahydrofolic acid (5-MTF), in serum samples (Fig. 16). In this figure it can be appreciated that FA and THF show an initially weak fluorescence, and that the intensity of FA and THF considerably increases as a function of reaction time, but at different rates. The strongly fluorescent photoproducts, after irradiation of FA and THF with UV light in acidic medium, have identical excitation and emission wavelengths, 270 nm and 445 nm, respectively. In contrast, the strong native fluorescence of 5-MTF decreased with reaction time. The method achieves selectivity from the different rates at which the corresponding photoproducts are formed and degraded, as a discriminatory parameter, allowing the successful determination of the three constituents by using a combination of U-PLS or N-PLS with RTL [79].

** Insert Fig. 15 **

** Insert Fig. 16 **

5.2. Chromatography/multivariate detection

One common example of four-way/third-order data involving chromatography is comprehensive two-dimensional gas chromatography followed by mass spectrometric detection (GC-GC-MS). As mentioned in Section 3.2., multi-way data of chromatographic origin usually lose multi-linearity in the time mode because of elution time changes from sample to sample. Consequently, in order to apply MCR-ALS to these data, it is necessary to first unfold the original third-order array for each sample (time1 × time2 × mass/charge ratio) into a matrix (unfolded time × mass/charge ratio second-order data). These matrices are then arranged into a bilinear augmented matrix and processed as shown in Fig. 5 and commented in the corresponding section.

Tauler et al. reported the resolution and quantitation of mixtures of polycyclic aromatic hydrocarbons (PAH) in heavy fuel oil samples by MCR-ALS modelling of data obtained with GC × GC-TOFMS [44]. Different three-way data corresponding to pure PAH standards and real samples were transformed into matrices and submitted to MCR-ALS analysis of the augmented matrix. This procedure allowed to obtain component elution profiles in the two chromatographic dimensions and their pure mass spectra.

The general procedure followed in this work for organizing the data and their subsequent MCR-ALS analysis can be appreciated in Fig. 5. The samples were injected into the first column, the eluted compounds were then pre-concentrated in the modulator, and then re-injected into the second column after a modulation period. Thus, the entire first column chromatogram was sliced into a series of high-speed short secondary chromatograms of a length equal to P_M, which were continuously recorded

by the TOFMS detector. Every slice produced a data matrix **X** of size ($J \times K$), where J is the number of collected data points (elution times, ^2t_R) in the second column, and K is the number of m/z values. The slices obtained at the different elution times in the first mode (^1t_R) were then stacked to form an augmented matrix of size ($LJ \times K$) as shown in Fig. 5, in which L is the number of collected data points taken from the first column (L). Here LJ represents the product of the data points taken from the first column by the data points taken from the second column ($^2t_R \times {}^1t_R$). After this phase, a super-augmented data matrix $\mathbf{D}_{\text{sup-aug}}$ of size [$LJ(I+1) \times K$] was built, in which I is the number of standards and "1" represents the unknown sample being analyzed. Owing to the fact that the number of m/z values is equal for all slices and for all chromatographic runs, the column-wise data arrangement shown in $\mathbf{D}_{\text{sup-aug}}$ is very flexible and adequate to bilinear modeling requirements. For example, the number of chromatographic ranges analyzed can be different in each slice, and the presence of time shifts among different slices and among different samples do not destroy the bilinear model assumption associated to the column-wise data augmentation strategy. On the other hand, each slice in the super-augmented data matrix can have different numbers of rows (elution times in the first and second modes). This is a very flexible property of matrix augmentation, which adapts very well to GC×GC-TOFMS data, because of ubiquitous elution time peak shifts of the eluted components from slice-to-slice (in the second column) and from sample-to-sample (in the first and second columns). Furthermore, with this super-augmented data arrangement, there is no limitation in the number of included sub-matrices (number of slices and samples). For quantitative analysis, it is possible to include the second mode slices taken from the first column for the different standard mixtures and unknown samples and their replicates in the same column-wise super-augmented data matrix, and perform their simultaneous analysis in one shot [44].

MCR-ALS analysis yields a super-augmented concentration matrix containing the pure second dimension elution profiles for different slices of different samples ($\mathbf{C}_{sup\text{-}aug}$), a matrix of pure mass spectra profiles ($\mathbf{S}^T$) for the N components, and a residual matrix ($\mathbf{E}_{sup\text{-}aug}$) containing noise and unresolved background (Fig. 5). The $\mathbf{C}_{sup\text{-}aug}$ matrix contains the second mode elution profiles in all $[L(I+1) \times K]$ slices for the N resolved components. The single pure mass spectral matrix $\mathbf{S}^T$ ($N \times K$) can be used for the identification of the resolved components. On the other hand, to get the first mode elution profiles for every component in each sample analyzed, the areas under the resolved second mode elution profiles corresponding to one sample are used (see Fig. 5). In this way, a matrix of size ($L \times N$) for every sample is obtained. In addition, areas under the resolved second dimension elution profiles can be used for quantitative purposes.

In another very recently reported application, a four-way multivariate calibration approach based on the combination of HPLC data and third-order algorithms has been described for the first time [80]. In this case, each sample was injected into the chromatograph eight times, each time exciting with a different excitation wavelength, and the emission spectra were recorded along the full chromatogram. The third-order data thus obtained were joined into a four-way data array, which was subsequently analyzed with PARAFAC, U-PLS/RTL and N-PLS/RTL, because no substantial changes in elution profiles were detected from sample to sample. The best algorithm to perform the multi-way calibration was U-PLS/RTL.

5.3. Five-way data

Five-way data can be obtained by joining fourth-order data for a sample set. The algorithm U-PLS/RQL has been developed as a new latent structured algorithm for the processing of these instrumental data. In order to check its analytical predictive ability, fluorescence excitation-emission-kinetic-pH data were measured and processed. The concentration of the fluorescent pesticide carbaryl was determined in the presence of uncalibrated interferents, in the first reported example of fourth-order multivariate calibration.

** Insert Fig. 17 **

The hydrolysis of the analyte was followed at different pH values using a fast-scanning spectrofluorimeter, recording the excitation-emission fluorescence matrices during its evolution to produce 1-naphthol, which does also emit fluorescence (Fig. 17) [47].

6. Multi-way analytical figures of merit

The development of new multi-way analytical methods demands to be able to estimate the corresponding analytical figures of merit, in order to compare with previously existing methodologies and to report detection capabilities and other important features [10]. In this context, the sensitivity is a crucial parameter, because: (1) it allows proper comparison among different methods, (2) it permits the estimation of other figures of merit, such as the response-independent analytical sensitivity [81], and (3) it is needed to compute prediction uncertainties and detection capabilities [82].

Pertinent questions in this area are the following: (1) is there any conceptual difference between classical univariate figures of merit and their multi-way analogues? and (2) can multi-way figures of merit be estimated for all calibration methodologies and models? The answers to these important questions are, fortunately, no and yes, respectively. Below we present a brief discussion connecting the classical approach with the advanced field of multi-way figures of merit.

According to the International Union of Pure and Applied Chemistry (IUPAC), the sensitivity is well-defined in some analytical calibration [83-85], such as univariate calibration, where it is the change in the instrument response for a unit change in the concentration of the analyte of interest [83]. The slope of the calibration graph is in this case a convenient measure of the sensitivity [83]. For first-order multivariate calibration, this definition has been generalized employing an intuitive analogy between instrumental signal and a so-called net analyte signal (NAS) [86]. The NAS is defined as the portion of the overall signal which can be uniquely ascribed to the analyte [86,87] and is completely general and applicable to all first-order calibration methods, including the popular partial least-squares (PLS) algorithm [87]. Mathematically, the NAS for a given analyte is defined as the projection of the sample signal orthogonal to the space spanned by the interferent agents (Fig. 18A). A plot of the length of NAS vectors as a function of the analyte concentrations should be linear, the slope being the length of the NAS vector at unit concentration, which is a good measure of the first-order sensitivity.

In the case of second-order multivariate calibration, there have been several proposals for estimating the sensitivity [88-90]. Some of them were based in extensions of the NAS concept, although difficulties appear because: (1) there are different NAS definitions, and it is difficult to understand their true meanings [91-93] and (2) extrapolation to orders higher than two leads to a significantly underestimation of true

sensitivity values [94]. We discuss as an example the most general second-order PARAFAC sensitivity, whose expression has been shown to be [90]:

$$\mathrm{SEN} = s_n\{[(\mathbf{B}_{cal}^{T}\,(\mathbf{I} - \mathbf{B}_{unx}\,\mathbf{B}_{unx}^{+})\,\mathbf{B}_{cal}) * (\mathbf{C}_{cal}^{T}\,(\mathbf{I} - \mathbf{C}_{unx}\,\mathbf{C}_{unx}^{+})\,\mathbf{C}_{cal})]^{-1}\}_{nn}^{-1/2} \qquad (4)$$

where s_n is the slope of the PARAFAC univariate plot, the index *n* identifies the analyte, the symbol '*' is the Hadamard matrix product, the subscript '*nn*' indicates the (*n*,*n*) diagonal element of a matrix, the matrices **B** and **C** collect the loadings, the subscripts 'cal' and 'unx' indicate calibration and unexpected respectively, and the symbol '+' implies the pseudo-inverse of a matrix.

A recent alternative definition is based on the concepts of input and output noise in a given system: the sensitivity (SEN) measures the ratio of output noise to input noise [95,96]. This approach allowed to develop sensitivity expressions for multi-way calibration based on PARAFAC [94], MCR-ALS [97] and PLS/RML (RML indicates residual multi-linearization, and includes RBL, RTL, RQL, etc.) [98]. Further work is needed, however, to validate these expressions and to include all of them into a generalized conceptual scheme.

It is important to notice, however, that the estimation of figures of merit is included in some of the available software packages for multi-way calibration, and thus by employing this software the analyst has access to all analytical figures to be reported along with the concentration of the analyte of interest: MVC2 and MVC3 for second- and third-order calibration respectively (see Table 2) [51,52].

** Insert Fig. 18 **

The sensitivity is important because it allows estimating additional figures of merit [10], such as the uncertainty in predicted concentrations. The best known approximation to concentration variance is the following expression:

$$\sigma_y^2 = \mathrm{SEN}^{-2}\,\sigma_x^2 + h\,\mathrm{SEN}^{-2}\,\sigma_x^2 + h\,\sigma_{y\mathrm{cal}}^2 \qquad (5)$$

where SEN is the sensitivity, σ_x^2 the variance in instrumental signals, *h* the sample leverage and σ_{ycal}^2 the variance in calibration concentrations. The three terms in equation (5) correspond to the propagation of uncertainties from: instrumental signals in the test sample, instrumental signals in the calibration, and calibration concentrations. The last two terms are scaled by the sample leverage *h*, a dimensionless parameter measuring the relative position of the sample in the calibration space.

The limit of detection (LOD) and quantitation (LOQ) are additional important figures of merit. They can be estimated based on IUPAC's recommendations on the so-called type I and II errors [82]. It is first required to define a critical concentration value, which is the level for the detection decision, involving a certain risk of type I errors (α is the probability of false positive). In this way, the decision "detected" or "not detected" is made by comparison of the estimated quantity ($\hat{L}$) with the critical value (L_C) of the respective distribution, such that the probability of exceeding L_C is no greater than α if the analyte is absent. The limit of detection is defined as the concentration (LOD) for which the risk of Type II errors has a probability β (β is the probability of false negative) given L_C or α. Both α and β are assigned reasonably values depending on the specific analytical application. If it can be assumed that the concentration uncertainty at zero analyte concentration is close to that at the LOD level, the latter can be approximated by $(t_{\alpha,\nu} + t_{\beta,\nu})\,\sigma_0$, where σ_0 is the prediction uncertainty for a blank sample, and $t_{\alpha,\nu}$ and $t_{\beta,\nu}$ are the *t*-coefficients for probabilities α and β with ν degrees of freedom. For 95% probabilities of coverage against both type I and II errors and a large number of degrees of freedom, the limit of detection can be estimated as the product of a specific coefficient and the standard deviation of the blank, when the uncertainty in

969 the mean (expected) value of the blank is negligible, and α and β are each equal 0.05,
970 and the estimated value is normally distributed with known, constant variance.

971 $$LOD = 3.3\ \sigma_0 \qquad (6)$$

972 The limit of quantitation, in turn, is estimated as:

973 $$LOQ = 10\ \sigma_0 \qquad (7)$$

974 which ensures a maximum of 10% relative error in prediction.

975 We now describe in detail a work which nicely illustrates the consequences of the
976 interplay among the different terms in the LOD expression. The main purpose of this
977 description is to alert readers on the fact that the significant experimental effort required
978 to achieve larger sensitivity may not be directly translated into a proportional decrease
979 in limit of detection. The work in question discusses the determination of a fluorescent
980 pesticide in the presence of other fluorescent interferents in water samples, which has
981 been accomplished using multi-way methodologies of increasing order [47]. The most
982 complex work involved the measurement of five-way data for a group of samples, i.e.,
983 fourth-order data based on the kinetic evolution of excitation-emission fluorescence
984 matrices as a function of pH as stated above [47]. The four data modes for each sample
985 were excitation wavelength, emission wavelength, reaction time and pH. It is important
986 to notice that these data are not quadri-linear, since the rate constant for the hydrolysis
987 of the analyte is pH-dependent, and hence two data modes are highly correlated. This
988 precludes, in principle, the use of PARAFAC, which requires quadri-linearity in the
989 data for each sample. However, concatenation of the reaction time and pH modes into a
990 single combined mode does allow for quadri-linear PARAFAC decomposition of a
991 four-way array whose modes are sample, excitation, emission and the concatenation of
992 pH and time.

Fourth-order data provide the second-order advantage and in principle increased sensitivity and improved detection capability. However, the second-order advantage could also be achieved by third-order data (measuring the same data set at a fixed pH, or at a fixed reaction time), or even by second-order data (measuring the excitation-emission matrices at a fixed reaction time and pH values). What is exactly gained in going from second- to third- to fourth-order data can be estimated from the calculation of the analytical figures of merit for each alternative. It is not the purpose of this work to comprehensively cover all the possibilities, which are too many, and hence we limit to only the available alternative measurement schemes.

Since the data set in its original structure is fourth-order and is not quadri-linear, the best data processing algorithm is U-PLS/RQL, because of its inherent flexibility towards non multi-linear data. The sensitivity can be precisely calculated for three data orders: (1) second-order fixing reaction time and pH, (2) third-order fixing pH, and (3) fourth-order [98]. The result is shown in Fig. 18B, where a clear gain in sensitivity is detected.

However, Fig. 18C shows the progression of the LOD value. As can be seen, the LOD seems to level off at third-order, this is because the contribution of the remaining terms in equation (5), which depend not only on the sensitivity but also on the sample leverage. The propagated calibration uncertainty, particularly the one on concentrations, which does not depend on the value of SEN, is mainly responsible for the leveling off action on increasing the data order. These considerations should become useful when planning multi-way experiments, since it is clear that an increase in sensitivity does not lead, per se, to a concomitantly decrease in LOD.

Finally, two other approaches to estimating LOD and LOQ should be mentioned. One of them involves the consideration of the univariate calibration plot as a true

single-component calibration, and estimation of these limits directly by the classical univariate approach. This has been done in the framework of PARAFAC and MCR-ALS [99-101]. In ref. [102] it is mathematically derived that the decision and detection limits as defined by IUPAC are the same if the regression concentration calculated versus nominal concentration is used instead of the signal versus nominal concentration. In this new equivalent formulation the definitions are applicable to any calibration model because, once built, it allows to compute the calculated concentration. Applications of this procedure can be seen in [103]. Further work is required to investigate the effect of potential interferents and the second-order advantage in the latter formulation.

Acknowledgments

The authors are grateful to the Ministerio de Economía y Competitividad of Spain (Project CTQ2011-25388) and the Gobierno de Extremadura, both co-financed by European FEDER Funds (Consolidation Project of Research Group FQM003, Project GR1003), Universidad Nacional de Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas (Projects PIP 1950 and PIP 455), Universidad Nacional del Litoral (Project CAI+D Nº 12-65) and to ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Projects PICT 2010-0084 and 2011-0005) for financially supporting this work.

1040 # References

[1] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J. Poppi, Trends Anal. Chem. 26 (2007) 752-765.

[2] A.C. Olivieri, G.M. Escandar, A. Muñoz de la Peña, Trends Anal. Chem. 30 (2011) 607-617.

[3] S. Mas, A. de Juan, R. Tauler, A.C. Olivieri, G.M. Escandar, Talanta 80 (2010) 1052-1067.

[4] E. Sanchez, B.R. Kowalski, J. Chemom. 2 (1988) 265-280.

[5] C.-N. Ho, G.G. Christian, E.R. Davidson, Anal. Chem. 50 (1978) 1108-1113.

[6] J.A. Arancibia, P.C. Damiani, G.M. Escandar, G.A. Ibañez, A.C. Olivieri, J. Chromatogr. B 910 (2012) 22-30.

[7] M.C. Ortiz, L. Sarabia, J. Chromatogr. A 1158 (2007) 94-110.

[8] V. Gómez, M.P. Callao, Anal. Chim. Acta 627 (2008) 169-183.

[9] C. Ruckebusch, L. Blanchet Anal. Chim. Acta 765 (2013) 28-36

[10] A.C. Olivieri, N.M. Faber, Validation and Error, in S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Elsevier, Amsterdam, 2009, Vol. 3, pp. 91-120.

[11] W.E. Van der Linden, Pure Appl. Chem. 61 (1989) 91-95.

[12] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A-791A

[13] J. Ferré, N. M. Faber, Chemom. Intell. Lab. Syst. 69 (2003) 123-126.

[14] N. M. Faber, Chemom. Intell. Lab. Syst. 50 (2000) 107-114.

[15] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149-171.

[16] A. Smilde, R. Bro, P. Geladi, Multi-way analysis: applications in the chemical sciences, Wiley, Chichester, 2004.

[17] R. Bro, Crit. Rev. Anal. Chem. 36 (2006) 279-293.

[18] P. M. Kroonenberg, Applied multiway data analysis, Wiley, Chichester, 2008.

[19] P.A. Paatero, Chemom. Intell. Lab. Syst. 38 (1997) 223-242.

[20] Z.P. Chen, H. L. Wu, J. H. Jiang, Y. Li, R. Q. Yu, Chemom. Intell. Lab. Syst. 52 (2000) 75-86.

[21] A.L. Xia, H.L. Wu, D. M. Fang, Y.J. Ding, L.Q. Hu, R.Q. Yu, J. Chemom. 19 (2005) 65-76.

[22] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496-499.

[23] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29-45.

[24] M. Linder, R. Sundberg, J. Chemom. 16 (2002) 12-27.

[25] A. de Juan, R. Tauler, J. Chemom. 15 (2001) 749-772.

[26] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133-146.

[27] A. de Juan, R. Tauler, Crit. Rev. Anal. Chem. 36 (2006) 163-176.

[28] A. de Juan, R. Tauler, Anal. Chim. Acta 500 (2003) 195-210.

[29] R. Tauler, M. Maeder, A. de Juan, Multiset Data Analysis: Extended Multivariate Curve Resolution, in S. Brown, R. Tauler, R. Walczak (Eds.), Comprehensive Chemometrics, Elsevier, Amsterdam, 2009, Vol. 2, pp. 473-505.

[30] W. Windig, J. Guilment, Anal. Chem. 63 (1991) 1425-1432.

[31] M.C. Antunes, J.E.J. Simao, A.C. Duarte, R. Tauler, Analyst 127 (2002) 809-817.

[32] H.A. L. Kiers, J.M.F. ten Berge, R. Bro, J. Chemom. 13 (1999) 275-294.

[33] J.M. Amigo, T. Skov, J. Coello, S. Maspoch, R. Bro, Trends Anal. Chem. 27 (2008) 714-725.

[34] R. Bro, C.A. Andersson, H.A. L. Kiers, J. Chemom. 13 (1999) 295-309.

[35] R. Tauler, I. Marqués, E. Casassas, J. Chemom. 12 (1998) 55-75.

[36] B.E. Wilson, W. Lindberg, B.R. Kowalski, J. Am. Chem. Soc. 111 (1989) 3797-3804.

[37] S. Wold, P. Geladi, K. Esbensen, J. Øhman, J. Chemom. 1 (1987) 41-56.

[38] R. Bro, J. Chemom. 10 (1996) 47-61.

[39] J. Öhman, P. Geladi, S. Wold, J. Chemom. 4 (1990) 79-90.

[40] A.C. Olivieri, J. Chemom. 19 (2005) 253-265.

[41] A.L. Xia, H.L. Wu, S.F. Li, S. H. Zhu, L.Q. Hu, R.Q. Yu, J. Chemom. 21 (2007) 133-144.

[42] H.Y. Fu, H.L. Wu, Y.J. Yu, L.L. Yu, S.R. Zhang, J.F. Nie, S.F. Li, R.Q. Yu, J. Chemom. 25 (2011) 408-429.

[43] H.P. Bailey, S.C. Rutan, Chemom. Intell. Lab. Syst. 106 (2011) 131-141.

[44] H. Parastar, J.R. Radovic, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Anal. Chem. 83 (2011) 9289-9297.

[45] P.C. Damiani, I. Durán Merás, A.G García Reiriz, A. Jimenez Girón, A. Muñoz de la Peña, A.C. Olivieri, Anal. Chem. 76 (2007) 6949-6958.

[46] J.A. Arancibia, A.C. Olivieri, D. Bohoyo Gil, A. Espinosa Mansilla, I. Durán Merás, A. Muñoz de la Peña, Chemom. Intell. Lab. Syst. 80 (2006) 77-86.

[47] R.M. Maggio, A. Muñoz de la Peña, A.C. Olivieri, Chem. Intell. Lab. Syst. 109 (2011) 178-185.

[48] MATLAB, The Mathworks, Inc., Natick, Massachussets, USA.

[49] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, Chemom. Intell. Lab. Syst. 76 (2005) 101-110.

[50] P.J. Gemperline, E. Cash, Anal. Chem. 75 (2003) 4236-4243.

[51] A.C. Olivieri, H.L. Wu, R.Q. Yu, Chemom. Intell. Lab. Syst. 96 (2009) 246-251.

[52] A.C. Olivieri, H.L. Wu, R.Q. Yu, Chemom. Intell. Lab. Syst. 116 (2012) 9-16.

[53] A. Espinosa-Mansilla, A. Muñoz se la Peña, H.C. Goicoechea, A.C. Olivieri, Appl. Spectrosc. 58 (2004) 83-90.

[54] H.C. Goicoechea, A.C. Olivieri, Appl. Spectrosc. 59 (2005) 926-933.

[55] A. Muñoz de la Peña, N. Mora Díez, D. Bohoyo Gil, E. Cano Carranza, J. Fluoresc. 19 (2009) 345-352.

[56] A. Muñoz de la Peña, A. Espinosa Mansilla, D. González Gómez, A.C. Olivieri, H.C. Goicoechea, Anal. Chem. 75 (2003) 2640-2646.

[57] A. Jimenez Girón, I. Durán Merás, A. Muñoz de la Peña, A. Espinosa Mansilla, F. Cañada Cañada, A.C. Olivieri, Anal. Bioanal. Chem. 391 (2008) 827-835.

[58] J.A. Porini, G.M. Escandar, Anal. Meth. 3 (2011) 1494-1500.

[59] http://www.models.kvl.dk/~pih/parafac/chap2parafac.htm

[60] G.N. Piccirilli, G.M. Escandar, Analyst 131 (2006) 1012-1020.

[61] D. Bohoyo Gil, A. Muñoz de la Peña, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, Anal. Chem. 78 (2006) 8051-8058.

[62] M.J. Culzoni, R.Q. Aucelio, G. M. Escandar, Talanta 82 (2010) 325-332.

[63] G.N. Piccirilli, G.M. Escandar, Analyst 135 (2010) 1299-1308.

[64] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, Anal. Chem. 80 (2008) 8276-8286.

[65] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, Environ. Sci. Technol. 45 (2011) 1513-1520.

[66] V.A. Lozano, R. Tauler, G.A. Ibañez, A.C. Olivieri, Talanta 77 (2009) 1715-1723.

[67] M.W. Dong, Modern HPLC for Practicing Scientists, Wiley, Synomics Pharmaceutical Services, LLC, Wareham, Massachusetts, 2006.

[68] M. Daszykowski, B. Walczak, Trends Anal. Chem. 25 (2006) 1081-1096.

[69] M.J. Culzoni, A.V. Schenone, N.E. Llamas, M. Garrido, M.S. Di Nezio, B.S. Fernández Band, H.C. Goicoechea, J. Chromatogr. A 1216 (2009) 7063-7070.

[70] H.C. Goicoechea, M.J. Culzoni, M.D. Gil García, M. Martínez Galera, Talanta 83 (2011) 1198-1107.

[71] T.G. Bloemberg, J. Gerretzen, A. Lunshof, R. Wehrens, L.M.C. Buydens, Anal. Chim. Acta 781 (2013) 14-32.

[72] 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, Anal. Bioanal. Chem. 399 (2011) 2123-2135.

[73] 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, Talanta 85 (2011) 2368-2374.

[74] A.C. Olivieri, J.A. Arancibia, A. Muñoz de la Peña, I. Durán Merás, A. Espinosa Mansilla, Anal. Chem. 76 (2004) 5657-5666.

[75] A.E. Sinha, B.J. Prazen, R.E. Synovec, Anal. Bioanal. Chem. 378 (2004) 1948-1951.

[76] D.S. Burdick, X.M. Tu, L.B. McGown, D.W. Millican, J. Chemom. 4 (1990) 15-28.

[77] H.C. Goicoechea, S. Yu, A.F.T. Moore, A.D. Campligia, Talanta 101 (2012) 330-336.

[78] H.C. Goicoechea, S. Yu, A.C. Olivieri, A.D. Campligia, Anal. Chem. 77 (2005) 2608-2616.

[79] A. Jiménez Girón, I. Durán Merás, A. Espinosa Mansilla, A. Muñoz de la Peña, F. Cañada Cañada, A. C. Olivieri, Anal. Chim. Acta 622 (2008) 94-103

[80] V. Lozano, A. Muñoz de la Peña, I. Durán Merás, A. Espinosa Mansilla, G.M. Escandar, Chemom. Intell. Lab. Syst. 125 (2013) 121-131.

[81] L. Cuadros Rodríguez, A.M. García Campaña, C. Jiménez Linares, M. Román Ceba, Anal. Lett. 26 (1993) 1243-1258.

[82] L.A. Currie, Pure Appl. Chem. 67 (1995) 1699-1723.

[83] K. Danzer, L.A. Currie, Pure Appl. Chem. 70 (1998) 993-1014.

[84] K. Danzer, M. Otto, L.A. Currie, Pure Appl. Chem. 76 (2004) 1215-1225.

[85] A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J. Kalivas, H. Mark, Pure Appl. Chem. 78 (2006) 633-661.

[86] A. Lorber, Anal. Chem. 58 (1986) 1167-1172.

[87] A. Lorber, K. Faber, B.R. Kowalski, Anal. Chem. 69 (1997) 1620-1626.

[88] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 52 (1980) 1071-1079.

[89] N.J. Messick, J.H. Kalivas, P.M. Lang, Anal. Chem. 68 (1996) 572-1579.

[90] A.C. Olivieri, N.M. Faber, J. Chemom. 19 (2005) 583-592.

[91] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11(1997) 419-461.

[92] A.C. Olivieri, Anal. Chem. 77 (2005) 4936-4946.

[93] A.C. Olivieri, Anal. Chem. 80 (2008) 5713-5720.

[94] A.C. Olivieri, N.M. Faber, Anal. Chem. 84 (2012) 186-193.

[95] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, Chem. Rev. 105 (2005) 2811-2828.

[96] N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, Chemom. Intell. Lab. Syst. 63 (2002) 107-116.

[97] C. Bauza, G.A. Ibañez, R. Tauler, A.C. Olivieri, Anal Chem 84 (2012) 8697-8706.

[98] F.A. Allegrini, A.C. Olivieri, Anal. Chem. 84 (2012) 10823-10830.

[99] J. Saurina, C. Leal, R. Compañó, M. Granados, M. Dolors Prat, R. Tauler, Anal. Chim. Acta 432 (2001) 241-251.

[100] M.J. Rodríguez-Cuesta, R. Boqué, F.X. Rius, Anal. Chim. Acta 491 (2003) 47-57.

[101] M.J. Rodríguez-Cuesta, R. Boqué, F.X. Rius, J.L. Martínez Vidal, A. Garrido Frenich, Chemom. Intell. Lab. Syst. 77 (2005) 251-260.

[102] M.C. Ortiz, L.A. Sarabia, A. Herrero, M.S. Sánchez, M.B. Sanza, M.E. Rueda, D. Giménez, M.E. Meléndez, Chemom. Intell. Lab. Syst. 69 (2003) 21-33.

[103] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, Anal. Chim. Acta 674 (2010) 123-142.

Table 1. Different arrays that can be obtained for a single sample and for a set of samples.

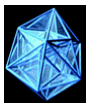
Data order	Array				Calibration
	One sample		A sample set		
Zero	Scalar		One-way		Univariate
First	Vector		Two-way		Multivariate
Second	Matrix		Three-way		Multi-way
Third	Three-way		Four-way		Multi-way
Fourth	Four-way		Five-way		Multi-way

Table 2. Free software for multi-way data processing.

Name	Algorithm(s)	Web page[a]
The N-way toolbox	PARAFAC PARAFAC2 GRAM DTLD U-PLS N-PLS	http://www.models.life.ku.dk/nwaytoolbox/download
MCR graphical interface	MCR-ALS	http://www.mcrals.info/
GUIPRO graphical interface	MCR-ALS	http://personal.ecu.edu/gemperlinep
MVC2 graphical interface	PARAFAC APTLD SWATLD BLLS/RBL U-PLS/RBL N-PLS/RBL	www.iquir-conicet.gov.ar/descargas/mvc2.rar
MVC3 graphical interface	PARAFAC APQLD AWRCQLD TLLS/RTL U-PLS/RTL N-PLS/RTL	www.iquir-conicet.gov.ar/descargas/mvc3.rar

[a] All pages were accessed in March 2013.

Figures and captions

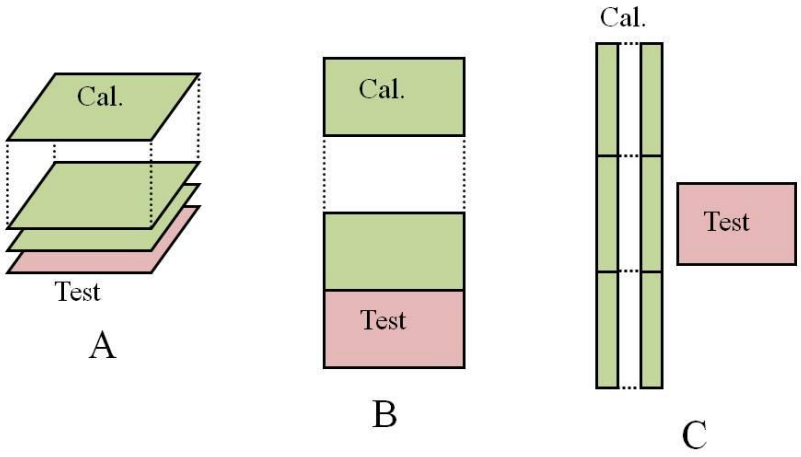


Fig. 1. Schematic representation of the way in which three different algorithms may organize second-order data. A) Joining the data matrices (calibration and test) into a three-way array. B) Placing data matrices adjacent to each other and creating a so-called augmented data matrix. C) Unfolding the calibration matrices into vectors, while keeping the matrix structure of the test sample data.

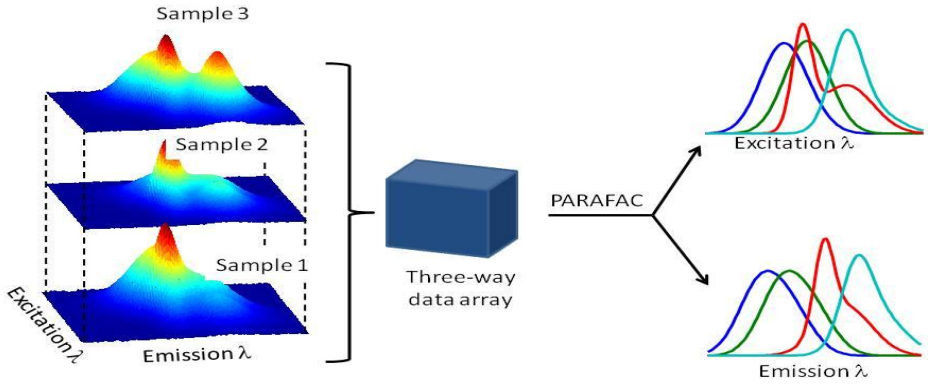


Fig. 2. Schematic representation of the PARAFAC operation, which builds a three-way data array with the different data matrices, and then decomposes the array into loadings in both instrumental data modes (in this case excitation and emission fluorescence loadings as a function of wavelength, contained in **B** and **C** matrices). Sample scores are also produced, containing information relative to constituent concentrations.

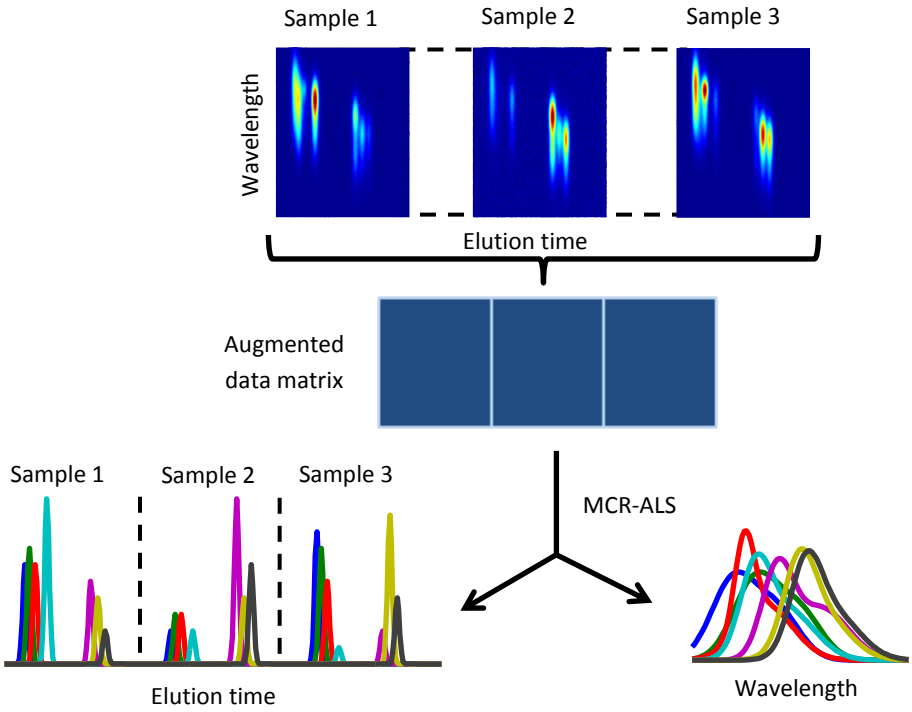


Fig. 3. Schematic representation of the operation of MCR-ALS. After building the augmented data matrix, the latter is decomposed into profiles in the augmented mode (the elution time mode in chromatographic-spectral data processing) and the spectra which are common to all samples.

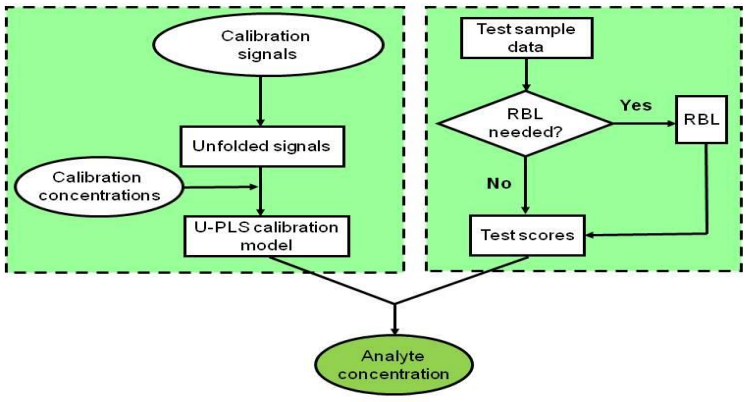


Fig. 4. Schematic representation of the U-PLS/RBL procedure.

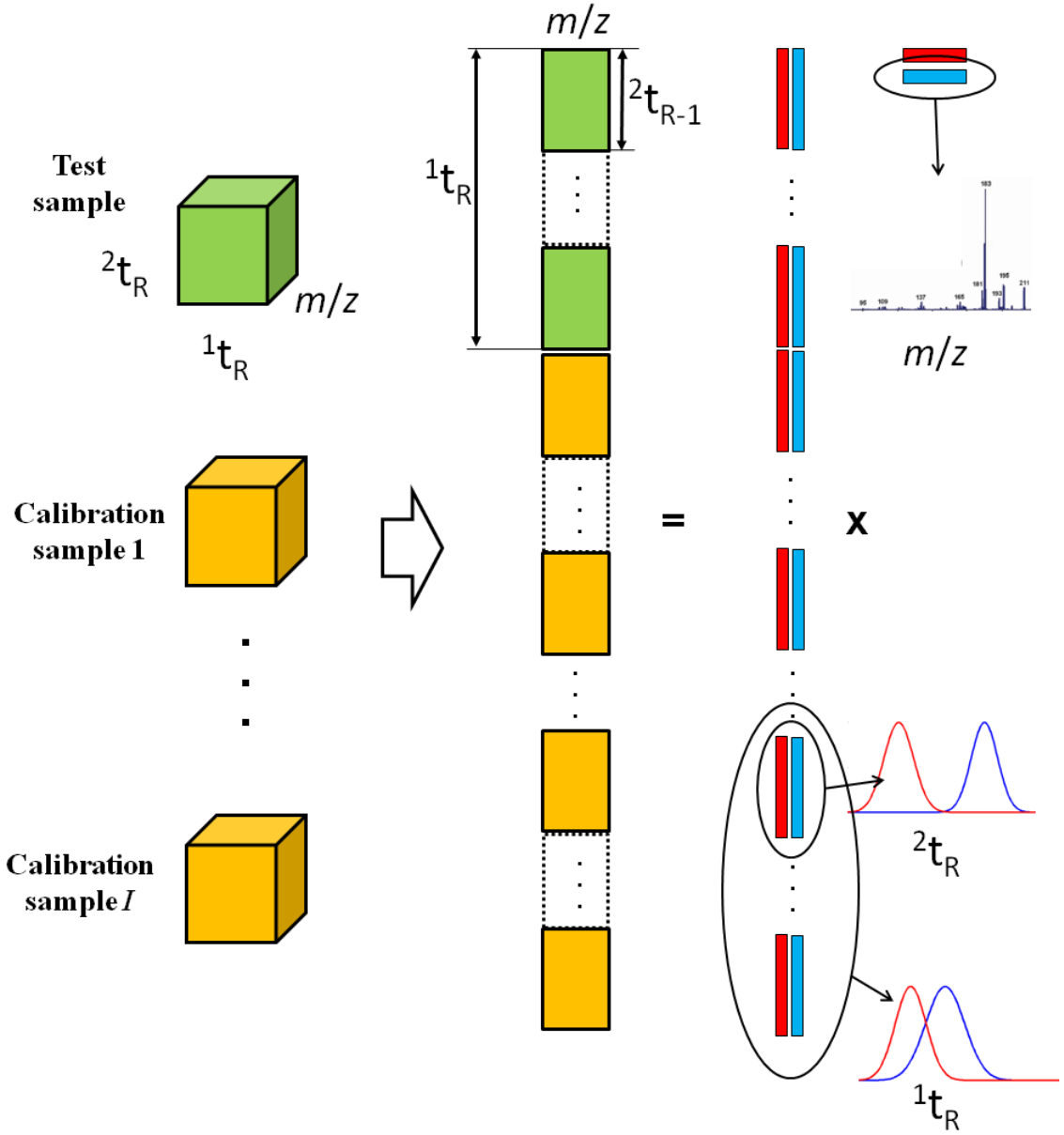


Fig. 5. Schematic representation of the MCR-ALS processing of GC-GC-MS third-order data in a two component system: each of the data arrays is first unfolded into a matrix, by concatenating the two temporal modes. The matrices are then joined into a single super-augmented matrix and processed, leading to spectra and unfolded elution time profiles. The latter can be refolded to get the individual elution time profiles along both temporal data modes. The second temporal mode is obtained directly from the retrieved temporal profiles, while the first temporal profile is obtained through the areas computed on the second-mode profile.

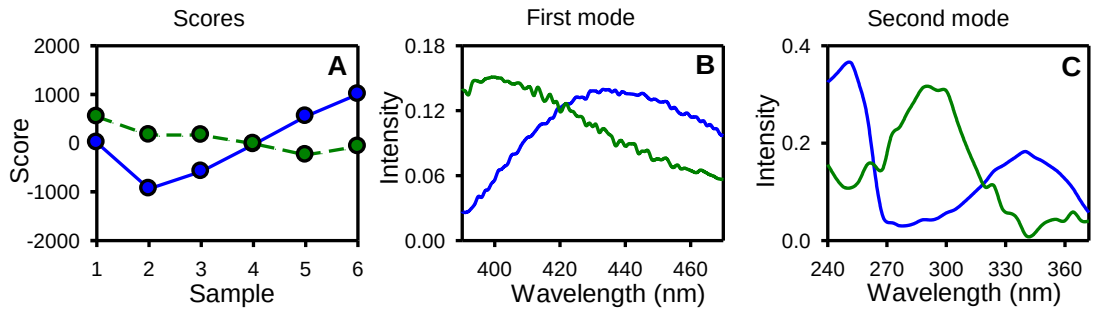


Fig. 6. PARAFAC scores (A) and loadings (B and C) for a natural stream sample added with bentazone. Blue and green lines and symbols correspond to constituents 1 and 2, respectively.

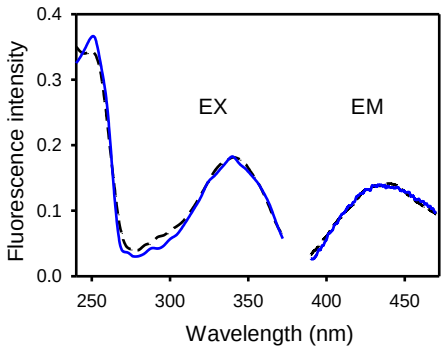


Fig. 7. Experimental excitation (EX) and emission (EM) spectra of bentazone (black dashed lines) and the corresponding PARAFAC loadings (blue solid lines). Loadings and spectra have been normalized to unit amplitude.

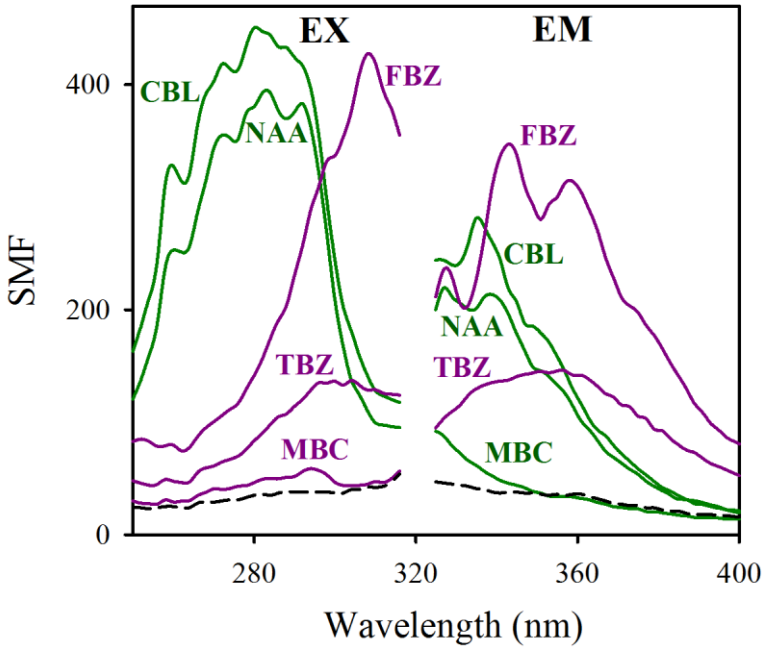


Fig. 8. Solid matrix fluorescence (SMF) excitation and emission spectra for thiabendazole (TBZ), fuberidazole (FBZ), carbaryl (CBL), 1-naphthalene acetic acid (NAA) and carbendazim (MBC) immobilized onto silica gel C18. The dashed black lines correspond to the background signals.

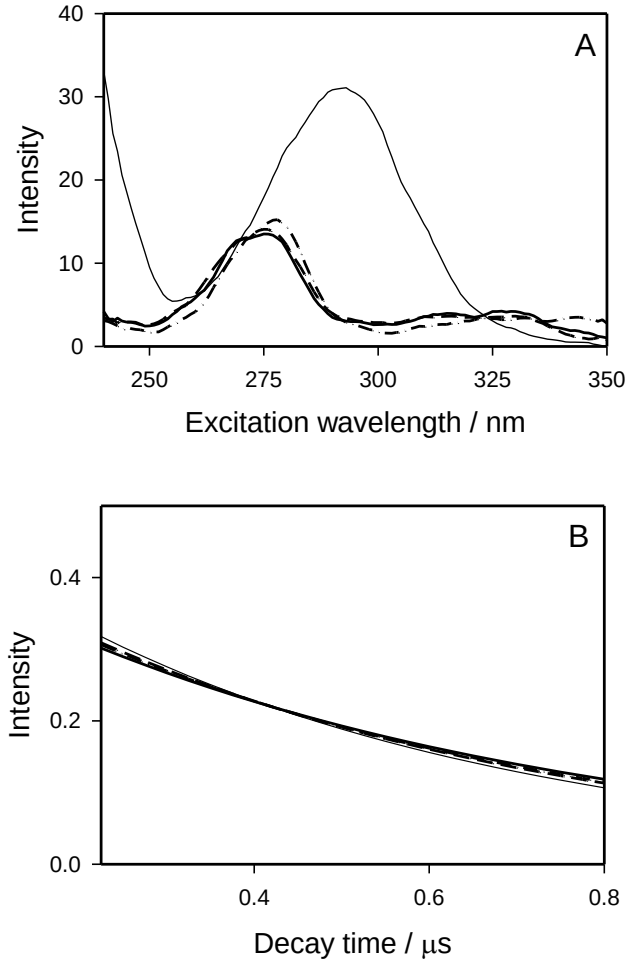


Fig. 9. A) Thick lines: excitation lanthanide-sensitized luminescence spectra for the terbium (III) complexes of the studied analytes in serum: ciprofloxacin, solid line, norfloxacin, dashed line, danofloxacin, dashed-dotted line. Thin line: the corresponding excitation spectrum for salicylate in serum. B) Time decay curves for all the constituents shown in part A), after normalization to unit length. Lines are denoted as in plot A). Reprinted with permission from [62]. Copyright 2009 Elsevier.

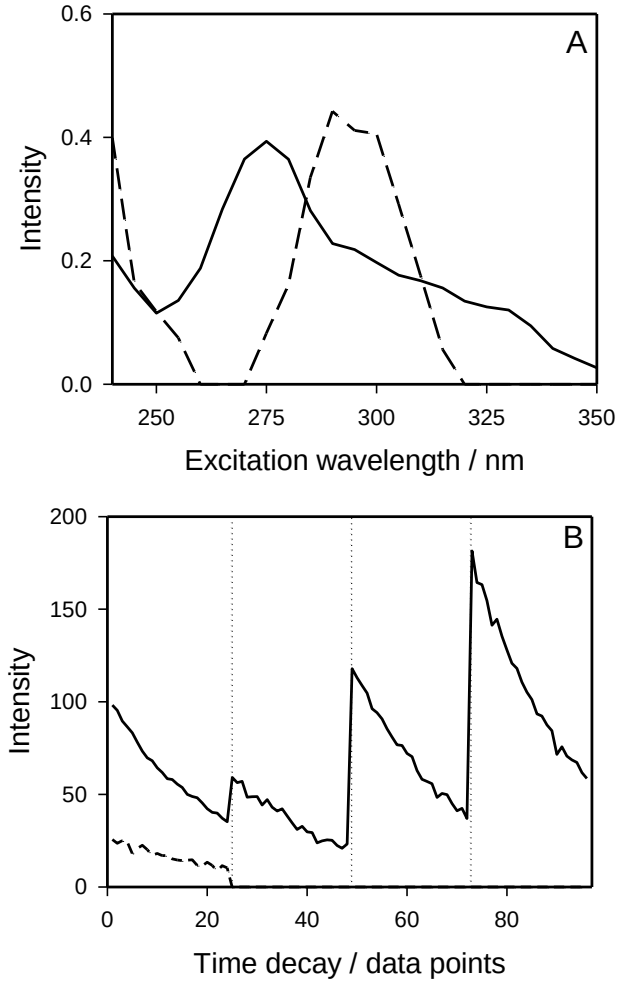


Fig. 10. MCR-ALS results for a serum sample spiked with ciprofloxacin and salicylate. A) Excitation lanthanide-sensitized luminescence profiles for the terbium (III) complexes of both sample constituents. B) Time-decay profiles for successive matrix samples in the standard addition mode. The dotted lines separate the different samples: from left to right, test sample and the results of subtracting the test sample data from the three standard additions. In all cases, the solid line indicates ciprofoxacin and the dashed line salicylate. The vertical scales are arbitrary. Reprinted with permission from [62]. Copyright 2009 Elsevier.

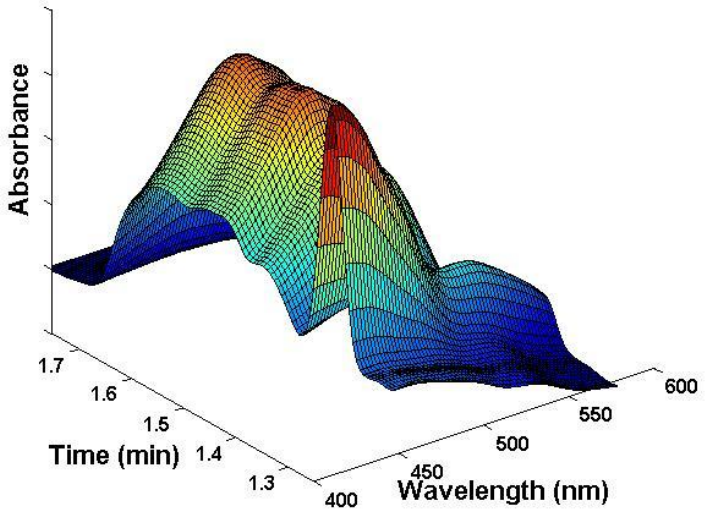


Fig. 11. Second-order data matrix generated by HPLC-DAD

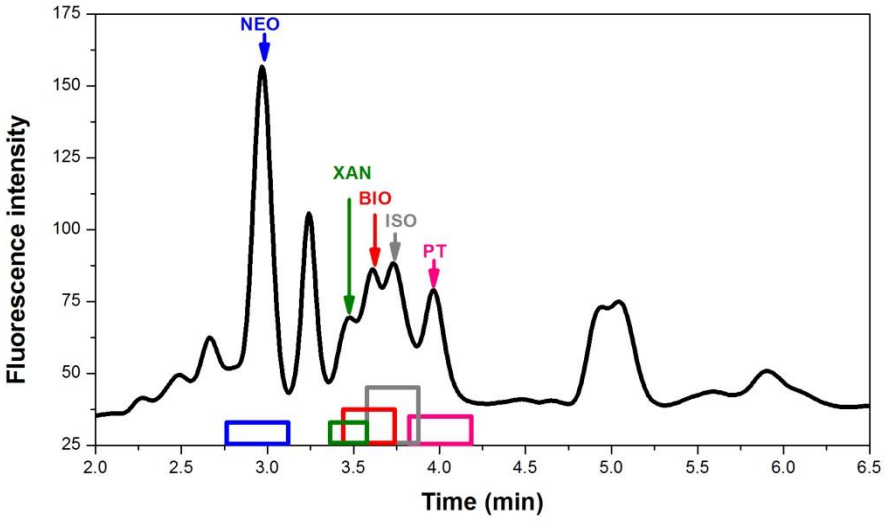


Fig. 12. Chromatogram registered at λ_{exc} = 272 nm, λ_{em} = 445 nm and elution time between 2.0 and 6.5 min for a urine sample after spiking with the five analytes. The boxes correspond to the five regions used the MCR-ALS analysis.

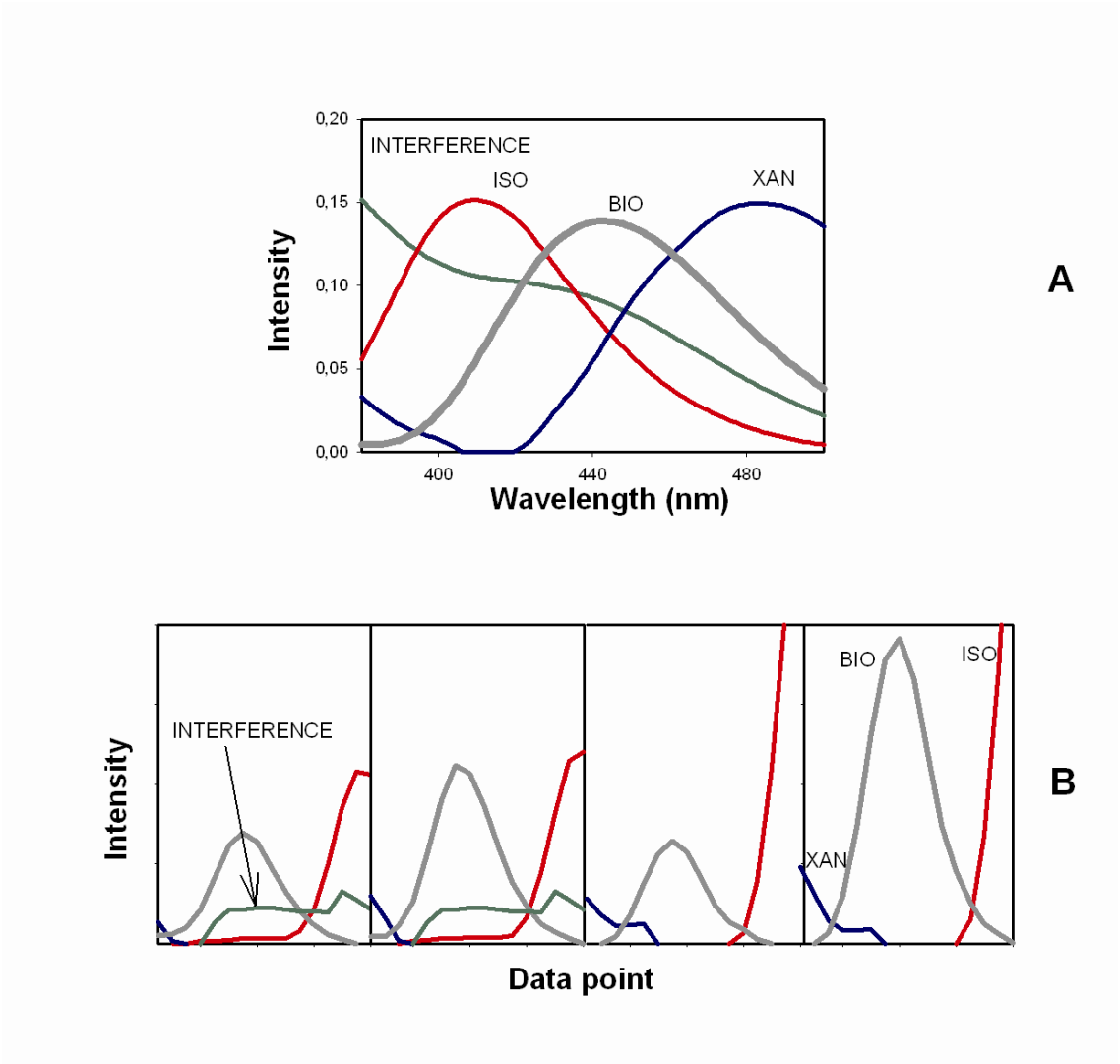


Fig. 13. A) MCR-ALS retrieved spectral profiles in region number 3: BIO, ISO, XAN and one interference, as indicated. B) Successive elution time profiles, corresponding to the MCR-ALS analysis of 4 samples in region 3, which includes the BIO peak in two urine samples from a healthy adult (two spiked concentrations), and two pure standard solutions (3.0 and 8.0 ng mL^{-1}, respectively). The remaining three profiles correspond to ISO, XAN, and one interferent.

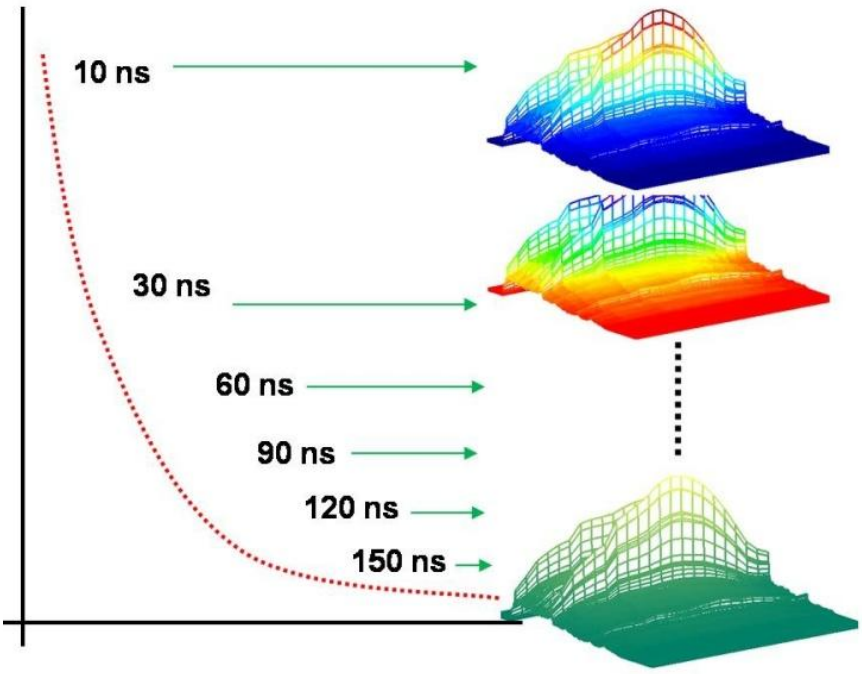


Fig. 14. Schematic representation of laser-excited time-resolved EEM fluorescence data collection.

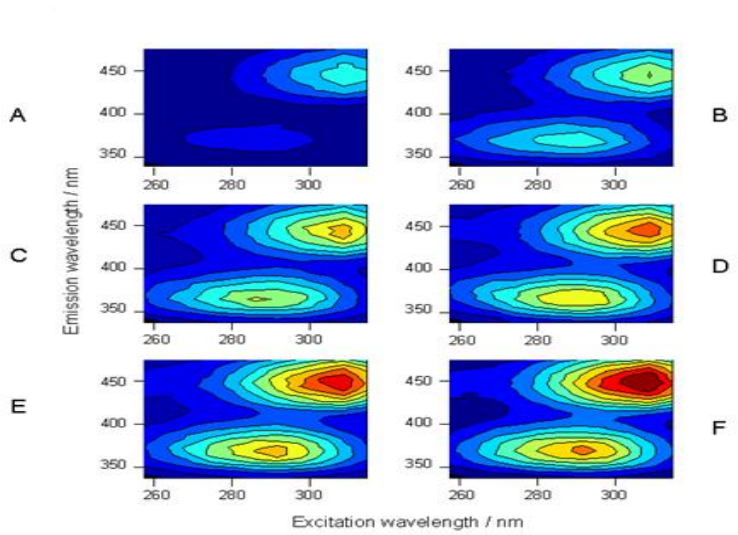


Fig. 15. Contour plots of the EEMs for an aqueous solution containing methotrexate and leucovorin as a function of the time of permanganate oxidation. The times selected for illustrating the kinetic evolution of the EEMs are (in min): A) 0, B) 2.4, C) 4.8, D) 7.2, E) 9.6 and F) 10.8. Fluorescence intensity has been coded in colors, with deep blue indicating the lowest value, and deep red the largest one. Reprinted with permission from [75]. Copyright 2004 American Chemical Society.

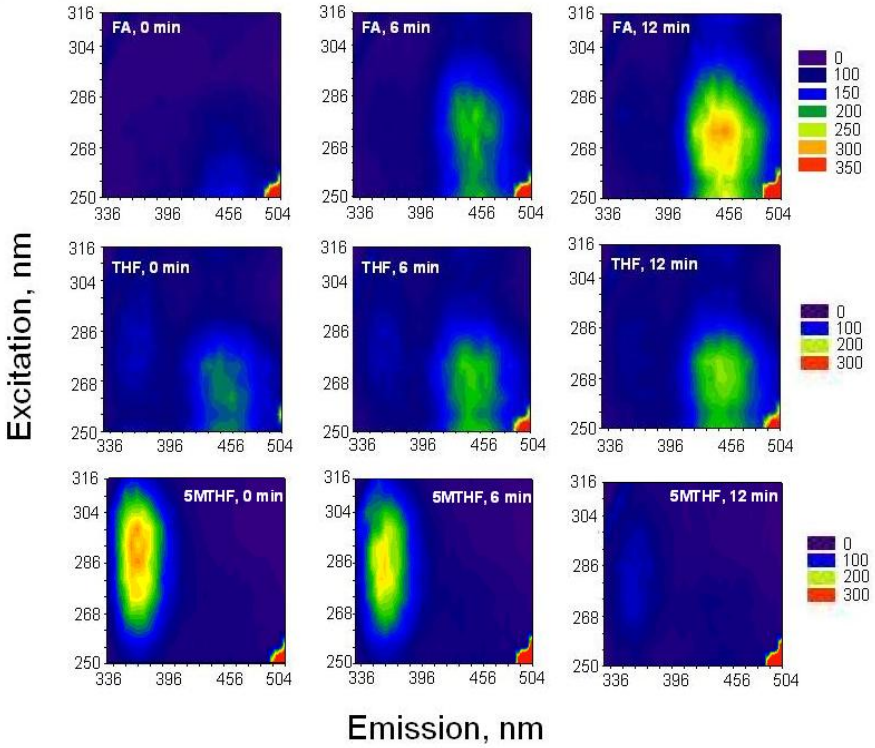


Fig. 16. Evolution of the contour plot of the excitation-emission matrices with the irradiation time, of solutions of folic acid (FA), tetrahydrofolic acid (THF) and 5-metryltetrahydrofolic acid (5-MTF). Reprinted with permission from [77]. Copyright 2008 Elsevier.

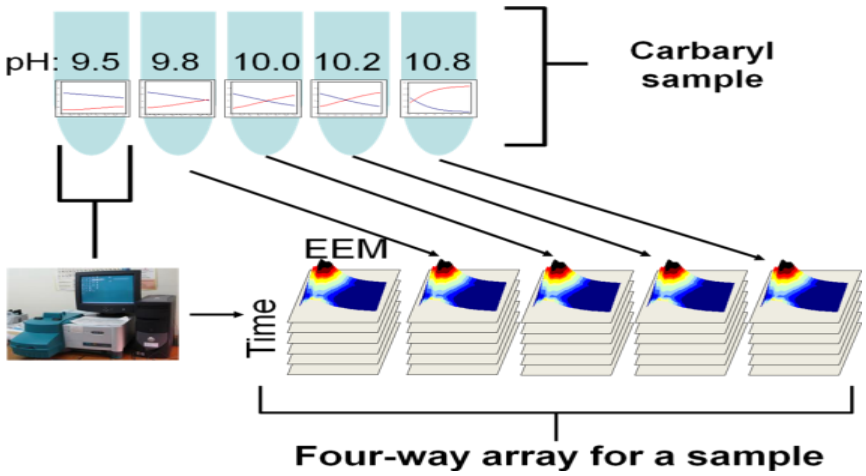


Fig. 17. Building of a four-way data array for a single sample of carbaryl.

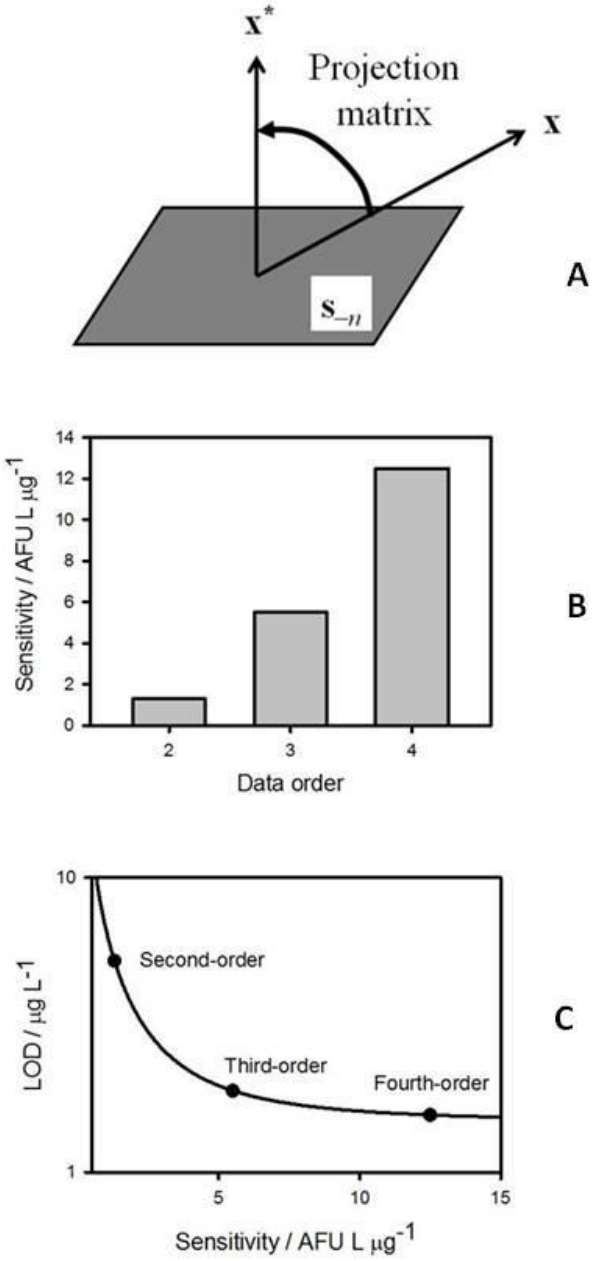


Fig. 18. A) Schematic illustration of the projection of a signal (**x**) orthogonal to the space spanned by the interferents, leading to the net analyte signal (**x***). B) Sensitivity as a function of order in an experimental system (see text). C) LOD as a function of sensitivity for various orders (see text).