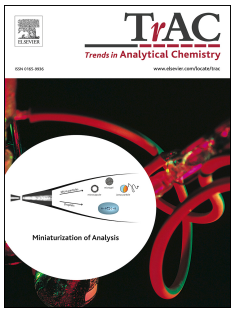
Accepted Manuscript



Third order chromatographic-excitation-emission fluorescence data: Advances, challenges and prospects in analytical applications

Milagros Montemurro, Gabriel G. Siano, Mirta R. Alcaráz, Héctor C. Goicoechea

PII: S0165-9936(17)30120-6

DOI: 10.1016/j.trac.2017.05.011

Reference: TRAC 14930

To appear in: *Trends in Analytical Chemistry*

Received Date: 31 March 2017

Revised Date: 12 May 2017

Accepted Date: 24 May 2017

Please cite this article as: M. Montemurro, G.G. Siano, M.R. Alcaráz, H.C. Goicoechea, Third order chromatographic-excitation-emission fluorescence data: Advances, challenges and prospects in analytical applications, *Trends in Analytical Chemistry* (2017), doi: 10.1016/j.trac.2017.05.011.

Third order chromatographic-excitation-emission fluorescence data: Advances, challenges and prospects in analytical applications

Milagros Montemurro[a,b], Gabriel G. Siano[a,b], Mirta R. Alcaráz[a,b,*] and Héctor C. Goicoechea[a,b,*]

[a] *Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, 3000 Santa Fe, Argentina*
[b] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 CABA (C1425FQB), Argentina.*

[*] Corresponding authors: Mirta R. Alcaráz, *E-mail*: malcaraz@fbcb.unl.edu.ar, *Tel.:* +54 342 4575206 x190
and
Héctor C. Goicoechea, *E-mail*: hgoico@fbcb.unl.edu.ar, *Tel.:* +54 342 4575206 x190

Abstract

Three analytical methodologies for the generation of third-order liquid chromatography-excitation-emission fluorescence matrix (LC-EEM) data are presented. Instrumental requirements were evaluated considering equipment complexity, costs and accessibility. A descriptive analysis of the generated data was done along trilinearity concept and chemometric resolution. For trilinear decomposition, PARallel FACtor Analysis (PARAFAC) model was utilized. Hence, possible effects that are caused in the resolution due to loss of trilinearity are detailed. Then, several data pre-processing and processing alternatives are proposed in order to successfully overcome the drawbacks that can be present in the chemometric resolution. Additionally, a literature analytical method for the determination of three analytes is presented to showcase the potential of the methodology to generate third-order LC-EEM data with quantitative aims. For data modelling, Augmented PARAFAC (*A*PARAFAC) and Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) were used. Both algorithms demonstrated to be able to bear non-quadrilinear data in a multi-set analysis.

Keywords: Third-order data; multi-way analysis; liquid chromatography; excitation-emission fluorescence matrix; trilinear decomposition.

1. Introduction

Over the last years, a remarkable growth in the number of chemometric applications in the analytical chemistry field has been noticed. The potential demonstrated for the combination of both disciplines has been accompanied by a tireless interest in the investigation of the advantages and benefits of multidimensional data analysis. In this matter, recently published works have proved that the increment in the number of instrumental modes represents a positive impact in the analytical properties of the methods, which is traduced into an improvement in the analytical figures of merit, essentially, in the sensitivity and selectivity in a multi-component system [1, 2].

For multivariate calibration, first- and second-order data have been extensively evaluated and countless analytical applications for a wide variety of multi-component systems have been reported. In this context, methods based on liquid chromatography (LC) with spectral detection coupled to second-order data modelling have proved to be an efficient and useful strategy for the analysis of complex samples in presence of several components [3]. One of the most remarkable benefits of second-order calibration methods is that tedious and long sample pre-processing steps are not strictly necessary due to the fact that second-order modelling can accomplish the so-called "second-order advantage" [4].

At present, there is an important number of ongoing investigations of multidimensional data analysis aiming to prove additional analytic advantages [5-8]. Thus, even though it is still in the beginning of its progress, high-order data analysis for analytical applications constitutes a field worth to be explored [9]. Although no agreement about its existence has been reached among the scientific community yet, some authors propose that additional advantages over the second-order advantage can be achieved in high-order multivariate calibration. Those additional advantages are characterized as the enhancement in sensitivity and selectivity, the possibility of relieving problems of collinearity and the feasibility of decomposing the data array for each sample individually, independent of other samples [10].

Although multidimensional instrumental signals are easy to be obtained with the available modern instrumentation, and several chemometric algorithms have been successfully developed to solve multi-way data problems, the way in which the data are generated may have a significant effect on the data structure and, in consequence, the final results. Hence, developing a method based on multidimensional data processing implies, among the development of the method itself, an in-depth study of the properties

and the characteristics of the obtained data in order to select both appropriate pre-processing strategies and the most suitable algorithms for the chemometric resolution. For this reason, it becomes crucial recognizing in advance the type of data by means of its mathematical properties and establishing the correct procedure for the analysis in order to achieve unequivocal results.

In univariate calibration, a very important concept to consider is the linearity, *i.e.*, the linear relationship between a dependent variable and an independent one. This concept is the basis of the validity of Beer-Lambert's law where the independent and dependent variables are the concentration and the measured signal, respectively [11, 12]. In this way, the first topic that must be considered for high-order data analysis is the multilinearity of the data. In third-order data analysis, in particular, it is important to know if the data array fulfils the concept of trilinearity, which must be evaluated in terms of the individual three-dimensional array for a single sample.

Trilinearity can be seen as an extension from the concept of linearity, where the linear relationship is given between a two independent variables and a dependent one. Then, trilinearity takes place when the three instrumental modes are independent of each other; therefore, if mutually dependent phenomena in more than two modes occur, the third-order array is a non-trilinear data [2, 13]. In sum, trilinearity is a concept that can be seen as an extension of the Beer-Lambert's law. As an example, it can be considered the second-order data generated by chromatography coupled to spectral detection, *e.g.*, three-way array built with several LC-DAD runs from different samples with the same composition. Here, a trilinear structure would indicate that the pure spectrum and the pure retention profile of an analyte remain invariant in the different experiments or runs. Considering that the experiments are performed under same experimental conditions, the spectrum of a pure compound does not change; however, lack of run-to-run reproducibility due to differences in peak shape and position of the pure retention profiles are usually observed. In consequence, lack of trilinearity occurs and the data must be considered as non-trilinear.

Furthermore, for four-way data generated from a set of data for multiple experiments, both trilinearity and quadrilinearity concepts for individual data cubes and multi-set data, respectively, ought to be evaluated. In this case, quadrilinearity can be seen as an extension of trilinear concept where the linear relationship is given between three independent variables and a dependent one. In case the individual data fulfils a trilinear model and no lack of quadrilinearity occurs in the four-way array, the data are

classified as quadrilinear. On the contrary, a further subdivision can be done considering the number of quadrilinearity-breaking modes [14]. Then, it is possible to distinguish 4 types of non-quadrilinear data, whose are schematically presented in the classification tree, which has been introduced by *Olivieri and Escandar* (Fig. 1).

Insert Fig. 1

The correct selection of the mathematical model and algorithm is influenced, in one sense, by the characteristics and the properties of the generated data. In the literature, there are a vast number of available algorithms that can be utilized for data processing. Algorithms based on Alternating Least Squares (ALS) are the most employed for second- and third-order data resolution, either for descriptive or predictive analysis, being PARallel FACtor Analysis (PARAFAC) [15] and Multivariate Curve Resolution (MCR) [16] the most representative ones. Besides, algorithms mainly used for quantitative purposes are based on Partial Least Squares (PLS) [17, 18] resolution, and the second order advantage is achieved by application of a Residual Bi-Linearization procedure (RBL) [19]. Unfolded and multi-way PLS coupled to RBL procedure (U-PLS/RBL and N-PLS/RBL) are examples of the latter. Finally, there is a family related with the Alternating Trilinear Decomposition (ATLD) algorithm, which was firstly developed by *Wu et al.* in 1998 [20]. ATLD is an iterative algorithm with similar characteristics to PARAFAC. It is commonly used by virtue of the advantages of being insensitive to excessive component number, fast convergence and fully exploiting the second-order advantage.

In this review, a comparative study of three different third-order liquid chromatography-excitation-emission fluorescence matrix (LC-EEM) data generation approaches was carried out. Moreover, three methods based on identical chromatographic conditions but coupled to different fluorescence excitation and emission detection systems for the quantitative analysis of antibiotics in aqueous matrices are here discussed.

2. Analytical procedures

The methodology generally used to generate third-order LC-EEM data consists on a chromatographic procedure coupled to excitation-emission data matrix detection. At present, to the best of our knowledge, only two strategies to generate third-order LC-EEM data have been reported. One of these approaches is based on the collection of discrete fractions at the end of the chromatographic procedure with the subsequent

excitation-emission data matrix registering of each collected fraction [21-23]. In the second procedure, multiple aliquots of a given sample are injected into the chromatograph and the retention time-emission spectra matrix of each injection is recorded using different excitation wavelength [24-26]. In both cases, the three instrumental modes are retention time, excitation and emission wavelengths.

Besides the aforementioned approaches, another way to generate third-order LC-EEM data is described in the present review, where a fast-scanning spectrofluorimeter with a flow-cell connected at the end of the LC instrument is utilized.

It is worthwhile mentioning that even though the first two strategies above-mentioned have been thoroughly described elsewhere [21-26], they were developed for different analytical purposes. Therefore, to make an appropriate comparison and reach reliable conclusions, it becomes necessary using an analytical system with similar particularities, which permits the evaluation of the instrumental characteristics and the generated data properties avoiding as much as it is possible the effects that can be caused by the inherent features of the system. In this regard, all the cases evaluated in the present review were carried out by using the same general chromatographic procedure, *i.e.*, same LC instrument under identical separation conditions (column and mobile phase composition), but changing the detection methodology. Then, solutions containing the same analytes were evaluated by using the three analytical procedures. (For a better understanding, some specific properties of the procedures will be depicted). It must be clarified that samples containing different number of analytes were used for each methodology due to the complexity of the generated data, which is further demonstrated.

2.1. Methodology I –Collection of fractions

The first methodology described (*MI*) was firstly proposed by *Bro* for a qualitative study [23] and it has been recently reported by *Alcaraz et al*. [21] for quantitative purposes. It consists on an instrumental analytical system that includes an automated custom-made device connected at the end of the chromatograph, allowing the collection of several discrete fractions in 96-well plates, whose are commonly used for ELISA test. Upon completing the chromatographic procedure and collecting all the fractions in the 96-well plate, the plate is placed into a spectrofluorimeter that is equipped with a plate reader. Thus, the excitation-emission matrices are separately measured, obtaining one matrix for each collected fraction [21].

Here, for the analysis of a ternary solution, containing ofloxacin (OFL), ciprofloxacin (CPF) and danofloxacin (DNF), 17 discrete fractions were sampled from the LC instrument. Each EEM was then measured in the range of 260-340 nm and 380-500 nm for excitation and emission spectra, respectively. Fig. 2 summarizes the data generation using *MI* methodology.

Insert Fig. 2

2.2. Methodology II – Multiple chromatographic runs

Two different applications using the methodology II (*MII*) for third-order LC-EEM data generation have been further reported. In general, this methodology consists in the injection of several aliquots of a given sample into a chromatograph. For each aliquot, the retention time-emission spectra data matrix is registered using different excitation wavelength.

Using this methodology, the analysis of green pigments in olive oil samples was performed by injecting 8 aliquots of a given sample [24], and 6 injections were utilized for the pesticides evaluation in fruits [26]. In the present review, and with the aim of making a fair comparative analysis, binary solutions containing OFL and CPF were employed. Then, 10 aliquots per sample were injected and the emission spectra were registered in the range of 380-500 nm at each retention time, using excitation wavelengths ranging from 260 nm to 305 nm. In Fig. 3, the data generation using *MII* methodology is shown.

Insert Fig. 3

2.3. Methodology III – On-line excitation-emission matrices

Methodology III (*MIII*) comprises the measurement of several consecutive excitation-emission matrices by using a chromatograph-spectrofluorimeter hyphenated system. Thus, neither flow interruption nor collection of fractions is required. For the fluorescence matrix registering, a fast-scanning spectrofluorimeter with a flow cell connected at the end of the LC instrument is used. Besides, in order to avoid time lags that may occur from triggering inaccuracies, a controller enabling the synchronization between instruments becomes necessary. It is important to highlight the fact that this approach, to the best of our knowledge, has not been employed for LC-based applications yet.

With the purpose of comparing methodologies, solutions containing CPF were analysed. Considering the fact that the spectrofluorimeter allows registering a complete excitation-emission matrix in a reasonably short time, an acceptable number of matrices (15) per sample were acquired, covering the excitation and emission range of 260-300 nm and 390-490 nm, respectively. The data generation using *MIII* methodology is represented in Fig. 4.

Insert Fig. 4

3. Descriptive evaluation: requirements, properties and data modelling

3.1. Instrumental requirements

In order to evaluate different strategies for multidimensional data generation and to analyse the properties of the data obtained, three instrumental arrangements based on chromatographic separation coupled to excitation-emission fluorescence matrix detection are proposed. In this section, a comparative study between the three instrumental approaches is presented, evaluating equipment complexity and the number of the required instruments for each arrangement. The time of analysis consumed per sample was also considered in this study.

3.1.1. Methodology I

To perform an analysis utilizing *MI*, a conventional LC instrument and a spectrofluorimeter equipped with a well plate reader are required. Additionally, an automated device for the collection of individual fractions in 96-well plates is demanded. For chromatographic separation, the flow rate must be properly selected in order to ensure the appropriate collection of the fractions, leading to an accurate volume distribution in the wells of the well plates. Furthermore, the time demanded for each fraction must represent a volume that guarantees both the chromatographic resolution previously achieved and the proper matrix reading in the spectrofluorimeter.

The first disadvantage that can be clearly noticed for *MI* is the use of a device for the collection of fractions in a multi-well plate. However, even though it would represent an instrumental restriction, an automatized custom-made device can be easily built in the laboratory, as it has been reported in previous works [21, 22].

The time consumed for the total analysis of a ternary solution was approx. 42 min, including both the chromatographic procedure (2 min) and the recording of 17 fluorescence matrices (40 min). As can be seen, the considerably long time demanded

for each sample makes *MI* an inappropriate alternative for the study of unstable analytes or volatile solutions. Even though it would be possible to reduce the time of the analysis by using a fast-scanning spectrofluorimeter for the matrix recording step, although the complexity and the cost of the equipment will be incremented. On the other hand, despite it is time-consuming, *MI* requires small amount of sample and solvents resulting in a method included within the framework of the green chemistry [27].

3.1.2. Methodology II

Only a LC instrument is required to perform an analysis with *MII*. Since several aliquots for a given sample are consecutively injected and the retention time-emission spectra matrices using different excitation wavelength are registered, an auto-sampler and a fast-scanning fluorescence detector (FSFD) modules for the LC instrument are thus needed. In this manner, despite only one instrument is required to obtain third-order LC-EEM data, the modules needed are not usually present in a conventional LC instrument.

In this work, the time spent for the evaluation of a CPF and OFL solution was approx. 40 min, remarking the fact that the chromatographic run for each aliquot took only 2 min. Therefore, *MII* is highly time-consuming and can only be improved in spite of a detriment in the excitation spectra quality, *i.e.*, loss of spectral resolution and/or reduction of the spectral range. Thus, same as *MI*, *MII* results unsuitable for the evaluation of unstable samples or volatile solutions. On the other hand, the multiple injections that are necessary for a given sample demand important amounts of sample and solvents, making *MII* an expensive alternative and a method that does not conform to the principles of green chemistry [27].

3.1.3. Methodology III

The new methodology here evaluated (*MIII*) comprises a combination of two analytical instruments in tandem, where a quartz flow-cell is connected at the end of a LC instrument and placed into a spectrofluorimeter, which must be able to accomplish real-time measurements at multiple wavelengths. It should be noticed that fluorescence matrices are taken in a finite time, which in chromatography means that the analyte concentration at the beginning of the matrix registering is different than at the end, as it happens, in a lower degree, for second-order LC-FSFD data generation [2]. In consequence, the emission and excitation spectra are dependent on the chromatographic

retention time. In this regard, in order to collect a complete fluorescence matrix in the shortest time possible as well as to diminish the effect of dependence modes phenomenon, a fast-scanning spectrofluorimeter is the principal requirement of this methodology. Additionally, a conventional LC instrument is used for the chromatographic procedure, where sophisticated detectors or auto-sampler module are not strictly necessary.

The first point to stress is that the time of the total analysis is defined by the performed chromatographic method due to the fact the fluorescence matrices are registered in parallel with the chromatographic procedure. Here, the evaluation of a solution containing one analyte was carried out in 5 min, obtaining a total of 15 complete fluorescence matrices. For these reasons, *MIII* is presented as an alternative that allows obtaining third-order LC-EEM data in a very short time, without requiring large amount of samples and reagents, as it happens with *MI*, which is one of the principles of green chemistry [27].

3.2. Data properties

In this section, a qualitative analysis of the data obtained with the three methodologies was carried out with the aim of evaluating whether the data for a single sample are trilinear or not. Moreover, different data processing strategies that can be applied to cope with the data obtained are depicted.

3.2.1. Methodology I

First, it must be considered that the collected fractions do represent the corresponding retention time of each analyte in the sample. Hence, to be able to rebuild the temporal profile, both the waiting time in each well and the initial collection time should be known.

A particularity of *MI* is the fact that the excitation-emission matrices registered for each well are independent of each other, which means that the emission and excitation spectra only depend on the analyte properties and its surrounding medium, and the intensities are given by the abundance of the analyte. So, considering a single substance and a chromatographic system operating in isocratic mode, the composition of the surrounding medium remains unchanged from the beginning to the end of the analysis and, in consequence, the emission and excitation spectra of the analyte will be identical in all the wells where it is present, but differing in its intensity as consequence of the

chromatographic dispersion. In this manner, and taking into account that the excitation-emission matrices (in absence of inner filter) are intrinsically bilinear, the third-order LC-EEM data obtained with *MI* are trilinear due to the fact that the three data modes (excitation wavelengths, emission wavelengths and retention times) are independent of each other. Fig. 5.A shows the LC-EEM data obtained for a ternary sample using *MI*.

On the other hand, an important issue to consider in multi-way data is the number of data points obtained in each instrumental mode. In this case, the third-order array comprises 17 × 17 × 25 data points for times, excitation and emission wavelengths, respectively. Although it can be considered as an array with balanced number of data points, only 17 discrete fractions were collected from the LC instrument, which leads to a low resolution in the retention time mode. However, time resolution could be improved minimizing the collection waiting times or using multi-well plates with a higher number of reduced volume wells.

3.2.2. Methodology II

The most important aspect needing to be addressed for *MII* is that the excitation spectra are result of the multiple aliquots injected for a given sample. Hence, the covered spectral region and the spectral resolution are directly dependent on the number of analysed aliquots. Thus, excitation spectra are obtained from the time-emission wavelength data matrices, meaning that it is possible to build a two-dimensional retention time-excitation wavelength matrix with the chromatographic profiles registered at the same emission wavelength (Fig. 5.B). However, this is only possible if the retention times among runs are reproducible, otherwise, a lack of run-to-run reproducibility would lead to misinterpretations of the excitation spectra. Besides, a lack of run-to-run reproducibility brings a loss of trilinearity in third-order data, phenomenon that derives from the fact that the times and excitation wavelength modes are mutually dependent. This fact can be analogously pictured as a three-way array built with LC-DAD second-order data corresponding to different samples, where sample-to-sample peak shifting are observed [28]. In sum, third-order data generated with *MII* are trilinear only if perfect reproducibility in peak times among runs are observed for a given sample, but also if the shape of the peaks remains invariant.

Finally, regarding the number of data points in each instrumental mode, for the present application example, only 10 wavelengths were registered in the excitation wavelength mode, while 150 and 45 times and emission wavelengths, respectively, were

recorded in the other modes. Thus, it is clearly shown the low resolution in the excitation wavelength mode, which could be a disadvantage for the analysis of multi-analyte systems with either highly overlapped fluorescence signals or strong differences between wavelengths of maximum fluorescence intensity. Moreover, it must be considered that an enhancement of the excitation spectrum quality requires an increment of the number of injections and, in consequence, an increment of the solvent and sample consumption as well as time of analysis.

3.2.3. Methodology III

The most noticeable advantage of *MIII* is that the EEM are recorded simultaneously with the LC procedure, entailing a drastically reduction of the total time of analysis. On the other hand, the first drawback to overcome is that, since the fluorescence matrices are registered in a finite time, both the emission and the excitation wavelength modes are dependent on the chromatographic retention time mode. However, due to the fact that emission wavelengths are scanned in a considerably short time (less than 1 s), the consequent effect of the dependence between emission wavelength and retention time modes is negligible. That is not the case for the excitation wavelength mode where the time required for a total spectrum scan may take on the order of seconds. Therefore, the third-order data obtained with *MIII* does not fulfil the concept of trilinearity.

In the light of the preceding, at least three strategies can be proposed to overcome the lack of trilinearity: 1) instrumental improvement: by using a spectrofluorimeter enabling faster fluorescence measurements; 2) pre-processing procedure: by applying mathematical procedures to transform the data into a trilinear data array; 3) data processing: by using chemometric algorithms that handle non-trilinear data. Unfortunately, none of these three approaches are suitable current options, since highly sophisticated equipment are not easily available in a routine laboratory and new chemometric algorithms have not been developed yet. In Fig. 5.C, LC-EEM data obtained for a pure analyte using *MIII* are depicted.

Regarding the number of data points in the instrumental modes, for this application, a total of 15 complete fluorescence matrices per sample were obtained. Moreover, compared with *MI*, smaller excitation and emission spectral ranges, as well as lower spectral resolution, were used in order to reduce the time required for the registering of a complete fluorescence matrix. As a result, although an array with balanced number of data points is obtained for each sample (15 × 15 × 28), both retention time mode and

excitation and emission wavelength modes show low resolution considering a chromatographic procedure and complete excitation-emission matrices. Nevertheless, retention time resolution can be enhanced in spite of a detriment in the spectral resolution, even if the latter can also be improved by using a spectrofluorimeter that would permit faster spectra scanning.

** Insert Fig. 5**

3.3. Data analysis

This section aims to chemometrically demonstrate the properties described in *data properties* section above. For this purpose, PARAFAC was employed as chemometric tool for the data modelling. PARAFAC is a trilinear decomposition algorithm (TLD) that, from the analytical chemistry standpoint, relies on the validity of Beer-Lambert's law of the investigated spectroscopic system. The decomposition of the data is made into trilinear components and it is achieved through alternating least-square procedure [15, 29]. This algorithm was selected because: 1) only trilinear data can be decomposed properly; 2) the retrieved profiles bear physically recognizable information; and 3) resolutions are often unique [30]. Hence, knowing in advance the real characteristics of the system, *i.e.*, excitation and emission spectra and chromatographic retention time of pure analytes, it would be possible to achieve reliable conclusions about the chemometric resolution. Although pre-processing procedures to cope with non-trilinear data are here described, the chemometric modelling was accomplished with non-pre-processed data in order to evaluate the effects on the results when lack of trilinearity, if present, is underestimated.

PARAFAC profiles retrieved from the decomposition of the third-order LC-EEM data obtained with the three methodologies are exposed in Fig. 6. For the modelling, initial estimates obtained by random initialization were used and only non-negativity constraint was applied (in the three modes) during optimization. The number of components was determined by core consistency diagnostic analysis (CORCONDIA) [31].

** Insert Fig. 6**

3.3.1. Methodology I

For all the samples, the number of components was 3, which agrees with the number of spectroscopically active compounds in the samples. Comparison analysis revealed

excellent agreement of the PARAFAC spectral profiles retrieved with the real spectra of the pure analytes. Additionally, peak times of each analyte obtained from PARAFAC retention time profile were correlated with DAD-UV reference chromatogram. This analysis showed a high degree of similarity between times, although slight differences were observed due to time lags among detection systems. On the basis of these results, it is possible to conclude that the third-order data array obtained with *MI* fulfil the trilinearity model. Fig. 6.A shows PARAFAC results retrieved from the decomposition of third-order LC-EEM data obtained with *MI.*

For multi-set analysis, the third-order data arrays obtained for each sample are usually arranged into a four-way data array. Thus, the quadrilinearity of four-way objects should be evaluated. In the presented case, loss of quadrilinearity was shown due to lack of reproducibility in retention times and the small differences between times of the collection of the fractions among samples. These facts lead to a non-quadrilinear data of type 1, according to the classification tree described by *Olivieri and Escandar* [14] (see Fig.1). Therefore, PARAFAC would not be the appropriate algorithm for the resolution. Instead, algorithms such U-PLS/RTL, MCR-ALS and *A*PARAFAC can be conveniently applied to unfolded bilinear data matrix or augmented trilinear three-dimensional data arrays [21, 22].

3.3.2. Methodology II

In chromatography, the ideal situation is when excellent reproducibility in peak times among runs is observed and also when the shape of the peaks remains invariant. In a real situation, these effects are not always accomplished, thus, the trilinearity of the third-order data array is not fulfilled. A way to overcome this drawback is utilizing mathematical procedures to turn the data into trilinear before performing data processing. In this regard, there are methods that digitally correct the chromatograms by correcting the chromatographic peaks into the same position and shape. Some of these methods, such as interval-correlation-shifting (*i*-coshift) [32], are capable of aligning peaks but not modifying peak shapes, whereas more sophisticated methods, *e.g.*, Correlation Optimized Warping (COW) [33], are able both to shift and stretch/compress peaks until best correlation between data is achieved. However, the available procedures at present cannot cope with the situation if long peak shifts or severe shape distortions occur. Additionally, the complexity of the system under study increases under high-overlapping condition or in presence of unexpected compounds [34, 35]. Recently, an

alternative data processing based on a combination of second order resolution algorithms coupled to a peak alignment procedure was proposed to tackle retention time shift problems in second-order data [36]. Even though this strategy was planned for second-order data resolution, it seems to be a clever alternative that can be applied for the resolution of non-trilinear three-dimensional data array with lack of retention time reproducibility.

For *MII* data set (obtained herein for binary samples), the number of components was ranging between 2 and 4. The difference between the numbers of components obtained (2-4) and the number the spectroscopically active compounds in the sample (2) lies in the effects generated by the lack of run-to-run reproducibility, *i.e.*, peak shifting. In Fig. 6.B, 3 components can be distinguished with marked features in the retention time and the excitation wavelength modes. However, 2 of the 3 profiles obtained for the emission wavelength mode show strong similarities. Additionally, excitation spectral profiles determined by PARAFAC do not match the spectra of the pure analytes. These unreliable solutions indicate a significant loss of trilinearity, which should be considered in advance for a successful resolution.

For quantitative analysis, N-PLS/RTL [24], U-PLS/RTL[24, 26] and MCR-ALS [26] algorithms have been utilized for chemometric resolution obtaining better results than those obtained by PARAFAC [24, 26]. In those reports, the authors have reached the conclusion that, for multi-set analysis, the better results are achieved due to the fact that the first-mentioned algorithms can tolerate times shifts among samples, whereas PARAFAC cannot cope with non-quadrilinearity data array in means of loss of sample-to-sample reproducibility [24, 26]. Then, the authors consider the data as non-quadrilinearity data of type 1. Also, it is interesting to note that, even though the same phenomenon occurs, lack of run-to-run reproducibility effect (for one sample) has not been evaluated, then, the extent artefacts that are introduced in the results due to the loss of trilinearity of the individual three-dimensional data objects have not been considered [14]. These observations lead to the conclusion that data set obtained with *MII* are included within the type 4 non-quadrilinearity class, instead of type 1, as they were considered. However, satisfactory results were achieved when U-PLS/RTL or MCR-ALS were used due to the low degree of non-trilinearity/quadrilinearity of the data array and the internal structure flexibility of the utilized algorithms.

3.3.3. Methodology III

As it was stated above for *MIII* data, there is a strong retention time mode-dependence with both spectral wavelength modes, not fulfilling the concept of trilinearity. This phenomenon is demonstrated, in principle, when the number of components is calculated, indicating that more than 1 component is necessary to explain the variance of the modelling when a pure analyte is analysed. In Fig. 6.C, it can be seen that, for a unique substance, 2 different temporal profiles and 2 excitation spectral profiles were obtained, while 2 identical emission spectral profiles were retrieved. This fact asserts the assumption that excitation mode is strongly dependent on the retention of the analyte, while the retention-dependence of the emission mode seems to be inconsequential. Additionally, for multi-set analysis, time shifting between samples leads to differences in the peak positions as well as in the features of the excitation profiles, showing a severe loss of quadrilinearity. Here, following the classification tree for four-way data for a set of samples [14], and considering the lack of trilinearity of the three-dimensional array for an individual sample, the generated data, like *MII*, are included in the category of non-quadrilinear data of type 4.

It is remarkable the high complexity of the third-order LC-EEM data generated with this methodology as consequence of the strong dependence of the instrumental modes. Unfortunately, no chemometric algorithms allowing a proper resolution of this kind of data have been developed yet, and no pre-processing tools to turn the data into trilinear have been further evaluated. Besides, it is noteworthy that same phenomenon occurs when fluorescence matrices are measured as function of reaction time. However, works published at the present do not report major inconvenient in the chemometric resolution mainly due to the low rates of the studied reactions in combination of the use of a fast-scanning spectrofluorimeter [37-42].

Accordingly, the development of new chemometric algorithms and the search of novel alternatives to cope with this kind of data represent an important and worthwhile challenge for chemometricians, as well as an exceptional step forward for chemometrics in the analytical chemistry field.

4. Analytical application

On the basis of the above-mentioned observations, it can be assumed that methodology I is the most feasible and efficient strategy for the generation of third-order LC-EEM data up to the present. Thus, with the goal of illustrating the capability of the *MI*-based analytical method for quantitative determinations, a recently published

work reporting an analytical method for the determination of 3 f-QUI in drinking water is here analysed [21, 22]. *A*PARAFAC and MCR-ALS have been chosen as chemometric data modelling algorithms and evaluation of algorithm performance has been accomplished. Additionally, second- and third-order data modelling was compared in terms of figures of merit and predictive ability [1, 2].

4.1. *MCR-ALS modelling*

MCR-ALS is a widespread and versatile soft-modelling technique that focuses on the mathematical resolution of the pure component signals of a data matrix [43, 44]. MCR-ALS enables decomposition of data matrices that can be described by a bilinear model, even when no prior information is available [45]. Its basic premise lies in the validity of Beer-Lambert's law of the investigated spectroscopic system, thus, profiles obtained for the pure components after resolution gain chemical meaning and they can be directly interpreted as abundance profile and spectra [46].

Bilinear model follow the expression that is shown in Eq. 1, where **X** is a two-way data matrix and **C** and **S** are the abundance distribution and spectra, respectively, of the *N* components involved in the system. Additionally, an **E** matrix comprising the residual variations of the data is obtained [43, 45, 46];

$$\mathbf{X} = \mathbf{C}\,\mathbf{S}^{\mathrm{T}} + \mathbf{E} \qquad \text{Eq. 1}$$

Multi-set data analysis, obtained from multiple experiments related to each other, can be accomplished through the extension of the model. Here, multi-set data are simultaneously analysed applying MCR-ALS to augmented data matrices [45, 47]. In this regard, MCR-ALS analysis is significantly improved and better description of the system can be done.

4.1.1. Data structure

For third-order data modelling, MCR-ALS resolution, showed in Fig. 7, is usually performed in the extended version using unfolded matrices as follows [21, 22, 25, 26, 48, 49]:

- Each EEM matrix $\mathbf{X}_{\mathrm{f}}$ ($K \times L$) corresponding to the collected fractions are unfolded generating row vectors $\mathbf{x}_{\mathrm{un,\,f}}{}^{\mathrm{T}}$ of dimension ($1 \times LK$). Then, the unfolded matrices, or row vectors, $\mathbf{x}_{\mathrm{un,\,f}}{}^{\mathrm{T}}$ are appended obtaining a bilinear matrix $\mathbf{X}_{\mathrm{unf}}$ ($J \times LK$) for each sample, with *J* fractions (retention times), *K* emission wavelengths and *L* excitation wavelengths. Therefore, all the obtained

$\mathbf{X}_{unf}$ matrices are then combined to a column-wise data array $\mathbf{X}_{aug}$ of size [(I + 1) $J \times LK$], in which I is the number of calibration samples and 1 represents the unknown, test or validation sample. In this regard, the augmented two-dimensional array conforms to the bilinear modelling requirements, since augmentation is done along the quadrilinearity-breaking mode, *i.e.*, column wise.

- Non-negativity, unimodality and correspondence between common species in different data matrices are the most used constraints applied to the retention time mode during ALS optimization, whereas only non-negativity constraint is generally implemented in the spectral mode.
- After chemometric modelling, the profiles corresponding to retention times ($\mathbf{C}_{aug}$) and fluorescence spectra ($\mathbf{S}$) for the N individual analytes are obtained, as well as a matrix $\mathbf{E}_{aug}$ that comprises the residuals of the modelling. On one hand, the information related to the contribution of the analytes is gathered from $\mathbf{C}_{aug}$ as the area under the sub-profiles in each of the samples, which is used for quantitative purposes. On the other hand, $\mathbf{S}$ comprises the unfolded fluorescence matrices of the individual analytes that can eventually be refolded to restore the two-dimensional fluorescence matrices. Hence, individual excitation and emission profiles of the N components in the samples are obtained, whose are then utilized for the identification of the resolved components.

** Insert Fig. 7**

4.2. APARAFAC modelling

*A*PARAFAC algorithm has been developed for the analysis of third-order data that do not fulfil a quadrilinear model, for example, in presence of retention times that change from sample to sample [25]. *A*PARAFAC model implies the construction of a trilinear augmented three-way array, where augmentation is done along the quadrilinearity-breaking mode. In principle, the application of *A*PARAFAC would only involve an initialization step and no constraints would be necessary due to the uniqueness property of the decomposition of a trilinear three-way data array [14, 22], analogous to the PARAFAC model for the modelling of a three-way data array. However, aiming to obtain profiles with chemically interpretable information, same MCR-ALS constraints are usually implemented.

*A*PARAFAC algorithm is based in three-way PARAFAC modelling and inspired by the augmentation philosophy applied in MCR-ALS analysis [25]. In this manner, *A*PARAFAC can be interpreted as an algorithm composed by the marriage of PARAFAC and MCR-ALS that collects the essential particularities of each individual model, *i.e.*, the ability to overcome the lack of quadrilinearity by virtue of its augmented structure, but maintaining the original three-dimensional structure of the data [22, 25]. Then, besides the ability to handle non-quadrilinear data, the most remarkable advantage of this modelling is that, since the original data structure is maintained, the statistical efficiency of decomposing a multiway array is higher in comparison with unfolding into arrays of lower dimensions, as it is required for the MCR-ALS analysis of four-way data.

*A*PARAFAC model can be represented by Eq. 2, where decomposition of the augmented three-way array $\boldsymbol{X}_{\text{aug}}$ retrieves three loading matrices, $\mathbf{A}_{\text{aug}}$, **B** and **C,** corresponding to retention times and excitation and emission spectral profiles respectively, for the *N* number of responsive components, as well as an $\mathbf{E}_{\text{aug}}$ matrix that comprises the model residuals;

$$\boldsymbol{X}_{\text{aug}} = \mathbf{A}_{\text{aug}}(\mathbf{B} \odot \mathbf{C})^{\text{T}} + \mathbf{E}_{\text{aug}} \qquad \text{Eq. 2}$$

"⊙" indicates the Khatri–Rao or column-wise Kronecker product [25]

4.2.1. Data structure

The algorithm *A*PARAFAC is implemented by building an augmented three-way array as follows [22, 25] (Fig. 8):

- For each sample, a three-way data object $\boldsymbol{X}_{\text{f}}$ is constructed with a size of ($J \times K \times L$), where *J*, *K* and *L* are, in this case, the collected fractions (retention times), emission wavelengths and excitation wavelengths, respectively. Then, an augmented three-way array $\boldsymbol{X}_{\text{aug}}$ is built by appending all the individual three-way arrays, generating a [(I + 1) $J \times K \times L$] object, in which *I* is the number of calibration samples and 1 represents the unknown, test or validation sample. In this regard, it is worth noticing that the augmented three-way object fulfils the trilinear modelling requirements, since augmentation is performed in the direction of the quadrilinearity-breaking mode.
- For ALS optimization, same constraints as those applied in MCR-ALS are implemented.

- At the end of the chemometric decomposition, retention time (**A**), excitation spectral (**B**) and emission spectral (**C**) profiles are acquired. Here, different to MCR-ALS, individual spectral profiles are obtained, *i.e.*, excitation and emission profiles are retrieved separately and no data post-processing is needed. However, similar to MCR-ALS, for quantitative purposes, the area under the sub-profiles comprised in **A** is related to the individual contribution of the analytes in each sample.

** Insert Fig. 8**

Fig. 9.A displays the results obtained from MCR-ALS resolution of a sample containing 3 analytes, as well as the individual excitation and emission profiles retrieved from the refolded fluorescence matrices. In Fig. 9.B, results retrieved from *A*PARAFAC modelling for a sample containing 3 analytes are shown.

Insert Fig. 9

4.3. Quantitative analysis and figures of merit

In order to compare the performance of the applied chemometric models for third-order data modelling, in terms of predictive ability and figures of merit, a recovery study in several validation and spiked drinking water samples reported by authors elsewhere [21, 22] was analysed. Table 1 and 2 summarize the prediction results corresponding to the application of MCR-ALS and *A*PARAFAC for validation and spiked drinking water, respectively, in presence of interferences. As can be seen, a satisfactory coincidence between predictions values corresponding to both models is demonstrated, and acceptable REP % values are obtained for both models.

Insert Table 1

Insert Table 2

Eventually, figures of merit were estimated for both models and a comparative analysis was performed. Additionally, second-order modelling was evaluated applying PARAFAC and MCR-ALS, and figures of merit were compared with those calculated for third-order modelling. It is important to highlight that, even though the estimations of figures of merit for an analytic method based on MCR-ALS model were obtained from well-stablished mathematic expressions [1], equations for a method based on *A*PARAFAC model have not been developed yet. Thus, an extension of derived expression from four-way calibration with PARAFAC has been utilized, despite

possible over estimations are introduced [5]. For second-order modelling, only OFL was considered as target analyte and the other components were considered as unexpected compounds.

To estimate the sensitivities in MCR-ALS and PARAFAC for three-way and four-way calibration, the following mathematical expressions were used:

$$\mathrm{SEN}_{\mathrm{MCR}} = s_n[J(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}]^{-1/2} \quad \text{Eq. 3}$$

where s_n is the slope of the MCR-ALS pseudo-univariate plot, J is the number of data points in each submatrix in the augmented mode, and $\mathbf{C}$ is a matrix containing the profiles for all sample components in the non-augmented direction [1, 50]; and

$$\mathrm{SEN}_{\mathrm{PARAFAC,3-way}} = s_n\left\{\left[\left(\mathbf{B}_{\mathrm{cal}}^{\mathrm{T}}\mathbf{P}_{\mathrm{B,unx}}\mathbf{B}_{\mathrm{cal}}\right) * \left(\mathbf{C}_{\mathrm{cal}}^{\mathrm{T}}\mathbf{P}_{\mathrm{C,unx}}\mathbf{C}_{\mathrm{cal}}\right)\right]^{-1}\right\}^{-1/2} \quad \text{Eq. 4}$$

$$\mathrm{SEN}_{\mathrm{PARAFAC,4-way}} = s_n\left\{\left[\left(\mathbf{B}_{\mathrm{cal}}^{\mathrm{T}}\mathbf{P}_{\mathrm{B,unx}}\mathbf{B}_{\mathrm{cal}}\right) * \left(\mathbf{C}_{\mathrm{cal}}^{\mathrm{T}}\mathbf{P}_{\mathrm{C,unx}}\mathbf{C}_{\mathrm{cal}}\right) * \left(\mathbf{D}_{\mathrm{cal}}^{\mathrm{T}}\mathbf{P}_{\mathrm{D,unx}}\mathbf{D}_{\mathrm{cal}}\right)\right]^{-1}\right\}^{-1/2} \quad \text{Eq. 5}$$

where s_n is the slope of the PARAFAC pseudo-univariate plot, $\mathbf{B}_{\mathrm{cal}}$, $\mathbf{C}_{\mathrm{cal}}$ and $\mathbf{D}_{\mathrm{cal}}$ collect the loading matrices for the calibrated analytes, * is the element-wise and $\mathbf{P}_{\mathrm{B,unx}}$, $\mathbf{P}_{\mathrm{C,unx}}$ and $\mathbf{P}_{\mathrm{D,unx}}$ are projection matrices given by $\mathbf{I}\text{-}\mathbf{B}_{\mathrm{unx}}\mathbf{B}_{\mathrm{unx}}^{+}$, $\mathbf{I}\text{-}\mathbf{C}_{\mathrm{unx}}\mathbf{C}_{\mathrm{unx}}^{+}$ and $\mathbf{I}\text{-}\mathbf{D}_{\mathrm{unx}}\mathbf{D}_{\mathrm{unx}}^{+}$, respectively, being $\mathbf{I}$ the identity matrices, $\mathbf{B}_{\mathrm{unx}}$, $\mathbf{C}_{\mathrm{unx}}$ and $\mathbf{D}_{\mathrm{unx}}$ collect the loading matrices for the unexpected samples constituents, and the superscript + indicates the generalized inverse operation.

For the estimation of the limit of detection (LOD) and limit of quantitation (LOQ), eq. 6 and eq. 7, respectively were utilized.

$$\mathrm{LOD} = 2 \times t_{0.05,\infty}\frac{s_{dtest}}{\mathrm{SEN}} = 3.3\frac{s_{dtest}}{\mathrm{SEN}} \quad \text{Eq. 6}$$

$$\mathrm{LOQ} = 10\frac{s_{dtest}}{\mathrm{SEN}} \quad \text{Eq. 7}$$

where $t_{0.05,\infty}$ is the one-tail t value assuming a large number of calibration samples and α value of 0.05, and s_{dtest} represents the standard deviation of the estimated net signal when its true value is zero [5, 50].

In Table 3, figures of merit obtained for third-order data modelling using both models are shown. Figures of merit computed for second- and third-order data modelling using MCR-ALS and PARAFAC are depicted in Table 4.

Insert Table 3

Insert Table 4

It is noticeable that there is an important improvement in the SEN, LOD and LOQ values obtained for third-order data modelling when *A*PARAFAC is used, in comparison to MCR-ALS, while a drastic reduction of LOD and LOQ values is shown

when the order or dimension of the data increases. However, figures of merit obtained for second- and third-order data using MCR-ALS modelling did not show significant differences. On the other hand, the strong difference observed in LOD and LOQ values when second-order data modelling is performed using MCR-ALS and PARAFAC lies, in principle, in the loss of trilinearity caused by the lack of sample-to-sample reproducibility, which can be overcome with MCR-ALS but not with PARAFAC.

The main basis of the aforementioned observations belongs in the assumption that third-order data modelled with *A*PARAFAC shows several advantages over MCR-ALS, stressing the possibility of processing the data in its original three-dimensional structure, instead of unfolding the data to arrays of lower dimensions, and the feasibility to overcome the lack of quadrilinearity, leading to an improvement in the figures of merit and prediction capability of the analytical method. Additionally, *A*PARAFAC exploits the second-order advantage even in presence of lack of sample-to-sample reproducibility, similar to MCR-ALS. In consequence, *A*PARAFAC is presented as an appropriate alternative for third-order LC-EEM data analysis achieving acceptable results in the analysis of multi-component samples in presence of uncalibrated components.

5. Conclusion

In the present review, three analytical methodologies for the generation of third-order LC-EEM data are reviewed. Methodology I, based on the collection of discrete fractions at the end of the chromatographic procedure, requires low complexity equipment, needing a device that enables the collection of fractions in multi-well plates. The time of analysis is limited by the detection procedure that strictly depends on the instrumental parameters and the characteristics of the used instrument. Generated data have shown perfect trilinearity as consequence of the independence between instrumental modes and the particular bilinearity/trilinearity properties of the EEM. The results obtained from trilinear decomposition were highly satisfactory, obtaining time and spectra profiles with strong similarities with the experimental chromatogram and the pure excitation and emission spectra, respectively.

Methodology II, although only one instrument is required, demands a chromatograph equipped with an auto-sampler and fast-scanning fluorescence detector. Besides, due to the fact a high number of injections is needed for each sample, the analysis is time-consuming, and it can only be improved in spite of a detriment of the spectral

information. Moreover, the high consumption of reagents and sample, as consequence of the multi-injections, involves a high environmental impact as well as an important increment in the total costs. On the other hand, regarding data properties, it has been shown that slight differences in the retention times among runs leads to modifications in the excitation spectra features. This fact indicates a direct dependence between time and excitation wavelength mode, which means a loss of trilinearity in the third-order LC-EEM data. Even though lack of trilinearity in the third-order data is a drawback to overcome to obtain reliable results, in the literature, it has not been evaluated the effects introduced in the results due to lack of trilinearity, whereas they report loss of quadrilinearity as a consequence of the same phenomena, *i.e.*, lack of run-to-run reproducibility [24, 26]. Finally, different alternatives to turn data into trilinear were here reported, including peak alignment algorithms.

The third methodology studied is presented as a new proposal for third-order LC-EEM data generation. It seems to be advantageous due to the short time of the analysis, the low consumption of solvents and sample and the low complexity of the required equipment. However, the generated data show an extreme complexity by virtue of the strong dependence between instrumental modes, leading to a severe loss of trilinearity. Unfortunately, no chemometric procedures able to resolve this kind of data have been developed yet. Also, no pre-processing procedures that would permit to turn data into trilinear have been found. Thus, the development of new chemometric algorithms to cope with this kind of data is a worthwhile challenge for chemometricians and analytical chemists.

In sum, on the basis of the above-mentioned observations, it can be assumed that methodology I is the most feasible and efficient current strategy for the generation of third-order LC-EEM data up to the present, which becomes promissory for further implementations.

Methodology I was then used for the determination of several analytes in drinking water samples. It has been demonstrated that in multi-set analysis, four-way arrays show loss of quadrilinearity due to differences in the retention times of the analytes among samples. However, *A*PARAFAC and MCR-ALS models proved to be able to bear non-quadrilinear data, and satisfactory results were achieved. Further, it was demonstrated that the so-called “third-order advantage” is successfully achieved when third-order data are analysed, representing an improvement of sensitivity and selectivity as well as the

possibility to resolve a complex problem with a unique data array, without needing additional information.

At last, it becomes crucial to remark the importance of doing an in-depth analysis of the system under study considering all the possible edges, from chemical to mathematical standpoints in order to obtain the most reliable and satisfactory results.

Acknowledgements

The authors express their gratitude to CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP-2015 Nº 0111) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2014-0347) for financially supporting this work. M.R.A and M.M gratefully acknowledge the postdoc and Ph.D. financial support, respectively, provided by CONICET.

References

[1] M.C. Bauza, G.A. Ibañez, R. Tauler, A.C. Olivieri, Sensitivity Equation for Quantitative Analysis with Multivariate Curve Resolution-Alternating Least-Squares: Theoretical and Experimental Approach, Anal. Chem., 84 (2012) 8697-8706.

[2] A.C. Olivieri, G.M. Escandar, Practical Three-Way Calibration, Elsevier, Waltham, USA, 2014.

[3] G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Pena, A.C. Olivieri, Second- and higher-order data generation and calibration: a tutorial, Anal. Chim. Acta, 806 (2014) 8-26.

[4] K.S. Booksh, B.R. Kowalski, Theory of analytical chemistry, Anal. Chem., 66 (1994) 782A-791A.

[5] A.C. Olivieri, N.M. Faber, New developments for the sensitivitz estimation in four-way calibration with the quadrilinear parallel factor moder, Anal. Chem., 84 (2012) 186-193.

[6] A.C. Olivieri, J.A. Arancibia, A. Muñoz de la Peña, I. Durán-Merás, A. Espinosa Mansilla, Second-Order Advantage Achieved with Four-Way Fluorescence Excitation−Emission−Kinetic Data Processed by Parallel Factor Analysis and Trilinear Least-Squares. Determination of Methotrexate and Leucovorin in Human Urine, Anal. Chem., 76 (2004) 5657-5666.

[7] C. Kang, H.-L. Wu, L.-X. Xie, S.-X. Xiang, R.-Q. Yu, Direct quantitative analysis of aromatic amino acids in human plasma by four-way calibration using intrinsic fluorescence: Exploration of third-order advantages, Talanta, 122 (2014) 293-301.

[8] C. Kang, H.-L. Wu, Y.-J. Yu, Y.-J. Liu, S.-R. Zhang, X.-H. Zhang, R.-Q. Yu, An alternative quadrilinear decomposition algorithm for four-way calibration with application to analysis of four-way fluorescence excitation–emission–pH data array, Anal. Chim. Acta, 758 (2013) 45-57.

[9] A.C. Olivieri, Analytical Advantages of Multivariate Data Processing. One, Two, Three, Infinity?, Anal. Chem., 80 (2008) 5713-5720.

[10] H.-L. Wu, C. Kang, Y. Li, R.-Q. Yu, Chapter 4 - Practical Analytical Applications of Multiway Calibration Methods Based on Alternating Multilinear Decomposition, in: A. Muñoz de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri (Eds.) Data Handling in Science and Technology, Elsevier2015, pp. 167-246.

[11] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom., 15 (2001) 749-771.

[12] J.M. Amigo, F. Marini, Chapter 7 - Multiway Methods, in: M. Federico (Ed.) Data Handling in Science and Technology, Elsevier2013, pp. 265-313.

[13] A.C. Olivieri, Recent advances in analytical calibration with multi-way data, Anal. Methods, 4 (2012) 1876-1886.

[14] A.C. Olivieri, G.M. Escandar, Chapter 11 - Third-order/Four-way Calibration and Beyond, Practical Three-Way Calibration, Elsevier, Boston, 2014, pp. 217-232.

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

[16] R. Tauler, Multivariate curve resolution applied to secon order data, Chemom. Intell. Lab. Syst., 30 (1995) 133-146.

[17] S. Wold, P. Geladi, K. Esbensen, J. Öhman, Multi-way principal components-and PLS-analysis, J. Chemom., 1 (1987) 41-56.

[18] R. Bro, Multiway calibration. Multilinear PLS, J. Chemom., 10 (1996) 47-61.

[19] A.C. Olivieri, On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization: Second-order advantage and precision properties, J. Chemom., 19 (2005) 253-265.

[20] H.-L. Wu, M. Shibukawa, K. Oguma, An alternating trilinear decomposition algorithm with application to calibration of HPLC–DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons, J. Chemom., 12 (1998) 1-26.

[21] M.R. Alcaráz, G.G. Siano, M.J. Culzoni, A. Munoz de la Pena, H.C. Goicoechea, Modeling four and three-way fast high-performance liquid chromatography with fluorescence detection data for quantitation of fluoroquinolones in water samples, Anal. Chim. Acta, 809 (2014) 37-46.

[22] M.R. Alcaráz, S. Bortolalo, H.C. Goicoechea, A.C. Olivieri, A new modeling strategy for third-order fast high-performance liquid chromatographic data with fluorescence detection. Quantitation of fluoroquinolones in water samples, Anal. Bioanal. Chem., 407 (2015) 1999-2011.

[23] R. Bro, Multi-way Analysis in the Food Industry, Chemometrics Group, Food Technology

Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Compenhagen, Denmark, 1997, pp. 290.

[24] V.A. Lozano, A. Muñoz de la Peña, I. Durán-Merás, A. Espinosa Mansilla, G.M. Escandar, Four-way multivariate calibration using ultra-fast high-performance liquid chromatography with fluorescence excitation–emission detection. Application to the direct

802 analysis of chlorophylls a and b and pheophytins a and b in olive oils, Chemom. Intell. Lab.
803 Syst., 125 (2013) 121-131.

804 [25] S.A. Bortolato, V.A. Lozano, A.M. de la Peña, A.C. Olivieri, Novel augmented parallel
805 factor model for four-way calibration of high-performance liquid chromatography–fluorescence
806 excitation–emission data, Chemom. Intell. Lab. Syst., 141 (2015) 1-11.

807 [26] M. Montemurro, L. Pinto, G. Véras, A. de Araújo Gomes, M.J. Culzoni, M.C. Ugulino de
808 Araújo, H.C. Goicoechea, Highly sensitive quantitation of pesticides in fruit juice samples by
809 modeling four-way data gathered with high-performance liquid chromatography with
810 fluorescence excitation-emission detection, Talanta, 154 (2016) 208-218.

811 [27] P.T. Anastas, J.C. Warner, Green Chemistry: Theory and Practice, Oxford University
812 Press1998.

813 [28] J.A. Arancibia, P.C. Damiani, G.M. Escandar, G.A. Ibañez, A.C. Olivieri, A review on
814 second- and third-order multivariate calibration applied to chromatographic data, J. Chromatogr.
815 B, 910 (2012) 22-30.

816 [29] H.-L. Wu, Y. Li, C. Kang, R.-Q. Yu, Chapter 3 - Multiway Calibration Based on
817 Alternating Multilinear Decomposition, in: A. Muñoz de la Peña, H.C. Goicoechea, G.M.
818 Escandar, A.C. Olivieri (Eds.) Data Handling in Science and Technology, Elsevier2015, pp. 83-
819 165.

820 [30] R. Boqué Martí, J. Ferré Baldrich, Chapter 1 - Fundamentals of PARAFAC, in: A. Muñoz
821 de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri (Eds.) Data Handling in Science and
822 Technology, Elsevier2015, pp. 7-35.

823 [31] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components
824 in PARAFAC models, J. Chemom., 17 (2003) 274-286.

825 [32] G. Tomasi, F. Savorani, S.B. Engelsen, icoshift: An effective tool for the alignment of
826 chromatographic data, J. Chromatogr. A, 1218 (2011) 7832-7840.

827 [33] N.-P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple
828 wavelength chromatographic profiles for chemometric data analysis using correlation optimised
829 warping, J. Chromatogr. A, 805 (1998) 17-35.

830 [34] A.C. Olivieri, G.M. Escandar, Chapter 7 - Parallel Factor Analysis: Nontrilinear Data of
831 Type 1, Practical Three-Way Calibration, Elsevier, Boston, 2014, pp. 109-125.

832 [35] A.C. Olivieri, G.M. Escandar, Chapter 13 - Application Example: MCR–ALS, Practical
833 Three-Way Calibration, Elsevier, Boston, 2014, pp. 251-271.

834 [36] H. Parastar, N. Akvan, Multivariate curve resolution based chromatographic peak
835 alignment combined with parallel factor analysis to exploit second-order advantage in complex
836 chromatographic measurements, Anal. Chim. Acta, 816 (2014) 18-27.

837 [37] Y.-C. Kim, J.A. Jordan, M.L. Nahorniak, K.S. Booksh, Photocatalytic Degradation-
838 Excitation−Emission Matrix Fluorescence for Increasing the Selectivity of Polycyclic Aromatic
839 Hydrocarbon Analyses, Anal. Chem., 77 (2005) 7679-7686.

840 [38] S.-H. Zhu, H.-L. Wu, A.L. Xia, J.-F. Nie, Y.-C. Bian, C.-B. Cai, R.-Q. Yu, Excitation-
841 emission-kinetic fluorescence coupled with third-order calibration for quantifying carbaryl and
842 investigating the hydrolysis in effluent water, Talanta, 77 (2009) 1640-1646.

[39] M.D. Carabajal, J.A. Arancibia, G.M. Escandar, Excitation-emission fluorescence-kinetic data obtained by Fenton degradation. Determination of heavy-polycyclic aromatic hydrocarbons by four-way parallel factor analysis, Talanta, 165 (2017) 52-63.

[40] A.P. Pagani, G.A. Ibañez, Four-way calibration applied to the processing of pH-modulated fluorescence excitation-emission matrices. Analysis of fluoroquinolones in the presence of significant spectral overlapping, Microchem. J., 132 (2017) 211-218.

[41] R.M. Maggio, P.C. Damiani, A.C. Olivieri, Four-way kinetic-excitation-emission fluorescence data processed by multi-way algorithms. Determination of carbaryl and 1-naphthol in water samples in the presence of fluorescent interferents, Anal. Chim. Acta, 677 (2010) 97-107.

[42] M. Montemurro, G.G. Siano, M.J. Culzoni, H.C. Goicoechea, Automatic generation of photochemically induced excitation-emission-kinetic four-way data for the highly selective determination of azinphos-methyl in fruit juices, Sensor Actuat. B-Chem., 239 (2017) 397-404.

[43] D.W. Cook, M.L. Burnham, D.C. Harmes, D.R. Stoll, S.C. Rutan, Comparison of multivariate curve resolution strategies in quantitative LCxLC: Application to the quantification of furanocoumarins in apiaceous vegetables, Anal. Chim. Acta, 961 (2017) 49-58.

[44] A. de Juan, S.C. Rutan, R. Tauler, 2.19 - Two-Way Data Analysis: Multivariate Curve Resolution – Iterative Resolution Methods, in: S.D. Brown, R. Tauler, B. Walczak (Eds.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 325-344.

[45] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: A review of advanced and tailored appilcations and challenges, Anal. Chim. Acta, 765 (2013) 28-36.

[46] R. Tauler, M. Maeder, 2.20 - Two-Way Data Analysis: Multivariate Curve Resolution – Error in Curve Resolution, in: S.D. Brown, R. Tauler, B. Walczak (Eds.) Comprehensive Chemometrics, Elsevier, Oxford, 2009, pp. 345-363.

[47] R. Tauler, A. de Juan, Chapter 5 - Multivariate Curve Resolution for Quantitative Analysis, in: A. Muñoz de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri (Eds.) Data Handling in Science and Technology, Elsevier2015, pp. 247-292.

[48] A. Malik, R. Tauler, Performance and validation of MCR-ALS with quadrilinear constraint in the analysis of noisy datasets, Chemom. Intell. Lab. Syst., 135 (2014) 223-234.

[49] A. Malik, R. Tauler, Extension and application of multivariate curve resolution-alternating least squares to four-way quadrilinear data-obtained in the investigation of pollution patterns on Yamuna River, India—A case study, Anal. Chim. Acta, 794 (2013) 20-28.

[50] A.C. Olivieri, Analytical Figures of Merit: From Univariate to Multiway Calibration, Chem. Rev., 114 (2014) 5358-5378.

Captions of figures and tables

Fig. 1. Classification tree for four-way data for a set of samples, according to whether the individual three-dimensional arrays data are trilinear or not, and to the number of quadrilinearity-breaking modes. Reprinted with permission of the authors of Ref [2]. Copyright 2014 Elsevier.

Fig. 2. General procedure for the third-order data generation by using *MI* methodology for a sample containing 3 compounds.

Fig. 3. General procedure for the third-order data generation by using *MII* methodology for a sample containing 2 compounds.

Fig. 4. General procedure for the third-order data generation by using *MIII* methodology for a sample containing a pure analyte.

Fig. 5. (A) Data generated using methodology I. Excitation-emission matrices, showed from the excitation mode, registered for all the collected fractions for a sample containing OFL (blue), CPF (green) and DNF (red). (**B)** Data generated using methodology II. Solid grey lines are the chromatograms corresponding to an emission wavelength = 350 nm obtained from the retention time-emission wavelength matrices registered at different excitation wavelength (260-305 nm e. 5 nm). (**C)** Data generated using methodology III. Consecutive excitation-emission matrices, showed from the excitation mode, registered for a sample containing CPF.

Fig. 6. A. Retention time (**1**), excitation spectra (**2**) and emission spectra (**3**) profiles obtained from PARAFAC resolution of the data generated for the corresponding sample using methodology I (**A**), II (**B**) and III (**C**). Dashed blue lines, solid green lines and dash-dotted red lines are OFL, CPF and DNF, respectively. Dotted yellow lines represent an unknown component obtained as a result of lack of trilinearity. Dotted green lines in **C** correspond to CPF profile, which is a consequence of the dependence between time and excitation wavelength modes.

Fig. 7. Schematic representation of MCR-ALS model to third-order LC-EEM data processing.

Fig. 8. Schematic representation of Augmented PARAFAC model to third-order LC-EEM data processing.

916

917 **Fig. 9.** MCR-ALS (**A**) and *A*PARAFAC (**B**) profiles obtained from the analysis of third-order
918 data obtained for a sample containing OFL (dashed blue), CPF (solid green) and DNF (dash-
919 dotted red). Temporal (**A.1** and **B.1**) as well as excitation (**A.3** and **B.2**) and emission spectral
920 (**A.4** and **B.3**) profiles are depicted. Unfolded fluorescence matrices obtained from MCR-ALS
921 resolution are shown in **A.2**

922

Table 1. Recovery study for 3 FQ in validation samples using MCR-ALS and *A*PARAFAC modelling. Reprinted with permission of the authors of Ref. [20].Copyright 2015 Springer.[a]

Sample	OFL			CPF			DNF		
	Nominal	Predicted		Nominal	Predicted		Nominal	Predicted	
		MCR-ALS	*A*PARAFAC		MCR-ALS	*A*PARAFAC		MCR-ALS	*A*PARAFAC
M01	20.0	21.1	20.1	90.0	99.3	92.5	25.0	27.1	23.2
M02	20.0	19.3	19.7	150.0	131.0	121.7	15.0	16.4	15.9
M03	60.0	51.1	52.1	30.0	44.9	58.5	5.0	5.3	5.1
M04	100.0	101.0	99.7	90.0	95.8	90.9	5.0	8.6	8.9
M05	60.0	68.1	70.1	150.0	144.8	147.0	25.0	28.4	28.8
M06	100.0	98.9	99.1	150.0	132.7	136.9	15.0	17.8	18.2
M07	100.0	104.1	101.0	30.0	21.0	22.0	5.0	7.4	7.6
M08	20.0	31.0	31.7	30.0	58.0	51.7	2.0	4.0	4.0
M09	60.0	45.3	55.2	30.0	19.8	25.6	8.0	9.6	9.4
M10	60.0	55.1	72.3	60.0	54.2	44.3	2.0	5.3	2.6
REP % [b]		**14.5**	**13.8**		**19.0**	**21.5**		**19.9**	**19.1**
$\bar{R}_{exp}$ [c]		**102.7**	**107.4**		**105.8**	**107.4**		**146.0**	**131.5**

[a] Concentrations are given in ng mL^{-1};

[b] REP %: relative error of prediction given in percentage and calculated as $REP\ \% = 100 \times \frac{\sqrt{\frac{1}{I}\sum_1^I (c_{nom} - c_{pred})^2}}{\bar{c}}$, for I= 10;

[c] $\bar{R}_{exp}$: average experimental recoveries given in percentage.

Table 2. Recovery study for 3 FQ in spiked drinking water samples using MCR-ALS and *A*PARAFAC modelling. Reprinted with permission of the authors of Ref. [20].Copyright 2015 Springer.[a]

Sample [b]	**OFL**			**CPF**			**DNF**		
	Taken	**Found**		**Taken**	**Found**		**Taken**	**Found**	
		MCR-ALS	***A*PARAFAC**		**MCR-ALS**	***A*PARAFAC**		**MCR-ALS**	***A*PARAFAC**
Mw_01	20.0	30.6	17.4	30.0	26.3	25.0	3.5	3.2	2.9
Mw_02	60.0	81.2	63.8	90.0	78.6	91.2	5.5	7.5	7.6
Mt_01	60.0	61.9	65.3	90.0	86.6	89.5	2.2	2.7	2.6
Mt_02	40.0	32.2	26.2	60.0	60.5	62.9	9.0	12.8	12.4
Mm_01	20.0	12.5	19.1	30.0	19.7	17.0	2.2	1.9	1.9
Mm_02	40.0	41.1	49.9	60.0	91.1	81.3	9.0	9.5	9.2
$\bar{R}_{exp}$[c]		**106**	**98**		**98**	**97**		**116**	**113**

[a] Concentrations are given in ng mL^{-1}. Each mean value is the average of three replicates;
[b] Mw: well water from Colastiné City (Santa Fe, Argentina); Mt: tap wáter form Santa Fe City (Santa Fe, Argentina); Mm: commercial mineral water;
[c] $\bar{R}_{exp}$, average experimental recoveries given in percentage.

Table 3. Figures of merit obtained for third-order data modelling, applying MCR-ALS and *A*PARAFAC chemometric models. Reprinted with permission of the authors of Ref. [20]. Copyright 2015 Springer.

Figure of merit	OFL		CPF		DNF	
	MCR-ALS	*A*PARAFAC	MCR-ALS	*A*PARAFAC	MCR-ALS	*A*PARAFAC
SEN	10.4	21.0	2.7	20.0	22.9	83.0
SEL	0.68	0.65	0.21	0.29	0.88	0.30
LOD	0.25	0.20	0.99	0.15	0.12	0.02
LOQ	0.75	0.60	2.97	0.47	0.36	0.08

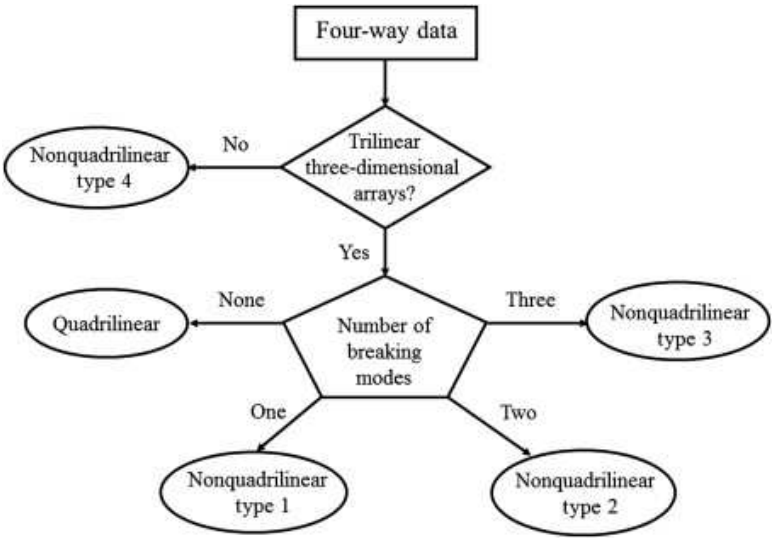
[a] SEN: sensitivity; SEL: selectivity; LOD: limit of detection and LOQ: limit of quantitation calculated according to Ref [1] and Ref [5] for MCR-ALS and *A*PARAFAC, respectively. LOD and LOQ are given in ng mL^{-1}.

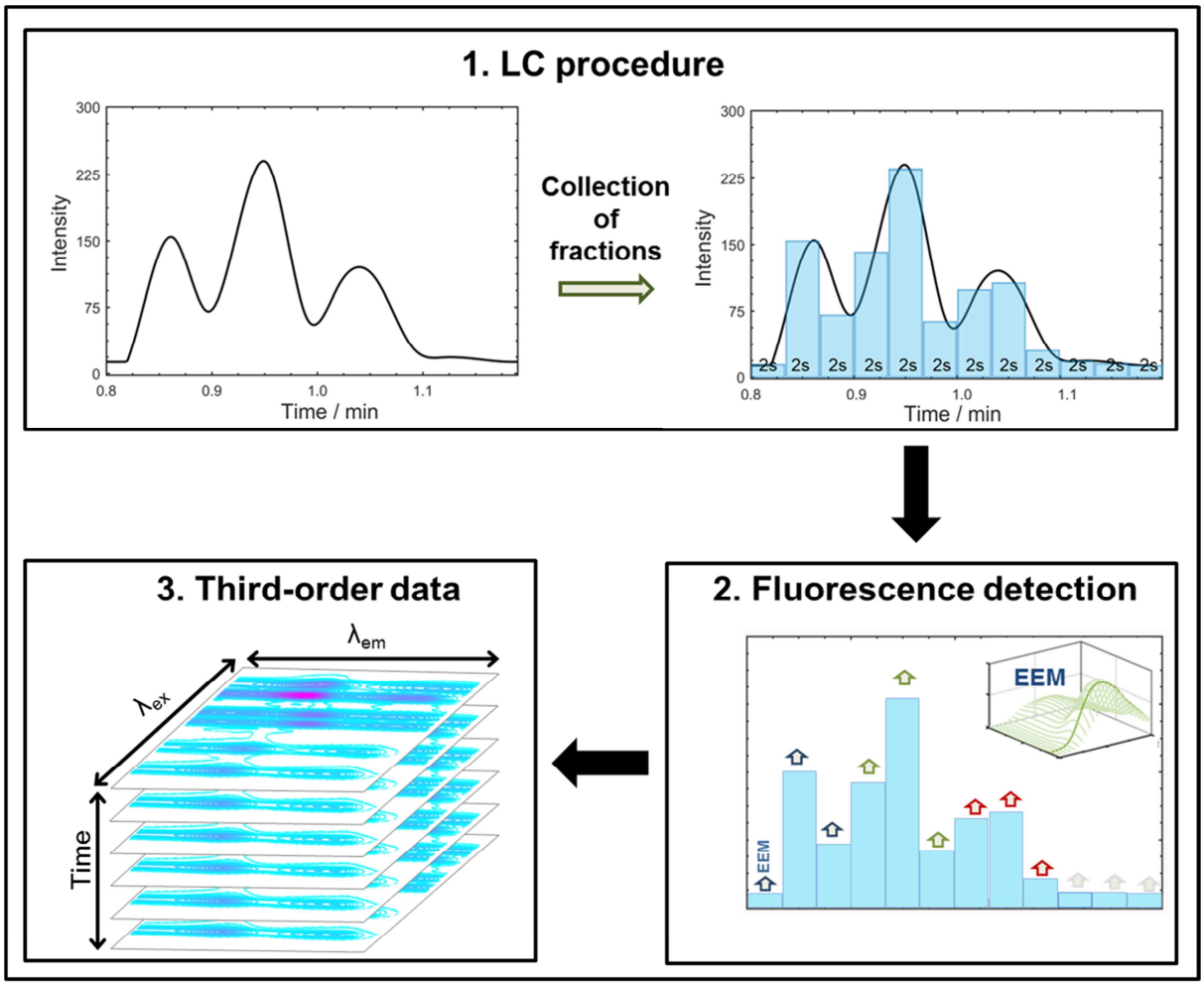
Table 4. Figures of merit obtained for OFL using second- and third-order data modelling, applying MCR-ALS and PARAFAC/*A*PARAFAC chemometric models. Reprinted with permission of the authors of Ref. [20]. Copyright 2015 Springer. [a]

Figure of merit	MCR-ALS		PARAFAC/APARAFAC [b]	
	Second-order	Third-order	Second-order	Third-order
SEN	5.2	10.4	7.6	21.0
SEL	0.23	0.68	0.25	0.65
LOD	0.4	0.25	6.9	0.20
LOQ	1.1	0.75	21.0	0.60

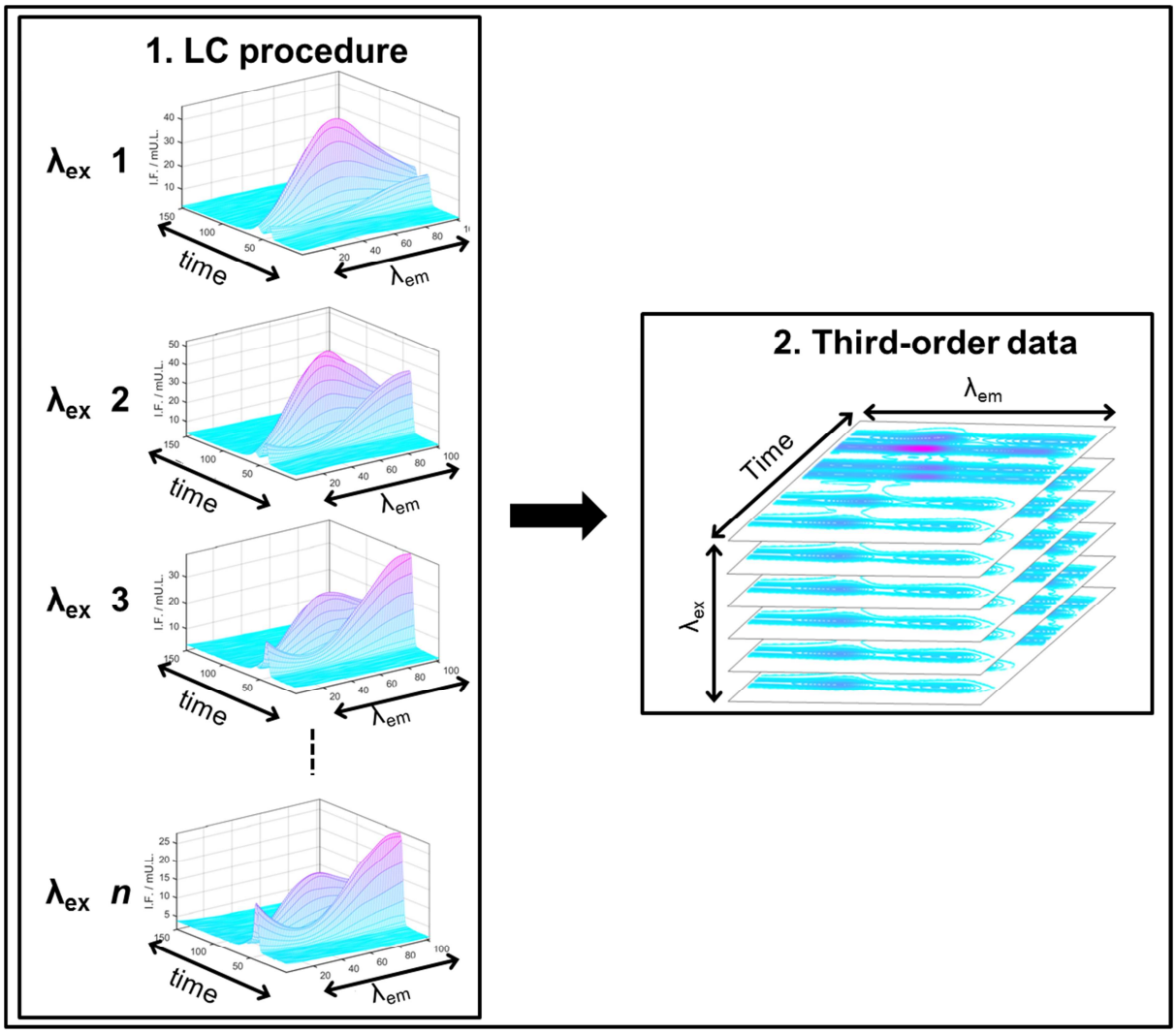
[b] SEN: sensitivity; SEL: selectivity; LOD: limit of detection and LOQ: limit of quantitation calculated according to Ref [1] and Ref [5] for MCR-ALS and *A*PARAFAC, respectively. LOD and LOQ are given in ng mL^{-1};

[a] For second-order data modelling, PARAFAC was applied, while for third-order data modelling *A*PARAFAC was used.

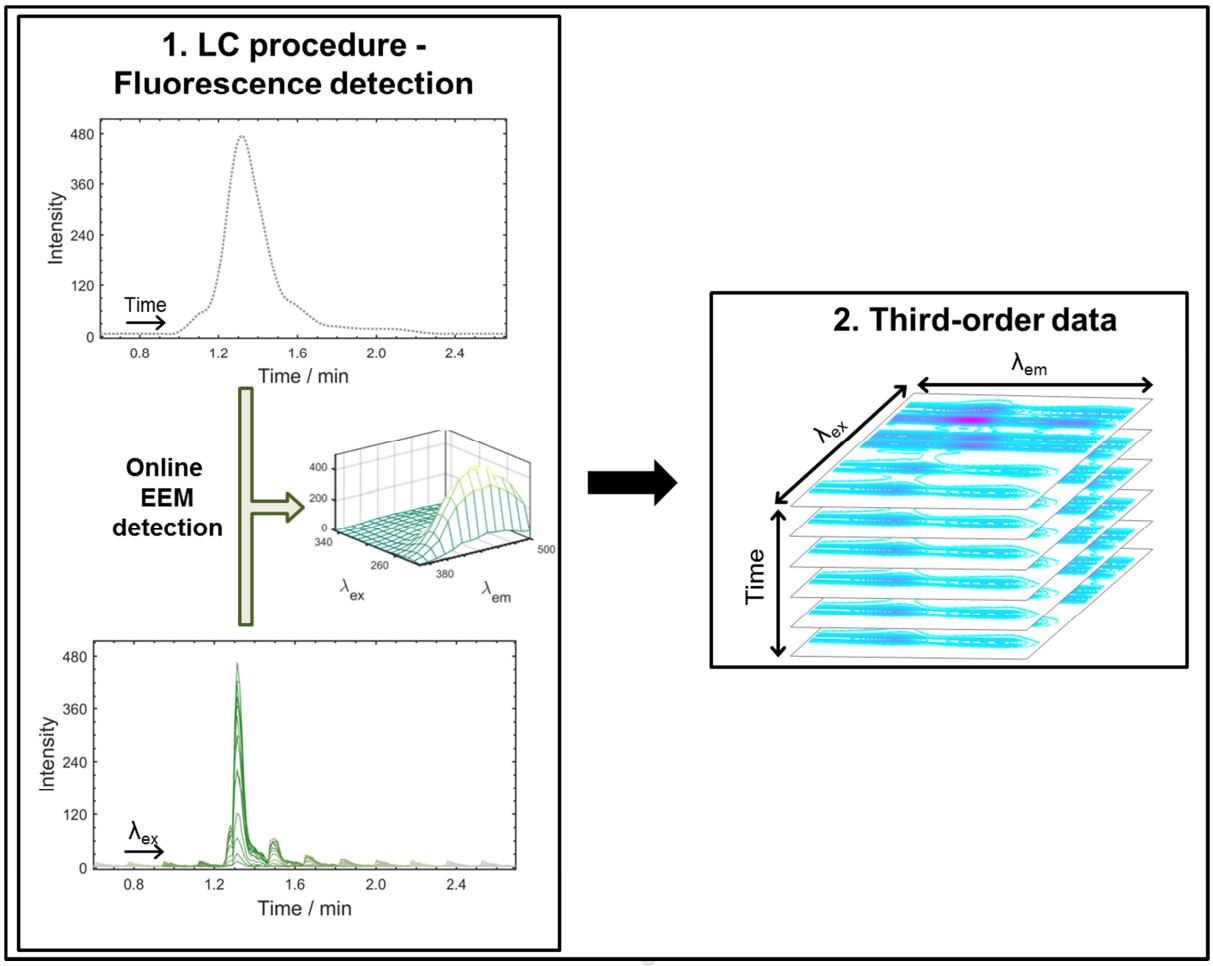




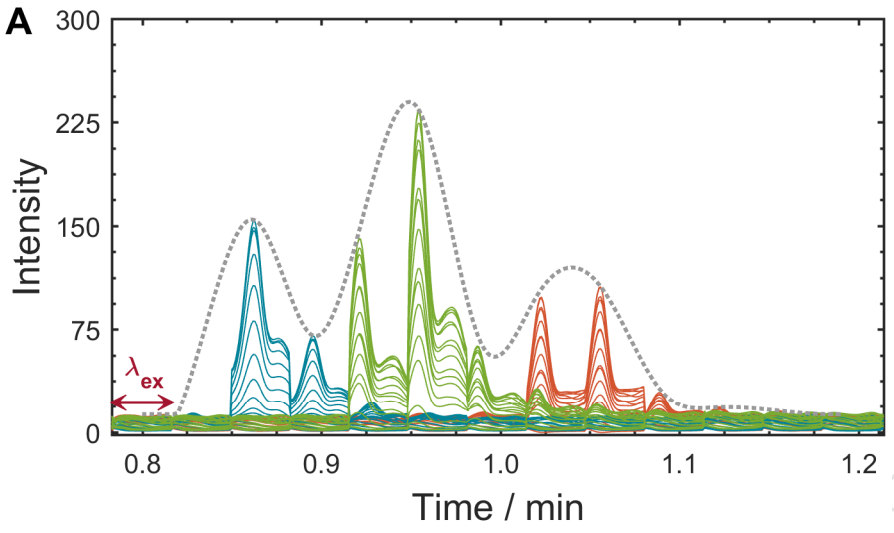
sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Montemurro, G. Siano, M. R. Alcaráz & H. C. Goicoechea; "Third order chromatographic-excitation–emission fluorescence data: Advances, challenges and prospects in analytical applications"
TrAC Trends in Analytical Chemistry - 2017, Vol. 93, pp. 119-133, 2017.

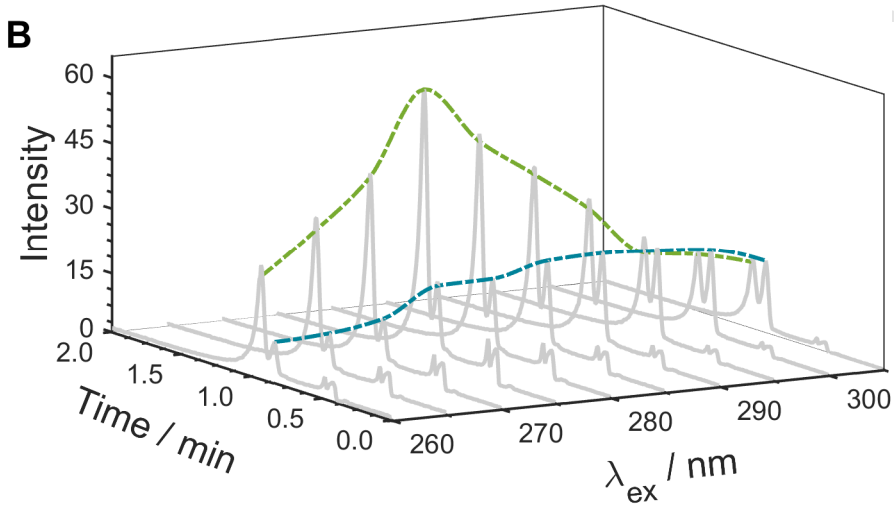


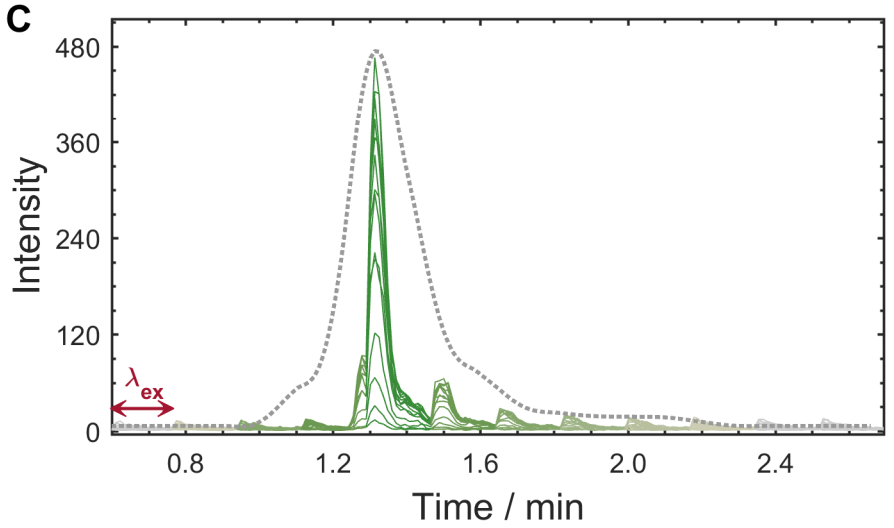
sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Montemurro, G. Siano, M. R. Alcaráz & H. C. Goicoechea; "Third order chromatographic-excitation–emission fluorescence data: Advances, challenges and prospects in analytical applications"
TrAC Trends in Analytical Chemistry - 2017, Vol. 93, pp. 119-133, 2017.

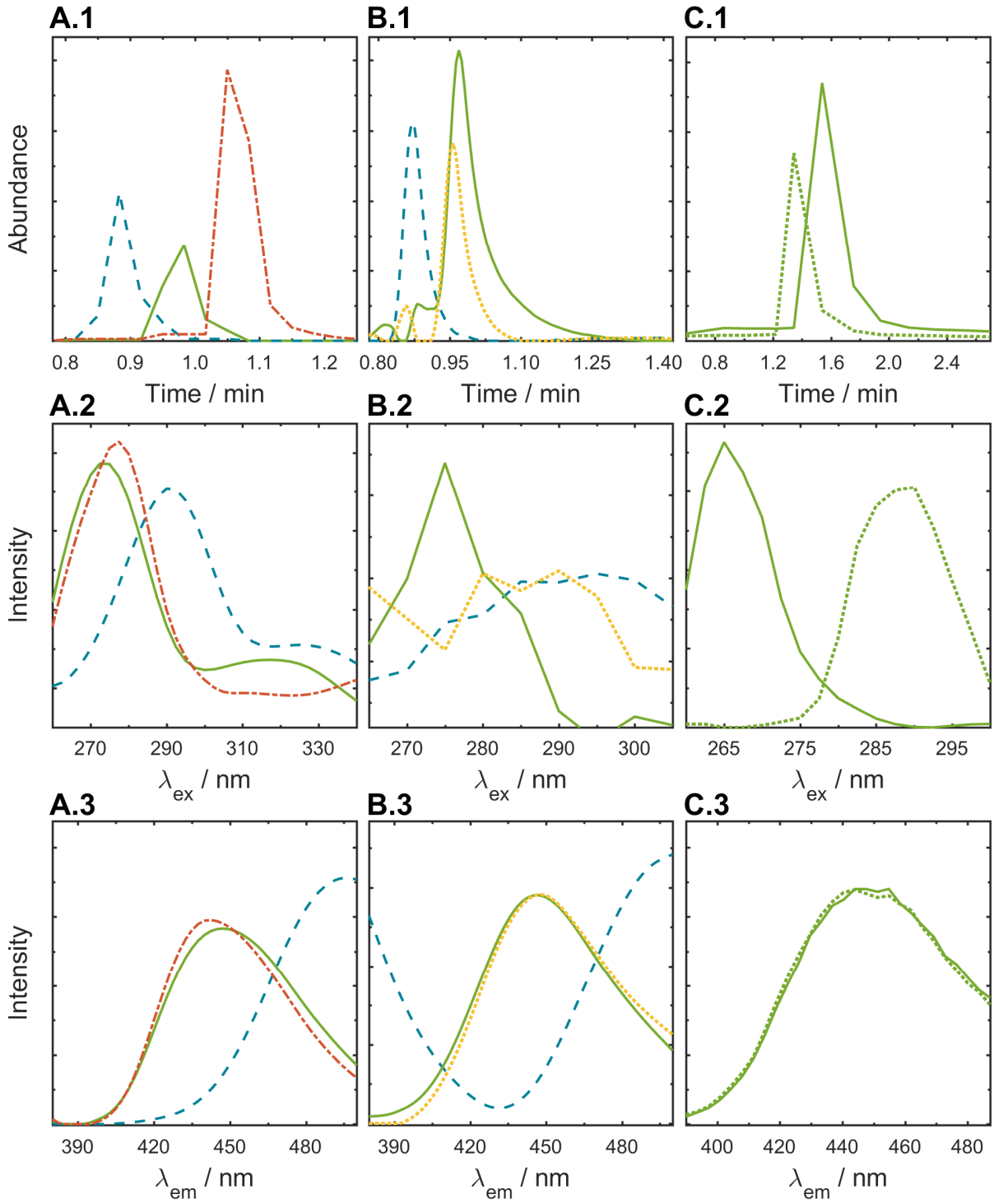


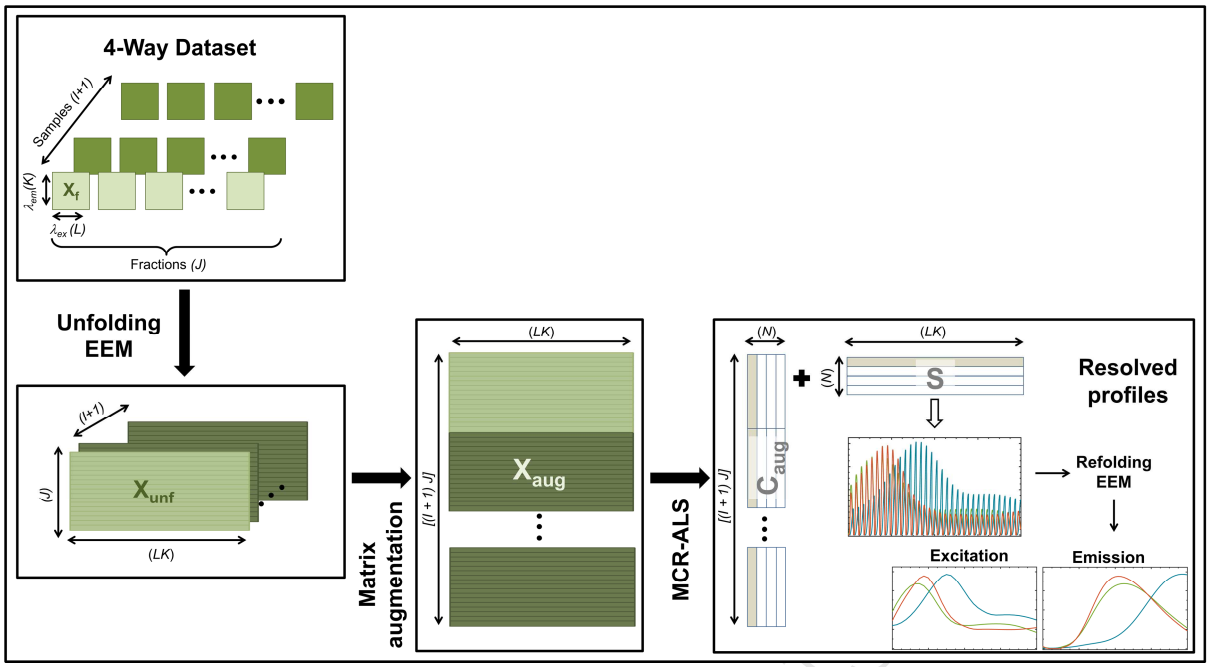
sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Montemurro, G. Siano, M. R. Alcaráz & H. C. Goicoechea; "Third order chromatographic-excitation–emission fluorescence data: Advances, challenges and prospects in analytical applications"
TrAC Trends in Analytical Chemistry - 2017, Vol. 93, pp. 119-133, 2017.

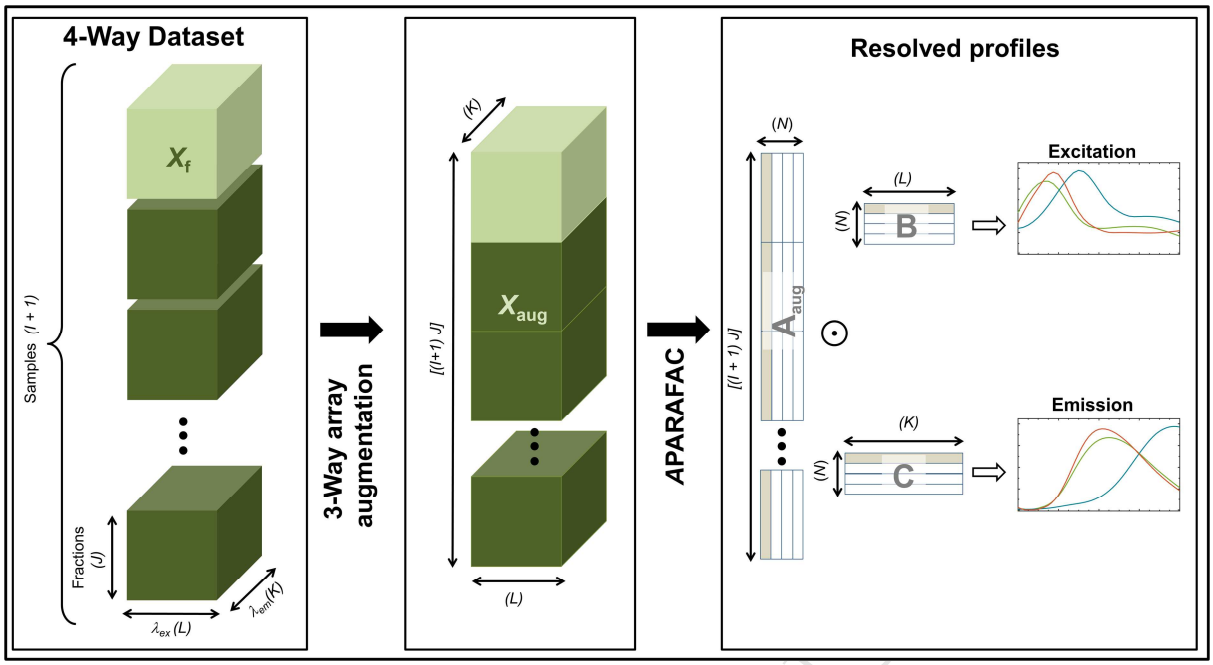


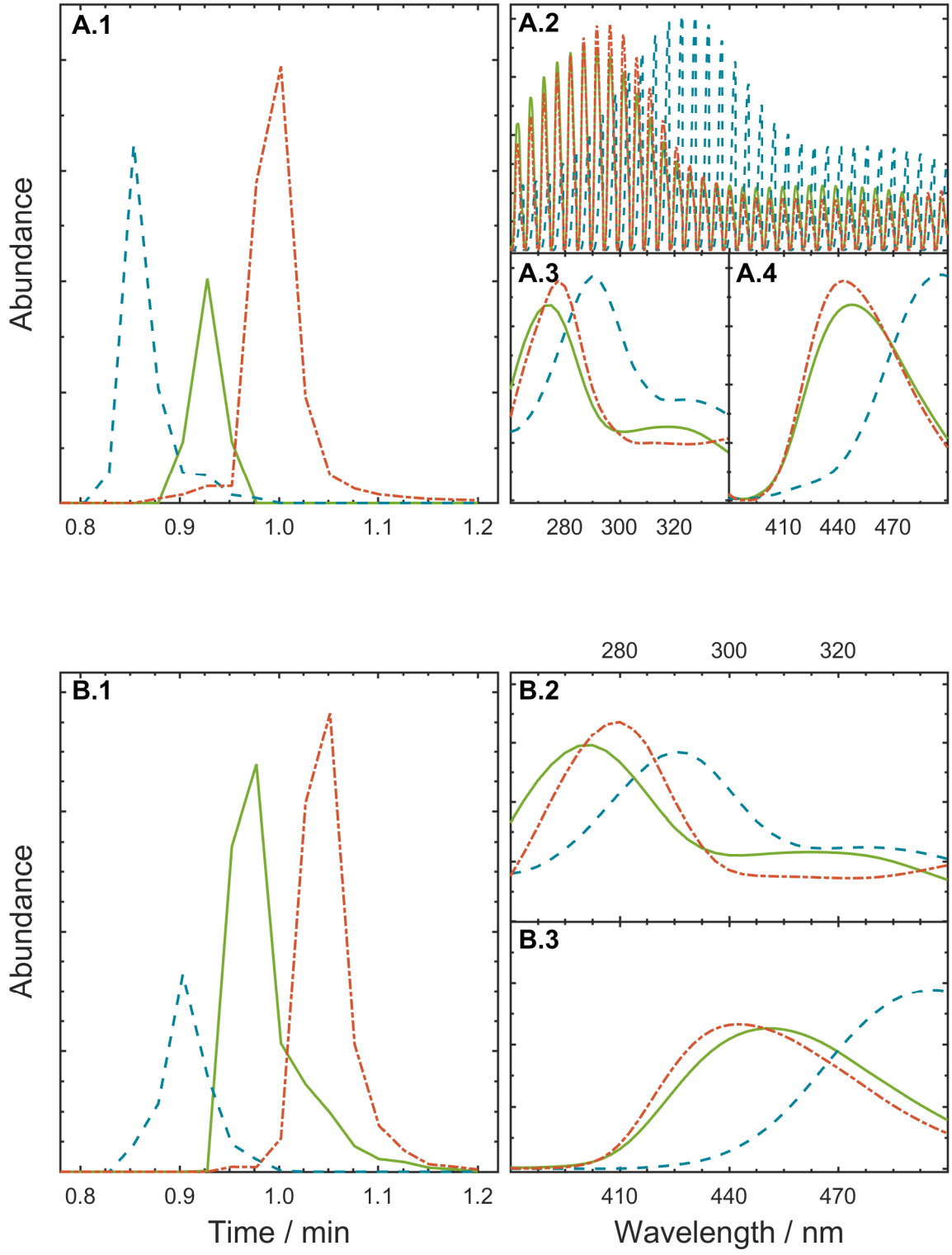












Highlights

- Third order chromatographic-excitation-emission fluorescence data are reviewed.
- Different instrumental setups for third-order data generation are compared.
- Data structure and chemometric modelling depends on the instrumental setup.
- Data pre-processing and processing alternatives are proposed.
- Analytical applications of four-way calibration are presented showing results.