Review

A review on second- and third-order multivariate calibration applied to chromatographic data☆

Juan A. Arancibia, Patricia C. Damiani, Graciela M. Escandar, Gabriela A. Ibañez, Alejandro C. Olivieri*

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

ABSTRACT

Quantitative analytical works developed by processing second- and third-order chromatographic data are reviewed. The various modes in which data of complex structure can be measured are discussed, with chromatographic separation providing either one or two of the data dimensions. This produces second-order data (matrices from uni-dimensional chromatography with multivariate detection or from two-dimensional chromatography) or third-order data (three-dimensional data arrays from two-dimensional chromatography with multivariate detection). The available algorithms for processing these data are classified and discussed, regarding their ability to cope with the ubiquitous phenomenon of retention time shifts from run to run. A summary of relevant works applying this combination of techniques is presented, with focus on quantitative analytical results. Special attention is paid to works achieving the full potentiality of the multidimensional data, i.e., the second-order advantage.

© 2012 Elsevier B.V. All rights reserved.

Contents

1. Introduction

The quality of modern life has notably improved due to several key factors such as significant advances in medicine and pharmacology, emphasis on the nutritional properties of food, and awareness of the need to protect the environment in which we live. From the perspective of the analytical laboratory, the direct consequence of this social consciousness is the marked progress in techniques that allow quantitative analysis of chemical compounds (either beneficial or harmful) at widely different concentrations, and in samples of ever increasing complexity. Most of the applied analytical methodologies when the analytes are organic molecules include chromatographic separations with different detection modes. In general, the sophistication of the

detector is directly related to the complexity of the sample and to the number and concentrations of the constituents to be analyzed.

For samples of complex composition, fully resolved chromatographic bands are not always achieved. The traditional approach to this ubiquitous problem is to vary the available experimental chromatographic parameters, separation materials and detecting systems in order to increase the resolving power and selectivity. This may however be prohibitive in terms of cost, or simply because the complexity of the sample is such that it resists all known separating efforts. In these cases, multivariate data analysis can be used for increasing the selectivity by mathematical means [1]. The coupling of chromatographic analysis with chemometric tools is an economical alternative to resolving this problematic situation without using sophisticated instruments, and represents an interesting strategy which is increasingly used in modern laboratories. In certain cases, this approach allows one to shorten the working time and/or to employ the isocratic mode, with the concomitant benefit of saving toxic solvents. The terms 'chromametrics' [2] and 'chro*math*ography' [3] have been suggested to describe the combination of chromatography and chemometrics/mathematics.

Regarding the specific data to be recorded and analyzed, second- and third-order data processed with suitable algorithms constitute the most useful approach to chromametrics, because it possesses the intrinsic advantage of dealing with the presence of interferences in real samples, in contrast to both zeroth- and first-order approaches [4]. The chemometric literature has coined the expression 'second-order advantage' to describe this interesting property of second- and higher-order multivariate calibration methodologies [1].

Chromatographic second-order data can be measured by coupling or 'hyphenating' two first-order instruments, each of which provides a given dimension or mode to the recorded data. For example, chromatography can be followed by detection based on UV–visible/infrared absorption, fluorescence excitation/emission or nuclear-magnetic resonance spectroscopies, mass spectrometry or voltammetry. In these cases, the chromatograph itself provides the first data dimension (the retention time), while the detector provides the second one (spectral wavelength or wavenumber, chemical shift, mass/charge ratio, electric voltage). Another possibility is two-dimensional liquid (LC–LC) [5] or gas chromatography (GC–GC) [6] with univariate detection, where each separating element provides an individual retention time dimension to the recorded second-order data. All these data can be arranged into a data table or matrix, where columns and rows correspond to each of the data dimensions. Some information about the application of multivariate calibrations and algorithms applied to quantitative chromatographic analysis can be found in recent reviews [7–15].

Higher complexity can be achieved by adding an extra dimension to the recorded data, for example, by performing two-dimensional chromatography (either LC–LC or GC–GC) followed by multivariate detection based on: (1) UV–visible absorption with a diode array detector (DAD) or (2) mass spectrometry (MS). This type of data, when measured for a single sample, can be arranged into a three-dimensional array, and suitable third-order multivariate calibration strategies can be applied to process them. In principle, an increasing number of multiple measurements, arranged in a mathematical object of more complex structure, should provide more selectivity and sensitivity to the analysis [4]. However, there may be additional analytical advantages in recording these complex data, which are still waiting to be uncovered by theoreticians [4].

In the present report, we briefly describe the available algorithms for processing second- and third-order chromatographic data, with focus on an important phenomenon which should be studied before some of these algorithms are applied: retention time shifts which may occur from run to run in chromatography. We then summarize the recent work on second- and higher-order

chromametrics, particularly in what concerns quantitative analytical determinations in complex samples achieving the second-order advantage.

2. Higher-order analytical calibration

2.1. Calibration scenarios

Various calibration schemes are possible depending on the recorded instrumental data. In univariate calibration, a single numerical value (scalar or zeroth-order datum) is measured per sample, whereas in multivariate calibration increasingly complex data arrays are measured per sample, allowing analytical quantitative estimations in systems of low intrinsic selectivity [1,4,11]. In first-order calibration methods, the analyzed response is a vector of numerical values (first-order data) per sample [1,4,11,13]. The use of this information provides the so-called first-order advantage, i.e., the possibility to quantify an analyte in the presence of interferences, as long as the interfering compounds are present in the calibration samples during the establishment of the calibration model. In second-order calibration methods, the analyzed response is a data matrix per sample (second-order data, e.g., a liquid chromatogram with diode array detection) [1,4,11,13]. The so-called second-order advantage is achieved, which implies that the analyte contribution can be appropriately modeled, quantitatively estimated and resolved in the presence of unknown interferences, absent in the calibration samples [1,4,11]. Higher-order data sets (e.g. two-dimensional gas chromatography with mass spectrometric detection) and their related calibration strategies (higher-order multivariate calibration methods) are possible and improve resolution power and quantitative estimations [1,4,11].

2.2. Nomenclature

In multivariate calibration the term 'order' is usually employed to denote the number of modes for the data array which is recorded for a *single sample*. The term 'way', on the other hand, is reserved for the number of modes of the mathematical object which is built by joining data arrays measured for a *group of samples*. In this sense, the classical univariate calibration, which operates using a single datum per sample, is a zeroth-order and also a one-way method. Correspondingly, first-order data per sample leads to two-way data sets, second-order data per sample to three-way data sets, third-order data per sample to four-way data sets, fourth-order data per sample to five-way data sets, etc.

The analytical community seems to prefer 'order' for distinguishing the various calibration scenarios, focusing on the data dimensions collected for each sample. This is also linked to the expression 'second-order advantage', which is popular in analytical chemistry works. On the other hand, in the context of multivariate data modeling, in non-analytical applications and in many basic chemometric works, the preferred expression is 'way'. However, this latter terminology should strictly be applied to truly multi-way algorithms, because in some multivariate calibration algorithms a multi-way data array is never built.

2.3. Data multi-linearity

Before applying a specific multivariate algorithm to second- and third-order chromatographic data, it is necessary to assess if the sample constituent profiles in the retention time dimension are constant or not. In chemometric terms, constancy of retention time profiles implies that the recorded data are multi-linear. In general, multi-linearity can be defined as the possibility of mathematically expressing a generic element of a multi-way data array as a linear function of component concentrations and profiles in the

data modes. When second- and third-order data fulfil this requirement, they are called 'trilinear' and 'quadrilinear' respectively. In the chromatographic field, multi-linearity implies that the shape of the profiles in the relevant retention time dimension(s) for a given component is the same.

There are basically two strategies for dealing with the phenomenon of varying retention times: (1) to restore the data multi-linearity, removing the effect introduced by sample-to-sample retention time shifts with a suitable algorithm for aligning or synchronizing the chromatograms, or (2) to process the data with algorithms allowing for varying retention time profiles across samples. For recent reviews on retention time shifts and their correction before chemometric processing see Refs. [3,10,12,14].

It may be noticed that, in general, additional causes of multi-linearity losses exist, as has been recently reviewed [15].

2.4. Second-order calibration algorithms

Second-order chromatographic data can be processed by a variety of algorithms for analyte quantitation. Those classified as trilinear are: (1) parallel factor analysis (PARAFAC) [16], (2) different versions of alternating trilinear decomposition (ATLD) [17], such as self-weighted ATLD (SWATLD) [18] and penalized ATLD (APTLD) [19], (3) generalized rank annihilation (GRAM) [20], (4) direct trilinear decomposition (DTLD) [21], and (5) bilinear least-squares combined with residual bilinearization (BLLS/RBL) [22,23]. All of these methods assume an intrinsic mathematical model in which the profiles of all components are the same in all samples, and thus they require that: (1) the retention time profiles are constant from sample to sample, or (2) the chromatograms are appropriately aligned before data processing. Each of these algorithms achieves the decomposition of an array of data built with second-order data for a group of samples in a specific way. In general, the consensus is that alternating least-squares (ALS) is the most efficient method for trilinear decomposition [11]. It is thus understandable that the most employed trilinear algorithm is PARAFAC, because of its recognized ALS efficiency and robustness, and also of its ability to process multiple calibration

samples. The available PARAFAC software permits a variety of constraints to be imposed during the ALS fit, which ensure reaching physically interpretable results [16]. Eigenvalue methods are less precise than PARAFAC [4], and hence GRAM and DTLD are less employed methodologies. Some authors, however, advocate for SWATLD and APTLD, claiming that they are faster and less sensitive to the tuning of some parameters in comparison with PARAFAC [18].

One common aspect to these trilinear algorithms is that they provide useful qualitative information, in the form of spectral (or other detection dimensions) and chromatographic profiles for individual sample components, whether they are calibrated or not. Each profile corresponds to the one which would have been recorded by physically isolating the constituent, hence the name 'virtual chromatography' to this property of second-order algorithms. Fig. 1 pictorially shows the construction of a three-dimensional data array from second-order chromatograms for several samples, and its decomposition with a trilinear model such as PARAFAC, rendering retention time and spectral profiles for individual sample components, as well as relative concentrations or scores. The three-way array is composed of the data matrices for all calibration samples and also for the unknown sample.

Usually both profiles shown in Fig. 1 are normalized to unit length, but the concentration scale is saved in the so-called scores or relative concentrations of each analyte in all samples (both calibration and unknown). The analyte scores for the calibration samples are then employed to construct a pseudo-univariate calibration plot, from which the concentration is predicted after interpolation of the analyte score in the unknown sample. Table 1 briefly describes the PARAFAC modeling of a multi-way data array collected for a group of samples, and the process of analyte calibration and prediction. Additional details can be found in the Supplementary Material.

On the other hand, there are algorithms allowing for deviations of multi-linearity, which may be able to model retention time shifts, such as: (1) multivariate curve resolution coupled to ALS (MCR-ALS) [24] and (2) PARAFAC2, a variant of PARAFAC which allows profile variations in one of the data dimensions from sample to sample
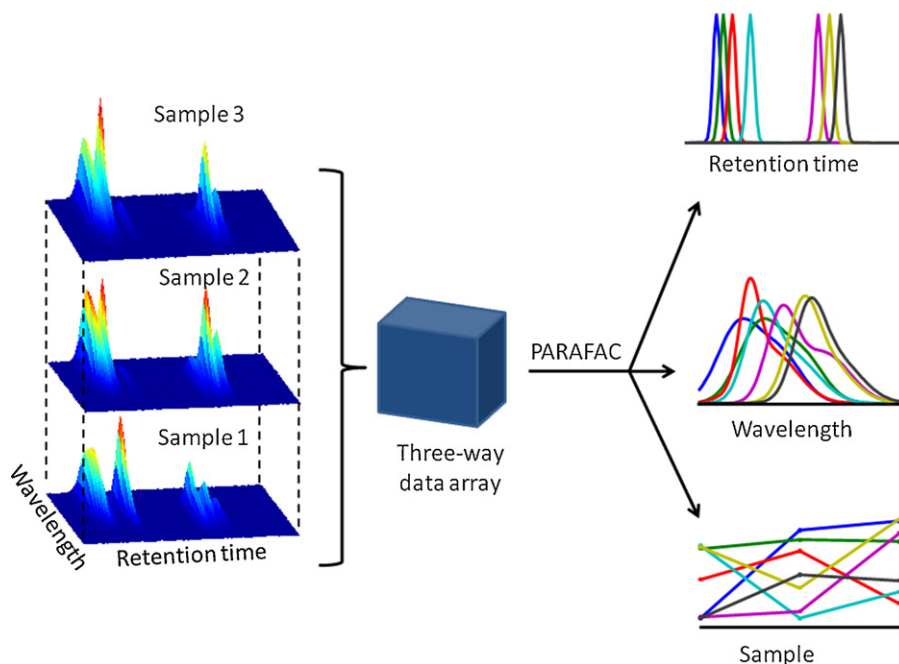

Fig. 1. Pictorial representation of second-order PARAFAC analysis. A three-way data array is built with matrix data for a set of samples (only three representative samples are shown). PARAFAC decomposition of the three-way array leads to retention time profiles, spectra and scores (relative concentrations) of all sample components. The time and spectral profiles are normalized, but the concentration information is retained in the scores, which are employed for analyte quantitation.

Table 1
Description of three types of algorithms employed for processing second-order data.

Algorithm	Comments[a]
PARAFAC	Assumes the data array built with second-order data for a group of samples follows a trilinear model, i.e., an element (i,j,k) is the sum of contributions of the form $(x_i \times y_j \times z_k)$, where x_i is the relative concentration of a component in the ith. sample, and y_j, z_k are the values of the time and spectral profiles at the jth, kth channel in each data dimension. Values of x_i are employed for analyte quantitation using a pseudo-univariate calibration graph.
MCR-ALS	Places second-order data for a group of samples adjacent to each other along the time dimension, and assumes that the augmented matrix follows a bilinear model, i.e., a matrix element (m,j) is the sum of contributions of the form $(x_m \times y_j)$, where x_m describes the profile for each sample in the augmented dimension, and y_j in the spectral dimension (m ranges from 1 to $I \times K$). For analyte quantitation, areas under each of the sample profiles in the augmented dimension are computed, and employed to build a pseudo-univariate calibration graph.
PLS/RBL	Calibrates a PLS model, and assumes the signals from the interferents in the test sample follow a bilinear model, i.e., an element (j,k) of the interferent array is the sum of contributions of the form $(y_j \times z_k)$, where y_j, z_k are abstract loadings in the spectral and time dimension. Analyte scores are produced by modeling the residuals of the fit of the test sample data array to the bilinear model, hence the name residual bilinearization.

[a] A three-way data array to be decomposed has a size $I \times J \times K$, and contains data for I samples, each as a $J \times K$ matrix. In MCR-ALS, this array is unfolded into a $J \times IK$ matrix.

[25]. In MCR-ALS, the basis of the successful data resolution is the building of an augmented matrix, placing all calibration and test data matrices adjacent to each other in the retention time direction. In this way, the time profile for a given component is allowed to vary from sample to sample. Fig. 2 shows how an augmented data matrix is created in the time dimension from individual sample matrices (both calibration and unknown), and how MCR-ALS

decomposition leads to individual component spectra and retention time profiles. The latter correspond to the augmented retention time dimension, with profiles for the various samples concatenated into single vectors. The area under a given component peak in a given sample defines the analyte score, which is employed for quantitative purposes in the same manner as in PARAFAC. This is also briefly explained in Table 1 and in the Supplementary Material. In comparison with PARAFAC, MCR-ALS needs initial estimates of either spectral or time profiles, while PARAFAC is usually automatically initialized. However, initial estimates for MCR-ALS can be efficiently computed by a variety of methods, such the computation of the so-called purest variables [26], or through evolving factor analysis (EFA) [27], which is particularly useful for chromatographic data. Most importantly, MCR-ALS does not require time synchronization of the chromatograms, which in practice may constitute a significant algorithmic advantage.

In the case of PARAFAC2, a relaxed PARAFAC model is employed which allows profiles to vary in one of the data dimensions (the retention time) from sample to sample. PARAFAC2 provides similar information to its trilinear counterpart, except that the retention time profile is not common to all samples (see Supplementary Material). What is important, however, is that it renders analyte scores which are also used to predict its concentration in the unknowns by pseudo-univariate calibration. PARAFAC2 admits lesser constraints to be imposed during the least-squares fit in comparison with the full range of constraints which are available for MCR-ALS in both data dimensions [25]. Thus, in certain complex cases the latter methodology may produce results which are better from the point of view of their physical interpretability, analytical accuracy and precision.

Other potentially useful second-order calibration methods are unfolded partial least-squares (U-PLS) [28] and multi-way PLS (N-PLS) [29], both of which should be combined with RBL if the second-order advantage is to be achieved [30–35]. These PLS algorithms intend to model the profile variations by incorporating a flexible latent structure in regressing the data (see Supplementary
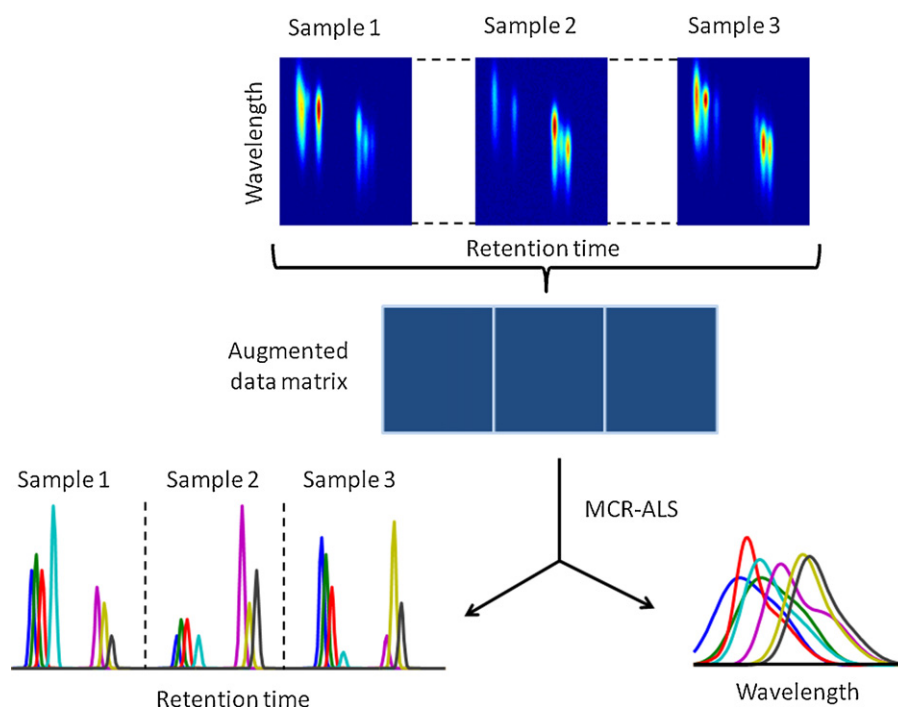

Fig. 2. Graphical representation of second-order MCR-ALS analysis. An augmented matrix is created with matrix data for a set of samples, placed adjacent to each other in the time dimension (only three representative samples are shown). MCR-ALS decomposition leads to spectra and retention time profiles of all sample components. These latter profiles describe the retention behavior in all samples. The area under each component peak defines the score, which is proportional to the component concentration.

Material). However, their full potentiality regarding the modeling of retention time variations is still a matter of debate. In any case, PLS models do not provide chemical recognizable information, because they internally work with abstract loadings, which are linear combinations of true profiles.

Table 1 intends to briefly describe the general manner in which three of the most employed data processing algorithms work on second-order data. The Supplementary Material provides detailed discussions concerning most of the above discussed algorithms for second-order multivariate calibration.

2.5. Third-order calibration algorithms

Third-order data are experimentally available from two-dimensional GC–GC or LC–LC chromatography with multivariate detection, requiring adequate third-order calibration models to be processed. They can also be classified according to whether the underlying models are quadrilinear or not [15].

Quadrilinear third-order models are: (1) PARAFAC, which is indeed suitable for any data order, (2) trilinear least-squares (TLLS) with residual trilinearization (RTL) [36], (3) alternating penalty quadrilinear decomposition (APQLD) [37] and (4) alternating weighted residue constraint quadrilinear decomposition (AWRCQLD) [38]. All of them require that the profiles of the sample components in both time dimensions remain constant from run to run, or are adequately corrected using a separate algorithm. As for second-order calibration, PARAFAC appears to be the model of choice in terms of efficiency, multiple-sample processing and accuracy in comparison with the remaining quadrilinear models [4]. However, since it is a multi-linear model, PARAFAC requires chromatographic synchronization, which may be a complex task in two-dimensional chromatography. Further details on the use of quadrilinear models for chromatographic data can be found in the Supplementary Material.

Fig. 3 shows the building of a four-way data array from a series of third-order arrays for a group of samples (both calibration and unknown), and the application of the PARAFAC model to retrieve

the constituent spectra, the retention time profiles of the sample components in both separation dimensions, and the relative concentrations or scores. As with second-order PARAFAC, profiles in the three dimensions are normalized, while the scores allow for analyte quantitation.

Flexible third-order multivariate models which may allow for temporal variation in profiles are those based on PLS regression combined with RTL [36] (i.e., U-PLS/RTL and N-PLS/RTL), although their applicability has not been tested to date in this regard.

A final data processing possibility for third-order data is to unfold them into matrices, concatenating the separation dimensions into a single mode, and applying a non-trilinear second-order methodology such as MCR-ALS or PARAFAC2 to the resulting second-order data. As mentioned above, MCR-ALS usually provides profiles with improved physical interpretability in comparison with PARAFAC2, due to the possibility of applying a variety of restrictions during the least-squares fit. Thus the MCR-ALS strategy is able to efficiently cope with the problem of retention time shifts. However, the price paid may be lower sensitivity and selectivity [39].

3. Discussion

3.1. Chromatographic alignment

Alignment or synchronization of chromatograms is an important chemometric activity which should be conducted before the application of multi-linear data processing algorithms. The basis of these techniques involves digitally moving (and/or stretching or compressing) a chromatogram until it matches a reference one, with certain objective function indicating the quality of the match (correlation coefficient, residual fit, similarity index, etc.). Some of the main differences among them are: (1) whether alignment is performed using the continuous signal or individually detected peaks, (2) whether or not signal intensity is used, and (3) whether or not scale changes are corrected. Most algorithms require a previously selected reference chromatogram, to which all the remaining
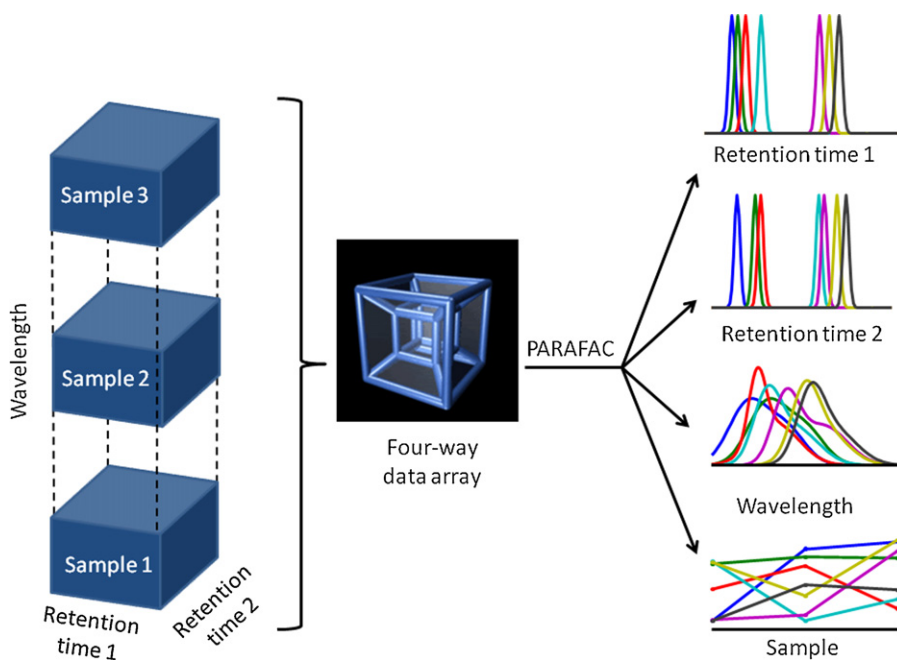

Fig. 3. Pictorial representation of third-order PARAFAC analysis. Third-order data for a set of samples (only three representative samples are shown) are joined to build a four-way data array, represented by the three-dimensional projection of a four-dimensional cube. PARAFAC decomposition leads to retention time profiles in both separation dimensions, spectra and scores (relative concentrations) of all sample components. The time and spectral profiles are normalized, but the concentration information is retained in the scores, as for second-order analysis.

ones are aligned. Suboptimal choice of the template could affect the alignment results, and thus it would be desirable to avoid this selection [40].

Two important alignment methods which have been employed for correcting retention time shifts prior to multivariate calibration analysis are: (1) rank alignment, based on principal component analysis of an augmented data matrix [41,42] and (2) correlation optimized warping (COW), based on piecewise linear stretching and compression of the time axis of the profiles [43]. Using artificial chromatograms, COW has been shown to be robust to varying peak numbers, heights and widths. Several refinements were implemented in COW and demonstrated to be useful for synchronizing both simulated and real LC–MS data before PARAFAC analysis [44]. In the presence of moderate time shifts, pre-processing with this COW-modified algorithm yielded reasonably trilinear data. Closely related to COW is dynamic time warping (DTW), which aligns a time profile to a specified reference one, except that every data point of a chromatographic-spectral data matrix can be moved [45]. In a specific case, COW and DTW subjected to some constraints were found to be superior to unconstrained DTW, which was unsuitable because it overcompensated the observed shifts [45].

In two-dimensional GC–GC or LC–LC separations, time shifts may occur in any of both time dimensions, creating the need for new alignment methods [46–48]. Results for the processing of time-shifted GC–GC–MS data were compared using PARAFAC on aligned data and PARAFAC2 on individual sample three-way GC–GC–MS arrays [49].

In the presence of uncalibrated interferences, the alignment of chromatograms becomes more difficult, and special methods have been developed to cope with this problem: (1) iterative target transformation factor analysis (ITTFA) [50], in which the calibration and test data matrices are independently decomposed into profiles and spectra using curve resolution techniques, and these profiles are digitally aligned before multivariate calibration is applied, and (2) decomposing a three-way array built with a test and a reference data matrix, using a suitably initialized and constrained PARAFAC model [51].

In this context, it should be mentioned that chromatographic alignment is profusely used as a pre-processing of GC–MS and LC–MS data in the framework of proteomic and metabolomic analysis, in order to facilitate the identification of proteins or metabolites within these data, which usually contain many individual signals [52,53]. Different algorithms for aligning GC–MS/LC–MS data for proteomic purposes have been developed [54–59]. ChromA, for example, provides useful visual representations of the synchronization process, with focus on differences and similarities between the chromatograms [60]. Although an initial selection of peaks for alignment is not strictly required, a priori knowledge can be used to improve and speedup the process. DTW has been successfully applied to LC–MS data to correct for chromatographic time shifts between replicate runs [61–63]. A variant of DTW has been employed to align multiple LC–MS analyses to a common template, allowing to detect differences between samples even with minimal chromatographic separation [64].

In the next sections, works on second- and third-order calibration of chromatographic data will be reviewed, indicating whether they were carried with previous time synchronization, or analyzed with algorithms modeling retention time shifts (MCR-ALS or PARAFAC2).

3.2. Second-order data

3.2.1. Liquid chromatography

Two pesticides (oxamyl and methomyl), resorcinol and phenol were determined in river and wastewater samples by processing LC–DAD data with GRAM, PARAFAC and MCR-ALS [65]. Before data processing with the trilinear models (GRAM and PARAFAC), time-shift correction was performed using ITTFA. Although the three algorithms provided similar figures of merit, MCR-ALS appeared to be more robust concerning the presence of chromatographic time shifts.

Synthetic mixtures of p-chlorobenzoic acid and benzoic acid, uracil and pyruvic acid, and fumaric, maleic and phenyl phosphoric acid were analyzed in the standard addition mode by two-dimensional LC–LC data (including an anion-exchange and a reverse-phase column with single-wavelength UV detection) [66]. The overlapping analyte signals were then processed by PARAFAC, after time shift correction using rank alignment and a new alignment method introduced by the authors, involving incrementally shifting the LC–LC matrices until the PARAFAC fit was optimal.

Pesticides such as simazine, carbaryl, carbendazim, methyl thiophanate and dimethoate, and two metabolites (phthalimide and 3,5-dichloroaniline) were determined in wine samples by LC–DAD data processed by both PARAFAC and BLLS [67]. The synchronization of the chromatographic matrices was done by rank alignment. BLLS presents some advantages such as not requiring initialization or constraints, but demands knowledge of all components in the calibration samples. Overall, the analytical performance of PARAFAC was superior.

LC–DAD data were processed with ATLD to simultaneously determine levodopa, carbidopa and methyldopa in human plasma [68], and with PARAFAC for the simultaneous determination of four aflatoxins in a set of spiked and naturally contaminated pistachio nuts in the presence of matrix interferences [69]. In both cases the effect of retention time shifts was corrected by rank alignment.

In certain LC experiments no significant shifts were detected, probably due to the small time window employed for data processing. This appeared to be the case in the quantitation of sulfamethoxypyridazine, sulfamethoxazole and sulfadimethoxine in porcine kidney by PARAFAC processing of LC–DAD data [70]. In the analysis of LC–DAD data for binary mixtures of lidocaine and prilocaine, however, the trilinear PARAFAC model was directly employed [71], but some of the analytical results were biased, and additional components to those expected were required to model the data. This was apparently due to retention time shifts, although no attempts were made to correct them.

Several agrochemicals were determined in environmental wastewaters and sediments in the presence of coelution phenomena and matrix interferences, processing LC–DAD [72], LC–MS [73] and fused LC–DAD–MS data [74]. The analytical results improved in going from DAD to MS to DAD–MS data, with final quantitation errors below 12% using MCR-ALS as chemometric algorithm.

The same MCR-ALS strategy was employed to process LC–DAD data for the quantitation of nine phenolic acids in both synthetic samples and strawberry samples [75], three synthetic dyes in non alcoholic beverages [76], seven non steroidal anti-inflammatory drugs and the anticonvulsant carbamazepine in river and wastewater [77], nine β-blockers and two analgesics (paracetamol and phenazone) in river water [78], pesticides in water samples [79], seven organic UV filters in effluent wastewaters [80] and four phenolic acids in olive oils [81]. In some of the above mentioned cases, MCR-ALS showed better predictive ability than other second-order algorithms such as U-PLS/RBL [76] or PARAFAC2 [81]. MCR-ALS was also employed to process data from LC with attenuated total reflection-FTIR detection for the determination of carbohydrates, alcohols and organic acids in red wine [82].

PARAFAC2 was used as an alternative for processing LC–DAD data aimed at the determination of six sulfamides after extraction from kidney [83].

Ten polycyclic aromatic hydrocarbons (PAHs) were determined in aqueous samples in the presence of interferences, processing second-order LC-fluorescence detection data [84]. Short analysis

times were achieved under isocratic conditions. Both MCR-ALS and PARAFAC2 were found to be useful to overcome the presence of uncalibrated interferences and retention time shifts. However, MCR-ALS was found to provide better recoveries.

Both aligned and unaligned LC-fluorescence emission data matrices were studied for the determination of eight fluoroquinolones in samples with and without interferences [85]. PARAFAC and N-PLS/RBL yielded good results for all the investigated systems, provided they were fed with suitably aligned chromatographic profiles using rank alignment. MCR-ALS did also produce reasonably accurate results, even if raw data were processed. However, serious spectral overlapping affected the MCR-ALS predictions in one of the analyzed cases.

Similar LC-fluorescence emission data matrices were processed using MCR-ALS, allowing for the determination of the marker pteridins neopterin, biopterin, pterin, xanthopterin and isoxanthopterin in urine samples in the presence of interferences, and the monitoring of the pteridine/creatinine ratio in pathological children urines [86].

As mentioned above, the latent structured PLS models have been scarcely applied to LC data with multivariate detection, and thus only few examples exist on the subject. In one of them, U-PLS/RBL was applied to LC–DAD data for the simultaneous determination of eight tetracyclines in wastewaters [87], with predictive results which were comparable to those from MCR-ALS. In the second, MCR-ALS and PLS were compared for the resolution of coeluted peaks of pyrocatechol, dopamine and epinephrine in LC with electrochemical detection [88]. Voltammetric detection coupled to MCR-ALS was preferred for high analyte concentrations, whereas amperometric detection combined with PLS was more adequate at lower levels.

3.2.2. Gas chromatography

Six essential oil markers were quantitated in perfumes using two-dimensional gas chromatography (GC–GC) with flame-ionization detection [89]. The results of PARAFAC, PARAFAC2 and N-PLS were compared. For correcting retention time shifts, COW was applied.

Fraga developed a chemometric approach for the determination of triethyl phosphate and 1,4-dithiacyclohexane in environmental samples, in a GC-selected ion monitoring (GC-SIM) mass spectrometer [90], which collects only the signals for ions having masses of interest rather than masses that span a wide range as in full scan mode. Run-to-run retention-time differences between GC-SIM data matrices for a sample and a standard addition were corrected using rank alignment. The trilinear GRAM model was subsequently applied to quantify the target analytes.

Clenbuterol was analyzed in several biological matrices by GC–MS using DTLD, PARAFAC, PARAFAC2 and N-PLS as processing algorithms [91], and non-steroidal anti-inflammatory drugs in bovine milk were determined by GC–MS data processed by PARAFAC or PARAFAC2, depending on whether the data were trilinear or not [92]. Several works have been published processing GC–GC data with flame ionization detection (FID) with N-PLS: determination of naphthene [93] and kerosene in gasolines [94], identification of gasoline adulteration [95], prediction of physicochemical properties of gasoline [96], and allergens in perfumes [97]. The same type of data were processed with MCR-ALS for the analysis of essential oils in perfumes [98].

GRAM and MCR-ALS were used to process GC–MS data for the quantitation of four unsaturated fatty acids in the presence of interfering components [99]. Unlike MCR-ALS, a retention time shift correction on GC profiles was necessary for GRAM [41]. Amigo et al. have shown the potential of PARAFAC2 for solving common GC–MS problems, using data from wine samples to illustrate the solutions [100].

3.3. Third-order data

Several published examples of third-order chromatographic data comprise two-dimensional chromatography (either GC–GC or LC–LC) with multivariate detection. When detection proceeds by time-of-flight mass spectrometry (TOFMS), the trilinear algorithms retrieve retention time profiles on both chromatographic columns, and a complete mass spectrum for each component. Examples of third-order GC–GC–TOFMS data processed by PARAFAC have been described for the resolution of four isomers (*iso*-butyl, *sec*-butyl, *tert*-butyl and *n*-butyl benzenes) [101], for the study of environmental samples containing fuel components, pesticides and natural products [102], for the determination of a neurodegenerative biomarker in human tissue [103] and impurities in a chemical weapon precursor [104].

An analytical procedure based on MCR-ALS processing of GC–GC–TOFMS for the simultaneous determination of 97 organic contaminants at trace concentration in river water was presented, including 13 pharmaceuticals, 18 plasticizers, 8 personal care products, 9 acid herbicides, 8 triazines, 10 organophosphorous compounds, 5 phenylureas, 12 organochlorine biocides, 9 PAHs and 5 benzothiazoles and benzotriazoles [105]. In the latter case, and in a recent analysis of mixtures of polycyclic aromatic hydrocarbons from GC–GC–MS data [106], the original third-order data were unfolded into matrices by concatenating the GC–GC chromatographic matrices into vectors at each detection channel. The resulting matrix data were processed using MCR-ALS.

In order to predict the composition of biodiesels and conventional diesels, PLS models were applied to data of varying degrees of complexity, obtained by GC with MS detection: (1) unidimensional GC data with single ion detection (first-order data), (2) two-dimensional GC–MS and GC–GC with single ion detection (second-order data) and (3) fully three-dimensional GC–GC–MS (third-order data) [107]. The analytical figures of merit significantly improved in going from first- to second-order calibration, but third-order data did not appear to produce further improvements.

In some cases, retention time corrections were found to be necessary. When studying metabolite extracts isolated from yeast cells with PARAFAC [108], an algorithm for retention time alignment was applied, based on peak matching at a single *m/z* value [109].

As an interesting development in this field, a comprehensive three-dimensional gas chromatograph (GC–GC–GC) has been described [110]. The resulting third-order data were shown to be adequately modeled by quadrilinear PARAFAC, providing considerably higher sensitivity in comparison with third-order GC–GC–MS data.

In the case of two-dimensional LC–LC, the problem of retention time shifts can be overcome by dividing the entire chromatogram in small regions, allowing for decomposition of data which are apparently quadrilinear. This has been done, for example, in metabolomic analysis carried out by processing third-order LC–LC–DAD data with PARAFAC, leading to the quantitative analysis of a number of indolic compounds in maize seedlings [111].

In a recent LC–LC–DAD metabolic study, the original third-order data were unfolded into second-order data, by concatenating the LC–LC matrices into vectors at each detection wavelength. The resulting matrices were processed by applying MCR-ALS, eliminating the problem caused by retention time shifts [112]. For initialization of MCR-ALS, it was found that iterative key set factor analysis (IKSFA) was better than other methods for finding initial estimates of component spectra.

It should be noticed that unfolding from third- to second-order data may lead to decreased sensitivity in comparison with the processing of the original data with a truly multivariate third-order methodology. The latter would require synchronization of two-dimensional chromatograms in both separation dimensions before

data processing. Clearly the field needs some additional research in the time alignment of multi-dimensional chromatograms.

4. Conclusions and outlook

Chromatographic separations followed by multivariate detection are producing increasingly complex data structures, whose appropriate chemometric processing opens new dimensions in analytical studies. Improved sensitivity and selectivity, and the possibility of analyte quantitation in the presence of uncalibrated interferents, are some of the advantages which can be achieved. Future work will certainly lead to data of even higher dimensions on the instrumental side, and to the development of new data processing tools on the chemometric one. Specifically, as the instrumentation produces data arrays of increasing number of dimensions and complex structures, it will be necessary to actively research in the following areas: (1) efficient synchronization methods for two-dimensional chromatographic data, in order to provide suitable multi-linear data for robust algorithms such as PARAFAC, (2) theoretical developments in sensitivity and other multi-way figures of merit, in order to be able to compare the relative performances of PARAFAC processing of the original (synchronized) multi-way data and MCR-ALS processing of the data unfolded to arrays of lower dimensions (matrices), (3) development, testing and comparison with classical methods of new latent-structured methods based on PLS/residual multi-linearization, which are capable of handling non-multilinear chromatographic data.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2012.02.004

References

[1] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[2] V.G. van Mispelaar, PhD Thesis, Amsterdam, The Netherlands, 2005.
[3] J.M. Amigo, T. Skov, R. Bro, Chem. Rev. 110 (2010) 4582.
[4] A.C. Olivieri, Anal. Chem. 80 (2008) 5713.
[5] I. François, K. Sandra, P. Sandra, Anal. Chim. Acta 641 (2009) 14.
[6] M. Adahchour, J. Beens, U.A.Th. Brinkman, J. Chromatogr. A 1186 (2008) 67.
[7] M. Daszykowski, B. Walczak, Trends Anal. Chem. 25 (2006) 1081.
[8] A.C. Duarte, S. Capelo, J. Liq. Chromatogr. Rel. Technol. 29 (2006) 1143.
[9] M. Wasim, R.G. Brereton, J. Chem. Inf. Model. 46 (2006) 1143.
[10] M.C. Ortiz, L. Sarabia, J. Chromatogr. A 1158 (2007) 94.
[11] 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.
[12] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, J. Chromatogr. A 1168 (2007) 3.
[13] V. Gómez, M.P. Callao, Anal. Chim. Acta 627 (2008) 169.
[14] T. Skov, R. Bro, Anal. Bioanal. Chem. 390 (2008) 281.
[15] A.C. Olivieri, G.M. Escandar, A. Muñoz de la Peña, Trends Anal. Chem. 30 (2011) 607.
[16] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[17] H.L. Wu, M. Shibukawa, K. Oguma, J. Chemometrics 12 (1998) 1.
[18] Z.P. Chen, H.L. Wu, J.H. Jiang, Y. Li, R.Q. Yu, Chemom. Intell. Lab. Syst. 52 (2000) 75.
[19] A.L. Xia, H.L. Wu, D.M. Fang, Y.J. Ding, L.Q. Hu, R.Q. Yu, J. Chemometrics 19 (2005) 65.
[20] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[21] E. Sanchez, B.R. Kowalski, J. Chemometrics 1 (1990) 29.
[22] M. Linder, R. Sundberg, Chemom. Intell. Lab. Syst. 42 (1998) 159.
[23] M. Linder, R. Sundberg, J. Chemometrics 16 (2002) 12.
[24] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133.
[25] H.A.L. Kiers, J.M.F. Ten Berge, R. Bro, J. Chemometrics 13 (1999) 275.
[26] W. Windig, J. Guilment, Anal. Chem. 63 (1991) 1425.
[27] M.C. Antunes, J.E.J. Simao, A.C. Duarte, R. Tauler, Analyst 127 (2002) 809.
[28] S. Wold, P. Geladi, K. Esbensen, J. Øhman, J. Chemometrics 1 (1987) 41.
[29] R. Bro, J. Chemometrics 10 (1996) 47.
[30] J. Öhman, P. Geladi, S. Wold, J. Chemometrics 4 (1990) 79.
[31] V.A. Lozano, G.A. Ibañez, A.C. Olivieri, Anal. Chim. Acta 610 (2008) 186.
[32] A.C. Olivieri, J. Chemometrics 19 (2005) 253.
[33] M.J. Culzoni, H.C. Goicoechea, A.P. Pagani, M.A. Cabezón, A.C. Olivieri, Analyst 131 (2006) 718.
[34] A. García Reiriz, P.C. Damiani, M.J. Culzoni, H.C. Goicoechea, A.C. Olivieri, Chemom. Intell. Lab. Syst. 92 (2008) 61.
[35] A. García Reiriz, P.C. Damiani, A.C. Olivieri, Chemom. Intell. Lab. Syst. 100 (2010) 127.
[36] 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.
[37] A.L. Xia, H.L. Wu, S.F. Li, S.H. Zhu, L.Q. Hu, R.Q. Yu, J. Chemometrics 21 (2007) 133.
[38] 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. Chemometrics 25 (2011) 408.
[39] X. Liu, S.D. Sidiropoulos, IEEE Trans. Signal Process. 49 (2001) 2074.
[40] J. Listgarten, A. Emili, Mol. Cell. Proteomics 4 (2005) 419.
[41] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Anal. Chem. 70 (1998) 218.
[42] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[43] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[44] D. Bylund, R. Danielsson, G. Malmquist, K.E. Markides, J. Chromatogr. A 961 (2002) 237.
[45] G. Tomasi, F. van den Berg, J. Chemometrics 18 (2004) 231.
[46] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, Anal. Chem. 77 (2005) 7735.
[47] K.J. Johnson, B.J. Prazen, D.C. Young, R.E. Synovec, J. Sep. Sci. 27 (2004) 410.
[48] R.C. Allen, S.C. Rutan, Anal. Chim. Acta 705 (2011) 253.
[49] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, J. Chromatogr. A 1216 (2009) 4020.
[50] E. Comas, R.A. Gimeno, J. Ferré, R.M. Marcé, F. Borrull, F.X. Rius, Anal. Chim. Acta 470 (2002) 163.
[51] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, Chemom. Intell. Lab. Syst. 101 (2010) 30.
[52] R.J.O. Torgrip, E. Alm, K.M. Åberg, Bioanal. Rev. 1 (2010) 105.
[53] C. Christin, H.C.J. Hoefsloot, A.K. Smilde, F. Suits, R. Bischoff, P.L. Horvatovich, J. Proteome Res. 9 (2010) 1483.
[54] P. Jonsson, A.I. Johansson, J. Gullberg, J. Trygg, A. Jiye, B. Grung, S. Marklund, M. Sjöström, H. Antti, T. Moritz, Anal. Chem. 77 (2005) 5635.
[55] P. Jonsson, E.S. Johansson, A. Wuolikainen, J. Lindberg, I. Schuppe Koistinen, M. Kusano, M. Sjöström, J. Trygg, T. Moritz, H. Antti, J. Proteome Res. 5 (2006) 1407.
[56] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[57] R.C.H. De Vos, S. Moco, A. Lommen, J.J.B. Keurentjes, R.J. Bino, R.D. Hall, Nat. Protocols 2 (2007) 778.
[58] O. Kohlbacher, K. Reinert, C. Gröpl, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, M. Sturm, Bioinformatics 23 (2007) 191.
[59] M. Sturm, A. Bertsch, C. Gröpl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert, O. Kohlbacher, BMC Bioinformatics 9 (2008) 163.
[60] N. Hoffmann, J. Stoye, Bioinformatics 25 (2009) 2080.
[61] A. Prakash, P. Mallick, J. Whiteaker, H. Zhang, A. Paulovich, M. Flory, H. Lee, R. Aebersold, B. Schwikowski, Mol. Cell. Proteomics 5 (2006) 423.
[62] R.G. Sadygov, F.M. Maroto, A.F. Huhmer, Anal. Chem. 78 (2006) 8207.
[63] J.T. Prince, E.M. Marcotte, Anal. Chem. 78 (2006) 6140.
[64] G.L. Finney, A.R. Blackler, M.R. Hoopmann, J.D. Canterbury, C.C. Wu, M.J. MacCoss, Anal. Chem. 80 (2008) 961.
[65] E. Comas, R.A. Gimeno, J. Ferré, R.M. Marcé, F. Borrull, F.X. Rius, J. Chromatogr. A 1035 (2004) 195.
[66] C.G. Fraga, C.A. Corley, J. Chromatogr. A 1096 (2005) 40.
[67] J.W.B. Braga, C.B.G. Bottoli, I.C.S.F. Jardim, H.C. Goicoechea, A.C. Olivieri, R.P. Poppi, J. Chromatogr. A 1148 (2007) 200.
[68] S.F. Li, H.L. Wu, Y.J. Yu, Y.N. Li, J.F. Nie, H.Y. Fu, R.Q. Yu, Talanta 81 (2010) 805.
[69] M. Vosougha, M. Bayata, A. Salemi, Anal. Chim. Acta 663 (2010) 11.
[70] I. García, L. Sarabia, M. Cruz Ortiz, J.M. Aldama, Analyst 129 (2004) 766.
[71] K. Wiberg, S.P. Jacobsson, Anal. Chim. Acta 514 (2004) 203.
[72] E. Peré Trepat, A. Hildebrandt, D. Barceló, S. Lacorte, R. Tauler, Chemom. Intell. Lab. Syst. 74 (2004) 293.
[73] E. Peré Trepat, S. Lacorte, R. Tauler, J. Chromatogr. A 1096 (2005) 111.
[74] E. Peré Trepat, R. Tauler, J. Chromatogr. A 1131 (2006) 85.
[75] S. Mas, G. Fonrodona, R. Tauler, J. Barbosa, Talanta 71 (2007) 1455.
[76] 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.
[77] M.D. Gil García, F. Cañada Cañada, M.J. Culzoni, L. Vera Candioti, G.G. Siano, H.C. Goicoechea, M. Martínez Galera, J. Chromatogr. A 1216 (2009) 5489.
[78] M. Martínez Galera, M.D. Gil García, M.J. Culzoni, H.C. Goicoechea, J. Chromatogr. A 1217 (2010) 2042.
[79] R.M. Maggio, P.C. Damiani, A.C. Olivieri, Talanta 83 (2011) 1173.
[80] M. Vosough, N.R. Mojdehi, Talanta 85 (2011) 2175.
[81] F. Marini, A. D'Aloise, R. Bucci, F. Buiarelli, A.L. Magrì, A.D. Magrì, Chemom. Intell. Lab. Syst. 106 (2011) 142.
[82] A. Edelmann, J. Diewok, J. Rodriguez Baena, B. Lendl, Anal. Bioanal. Chem. 376 (2003) 92.
[83] I. García, M.C. Ortiz, L. Sarabia, J.M. Aldama, Anal. Chim. Acta 587 (2007) 222.

[84] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, Anal. Chem. 81 (2009) 8074.
[85] F. Cañada Cañada, J.A. Arancibia, G.M. Escandar, G.A. Ibañez, A. Espinosa Mansilla, A. Muñoz de la Peña, A.C. Olivieri, J. Chromatogr. A 1216 (2009) 4868.
[86] 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.
[87] M.D. Gil García, M.J. Culzoni, M.M. De Zan, R. Santiago Valverde, M. Martínez Galera, H.C. Goicoechea, J. Chromatogr. A 1179 (2008) 115.
[88] O. González García, C. Ariño, J.M. Díaz Cruz, M. Esteban, Chemom. Intell. Lab. Syst. 93 (2008) 49.
[89] V.G. van Mispelaar, A.C. Tas, A.K. Smilde, P.J. Schoenmakers, A.C. van Asten, J. Chromatogr. A 1019 (2003) 15.
[90] C.G. Fraga, J. Chromatogr. A 1019 (2003) 31.
[91] I. García, L. Sarabia, M. Cruz Ortiz, J.M. Aldama, Anal. Chim. Acta 515 (2004) 55.
[92] D. Arroyo, M. Cruz Ortiz, L.A. Sarabia, J. Chromatogr. A 1218 (2011) 4487.
[93] B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Anal. Chem. 73 (2001) 5677.
[94] L.A.F. de Godoy, E.C. Ferreira, M.P. Pedroso, C.H.V. Fidélis, F. Augusto, R.J. Poppi, Anal. Lett. 41 (2008) 1603.
[95] M.P. Pedroso, L.A.F. de Godoy, E.C. Ferreira, R.J. Poppi, F. Augusto, J. Chromatogr. A 1201 (2008) 176.
[96] L.A.F. de Godoy, M.P. Pedroso, E.C. Ferreira, F. Augusto, R.J. Poppi, J. Chromatogr. A 1218 (2011) 1663.
[97] L.A.F. de Godoy, M.P. Pedroso, L. Wang, F. Augusto, R.J. Poppi, Talanta 83 (2011) 1302.
[98] L.A.F. de Godoy, L. Wang, M. Pedroso, R.J. Poppi, F. Augusto, Anal. Chim. Acta 699 (2011) 120.
[99] M. Vosough, A. Salemi, Talanta 73 (2007) 30.
[100] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Trends Anal. Chem. 27 (2008) 714.
[101] A.E. Sinha, J.L. Hope, B.J. Prazen, C.G. Fraga, E.J. Nilsson, R.E. Synovec, J. Chromatogr. A 1056 (2004) 145.
[102] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, J. Chromatogr. A 1027 (2004) 269.
[103] L.R. Snyder, J.C. Hoggard, T.J. Montine, R.E. Synovec, J. Chromatogr. A 1217 (2010) 4639.
[104] J.C. Hoggard, J.H. Wahl, R.E. Synovec, G.M. Mong, C.G. Fraga, Anal. Chem. 82 (2010) 689.
[105] V. Matamoros, E. Jover, J.M. Bayona, Anal. Chem. 82 (2010) 699.
[106] H. Parastar, J.R. Radovic, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Anal. Chem. 83 (2011) 9289.
[107] K.M. Pierce, S.P. Schale, Talanta 83 (2011) 1254.
[108] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Anal. Chem. 78 (2006) 2700.
[109] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, J. Chromatogr. A 996 (2003) 141.
[110] N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Anal. Chem. 79 (2007) 8280.
[111] S.E.G. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Anal. Chem. 78 (2006) 5559.
[112] H.P. Bailey, S.C. Rutan, Chemom. Intell. Lab. Syst. 106 (2011) 131.