Trends in Analytical Chemistry 137 (2021) 116207

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

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

Untangling comprehensive two-dimensional liquid chromatography data sets using regions of interest and multivariate curve resolution approaches

Miriam Pérez-Cova [a, b], Joaquim Jaumot [a], Romà Tauler [a, *]

[a] Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain
[b] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Diagonal 647, 08028, Barcelona, Spain

ARTICLE INFO

Article history:
Available online 28 January 2021

Keywords:
LC × LC
MCR-ALS
ROIMCR
Chemometrics
Data analysis

ABSTRACT

Data analysis remains a major challenge in the global application of comprehensive two-dimensional liquid chromatography (LC × LC). Advanced chemometric tools have been proposed to reduce the complexity of LC × LC datasets. In this work, key aspects of LC × LC are summarized from a chemometrics perspective. In particular, the recently developed ROIMCR method is proposed and adapted for LC × LC data analysis. First, this strategy consists of selecting of the Regions of Interest (ROI), in which data are filtered and compressed. Second, the resolution of the elution profiles of the sample constituents using the Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) method. A detailed overview of this recently developed tool and examples of its application in LC × LC are given, as well as pre-processing and post-processing tools to facilitate and complement the analysis of LC × LC data and the optimal interpretation of results.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Multidimensional separations have gained popularity in the last decades as a response to the need for increasing separation power [1,2] in the analysis of complex natural samples with matrix effects. The combination of different analytical platforms in the same study overcomes the limitations of one-dimensional approaches, i.e., resolution of overlapping peaks. Coelution of multiple analytes in one-dimensional liquid chromatography (1D-LC) might difficult its resolution, whereas the aid of an extra separation dimension can potentially separate, identify and quantify them. For instance, in comprehensive two-dimensional liquid chromatography (LC × LC), the effluent coming out from the first dimension (^{1}D) column is collected and divided into fractions at fixed periods (so-called modulations), which are then further separated in the second dimension (^{2}D) column. However, the large number of detected signals in these experiments, especially when coupled to high resolution mass spectrometry (HRMS), makes the manual inspection of chromatograms not feasible. In addition, beyond the analytical challenges, a significant increase in the data size and complexity is produced when considering multidimensional chromatography [2,3].

Thus, there are two main challenges of LC × LC data analysis compared to 1D-LC. The first is the higher complexity of the obtained data structures, e.g., multiple chromatographic peaks are obtained for a single analyte in the second column [2]. The second is the need for data analysis software. Some examples of commercial software available are GC Image LC × LC Edition Software from GC Image™, AnalyzerPro® XD from SpectralWorks, and Chromsquare from Shimadzu. However, they present some limitations, as only basic pre-processing steps are commonly included in their workflows. Besides, some of them present vendor-specificity, which difficulties general strategies applicable independently of how data have been acquired [2]. Consequently, the resolution of these complex mixtures and the extraction of the maximum amount of information from them is still a major challenge nowadays [2,4]. Hence, the lack of well stablished data-analysis protocols and software [3,5] emerges as the main drawback of multidimensional separations, and specifically of LC × LC.

Chemometrics is the discipline that applies mathematical and statistical methods to chemical systems with two main goals [6,7]. On the one hand, to improve the measurement process, in order to obtain optimal procedures and experiments. On the other hand, to extract maximum relevant qualitative and quantitative information from the chemical measurements. Consequently, chemometrics

* Corresponding author.
E-mail address: roma.tauler@idaea.csic.es (R. Tauler).

https://doi.org/10.1016/j.trac.2021.116207
0165-9936/© 2021 Elsevier B.V. All rights reserved.

List of acronyms

^{1}D	first dimension
^{2}D	second dimension
2DALC	2D assisted LC
ACD	at-column dilution
ANOVA	analysis of variance
ANOVA TP	ANOVA by target projection
ANOVA-PCA	ANOVA-principal component analysis
APTLD	alternating penalty trilinear decomposition
ASCA	ANOVA-simultaneous component analysis
ASM	active solvent modulation
ATLD	alternating trilinear decomposition
COShift	correlation-optimized shifting
COW	correlation optimized warping
DAD	diode array detector
DOE	experimental design
DWT	dynamic time warping
GASCA	group-wise ANOVA simultaneous component analysis
GC × GC	comprehensive 2D GC
IKSFA	iterative key set factor analysis
IOPA	iterative orthogonal projection approach
LC × LC	comprehensive 2D LC
LC × LC-HRMS	comprehensive 2D LC-high resolution MS
LC-LC	heart-cutting 2D LC
MANOVA	multivariate analysis of variance
MCR-ALS	multivariate curve resolution-alternating least squares
mLC-LC	multiple heart-cutting 2D LC
NEB	normal-exponential bernoulli
NG	normal-gamma
NGB	normal-gamma-bernouilli
NN	neural networks
OPLS-DA	orthogonal partial least squares-discriminant analysis
PAFFT	peak alignment fast fourier transform
PARAFAC	Parallel Factor Analysis
PARAFAC2	Parallel Factor Analysis 2
PCA	principal component analysis
PFP	pentafluorophenyl phase
PLS-DA	partial least squares-discriminant analysis
RF	random forest
ROI	regions of interest
ROIMCR	regions of interest-multivariate curve resolution
SCA	simultaneous component analysis
SIMCA	soft independent modeling of class analogies
sLC × LC	selective comprehensive 2D LC
SOM	self-organizing maps
SPAM	stationary-phase-assisted modulation
SVD	singular value decomposition
SVM	support vector machines
SWALTD	self-weighted alternating trilinear decomposition
TAGs	triacylglycerides
VEM	vacuum-evaporation modulation
VIPs	variables important in projection

includes, for instance, experimental design, multivariate calibration, pattern recognition and classification. Advanced chemometric methods go a step further from basic data analysis tools, also embracing methods for more sophisticated data pre-processing and model improvement, variable selection or resolving very complex mixtures among others. Thus, cutting-edge chemometrics has arisen as powerful tools able to shed some light into this issue and get through this bottleneck, offering solutions that can facilitate data compression, resolution and eventually interpretation of results. These goals are often achieved by combining different strategies. First, there are methods which aim to pre-process the experimental data to improve their quality (i.e., methods for data compression, baseline correction, elimination of background signals, alignment of chromatographic peaks within modulations of the same or among multiple samples). Second, other methods pursue to extract the sought analytical information from the data, including the discovery, resolution and quantitation of the components present in the analyzed samples. Hence, the usual output of this step is a table or matrix of the peak areas or concentrations of the resolved components. Third, there are additional post-processing steps, which include multiple types of multivariate data analysis methods. Their objective is to explore the patterns present in these tables or matrices and classify them into different groups (in the case where several types of samples are compared). Statistical analysis can also be performed to assess the effects of the experimental factors in designed experiments or to discover the most important variables (i.e., biomarkers) defining the investigated processes, often from a multivariate point of view. Although the third step can be applied directly on the pre-processed chromatograms from step one, only major differences between samples will be observed. It might be useful for a quick and visual overview of patterns in the data. For a more in-depth analysis, the second step is highly recommended, as it provides quantitative information. Besides, if the aim is, for instance, to identify variables responsible for the differences between the samples, performing this second step is necessary.

The aims of the present work are, on one side, to briefly describe the state of the art of LC × LC, and on another side, to go in-depth into the chemometric point of view, focusing on the structure of the datasets (i.e., possible multilinear behavior), and on the best strategies to analyze these huge data sets. In this sense, we propose the use of the spectral compression strategy based on the searching of the Regions of interests (ROI) [8], and its combination with the Multivariate Curve Resolution − Alternating Least Squares (MCR-ALS) [8,9] in the ROIMCR method for the optimal analysis of LC × LC data sets. In addition, some pre-processing and post-processing strategies of multivariate analysis of LC × LC data are also discussed, and some recent data examples of application are described to illustrate these data analysis strategies.

2. 2D-LC state of the art and recent developments

Two-dimensional chromatography (2D-LC) aims to increase peak capacity and peak production rates compared to 1D-LC when analyzing complex mixtures of compounds, often with strong peak coelutions and matrix effects. The second column or dimension, ^{2}D, adds extra selectivity to the separation, thanks to the possibility of combining two different retention mechanisms, ideally non-correlated (i.e., orthogonal) [1].

Here, only online chromatographic analysis is considered, where both dimensions are connected through a high-pressure valve that acts as the interface. However, other strategies, such as offline and stop-flow modes, have been also proposed [10]. Online setup allows a full automatization of the separation. There

are multiple types of online 2D-LC according to the number of fractions from the first column, ^{1}D, transferred to the second column, ^{2}D. Heart-cutting (LC-LC) implies that only one fraction from ^{1}D is collected and analyzed in the ^{2}D. In multiple heart-cutting (mLC-LC), several fractions of the ^{1}D separation are selected and stored in multiple loops before their further separation in the ^{2}D; whereas in selective comprehensive (sLC × LC), certain specific and successive regions of the ^{1}D separation are subsequent analyzed in the ^{2}D separation. Finally, comprehensive mode, LC × LC, implies the complete analysis of the effluent from the ^{1}D column also in the ^{2}D column [11]. In all cases, the separation already obtained in the ^{1}D needs to be maintained. Otherwise, undersampling will occur, which means that separated peaks are merged in the ^{2}D [12].

LC × LC is particularly powerful in untargeted analysis, where the comprehensive screening and profiling of all the constituents of a sample or set of samples are possible in one single chromatographic run. Non-destructive detectors (UV, Fluorescence) can be added in the middle of the two separations if needed, although it is not frequent. Nevertheless, destructive detectors (e.g., MS), can only be used at the end of both separations. Most of the examples of LC × LC applications are coupled to UV–Vis [13,14], MS [15,16] or both [17–19]. An extra dimension can be added if, for instance, LC × LC is coupled to ion mobility, as some other recent examples have proven [20–22].

One of the technical aspects that have suffered major improvements nowadays in LC × LC is the modulation strategy, i.e., how the interface between the two chromatographic separations is performed. A considerable effort has been spent in increasing solvent compatibility between dimensions. In passive modulation mode, there is no modification in the fractions from the ^{1}D prior to its further analysis in the ^{2}D column. Therefore, the ^{2}D separation may be highly affected by the ^{1}D effluent. Solvent strength mismatch should be avoided since it can lead to peak distortion and breakthrough, especially when peaks are not enough retained in the stationary phase due to the injection of a strong solvent coming from the previous separation. As a consequence, some of the sample constituents are eluted in the dead volume, instead of at their usual retention times [1]. As breakthrough issues arise as a major concern in LC × LC, special emphasis in its understanding has been made lately, to avoid or minimize it as much as possible. More information can be found, for instance, in the publications by Moussa et al. [23], about breakthrough from sampling loops, and van der Ven et al. [24], about how to improve the analysis of water-soluble biopolymers.

Active modulation strategies have been proposed to deal with solvent strength mismatch, breakthrough, and decreased detection sensitivity, caused by the dilution suffered by the sample in the LC × LC system. The main approaches developed to solve this problem are Active Solvent Modulation (ASM) [25], Stationary-Phase-Assisted Modulation (SPAM) [26], Vacuum-Evaporation Modulation (VEM) and at-column dilution (ACD) [27]. In ASM, the effluent from the ^{1}D, already stored in the sample loop, is diluted with weak solvent before this fraction exits the valve and enters the ^{2}D column. SPAM replaces standard storage loops by low-volume trapping columns, known as “traps”, whose stationary phase composition is similar to the ^{2}D column. VEM applies heat under vacuum conditions in the ^{1}D effluent to deposit the analytes in the loop, which are re-dissolved prior to their introduction in the ^{2}D column. A more detailed description of these three approaches can be found in Ref. [1]. ACD is the most recent approach and aims to automatically regulate dilution factor by adjusting flows from a transfer pump and from the ^{2}D gradient, which are joint in a mixer before reaching the ^{2}D column. The result is modulation based on a dilution at-column [28].

Optimization of the separation conditions in LC × LC is not straightforward. There are many parameters and considerations to be taken into account. The first step is the choice of the retention mechanisms for both dimensions, and therefore the selection of the appropriate stationary phases. This selection would require chromatographic expertise and previous knowledge about the chemical properties of the sample and targeted analytes. Then, mobile phase solvents, composition and modifiers need to be selected. In order to calculate retention parameters, statistical screening tests are useful as an input for the design of the experiments, optimal retention and chemometric modeling. Preliminary predictions of the retention times and simulation of elution profiles of the different analytes will reduce experimental effort and speed up the whole optimization process [29], especially due to the usually long LC × LC runs. Besides, automated multicolumn LC × LC workflows can accelerate method development, as they allow fast identification of the best combinations of columns and mobile phases compositions, for both targeted and untargeted analysis [30,31].

Choosing what quality descriptors (i.e., orthogonality, resolution) are required is a relevant step prior to objectively evaluate the quality of the separation under multiple possible scenarios. Another parameter to consider is how the gradient is performed in the ^{2}D, as different gradients can be applied. The main three are full, shifted and parallel gradients. A recent comparison of the three options can be found in Ref. [32]. In short, full gradients are the most common. They widely vary mobile phase composition in a very short period of time, and need re-equilibration time at the end, which limits the space in the 2D separation, as the modulation time is not fully employed in the separation itself. In contrast, in shifted gradients, the mobile phase composition range in every modulation is narrower and changes in agreement with the retention of the compounds in the ^{1}D column. However, they require specific software and hardware. Parallel gradients seem an attractive alternative in the case that retention mechanism of ^{1}D and ^{2}D are correlated, as in the case of employing reverse phase in both, RP × RP. In this option, the ^{2}D gradient is practically isocratic. The main advantage is that peak capacity increases, as the space available for the separation is higher than in the case of full gradient, for instance. Orthogonality is also improved when parallel gradients are used in both dimensions.

Evolutionary algorithms have been employed for optimizing gradient separations in LC × LC. These algorithms can be applied to method development and retention modeling, but also to molecular design or molecular modeling [33,34]. A comparison between genetic algorithms, non-adaptive evolution strategies and the covariance matrix adaptation evolution strategy has recently been published [35].

More detailed information about the use of LC × LC methods, practical aspects and data analysis can be found in previous reviews. For instance, the evaluation of the combination of different retention mechanisms in LC × LC was studied by Pirok et al. in Ref. [36]. In addition, recent developments in LC × LC and their new applications can be found in another publication by Pirok et al. [1]. A summary on the peak detection and profiling strategies for multidimensional chromatography was published by Navarro-Reig et al. in Ref. [4]. More information about retention modeling is reviewed in an article by den Uijl et al. [37]. Lastly, the chemometric analysis in one and two-dimensional chromatography was reviewed by Bos et al. in Ref. [2].

3. Multiway structure of LC × LC data

One important consideration to keep in mind when analyzing LC × LC data is how the data structure complexity increases when moving from one dimensional to multidimensional separations. For

instance, in one-dimensional liquid chromatography coupled to mass spectrometry (LC-MS), data are usually arranged in matrices with retention times in the rows and mass-to-charge ratios values (m/z values) in the columns. In chemometrics, this means that a single sample provides a two-way data set. Other common multi-wavelength detectors (e.g., LC-UV-Vis, using diode array detector or DAD), provide equivalent data structures. Analogously, when several samples are analyzed, three-way data sets are obtained, as shown in part **A** of Fig. 1. In multidimensional chromatography, more complex multiway data sets are obtained. For instance, in two-dimensional separations, a single sample gives a three-way data set, whose ways (modes/directions) are: ^{1}D retention times, ^{2}D retention times, and the multivariate response of the detector. In LC × LC-MS, this third mode is composed of the m/z values of the measured mass spectra. Therefore, the analysis of multiple samples in this situation will lead to a four-way dataset, exemplified in part **B** of Fig. 1. When combining different detectors at the same time, (i.e., LC × LC-DAD-MS), a data fusion step is required in order to horizontally concatenate the spectral data from each detector.

Another key aspect that should be considered despite data structure is what type of mathematical model suits the most for the analysis of the multiway data structure. According to the data set complexity, several multilinear models can be employed. In some cases, the choice can be straightforward, e.g., linear models for one-way data, bilinear models for two-way data, trilinear models for three-way data, and so on. However, this is not always necessarily true. For instance, there are some cases where the application of a trilinear model is not adequate for three-way data. Then, a lower complexity multilinear model, as the bilinear model, could be a better option.

Hence, the choice of the multivariate data analysis method depends on the multilinear behavior of the measured data. The fulfillment of trilinear models can be assessed when every chemical constituent can be expressed with a unique profile in every one of the three data ways, e.g., along the ^{1}D, ^{2}D and spectral dimensions for all samples analyzed, and they are invariant through the experimental conditions of the study [38–41].

Trilinearity loss can be caused by retention time shifts or changes in the peak shapes. This means that the same peak in different chromatographic runs may not appear in the same position and/or the same elution profile, which can be caused, for example, by fluctuations in pressure, flow, mobile phase composition and/or temperature [42,43]. These temporal misalignments in the data are often found in chromatography. For instance, although data from two-dimensional gas chromatography coupled to mass spectrometry (GC × GC-MS) have been analyzed under the

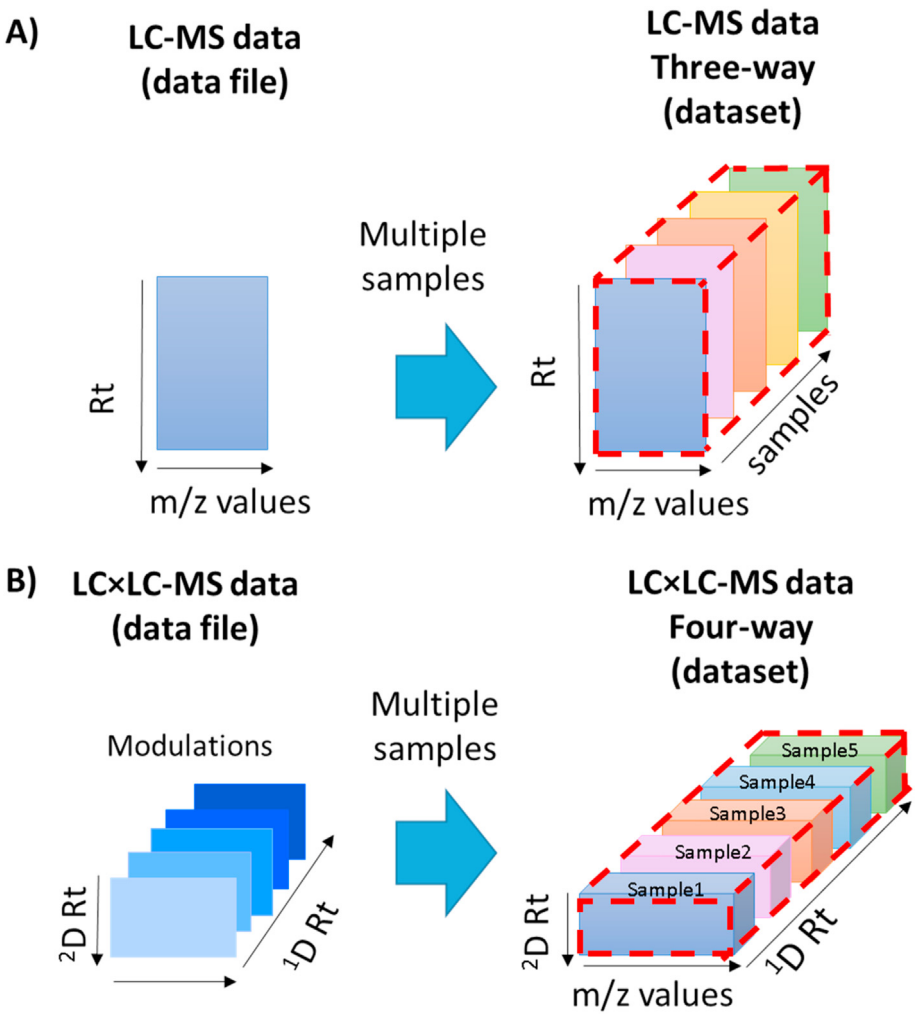

Fig. 1. Comparison of data arrangements in LC-MS (part A) and LC × LC-MS (part B). **A)** Results of LC-MS analysis of a single sample give a data matrix, whereas simultaneous LC-MS analysis of multiple samples provides a three-way data set. **B)** A single LC × LC-MS run is composed by the different modulations in which the ^{1}D separation is fractioned giving a three-way data set, and multiple LC × LC-MS runs provide a four-way data set.

assumption of the trilinear model [44,45], small deviations of this model have been encountered [46,47]. This issue is clearer in the case of liquid chromatography, even for 1D-LC, and it becomes critical in 2D-LC. Previous studies performed with LC × LC-MS and LC × LC-DAD, confirmed that deviations from the trilinear behavior were present [5,48,49]. The reason for this is because in LC × LC the total reproducibility of the elution profiles of the same component is not fulfilled, not only between the different chromatographic runs, but also within modulations. This variation is caused by both time shifts and changes in peak shapes. Therefore, in LC × LC, it is especially important to check if there are retention time shifts between modulations, and also between different samples, a problem which is commonly encountered in practice.

There are two manners of dealing with non-trilinear chromatographic data. The first implies the data pre-processing to achieve a proper alignment of peaks (see below, at the end of Section 4). Consequently, trilinearity is restored before employing methods based on trilinear models, such as Parallel Factor Analysis (PARAFAC). However, some difficulties may arise (e.g., increased complexity, the apparition of unexpected constituents, changes on chromatographic peak shapes when they are strongly coeluted, etc). A potential tool able to solve some of these difficulties is the Tchebichef moments' approach. This image moments method has demonstrated some advantages as pre-processing tools in the analysis of multiway chromatographic data with overlapping peaks and peak drifts [50]. However, this will be not possible when changes in the shapes of the chromatographic profiles of the same component in the different chromatographic runs are also produced.

The second involves the use of flexible algorithms which allows having different profiles in the samples mode for the same component, such as MCR-ALS or Parallel Factor Analysis 2 (an extension of PARAFAC that allows small deviations in its multilinear behavior). PARAFAC2 allows small time shift departures between the elution peaks of the same component in the different runs and modulations (a problem commonly encountered, especially when using parallel gradients). But when coelution occurs, this approach also fails, especially because of changes in the shapes of peak profiles.

A comparison of the effects in peak shapes and shifts in the trilinear structure of the data has been shown elsewhere [51,52]. These works discuss the limitations of PARAFAC2 and the recommendation of the use of MCR-ALS in the analysis of complex samples where the fulfillment of the trilinear model is not achieved. Thus, when trilinearity is not accomplished, the application of trilinear or higher multilinear models is not recommended, and simpler bilinear models should be applied instead [49,52–54].

In the case of LC × LC data, multiple strategies have been proposed to evaluate the adequacy of the application of the trilinear model and its extension in the analysis of a particular data set. For instance, one possibility is the comparison of the singular value decomposition (SVD) of the concatenated data matrices in their different augmentation modes (column- or row-wisely), obtained when the LC × LC data three-way array is unfolded [55]. Other options are the evaluation of the core consistency diagnostic of the PARAFAC decomposition [56], and the comparison of the data fitting obtained when bilinear and trilinear models are applied in multivariate curve resolution [49,57]. In the latter case, if the data behave following a trilinear model, data fitting results with the two models should be similar, apart from the effects of lower of degrees of freedom and some overfitting in the case of the softer bilinear model. These three strategies, together with the examination of the reliability of the resolved profiles in each case, help to investigate whether the trilinear model is adequate for the analysis of LC × LC-MS and LC × LC-DAD datasets, or if on the contrary, the use of trilinear and higher multilinear models should be avoided [49].

4. Pre-processing, data compression and feature selection

Multiple pre-processing methods can be proposed for the analysis of experimental chromatographic data to enhance their quality, e.g., eliminate noise and baseline contributions, signal smoothing, peak alignment or modeling. Most of them represent the adaptation from those approaches used in one-dimensional liquid chromatography. A brief summary of different pre-processing strategies is included below, with special emphasis to the Regions of Interest, ROI, approach.

The first type of pre-processing tools is focused on data compression. The size of LC × LC-MS data files is considerably large, especially if HRMS is employed, as intensity signals at thousands of m/z values can be acquired in a single scan. Besides, long analysis runs are also common in these multidimensional separations (i.e., more than one hour per chromatogram). This means that compression and filtering are crucial steps to reduce data dimensionality, filter noisy signals and make them more manageable, especially when multiple datasets (samples) are analyzed. Simultaneously, data compression and filtering can be performed in the spectral dimension (columns direction) or the chromatographic dimension (rows direction). The first type of strategy focuses on searching for the most relevant and significant m/z values, above an intensity threshold in the spectral direction, i.e., above noise-related instrumental signals. In contrast, the compression in the chromatographic direction can be especially useful when the focus of the study is on just one specific region of the 2D plot, i.e., on a specific range of retention times in the first or second dimensions.

There are several manners to reduce the size of LC × LC data sets. For instance, a classical approach is binning, which converts raw mass spectra into a matrix representation. The m/z axis is divided into equidistant pieces according to a specific bin size, related to the MS resolution [58]. However, some of its disadvantages include how difficult it is to choose the bin size in each data set, to avoid, for example, joining several peaks into the same bin, or the contrary, peak splitting into different bins. In contrast, wavelet compression reduces the data chromatographic dimension size without significant loss of information, while the effects of noise are minimized. Wavelets decompose the chromatographic data according to their frequency into a new reduced scale, but preserving the spatial location of the chromatographic peaks, their intensities and shapes [59]. It is also useful as a denoising tool, as wavelets are automatically adapted to remove noise-dependent high frequencies of a signal as well as to preserve low-frequency components [60,61]. Another strategy, known as time-windowing, aims to accelerate calculations while reducing storage is to divide the chromatogram into time sections or windows, which can be afterwards analyzed individually [62]. It is also worth to mention the approach proposed by Sinanian et al. consisting of a series of feature extraction processes (i.e., using resolution methods as described below) from low to high spectral resolution. This method allows evaluating the entire dataset in a fast and straightforward manner. However, there is a risk of missing low-concentration compounds (i.e., low explained variance) in the first steps of the analysis due to the large bins employed in the initial binning [63].

The Regions of Interest (ROI) strategy has been proposed to handle the large data sets obtained by HRMS analysis. The application of the ROI procedure allows a compression in the spectral dimension (column direction), filtering and pre-processing MS data in a straightforward manner in a single step, searching for the more relevant data regions and signals [64]. The main advantage of ROI compared to other procedures used for data compression and

matrix transformation, (i.e., binning or bucketing, [58]), is that its application is performed without any loss of the spectral accuracy of the raw measured MS data [64], an aspect which is fundamental for identification purposes and structure elucidation. The ROI algorithm proposed in this work is similar to the one implemented in the centWave of the popular XCMS metabolomic platform [58,65]. More information about practical aspects on how to process ROI in MATLAB or R can be encountered in Refs. [64,66].

In LC × LC, prior to analysis, some additional preliminary data rearrangements are required. In the case of considering a single sample, the data matrices from subsequent modulations should be concatenated vertically, one below the other, creating a column-wise augmented data matrix, where *m*/*z* values are in the columns. When several samples are analyzed at a time, a column-wise super-augmented data matrix is generated. This matrix includes the different modulations concatenated vertically, one below the other, repeatedly for all the samples, which are displayed analogously, sample two under sample one, etc. Fig. 2 depicts these two possible scenarios and how the column-wise augmented data matrices are built.

Therefore, a column-wise ROI super-augmented data matrix is generated with the common *m*/*z* values selected in its columns and filtered according to three main parameters: the MS signal intensity threshold, the mass accuracy (deviation error), and the minimum number of occurrences. The first parameter is an intensity threshold value, which should be above the noise baseline level. It is crucial not to select a threshold value too high, otherwise relevant low intensity signals from low concentration compounds can be lost. On the contrary, if the threshold is set too low, too much background noise will be included into the data analysis. In practice, this threshold value is usually established between 0.1% and 1% of the maximum intensity of the measured signal. It is also possible to combine this threshold value with a multiplication factor (so-called minmax factor), which allows setting a low threshold, but at the same time only consider these ROIs whose intensities are above the product of the threshold and the factor (e.g., intensity threshold x 3). The second parameter is the mass error uncertainty (mass accuracy) associated with the MS instrument, and should be proportional to its spectral resolution. HRMS measures have a very low mass error (high mass accuracy), but the specificities of every employed instrument are needed to be taken into account. This mass accuracy can be defined in terms of absolute mass units (daltons) or alternatively in relative ppm units, and it should be set at a multiple number

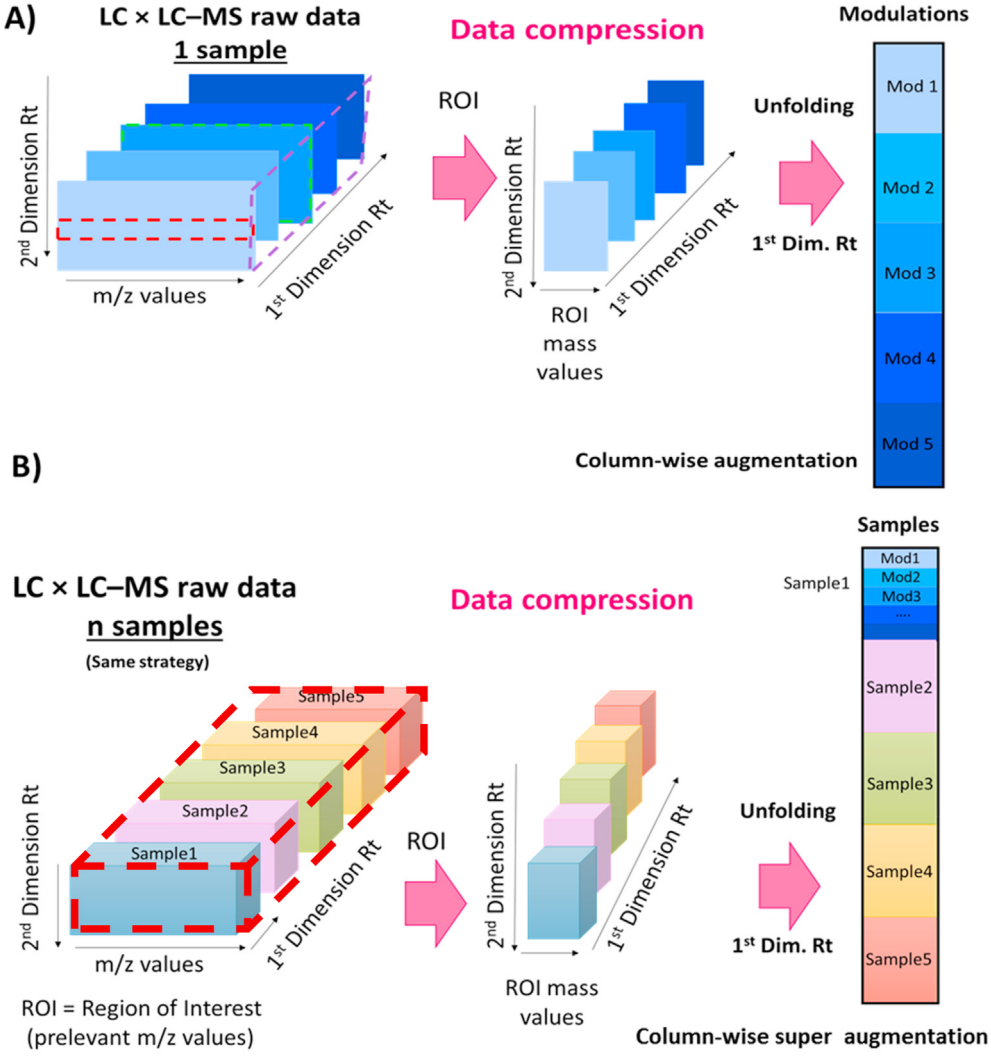

Fig. 2. Arrangement and ROI data compression of LC × LC-MS data (one or multiple samples) in a column-wise augmented ROI data matrix. **A)** All modulations from the LC × LC-MS analysis of the same sample are concatenated vertically, one below the other in a column-wise augmented data matrix. **B)** LC × LC-MS column-wise data matrices from different samples are further concatenated vertically in a new single column-wise super-augmented data matrix.

(i.e., 3 or 5 times) of the expected mass spectral resolution of the considered instrument. Consequently, features with very close m/z values within the selected mass accuracy are joined in one single m/z ROI, or they kept separately, if they differ more than the tolerated mass accuracy. The third parameter in the ROI evaluation is the minimum number of occurrences that a particular ROI signal should occur to be reliable. In chromatography, this value is related to the minimum number of retention times that are necessary to depict an elution peak correctly and to the acquisition frequency of the equipment. In the particular case of LC × LC, this value needs to take into consideration the width of retention times associated with the different peaks that belong to the same compound, which can be present in the subsequent modulations.

Although the ROI method was initially designed for the untargeted analysis of all significative masses of a particular sample [58], the ROI method can be adapted to the study of targeted ions or particular spectral regions, where only certain m/z values corresponding to *a priori* known compounds are selected and analyzed. This option also allows setting specific ranges of retention times of interest for its further analysis. Advantageously, the evaluation of the ROIs obtained provides information about the distribution of m/z values present in the same ROI, about the peak shape of the elution profiles assigned to the ROI in the samples, and about the relative peak areas of the same compound eluted in the subsequent modulations. The individual 2D plot representation of each LC × LC ROI is also useful and informative.

The ROI procedure gives two main outputs, as shown in Fig. 3. On the one hand, a vector including the list of ROI m/z values considered relevant according to the previously selected parameters. On the other hand, a data matrix containing the MS signal intensities at the selected ROIs, for all retention times and all the simultaneously analyzed samples. As detailed in Fig. 3, this data matrix will be arranged in a column-wise super-augmented data matrix where the intensities of the m/z values for each modulation, (individual mass spectra scans), are concatenated one below the other, for all first column retention times and all samples. The intensities at all retention times from both dimensions (rows) for the selected ROI m/z values (columns) are gathered. Then, this ROI augmented data matrix can be analyzed by the MCR-ALS method, as discussed in the following section.

Other pre-processing methods are focused on eliminating background noise, perform instrumental drift correction and signal smoothing. Despite the noise filtering step provided by the wavelets approach, modeling baseline and background contributions is recommended, using one or more additional components in the bilinear factor decomposition type of models, like in MCR methods [46] (see below). Other strategies that have been employed in GC × GC include removing the background contributions from the chromatographic 2D signals estimated using their structural and statistical properties [67], or by curve fitting and linear interpolation techniques [46,68]. Specifically, in LC × LC, the trilinear model decomposition based on the alternating trilinear decomposition (ATLD) method has been proposed to eliminate the background signal drifts without the need of running a blank sample or having previous knowledge about the sample composition [69]. Other methods, like PARAFAC, self-weighted alternating trilinear decomposition (SWALTD), or alternating penalty trilinear decomposition (APTLD) have also been proposed for the same purposes [70,71].

Alignment of chromatographic peaks is also a critical step in many methods analyzing LC × LC data [72,73]. In this case, retention time shifts within a sample are corrected among the subsequent ^{2}D modulations where the same sample constituent elutes. It is also important to remember that, in LC × LC, shifts in the chromatographic peaks between different samples are even more likely

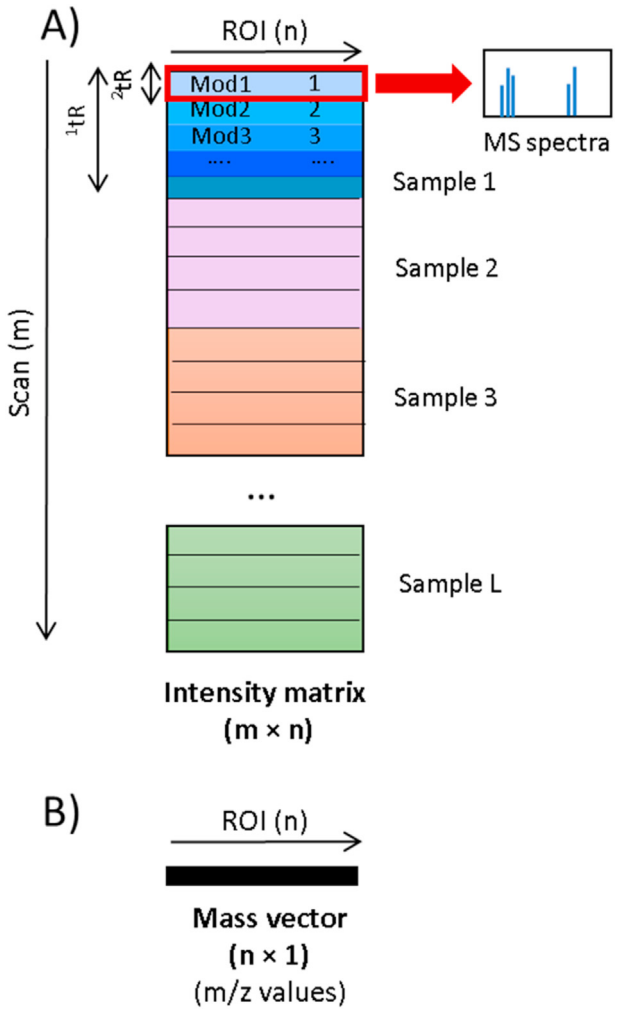

Fig. 3. Description of the two main outputs obtained by the ROI procedure: **A)** Data matrix with the MS signal intensities at the selected m/z ROIs for all modulations and samples analyzed; **B)** Data vector with the list of the m/z values selected by the ROI searching and filtering procedure.

than in LC, due to the long time of the ^{1}D, in contrast to the continuous and short times of analysis in the ^{2}D. Different methods have been proposed to perform this correction in multidimensional separations, such as rank alignment [74], correlation-optimized shifting (COShift) [75], 2D correlation optimized warping (COW) [76], dynamic time warping (DTW) [77], or peak alignment fast fourier transform (PAFFT) method [78]. However, in general, other related phenomena, e.g., peak swapping or absent peaks, are not fixed with peak alignment strategies. It is worth to emphasize here that in the case of the MCR based methods, chromatographic peak alignment correction is not needed since the only requirement for the data analysis is their alignment in the spectral direction, which is easily achieved, in contrast to the chromatographic peak alignment [46,79,80].

Finally, peak detection methods provide information about the different sample constituents resolved during the chromatographic separation in both dimensions. Usually, these methods include peak modeling and signal smoothing tools as a part of the detection process. Recently developed methods for peak detection in multidimensional chromatography include the Normal-Exponential Bernoulli (NEB) algorithm and mixture probability models [81], both procedures included in the *msPeak* R package [81]. In this case, baseline correction, background subtraction,

recognition of potential peak regions, peak picking, peak areas integration, and final peak detection by mass spectral similarity are part of the same software package. Similar adaptations of this method include Normal-Gamma (NG) [82], and Normal-Gamma-Bernouilli (NGB) [83]. Other peak detection alternatives have been also proposed [84]. In this case, initially, the method only considers the first chromatographic dimension, but then in a second step, peaks from the same compound are jointed in a two-dimensional peak, using a merging algorithm based on Bayesian inference [84]. More detailed descriptions about pre-processing methods for LC × LC data can be found elsewhere [2,85]. Apart from the pre-processing and feature selection methods briefly summarized above, multivariate curve resolution methods can be applied in order to recover the elution and spectra profiles of the different constituents present in the analyzed samples. In the following section, the MCR-ALS method will be discussed in more detail.

5. Multivariate Curve Resolution Alternating Least Squares (MCR-ALS)

Multivariate Curve Resolution methods resolve the constituents of unknown mixtures component by component using a bilinear model. It is important to notice that, despite its similarities, resolution and deconvolution are not synonyms. Whereas the deconvolution term is commonly used in the context of univariate signals (like in total ion or in single wavelength UV chromatograms), the term resolution is more appropriate in the context of multivariate signals, like in full scan LC-MS or multiwavelength LC-UV. In the first case, the deconvolution techniques used are usually model-based curve-fitting approaches (e.g., gaussian elution peak shapes). However, in the second case, the multivariate resolution techniques are model free, i.e., they do not need the postulation of any model to describe the shape of the profiles. They are derived directly from the bilinear multivariate data model. Thus, unlike deconvolution algorithms used with single channel univariate detection methods, Multivariate Curve Resolution methods deal with the whole response of multivariate detection methods (e.g., at multiple wavelengths, *m/z*, …) and do not require the postulation of a signal shape type of model.

When MCR methods are applied to spectroscopically (multichannel, multivariate) hyphenated chromatography, they allow the direct mathematical resolution of the chromatographic overlapped peaks of the coeluted sample constituents as well as the resolution of their pure multichannel responses (spectra) of the analyzed constituents of the sample. In the case that MS is employed as a multichannel detector, the resolved mass spectra contain rich qualitative information useful for identifying of the different constituents of the analyzed mixtures. At the same time, from the resolved elution profiles, it is possible to obtain quantitative information, from their peak areas or their peak heights.

Mathematically, MCR methods decompose the experimental datasets according to a bilinear additive factor decomposition model which naturally corresponds to the generalization of Lambert-Beer's law of molecular absorption. This bilinear model can be described using linear algebra data matrix notation as:

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^{\mathbf{T}} + \mathbf{E} \tag{1}$$

where **D** is the data matrix with the experimental measures, which in the case of a chromatographic run monitored spectroscopically, have the spectra at the different retention times in the rows ($i = 1, \ldots, I$), and the chromatograms at several spectral channels/variables (wavelengths, *m/z*, …, etc.) in the columns ($j = 1, \ldots, J$). The MCR bilinear decomposition provides the concentration/elution profiles of the different components ($n = 1, \ldots, N$) of the analyzed sample/mixture in the columns of the matrix **C**, and the spectra of these components in the rows of the matrix $\mathbf{S}^{\mathbf{T}}$. Matrix **E** in this Equation (1) represents the obtained residuals and accounts for the variance not explained by the bilinear model. When multiple hyphenated chromatographic runs are simultaneously analyzed, this bilinear model can be easily extended according to Equation (2):

$$\mathbf{D_{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \ldots \\ \mathbf{D_L} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \ldots \\ \mathbf{C_L} \end{bmatrix} \mathbf{S}^{\mathbf{T}} + \begin{bmatrix} \mathbf{E}_1 \\ \ldots \\ \mathbf{E_L} \end{bmatrix} = \mathbf{C_{aug}}\mathbf{S}^{\mathbf{T}} + \mathbf{E_{aug}} \tag{2}$$

where now the multiple chromatographic runs of different samples ($l = 1, \ldots, L$) analyzed with the same spectroscopic detector (UV, MS, etc.), each one of them giving an individual data matrix ($\mathbf{D_1}, \ldots \mathbf{D_L}$), are vertically concatenated in the column-wise augmented data matrix $\mathbf{D_{aug}}$, with $I \times L$ rows and (J) columns. $\mathbf{C_{aug}}$ has the concentration/elution profiles of the (N) components present in the (L) analyzed samples and $\mathbf{S}^{\mathbf{T}}$ the spectra of these components. Observe that, in this case, the bilinear model implies that the concentration/elution profiles of the same components in the simultaneously analyzed runs/samples. i.e., in $\mathbf{C_1}, \ldots \mathbf{C_L}$ can be different, whereas their spectra are the same in $\mathbf{S}^{\mathbf{T}}$. The non-explained variances are now in the $\mathbf{E_{aug}}$ residual matrix.

A similar extension of the MCR bilinear model can be performed in the case of 2D chromatography. The data from one run/sample analyzed by 2D can also be arranged in an augmented data matrix $\mathbf{D_{aug}}$, with $I \times K$ rows and J columns, where (I) refers to the number of retention times in the second column, and (K) refers to the number of modulations, as shown in Equation (3).

$$\mathbf{D_{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \ldots \\ \mathbf{D_K} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \ldots \\ \mathbf{C_K} \end{bmatrix} \mathbf{S}^{\mathbf{T}} + \begin{bmatrix} \mathbf{E}_1 \\ \ldots \\ \mathbf{E_K} \end{bmatrix} = \mathbf{C_{aug}}\mathbf{S}^{\mathbf{T}} + \mathbf{E_{aug}} \tag{3}$$

When this strategy is further extended to the simultaneous analysis of several LC × LC chromatographic runs (different second column modulations from multiple samples), the bilinear model applied to the column-wise super-augmented data matrices can be written as:

$$\mathbf{D_{saug}} = \begin{bmatrix} \mathbf{D_{aug}}1 \\ \ldots \\ \mathbf{D_{augL}} \end{bmatrix} = \begin{bmatrix} \mathbf{C_{aug}}1 \\ \ldots \\ \mathbf{C_{augL}} \end{bmatrix} = \mathbf{C_{saug}}\mathbf{S}^{\mathbf{T}} + \mathbf{E_{saug}} \tag{4}$$

As also detailed in Fig. 4, $\mathbf{D_{saug}}$ (size of $I \times K \times L$ rows, J columns) is the LC × LC column-wise super-augmented data matrix, where I corresponds to the number of retention times in the (K) modulations of the second chromatographic dimension, for the (L) samples, and (J) corresponds to the number of spectral channels (wavelengths in DAD and *m/z* values in MS). $\mathbf{C_{saug}}$ ($I \times K \times L$, N) is the matrix containing the resolved 2D chromatographic profiles in the (L) samples, in the (K) modulations, for the (N) components, and $\mathbf{S}^{\mathbf{T}}$ (N, J) are their corresponding spectra. Finally, $\mathbf{E_{saug}}$ ($I \times K \times L$, J) contains the residuals not explained by the bilinear model using this number of components.

The Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is a popular version of MCR method which uses an iterative alternating linear least squares (ALS) optimization to resolve the component profiles. ALS is a matrix factorization

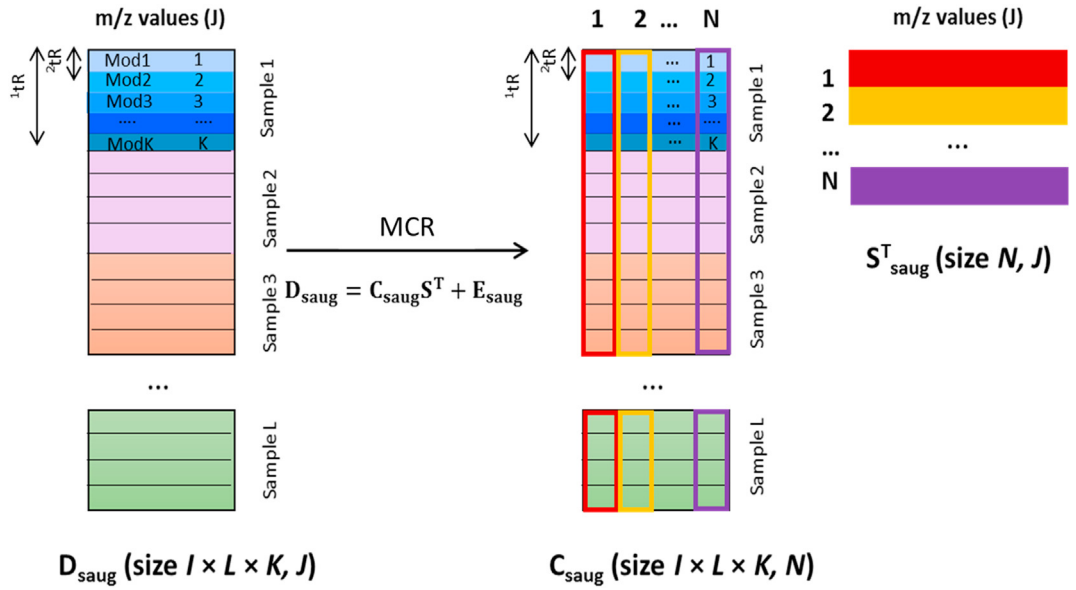

Fig. 4. MCR resolution of a single LC × LC data set. The column-wise augmented data matrix, $\mathbf{D}_{saug}$, is decomposed into two factor matrices: $\mathbf{C}_{saug}$ and $\mathbf{S}^T$, with the elution profiles and pure mass spectra of the resolved components.

algorithm. It includes regression steps to find solutions to local linear models where alternatively **C** or $\mathbf{S}^T$ are fixed (Equation (1)). The MCR-ALS method has been described in detail elsewhere [9,86], and it can be applied independently of the detector employed (e.g., MS, DAD, Fluorescence). The first step of the MCR-ALS procedure is selecting of the number of components, *N*, which is initially roughly estimated from the number of singular values [87] larger than experimental noise. In addition, other criteria that can be used in this selection are the evaluation of the amount of variance explained by each component, and the visual assessment of the reliability of the resolved chromatographic and spectral profiles associated with these components. The goal is that this number of components explains enough data variance, while avoiding the possibility of data overfitting with too many components explaining just experimental noise. In the case of MS, this step is more difficult, and may require a more detailed analysis and selection, specially to detect minor components contributing very little to the total measured signal. However, also in this case, the high selectivity of the MS detector facilitates this selection, specifically for the components contributing significantly (i.e., more than the noise) to the measured signal.

Next step is to get initial estimates (concentration profiles or spectra) for the selected number of components. Pure variable detection methods are usually a good choice in the LC × LC case since they can provide an initial estimation of the elution or spectral profiles for the selected number of components in a single step. These methods find the most dissimilar chromatographic elutions or mass spectra directly from the measured chromatographic raw data. When the overlapping in the chromatographic direction is lower than in the spectral direction, it is recommended to select the purest elution times and work with the initial estimates of the spectra measured at these selected times [9]. This is the case, for instance, of DAD detection. However, in the case of MS detection, the situation is different because the spectral direction is more selective. Therefore, the search for the elution profiles at the purest *m*/*z* channels is suggested.

With these initial estimations, the ALS iterative optimization starts. The ALS iterative process ends when the convergence criterion is achieved, i.e., when the relative change of the standard deviation of the residuals between consecutive iterations is lower than a preselected threshold value [9], or the maximum number of iterations selected is reached.

Since solving the bilinear MCR model does not assure unique solutions, different MCR solutions with same data fit but with distinct concentration, and spectral profiles are feasible. This number of feasible solutions, generally known as ambiguity, is one of the main concerns when working with MCR-based methods. There are different types of ambiguities but the more relevant are rotational ambiguities (due to the existence of rotation matrices that can affect the shape and intensity of the resolved profiles without changing data fitting and fulfilling the applied constraints). Application of inherent data properties can diminish the impact of these rotational ambiguities in the obtained solutions. For instance, sparse data (like in MS-generated datasets) show a minor effect of rotational ambiguities in the resolved elution profiles due to the high selectivity of MS signals. In addition, different data augmentation strategies have been proposed to minimize (and eventually eliminate) the effect of these rotational ambiguities. Finally, some methods have been recommended to quantify this rotational ambiguities impact [88,89].

During the ALS optimization, chemical, physical or mathematical information in the form of constraints are applied to reduce the ambiguity in the MCR bilinear solutions. Among the constraints imposed in MCR-ALS, the most commonly used in the case of analyzing LC × LC data is the non-negativity constraint, which is applied to both elution and spectral profiles in the respective factor matrices, $\mathbf{C}_{saug}$ and $\mathbf{S}^T$ (see Equation (4)) [9]. Due to the selectivity of the chromatographic separation and specially of the MS detection, the use of non-negativity constraints together with the simultaneous analysis of multiple chromatographic runs and samples with common constituents usually provide feasible solutions with very little or no ambiguity associated, giving unimodal elution profiles. Thus, other natural constraints such as unimodality are not needed when MS is employed, in contrast, to its common use in LC × LC-UV [90–93]. Alternatives to classical ALS optimization have also been considered. For instance, Cook et al. proposed an additional sparse regression step using an elastic net (MCR-ENALS) [94] to increase the algorithm's performance when analyzing MS data. This

approach takes advantage of the inherent sparseness of MS data to detect low intensity signals.

One major advantage of the application of MCR-ALS to 2D chromatographic systems is that it allows the resolution of the elution profiles of the constituents in the analyzed mixtures without the application of any peak alignment nor peak modeling pre-processing step. This is caused by the intrinsic flexibility provided by the bilinear modeling (see Equation (4) and Fig. 4 above) of the chromatographic peaks in both chromatographic dimensions (n $\mathbf{C}_{\mathbf{saug}}$ matrix of Equation (4)). Matching of elution times among chromatographic peaks of the same component in the different ^{1}D and ^{2}D columns is not required. Data alignment is only needed in the spectral (column) direction, in the columns (wavelengths or *m*/*z* values) of $\mathbf{S}^{\mathbf{T}}$. Hence, chromatographic time shifts within and between samples do not affect the MCR-ALS final results. Consequently, the spectral and chromatographic elutions of different components in both dimensions can be resolved independently.

As stated above, MCR-ALS resolved elution and spectra profiles can be used to get qualitative and quantitative information, respectively. From the resolved spectra (either MS, UV, fluorescence, etc), it is possible to retrieve qualitative information about the sample constituents and perform their identification. On the other side, from the peak areas or heights of the elution profiles, relative quantitative information of the different sample components can be obtained. From this relative quantitative information, pattern recognition and classification of the analyzed samples (differences and similarities) are possible, as it is shown in Fig. 5.

In the case of LC × LC analysis, from the resolved second-dimension elution profiles, it is also possible to perform the relative quantitation of the chemical constituents of the analyzed samples. Since the first-dimension elution profile (^{1}D peak) is split into the consecutive injections that are further separated into the ^{2}D column, it is possible to obtain the total peak contribution for a specific resolved component by summing all the peak areas of the same component in the second dimension. From the comparison of the peak areas of the components in the different samples, relative quantitative information can be obtained. Various integration strategies applied to LC × LC are described elsewhere [48,92]. Manual integration is recommended in the cases where signal to noise ratios are small, due to low concentrations of analytes in the samples, as residual background can have a significant effect on the areas when peak integrations are performed by an automatic summation.

It has also been proved that MCR-ALS can build good calibration models for LC analysis to quantify the concentrations of the different compounds present in the sample, if absolute reference concentrations are available or multiple standards are used [95]. When validation samples are employed, the quantitative results in the samples not used during the calibration process can also be calculated as well as their relative errors [46]. In the case of LC × LC, it is possible to calibrate and quantify employing the resolved ^{1}D chromatograms. Compared to other conventional calibration procedures, MCR-ALS calibrations can efficiently remove the background and interference contributions from the analyte signals, while resolving baseline contributions and compound coelutions separately. This has been demonstrated in previous works, like in GC × GC analysis of polyaromatic hydrocarbons in oil samples [96].

Recently, a new restriction called area correlation constraint has been proposed for second order quantitative calibration with MCR-ALS [97,98], to estimate the absolute concentrations of the constituents of a set of samples simultaneously analyzed by a chromatographic method in the presence of unknown interferences. This constraint is applied at each iteration step of the ALS process by regressing the peak areas or heights of the concentration profiles of the analytes in every analyzed sample against their known concentrations. In practical situations, the known and unknown samples are simultaneous analyzed, and this correlation constraint is applied to build the calibration curve with the known "calibration" samples, whereas the rest are used for validation and prediction of the unknown samples. This approach can also be employed using different calibration strategies, i.e., using external or internal standards and the standard addition method [41,95].

Besides, from the peak areas of the resolved elution profiles of the different sample constituents, other post-processing methods can be applied, from which it is possible to evaluate differences among samples and identification of potential markers, as it will be explained below in the following section.

6. Post-processing data analysis methods

Apart from quantitation, the table (matrix) of the peak areas from the resolved elution profiles of the different constituents of the analyzed samples (for instance obtained by MCR-ALS, or as by other feature detection methods), can be post-processed using several types of multivariate data analysis methods to gain additional information about the analyzed systems. As shown in Fig. 5, five main types of analysis can be performed with the output results obtained after application of MCR methods: exploratory, clustering, statistical, classification and machine learning-based analysis. A brief comment is given here on some examples of application of these multivariate analysis methods to extract additional information from the results of the analysis of LC × LC datasets, especially when different types of samples are analyzed simultaneously.

Among exploratory data analysis methods, the most common one is Principal Component Analysis (PCA) [99]. PCA is a non-supervised data analysis method that provides an overview of the number and nature data variance sources. In PCA, a reduced dimensional space (whose variables are known as principal components) explains the most relevant information about the major sources of data variance. On one side, sample scores plots allow the unsupervised visualization of the analyzed samples in the principal components vector space, where the different groups or trends in these samples can be distinguished in plots of reduced dimensions. On the other side, from the loadings plot, it is possible to detect the most relevant variables in the definition of each principal component, i.e., giving details of the nature of the major sources of data variance.

One example of the application of PCA in 2D-LC is, for instance, the analysis of the triacylglycerides (TAGs) composition of several oils [100,101]. The PCA scores plot enabled unsupervised differentiation among various types of oil samples. Several kinds of oils with dissimilar compositions were distinguished, whereas samples with similar TAG compositions were clustered together. In addition, from the loadings plot, it was possible to distinguish and identify the most significant TAG present in each sample type.

The PCA study of the selectivity of different 1D-LC systems allowed the classification and comparison of several types of LC columns [102,103]. Thus, another use of PCA, beyond the exploratory analysis of the samples, has been the determination of orthogonal column combinations for LC × LC, a key aspect to be taken into account in multidimensional chromatography. For instance, Græsbøll et al. developed a method to help in the selection of LC columns for its application in LC × LC. This work was based on the hydrophobic subtraction model and different approaches were tested to assess the orthogonality of the columns [104].

Clustering methods, either hierarchical or non-hierarchical, group objects and samples according to their similarities and

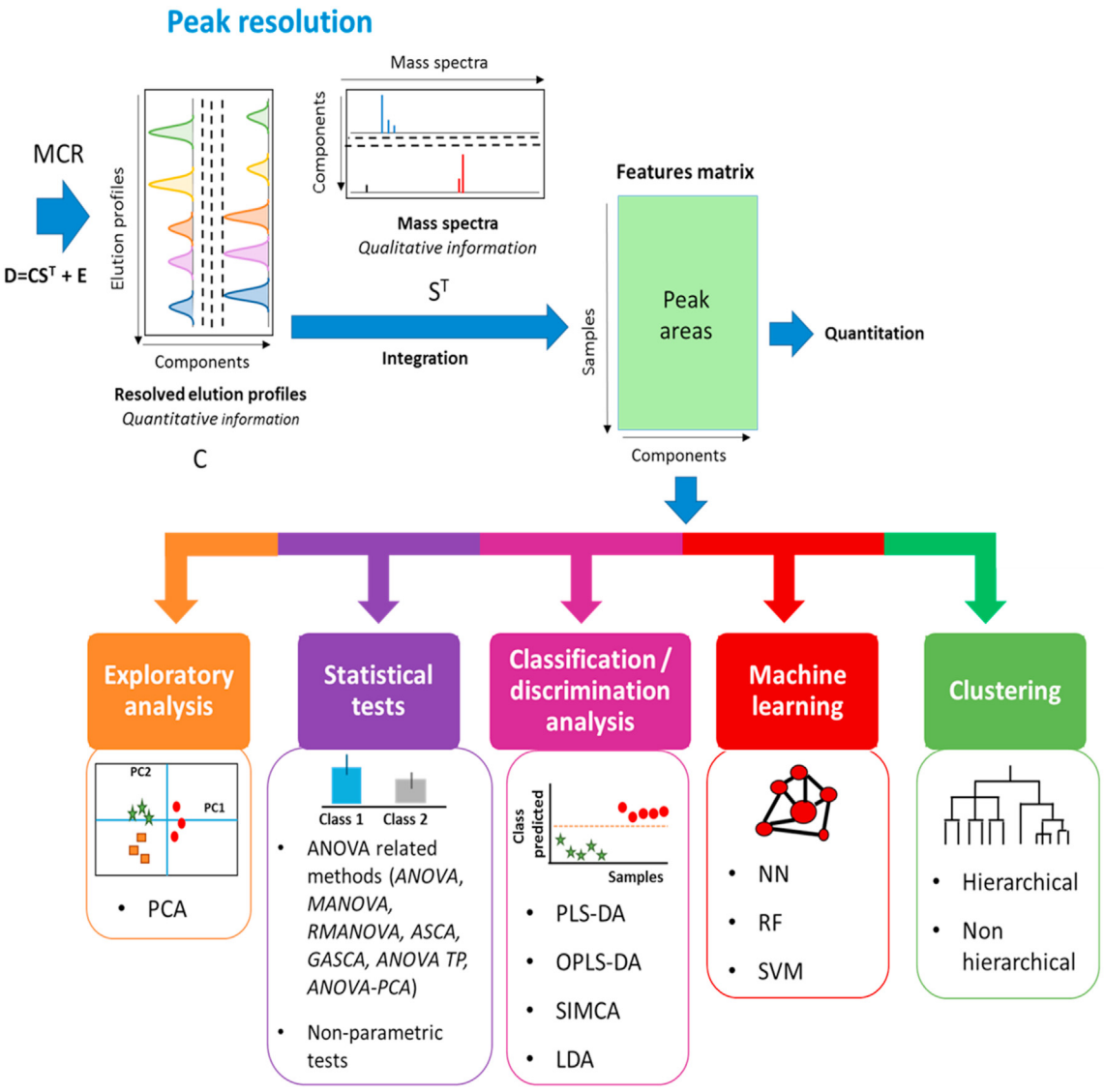

Fig. 5. Post-processing of the table of peak areas of the MCR resolved compounds. Five different types of analysis can be performed: exploratory, classification, statistical, machine learning-based and clustering.

differences within and between groups [105]. These methods are widely used in omics sciences, where some examples of applications in GC × GC [106], and in 2D-LC [101,107], have been already reported in combination with exploratory and classification methods. Multivariate peak area tables can be further analyzed using multiple statistical approaches. In the case of the simultaneous analysis of multiple variables per sample (i.e., when the tables of peak areas for several analytes and samples), different extensions of the ANOVA method have been proposed like the multivariate analysis of variance (MANOVA) [108], the ANOVA simultaneous component analysis (ASCA) [109], the group-wise ANOVA simultaneous component analysis (GASCA) [110], the ANOVA-PCA [111], the regularized MANOVA [112], or the ANOVA by target projection (ANOVA TP) [113]. In addition, linear mixed models have been generalized and adapted as an alternative to ANOVA based methods covering all possible experimental designs involving fixed factors [114,115]. For instance, in ASCA, which is based in the combination of ANOVA and Simultaneous Component Analysis (SCA), a method similar to PCA, individual factors as well as two or three-ways interactions between different factors can be analyzed, to determine whether they are significant or not [109]. ASCA has been applied to peak area tables obtained after MCR-ALS to statistically assess the significance of the different factors used in the design of the LC × LC-HRMS experiments [62].

Supervised classification analysis includes a group of methods that allow the class modeling and discrimination of samples in previously established groups before the analysis. The Soft Independent Modeling of Class Analogies (SIMCA) [116] method allows the assignment of a particular sample/observation to a well-defined class. This classification is performed through the size of the residuals of the disjoint PCA (a PCA model is generated for every class present in the dataset). In contrast, Partial Least Squares-Discriminant Analysis (PLS-DA) [117], has been proven to provide equal or even better performance in discrimination tasks (two-class type of problems). In PLS-DA, class distinction is established *a priori*, according to previous information known from the experimental design (DOE) about the different samples, which is then introduced in the model building. From the class assignment, contingency tables can be obtained where the quality of the model can be assessed, taking into account figures of merit

such as selectivity and sensitivity. In addition, PLS-DA results also allow identifying the features responsible for the differences between classes, for instance, through examining the changes in the size of the variables between classes, which in the case of MS data, can be easily associated with m/z values. The Variables Important in Projection (VIPs) method give the highest score values for the most significant variables (m/z values) explaining the differences between the classes. PLS-DA has also been employed to look for biomarkers through the VIPs, comparing healthy samples (or control samples) with samples from patients with a certain disease [118]. PLS-DA can also be useful in the analysis of 2D-LC data sets [62,101,107,119]. Other variants of PLS-DA methods are widely employed, such as the OPLS-DA (orthogonal PLS-DA) method [120], with classification models are often found easier to interpret.

Besides SIMCA and PLS-DA, other machine learning based methods can also be employed to explore and classify LC × LC data sets, although they have barely been explored until now. In this context, different variants of Neural Networks (NN), Random forests (RF) and Support vector machines (SVM) should be highlighted [121,122]. Examples of recent applications of machine learning methods in LC-MS can be found, especially in metabolomics [123–125], but also in method optimization, retention time predictions or in silico quantifications [126–128].

7. Examples of application of the MCR-ALS and ROIMCR methods to LC × LC data

In this section, some examples of the application of the MCR-ALS and the ROIMCR methods in the analysis of LC × LC data, taken from previous studies are succinctly described.

Several applications of MCR-ALS to LC × LC- UV–Vis (i.e., using DAD) data, have been described in detail in the works of Rutan et al., focused on two data analysis challenges, the evaluation of the structure of the LC × LC data, and the quantitative aspects of this type of analysis. For instance, Bailey et al. analyzed urine samples with the so-called IKFSA-ALS-ssel procedure [48]. Initial guesses for MCR-ALS analysis were obtained using the iterative key set factor analysis (IKSFA) method, in a similar way as in the application of other purest variable detection methods [105,129]. Non-negativity and selectivity constraints were also applied during the optimization. Relative concentrations of the sample constituents were estimated from the sum of the second-dimension peak areas that correspond to the same compound, after their resolution with the IKFSA/MCR-ALS procedure and manual baseline correction. Better integration results (lower % RSD) were obtained when this manual baseline correction of the MCR resolved elution peaks was applied, compared to the direct raw peak data integration using commercial software.

In another study about the LC × LC analysis of maize seedling samples, the PARAFAC-ALS and the IKSFA/MCR-ALS methods were compared. The authors concluded that retention time shifts were responsible for the worse results of PARAFAC-ALS (lower precision), whereas the IKSFA/MCR-ALS method gave better results and was recommended. Also, the IKFSA/MCR-ALS-ssel method coupled with a manual baseline subtraction quantification method was used in the analysis of wine samples using LC × LC-DAD [130]. In this work, different chemometric strategies (such as the Fisher Ratio or the Similarity Index methods) were evaluated as rapid screening methods in the LC × LC analysis of complex samples. In all these strategies, the focus was especially on determining of those sample constituents whose chromatographic peak areas (concentrations) differed among the investigated samples, rather than on the determination of all the sample constituents detected in the chromatograms. Thus, the time of the analysis was drastically reduced.

MCR-ALS has also been proven to be a useful tool combined with 2D assisted liquid chromatography (2DALC), where two DAD detectors were placed after each column. This set up allowed obtaining simultaneously ^{1}D and 2D chromatograms and improved the quantitation of targeted analytes [90]. 2DALC, in combination with the MCR-ALS method, presented some advantages over the traditional LC × LC-DAD method, since it could perform a proper background correction due to dilution in LC × LC, as well as avoiding, at least partially, the loss of resolution caused by the undersampling of the ^{1}D elution profile. This strategy takes advantage of the pure spectra of the components resolved in the ^{2}D by MCR-ALS, which can be used then for the MCR-ALS analysis of the ^{1}D chromatographic data. This method allowed the calibration and quantification of the different sample constituents resolved in the ^{1}D chromatograms.

Other strategies combining MCR-ALS and LC × LC data (including 2DALC, and its adapted algorithm, called combined-2DALC) have been proposed for the quantitative analysis of furanocoumarins in apiaceous vegetable samples [92], which were separated using an RP × RP-DAD chromatographic method (both column dimensions employed reverse phase as the retention mechanism, a pentafluorophenyl phase (PFP) and a C18 phase as ^{1}D and ^{2}D columns, respectively). In this example, initial estimates of the spectral profiles were obtained employing the iterative orthogonal projection approach (IOPA), in which the number of spectra extracted was preselected by the user from his previous knowledge of the sample complexity. Non-negativity and local rank constraints were applied in the chromatographic and spectral directions. Local rank constraints were applied in this case to force certain spectral regions of the resolved components to zero-intensity (i.e., intensities associated with wavelengths out of the range of absorption of the target analytes). Two integration strategies were also compared, one based on a manual selection of peak boundaries (including background subtraction) and another based on an automatic summation. The first strategy was especially recommended for samples at low concentrations of the analytes showing low signal-to-background ratios.

Other recent applications of MCR-ALS to LC × LC-MS have also been reported. For instance, mixtures of highly similar TAG structural isomers present in vegetable oil samples were separated with an IEX × RP-MS (ion exchange as ^{1}D and reverse phase as ^{2}D) and completely resolved by MCR-ALS, as shown in Fig. 6 [54]. Navarro-Reig et al. performed the comparison between MCR-ALS bilinear, MCR-ALS trilinear, MCR-ALS trilinear allowing time shifting, PARAFAC and PARAFAC2 methods [54]. It was demonstrated that the LC × LC data did not fulfill the trilinearity model requirements due to retention time shifts and to changes in the shapes of the elution peak profiles. Therefore, the MCR-ALS bilinear method was the most suitable choice for the analysis of this type of data. Non-negativity constraints were applied to both elution and spectra profiles, as well as spectral normalization (either of peak heights or peak areas). Regions where triacylglycerides (TAGs) isomers coeluted in both chromatographic dimensions were specifically studied to check the limits of the resolution potential of the proposed method. In this work, MCR-ALS was also confirmed to resolve and identify separately the different positional and chain TAG isomers despite their very strong coelution (with embedded peaks) in both dimensions.

Next, some other examples applying the combination of the ROI method with the MCR-ALS method in the recently proposed ROIMCR method to LC × LC data are briefly described. Examples of the application of different post-processing multivariate analysis of the peak areas or heights of the MCR-ALS resolved elution profiles of the analyzed sample constituents are given.

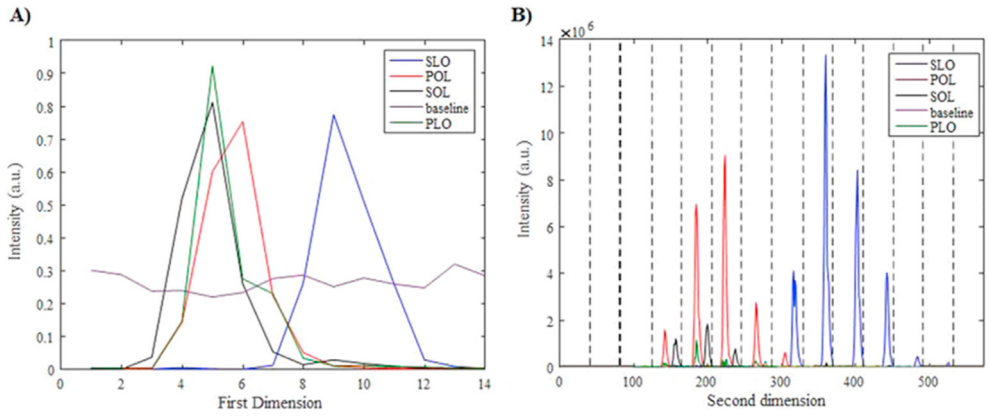

Fig. 6. [1]D and [2]D column elution profiles of a chromatographic region with highly overlapped triacylglycerol structural isomers (TAGs) resolved by MCR-ALS "Reprinted from Talanta, 160, M. Navarro-Reig et al., Chemometric analysis of comprehensive LC × LC-MS data: Resolution of triacylglycerol structural isomers in corn oil, 624–635, Copyright (2016), with permission from Elsevier."

In a recent publication, the ROIMCR method was applied to LC × LC-MS, LC × LC-UV and LC × LC-UV-MS data from the analysis of a mixture of 31 pharmaceutical compounds [49] using a combined RP × HILIC (an RP, reverse phase, column as ^{1}D, and a HILIC, hydrophilic interaction column, as ^{2}D). In this work, the selected ROI parameters employed for the analysis of the LC × LC-MS data were a signal threshold value of 0.1% of the maximum signal intensity, mass accuracy of 0.5 Da (a low resolution MS detector was employed), and a minimum number of *m*/*z* occurrences of 25. Non-negativity constraints were applied in the chromatographic direction (elution profiles of the two columns) and, also, in the spectral direction (mass spectra). Three different approaches assessed the trilinear behavior of the data. The first includes the comparison of the singular values of the data augmented matrices obtained by concatenation of consecutive ^{2}D modulations in their different directions (vertically, horizontally and slice by slice). The second implied the calculation of the core consistency in the PARAFAC decomposition of the data three-way array. The third compared the data fitting results obtained assuming either bilinear or trilinear decomposition models in MCR-ALS. Significant differences in the SVD decomposition of the augmented data matrices in their different directions or modes, poor trilinear core consistency values, and large data fitting deviations between bilinear and trilinear models were found for both types of data (MS or DAD). Consequently, the MCR-ALS bilinear model approach was preferred and recommended for the analysis of this LC × LC data set. In addition, the combined analysis of the data from the two detectors (DAD and MS) building up a 'fused multidetector' multiset data structure, allowed resolving more components than from individual detector analysis.

ROIMCR has also been used in other LC × LC-MS metabolomics studies, such as in the untargeted analysis of the metabolites present in a particular sample. For instance, the changes in rice metabolism caused by environmental factors such as watering and harvesting time, were studied by LC × LC-MS [62,119] using a combination of HILIC × RP-HRMS separation methods in the analysis of the polar metabolites of the rice samples, using the hydrophilic interaction column as ^{1}D and the reverse phase column as ^{2}D. The strategy proposed in this work included two compression steps: a first spectral compression using the ROI strategy, and a second compression in the time direction, using wavelets. An additional windowing approach was used to divide the chromatograms into three different chromatographic windows. In this way, the size of the analyzed data sets and the global time of analysis were feasible and significantly reduced. Possible instrumental drifts were corrected using an internal normalization based on dividing all the metabolite peak areas by the peak area of a standard compound added to each sample. ASCA investigated the effects of the watering and the harvesting time factors. Peak areas of all MCR-ALS resolved metabolites were arranged in a data table/matrix, and the effects of both factors were evaluated statistically, as well as the interaction between them. In addition, the application of PLS-DA allowed discerning what metabolites were able to separate the effects of watered and non-watered samples through the calculation of VIPs as shown in Fig. 7. MCR-ALS was further applied in a post-processing step to study the evolution of the metabolic concentration profiles (peak areas) over time, and to investigate in this way the effects of the harvesting time factor, for both water and non-watered samples. In this extensive study, the proposed ROIMCR chemometrics strategy allowed the simultaneous untargeted direct resolution of 150 metabolites from 15 different families, with 134 of them identified by their HRMS spectra.

In another recent study, RP × HILIC-MS and MS/MS untargeted throughout analysis of the lipids present in rice samples under arsenic (As) exposure were presented. In this case, only the spectral compression was necessary to reduce the LC × LC-MS data size, which was also analyzed using the ROIMCR strategy. ROI parameters employed were similar to those mentioned before for the analysis of pharmaceuticals explained above, since the same settings of the instrumental equipment were used. Arsenic effects were investigated on two plant tissues: aerial parts and roots. MCR-ALS was able to resolve the elution profiles and mass spectra of the lipids present in the rice samples and allowed their identification. Complementary, ASCA analysis revealed that As exposure affected the rice lipidome significantly. PCA analysis of the changes on the peak areas of the MCR-ALS resolved elution profiles of the lipids detected in positive ionization mode (see Fig. 8) showed that the rice samples were well separated at the two As concentrations levels in aerial rice samples. In contrast, the separation from control and the lowest concentration level was not obvious in root rice samples. In the lipidomic analysis using negative ionization mode, aerial and root rice samples exposed at the two As concentrations were not discriminated from control rice samples. PLS-DA was applied to identify the potential lipid markers of As exposure effects at two concentration levels. Identification of these markers was a challenge, especially due to the low sensitivity of the detector and to the lack of theoretical MS/MS lipid spectra databases. Overall, this study confirmed that LC × LC combined with the ROIMCR data analysis is an excellent tool in omics studies. Other applications can

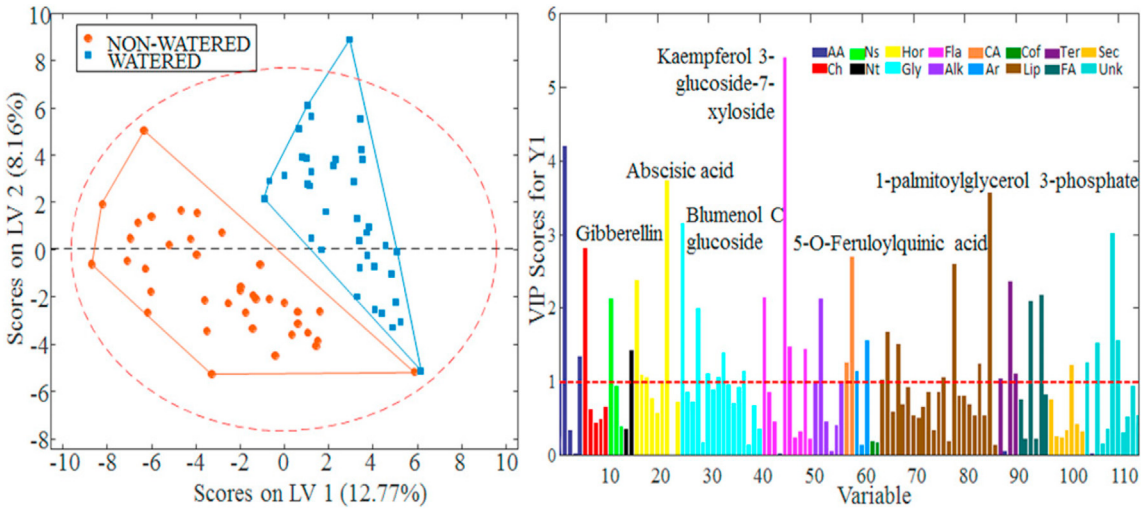

Fig. 7. PLS-DA scores and loadings plots for watered and non-watered rice samples analyzed by LC × LC-HRMS. "Reprinted from Analytical Chemistry, 89, M. Navarro-Reig et al., Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution, 7675−7683, Copyright (2017), with permission from American Chemical Sprociety."

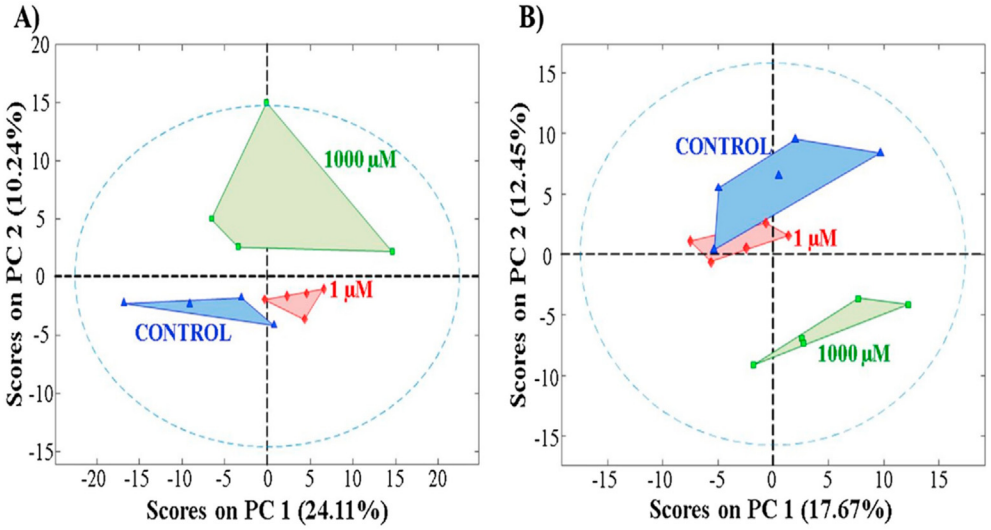

Fig. 8. Scores plot obtained for the lipids analyzed in aerial parts and roots, respectively, of rice samples treated with As. "Reprinted with permission from Journal of Chromatography A, 1568, M. Navarro-Reig et al., An untargeted lipidomic strategy combining comprehensive two-dimensional liquid chromatography and chemometric analysis, 80−90, Copyright (2018) Elsevier".

be envisioned, like in untargeted analysis of environmental monitoring samples, where huge data sets are also generated.

8. Conclusions

Although one of the main drawbacks of LC × LC coupled to multichannel detection (mainly MS and DAD is the difficulty of performing a thorough analysis of the large data sets generated, recent chemometric advances provide a good option to untangle the complexity of this type of data and facilitate the extraction of the sought information in different types of studies such as in metabonomics. Since data from LC and LC × LC analysis usually present significative deviations from the ideal trilinear behavior, bilinear model-based methods are preferred in general for their analysis.

Among the recently proposed chemometric strategies, the ROIMCR method is a powerful tool to analyze LC × LC data, thanks to the combination of compression and resolution steps in both targeted and untargeted analysis. The possibility of filtering, compressing and arranging large datasets without loss of spectral accuracy is the main achievement of the ROI approach. With the application of this strategy, resolution of the LC × LC elution profiles of all the sample constituents of the analyzed samples in the two dimensions as well as their mass spectra can be achieved, providing simultaneously quantitative and qualitative (identification) information. The application of MCR-ALS to the ROI data adds further advantages such as the flexibility in modeling the complex structure of the data (bilinear or trilinear) which allows the proper analysis of the different types of datasets encountered in practice. MCR-ALS does not require the modeling or alignment of the chromatographic peak signals resolved in different multiple chromatographic runs in their simultaneous analysis. MCR-ALS can also be used as a preliminary pre-processing step in the application of other multivariate data analysis methods. Thus, in conclusion, the

ROIMCR method is confirmed to be a promising approach for integrating pre-processing and exhaustive data analysis of complex LC × LC-HRMS datasets.

CRediT author statement

Miriam Pérez-Cova: Conceptualization, Writing-Original Draft preparation, Writing-Reviewing and Editing. **Joaquim Jaumot**: Conceptualization, Writing-Reviewing, and Editing, Funding Acquisition. **Romà Tauler**: Conceptualization, Writing-Reviewing and Editing, Funding Acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research leading to these results has received funding from the Spanish Ministry of Science and Innovation (MCI, Grants CTQ2017-82598-P and PID2019-105732GB-C21). The authors also want to grant support from the Catalan Agency for Management of University and Research Grants (AGAUR, Grant 2017SGR753) and the Spanish MCI (Severo Ochoa Project CEX2018-000794-S). MPC acknowledges a predoctoral FPU 16/02640 scholarship from the Spanish Ministry of Education and Vocational Training (MEFP).

References

[1] B.W.J. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent developments in two-dimensional liquid chromatography – Fundamental improvements for practical applications, Anal. Chem. 91 (2018). https://doi.org/10.1021/acs.analchem.8b04841. acs.analchem.8b04841.

[2] T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, J. Sep. Sci (2020) 1–50. https://doi.org/10.1002/jssc.202000011.

[3] M. Navarro-Reig, J. Jaumot, A. Baglai, G. Vivó-Truyols, P.J. Schoenmakers, R. Tauler, Untargeted comprehensive two-dimensional liquid chromatography coupled with high-resolution mass spectrometry analysis of rice Metabolome using multivariate curve resolution, Anal. Chem. (2017). https://doi.org/10.1021/acs.analchem.7b01648.

[4] M. Navarro-Reig, C. Bedia, R. Tauler, J. Jaumot, Chemometric strategies for peak detection and profiling from multidimensional chromatography, Proteomics 18 (2018) 1–12. https://doi.org/10.1002/pmic.201700327.

[5] M. Navarro-Reig, J. Jaumot, T.A. van Beek, G. Vivó-Truyols, R. Tauler, Chemometric analysis of comprehensive LC×LC-MS data: resolution of triacylglycerol structural isomers in corn oil, Talanta 160 (2016) 624–635. https://doi.org/10.1016/j.talanta.2016.08.005.

[6] B.R. Kowalski, Chemometrics, Anal. Lett. 11 (1978) 11–13. https://doi.org/10.1080/00032717808059728.

[7] H.T.L. dos Santos, A.M. de Oliveira, P.G. de Melo, W. Freitas, A.P.R. de Freitas, Chemometrics: theory and application, in: Multivar. Anal. Manag. Eng. Sci., 2013. https://doi.org/10.5772/53866.

[8] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinf. 20 (2019) 1–17. https://doi.org/10.1186/s12859-019-2848-8.

[9] A. De Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. https://doi.org/10.1039/c4ay00571f.

[10] K.M. Kalili, A. De Villiers, Systematic optimisation and evaluation of on-line, off-line and stop-flow comprehensive hydrophilic interaction chromatography × reversed phase liquid chromatographic analysis of procyanidins, Part I: theoretical considerations, J. Chromatogr. A 1289 (2013) 58–68. https://doi.org/10.1016/j.chroma.2013.03.008.

[11] D.R. Stoll, P.W. Carr, Two-dimensional liquid chromatography: a state of the art tutorial, Anal. Chem. (2017). https://doi.org/10.1021/acs.analchem.6b03506.

[12] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of sampling rate on resolution in comprehensive two-dimensional liquid chromatography, Anal. Chem. 70 (1998) 1585–1594. https://doi.org/10.1021/ac971184b.

[13] M. Muller, A.G.J. Tredoux, A. de Villiers, Predictive kinetic optimisation of hydrophilic interaction chromatography × reversed phase liquid chromatography separations: experimental verification and application to phenolic analysis, J. Chromatogr. A 1571 (2018) 107–120. https://doi.org/10.1016/j.chroma.2018.08.004.

[14] U. Woiwode, R.J. Reischl, S. Buckenmaier, W. Lindner, M. Lämmerhofer, Imaging peptide and protein chirality via amino acid analysis by chiral × chiral two-dimensional correlation liquid chromatography, Anal. Chem. 90 (2018) 7963–7971. https://doi.org/10.1021/acs.analchem.8b00676.

[15] M.H. Blokland, P.W. Zoontjes, L.A. Van Ginkel, M.G.M. Van De Schans, S.S. Sterk, T.F.H. Bovee, Multiclass screening in urine by comprehensive two-dimensional liquid chromatography time of flight mass spectrometry for residues of sulphonamides, beta-agonists and steroids, Food Addit. Contam. Part A Chem. Anal. Control. Expo. Risk Assess. 35 (2018) 1703–1715. https://doi.org/10.1080/19440049.2018.1506160.

[16] M. Xu, J. Legradi, P. Leonards, Evaluation of LC-MS and LC×LC-MS in analysis of zebrafish embryo samples for comprehensive lipid profiling, Anal. Bioanal. Chem. 412 (2020) 4313–4325. https://doi.org/10.1007/s00216-020-02661-1.

[17] F. Cacciola, D. Mangraviti, F. Rigano, P. Donato, P. Dugo, L. Mondello, H.J. Cortes, Novel comprehensive multidimensional liquid chromatography approach for elucidation of the microbosphere of shikimate-producing Escherichia coli SP1.1/pKD15.071 strain, Anal. Bioanal. Chem. 410 (2018) 3473–3482. https://doi.org/10.1007/s00216-017-0744-5.

[18] L. Montero, V. Sáez, D. von Baer, A. Cifuentes, M. Herrero, Profiling of Vitis vinifera L. canes (poly)phenolic compounds using comprehensive two-dimensional liquid chromatography, J. Chromatogr. A 1536 (2018) 205–215. https://doi.org/10.1016/j.chroma.2017.06.013.

[19] R. Karongo, T. Ikegami, D.R. Stoll, M. Lämmerhofer, A selective comprehensive reversed-phase×reversed-phase 2D-liquid chromatography approach with multiple complementary detectors as advanced generic method for the quality control of synthetic and therapeutic peptides, J. Chromatogr. A 1627 (2020) 461430. https://doi.org/10.1016/j.chroma.2020.461430.

[20] P. Venter, M. Muller, J. Vestner, M.A. Stander, A.G.J. Tredoux, H. Pasch, A. De Villiers, Comprehensive three-dimensional LC × LC × ion mobility spectrometry separation combined with high-resolution MS for the analysis of complex samples, Anal. Chem. 90 (2018) 11643–11650. https://doi.org/10.1021/acs.analchem.8b03234.

[21] S. Stephan, J. Hippler, T. Köhler, D. Brecht, O.J. Schmitz, A powerful four-dimensional separation method for complex samples, J. Anal. Test. 1 (2017) 1–9. https://doi.org/10.1007/s41664-017-0004-x.

[22] H. Zhang, J.M. Jiang, D. Zheng, M. Yuan, Z.Y. Wang, H.M. Zhang, C.W. Zheng, L.B. Xiao, H.X. Xu, A multidimensional analytical approach based on time-decoupled online comprehensive two-dimensional liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry for the analysis of ginsenosides from white and red ginsengs, J. Pharm. Biomed. Anal. 163 (2019) 24–33. https://doi.org/10.1016/j.jpba.2018.09.036.

[23] A. Moussa, T. Lauer, D. Stoll, G. Desmet, K. Broeckhoven, Numerical and experimental investigation of analyte breakthrough from sampling loops used for multi-dimensional liquid chromatography, J. Chromatogr. A 1626 (2020) 461283. https://doi.org/10.1016/j.chroma.2020.461283.

[24] H.C. van de Ven, J. Purmova, G. Groeneveld, T.S. Bos, A.F.G. Gargano, S. van der Wal, Y. Mengerink, P.J. Schoenmakers, Living with breakthrough: two-dimensional liquid-chromatography separations of a water-soluble synthetically grafted bio-polymer, Separations 7 (2020) 1–16. https://doi.org/10.3390/separations7030041.

[25] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active solvent modulation: a valve-based approach to improve separation compatibility in two-dimensional liquid chromatography, Anal. Chem. 89 (2017) 9260–9267. https://doi.org/10.1021/acs.analchem.7b02046.

[26] R.J. Vonk, A.F.G. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. De Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of saccharomyces cerevisiae, Anal. Chem. 87 (2015) 5387–5394. https://doi.org/10.1021/acs.analchem.5b00708.

[27] H. Tian, J. Xu, Y. Xu, Y. Guan, Multidimensional liquid chromatography system with an innovative solvent evaporation interface, J. Chromatogr. A 1137 (2006) 42–48. https://doi.org/10.1016/j.chroma.2006.10.005.

[28] Y. Chen, J. Li, O.J. Schmitz, Development of an at-column dilution modulator for flexible and precise control of dilution factors to overcome mobile phase incompatibility in comprehensive two-dimensional liquid chromatography, Anal. Chem. 91 (2019) 10251–10257. https://doi.org/10.1021/acs.analchem.9b02391.

[29] S.L. Weatherbee, T. Brau, D.R. Stoll, S.C. Rutan, M.M. Collinson, Simulation of elution profiles in liquid chromatography – IV: experimental characterization and modeling of solute injection profiles from a modulation valve used in two-dimensional liquid chromatography, J. Chromatogr. A 1626 (2020) 1–10. https://doi.org/10.1016/j.chroma.2020.461373.

[30] H. Wang, H.R. Lhotka, R. Bennett, M. Potapenko, C.J. Pickens, B.F. Mann, I.A. Haidar Ahmad, E.L. Regalado, Introducing online multicolumn two-dimensional liquid chromatography screening for facile selection of stationary and mobile phase conditions in both dimensions, J. Chromatogr. A 1622 (2020) 460895. https://doi.org/10.1016/j.chroma.2020.460895.

[31] C.J. Pickens, I.A. Haidar Ahmad, A.A. Makarov, R. Bennett, B.F. Mann, E.L. Regalado, Comprehensive online multicolumn two-dimensional liquid chromatography-diode array detection-mass spectrometry workflow as a framework for chromatographic screening and analysis of new drug

substances, Anal. Bioanal. Chem. 412 (2020) 2655–2663. https://doi.org/10.1007/s00216-020-02498-8.
[32] A.A. Aly, M. Muller, A. de Villiers, B.W.J. Pirok, T. Górecki, Parallel gradients in comprehensive multidimensional liquid chromatography enhance utilization of the separation space and the degree of orthogonality when the separation mechanisms are correlated, J. Chromatogr. A 1628 (2020) 461452. https://doi.org/10.1016/j.chroma.2020.461452.
[33] A. Niazi, R. Leardi, Genetic algorithms in chemometrics, J. Chemom. 26 (2012) 345–351. https://doi.org/10.1002/cem.2426.
[34] T. Alvarez-Segura, S. López-Ureña, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Multi-scale optimisation vs. genetic algorithms in the gradient separation of diuretics by reversed-phase liquid chromatography, J. Chromatogr. A 1609 (2020). https://doi.org/10.1016/j.chroma.2019.460427.
[35] B. Huygens, K. Efthymiadis, A. Nowé, G. Desmet, Application of evolutionary algorithms to optimise one- and two-dimensional gradient chromatographic separations, J. Chromatogr. A 1628 (2020) 461435. https://doi.org/10.1016/j.chroma.2020.461435.
[36] B.W.J. Pirok, A.F.G. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, J. Sep. Sci. 41 (2018) 68–98. https://doi.org/10.1002/jssc.201700863.
[37] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. Bommel, Recent applications of retention modelling in liquid chromatography, J. Sep. Sci. (2020) 1–27. https://doi.org/10.1002/jssc.202000905.
[38] R. Tauler, Multivariate curve resolution of multiway data using the multilinearity constraint, J. Chemom. (2020) 1–24. https://doi.org/10.1002/cem.3279.
[39] R. Bro, PARAFAC. Tutorial and applications, Chemometr. Intell. Lab. Syst. 38 (1997) 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4.
[40] J.D. Carroll, J.J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition, Psychometrika 35 (1970) 283–319. https://doi.org/10.1007/BF02310791.
[41] A.C. Olivieri, G.M. Escandar, A.M. de la Peña, Second-order and higher-order multivariate calibration methods applied to non-multilinear data using different algorithms, TrAC - Trends Anal. Chem. 30 (2011) 607–617. https://doi.org/10.1016/j.trac.2010.11.018.
[42] S.A. Bortolato, J.A. Arancibia, G.M. Escandar, A.C. Olivieri, Time-alignment of bidimensional chromatograms in the presence of uncalibrated interferences using parallel factor analysis. Application to multi-component determinations using liquid-chromatography with spectrofluorimetric detection, Chemometr. Intell. Lab. Syst. 101 (2010) 30–37. https://doi.org/10.1016/j.chemolab.2009.12.001.
[43] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, J. Chromatogr. A (2012). https://doi.org/10.1016/j.chroma.2012.05.050.
[44] S.E. Prebihalo, D.K. Pinkerton, R.E. Synovec, Impact of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry experimental design on data trilinearity and parallel factor analysis deconvolution, J. Chromatogr. A 1605 (2019) 460368. https://doi.org/10.1016/j.chroma.2019.460368.
[45] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709. https://doi.org/10.1021/ac052106o.
[46] H. Parastar, J.R. Radović, J.M. Bayona, R. Tauler, Solving chromatographic challenges in comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry using multivariate curve resolution-alternating least squares ABC Highlights: authored by Rising Stars and Top Experts, Anal. Bioanal. Chem. 405 (2013) 6235–6249. https://doi.org/10.1007/s00216-013-7067-y.
[47] Y. Izadmanesh, E. Garreta-Lara, J.B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data, J. Chromatogr. A 1488 (2017) 113–125. https://doi.org/10.1016/j.chroma.2017.01.052.
[48] H.P. Bailey, S.C. Rutan, Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine, Chemometr. Intell. Lab. Syst. 106 (2011) 131–141. https://doi.org/10.1016/j.chemolab.2010.07.008.
[49] M. Pérez-Cova, R. Tauler, J. Jaumot, Chemometrics in comprehensive two-dimensional liquid chromatography: a study of the data structure and its multilinear behavior, Chemometr. Intell. Lab. Syst. 201 (2020). https://doi.org/10.1016/j.chemolab.2020.104009.
[50] B.Q. Li, J. Chen, J.J. Li, X. Wang, H.L. Zhai, The application of a Tchebichef moment method to the quantitative analysis of multiple compounds based on three-dimensional HPLC fingerprint spectra, Analyst 140 (2015) 630–636. https://doi.org/10.1039/c4an01736f.
[51] M.B. Anzardi, J.A. Arancibia, A.C. Olivieri, Interpretation of matrix chromatographic-spectral data modeling with parallel factor analysis 2 and multivariate curve resolution, J. Chromatogr. A 1604 (2019). https://doi.org/10.1016/j.chroma.2019.460502.
[52] S.A. Bortolato, A.C. Olivieri, Chemometric processing of second-order liquid chromatographic data with UV-vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2, Anal. Chim. Acta 842 (2014) 11–19. https://doi.org/10.1016/j.aca.2014.07.007.
[53] G.M. Escandar, A.C. Olivieri, A road map for multi-way calibration models, Analyst 142 (2017) 2862–2873. https://doi.org/10.1039/c7an00822h.
[54] M. Navarro-Reig, J. Jaumot, T.A. van Beek, G. Vivó-Truyols, R. Tauler, Chemometric analysis of comprehensive LC×LC-MS data: resolution of triacylglycerol structural isomers in corn oil, Talanta (2016). https://doi.org/10.1016/j.talanta.2016.08.005.
[55] R. Tauler, I. Marqués, E. Casassas, Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths, J. Chemom. 12 (1998) 55–75. https://doi.org/10.1002/(sici)1099-128x(199801/02)12:1<55::aid-cem501>3.0.co;2-%23.
[56] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom. 17 (2003) 274–286. https://doi.org/10.1002/cem.801.
[57] A. Malik, R. Tauler, Performance and validation of MCR-ALS with quadrilinear constraint in the analysis of noisy datasets, Chemometr. Intell. Lab. Syst. 135 (2014) 223–234. https://doi.org/10.1016/j.chemolab.2014.04.002.
[58] R. Tautenhahn, C. Bottcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinf. 9 (2008) 1–16. https://doi.org/10.1186/1471-2105-9-504.
[59] A.W. Dowsey, J.A. English, F. Lisacek, J.S. Morris, G.Z. Yang, M.J. Dunn, Image analysis tools and emerging algorithms for expression proteomics, Proteomics 10 (2010) 4226–4257. https://doi.org/10.1002/pmic.200900635.
[60] C.R. Mittermayr, S.G. Nikolov, H. Hutter, M. Grasserbauer, Wavelet denoising of Gaussian peaks: a comparative study, Chemometr. Intell. Lab. Syst. 34 (1996) 187–202. https://doi.org/10.1016/0169-7439(96)00026-3.
[61] M. Daszykowski, B. Walczak, Use and abuse of chemometrics in chromatography, TrAC - Trends Anal. Chem. 25 (2006) 1081–1096. https://doi.org/10.1016/j.trac.2006.09.001.
[62] M. Navarro-Reig, J. Jaumot, A. Baglai, G. Vivó-Truyols, P.J. Schoenmakers, R. Tauler, Untargeted comprehensive two-dimensional liquid chromatography coupled with high-resolution mass spectrometry analysis of rice Metabolome using multivariate curve resolution, Anal. Chem. 89 (2017) 7675–7683. https://doi.org/10.1021/acs.analchem.7b01648.
[63] M.M. Sinanian, D.W. Cook, S.C. Rutan, D.S. Wijesinghe, Multivariate curve resolution-alternating least squares analysis of high-resolution liquid chromatography-mass spectrometry data, Anal. Chem. 88 (2016) 11092–11099. https://doi.org/10.1021/acs.analchem.6b03116.
[64] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinf. 20 (2019) 1–17. https://doi.org/10.1186/s12859-019-2848-8.
[65] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779–787. https://doi.org/10.1021/ac051437y.
[66] R. Tauler, E. Gorrochategui, J. Jaumot, R. Tauler, A protocol for LC-MS metabolomic data processing using chemometric tools, Protoc. Exch. (2015). https://doi.org/10.1038/protex.2015.102.
[67] S.E. Reichenbach, M. Ni, D. Zhang, E.B. Ledford, Image background removal in comprehensive two-dimensional gas chromatography, J. Chromatogr. A 985 (2003) 47–56. https://doi.org/10.1016/S0021-9673(02)01498-X.
[68] Z. Da Zeng, S.T. Chin, H.M. Hugel, P.J. Marriott, Simultaneous deconvolution and re-construction of primary and secondary overlapping peak clusters in comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1218 (2011) 2301–2310. https://doi.org/10.1016/j.chroma.2011.02.028.
[69] Y. Zhang, H.L. Wu, A.L. Xia, L.H. Hu, H.F. Zou, R.Q. Yu, Trilinear decomposition method applied to removal of three-dimensional background drift in comprehensive two-dimensional separation data, J. Chromatogr. A 1167 (2007) 178–183. https://doi.org/10.1016/j.chroma.2007.08.055.
[70] Z.P. Chen, H.L. Wu, J.H. Jiang, Y. Li, R.Q. Yu, A novel trilinear decomposition algorithm for second-order linear calibration, Chemometr. Intell. Lab. Syst. 52 (2000) 75–86. https://doi.org/10.1016/S0169-7439(00)00081-2.
[71] A.L. Xia, H.L. Wu, D.M. Fang, Y.J. Ding, L.Q. Hu, R.Q. Yu, Alternating penalty trilinear decomposition algorithm for second-order calibration with application to interference-free analysis of excitation-emission matrix fluorescence data, J. Chemom. 19 (2005) 65–76. https://doi.org/10.1002/cem.911.
[72] J.M. Amigo, T. Skov, R. Bro, ChroMATHography: solving chromatographic issues with mathematical models and intuitive graphics, Chem. Rev. 110 (2010) 4582–4605. https://doi.org/10.1021/cr900394n.
[73] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, Trends in data processing of comprehensive two-dimensional chromatography: state of the art, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 910 (2012) 31–45. https://doi.org/10.1016/j.jchromb.2012.06.039.
[74] C.G. Fraga, B.J. Prazen, R.E. Synovec, Objective data alignment and chemometric analysis of comprehensive two-dimensional separations with run-to-run peak shifting on both dimensions, Anal. Chem. 73 (2001) 5833–5840. https://doi.org/10.1021/ac010656q.
[75] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, Chemometr. Intell. Lab. Syst. 117 (2012) 80–91. https://doi.org/10.1016/j.chemolab.2012.02.003.
[76] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Two-dimensional correlation optimized warping algorithm for aligning GCxGC-MS data, Anal. Chem. 80 (2008) 2664–2671. https://doi.org/10.1021/ac7024317.

[77] J. Vial, H. Noçairi, P. Sassiat, S. Mallipatu, G. Cognon, D. Thiébaut, B. Teillet, D.N. Rutledge, Combination of dynamic time warping and multivariate analysis for the comparison of comprehensive two-dimensional gas chromatograms. Application to plant extracts, J. Chromatogr. A 1216 (2009) 2866–2872. https://doi.org/10.1016/j.chroma.2008.09.027.
[78] S. Furbo, A.B. Hansen, T. Skov, J.H. Christensen, Pixel-based analysis of comprehensive two-dimensional gas chromatograms (color plots) of petroleum: a tutorial, Anal. Chem. 86 (2014) 7160–7170. https://doi.org/10.1021/ac403650d.
[79] R. Tauler, A. de Juan, Chapter 5 - multivariate curve resolution for quantitative analysis, in: A.M. de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri (Editors), Fundam. Anal. Appl. Multiw. Calibration, Elsevier, 2015, pp. 247–292.
[80] A. de Juan, R. Tauler, Factor analysis of hyphenated chromatographic data. Exploration, resolution and quantification of multicomponent systems, J. Chromatogr. A 1158 (2007) 184–195. https://doi.org/10.1016/j.chroma.2007.05.045.
[81] S. Kim, M. Ouyang, J. Jeong, C. Shen, X. Zhang, A new method of peak detection for analysis of comprehensive two-dimensional gas chromatography mass spectrometry data, Ann. Appl. Stat. 8 (2014) 1209–1231. https://doi.org/10.1214/14-AOAS731.
[82] S. Plancade, Y. Rozenholc, E. Lund, Generalization of the normal-exponential model: exploration of a more accurate parametrisation for the signal distribution on Illumina BeadArrays, BMC Bioinf. 13 (2012) 1–16. https://doi.org/10.1186/1471-2105-13-329.
[83] S. Kim, H. Jang, I. Koo, J. Lee, X. Zhang, Normal–Gamma–Bernoulli peak detection for analysis of comprehensive two-dimensional gas chromatography mass spectrometry data, Comput. Stat. Data Anal. 105 (2017) 96–111. https://doi.org/10.1016/j.csda.2016.07.015.
[84] G. Vivó-Truyols, Bayesian approach for peak detection in two-dimensional chromatography, Anal. Chem. 84 (2012) 2622–2630. https://doi.org/10.1021/ac202124t.
[85] M. Navarro-Reig, C. Bedia, R. Tauler, J. Jaumot, Chemometric strategies for peak detection and profiling from multidimensional chromatography, Proteomics 18 (2018) 1–12. https://doi.org/10.1002/pmic.201700327.
[86] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: a review of advanced and tailored applications and challenges, Anal. Chim. Acta 765 (2013) 28–36. https://doi.org/10.1016/j.aca.2012.12.028.
[87] G.H. (Gene H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, 1989.
[88] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, R. Tauler, A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data, Anal. Chim. Acta 911 (2016) 1–13. https://doi.org/10.1016/j.aca.2016.01.011.
[89] R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, J. Chemom. 15 (2001) 627–646. https://doi.org/10.1002/cem.654.
[90] D.W. Cook, S.C. Rutan, D.R. Stoll, P.W. Carr, Two dimensional assisted liquid chromatography - a chemometric approach to improve accuracy and precision of quantitation in liquid chromatography using 2D separation, dual detectors, and multivariate curve resolution, Anal. Chim. Acta 859 (2015) 87–95. https://doi.org/10.1016/j.aca.2014.12.009.
[91] C. Tistaert, H.P. Bailey, R.C. Allen, Y. Vander Heyden, S.C. Rutan, Resolution of spectrally rank-deficient multivariate curve resolution: alternating least squares components in comprehensive two-dimensional liquid chromatographic analysis, J. Chemom. 26 (2012) 474–486. https://doi.org/10.1002/cem.2434.
[92] 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 (2017). https://doi.org/10.1016/j.aca.2017.01.047.
[93] H.P. Bailey, S.C. Rutan, P.W. Carr, Factors that affect quantification of diode array data in comprehensive two-dimensional liquid chromatography using chemometric data analysis, J. Chromatogr. A 1218 (2011) 8411–8422. https://doi.org/10.1016/j.chroma.2011.09.057.
[94] D.W. Cook, S.C. Rutan, Analysis of liquid chromatography-mass spectrometry data with an elastic net multivariate curve resolution strategy for sparse spectral recovery, Anal. Chem. 89 (2017) 8405–8412. https://doi.org/10.1021/acs.analchem.7b01832.
[95] E. Peré-Trepat, S. Lacorte, R. Tauler, Alternative calibration approaches for LC-MS quantitative determination of coeluted compounds in complex environmental mixtures using multivariate curve resolution, Anal. Chim. Acta 595 (2007) 228–237. https://doi.org/10.1016/j.aca.2007.04.011.
[96] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and quantification of complex mixtures of polycyclic aromatic hydrocarbons in heavy fuel oil sample by means of GC × GC-TOFMS combined to multivariate curve resolution, Anal. Chem. 83 (2011) 9289–9297. https://doi.org/10.1021/ac201799r.
[97] M. Bayat, M. Marín-García, J.B. Ghasemi, R. Tauler, Application of the area correlation constraint in the MCR-ALS quantitative analysis of complex mixture samples, Anal. Chim. Acta 1113 (2020) 52–65. https://doi.org/10.1016/j.aca.2020.03.057.
[98] A.C. de O. Neves, R. Tauler, K.M.G. de Lima, Area correlation constraint for the MCR–ALS quantification of cholesterol using EEM fluorescence data: a new approach, Anal. Chim. Acta 937 (2016) 21–28. https://doi.org/10.1016/j.aca.2016.08.011.
[99] I.T. Joliffe, B. Morgan, Principal component analysis and exploratory factor analysis, Stat. Methods Med. Res. 1 (1992) 69–95. https://doi.org/10.1177/096228029200100105.
[100] F. Wei, S.X. Ji, N. Hu, X. Lv, X.Y. Dong, Y.Q. Feng, H. Chen, Online profiling of triacylglycerols in plant oils by two-dimensional liquid chromatography using a single column coupled with atmospheric pressure chemical ionization mass spectrometry, J. Chromatogr. A 1312 (2013) 69–79. https://doi.org/10.1016/j.chroma.2013.09.005.
[101] X.Y. Dong, J. Zhong, F. Wei, X. Lv, L. Wu, Y. Lei, B.S. Liao, S.Y. Quek, H. Chen, Triacylglycerol composition profiling and comparison of high-oleic and normal peanut oils, JAOCS, J. Am. Oil Chem. Soc. 92 (2015) 233–242. https://doi.org/10.1007/s11746-014-2580-5.
[102] F. Andrić, K. Héberger, How to compare separation selectivity of high-performance liquid chromatographic columns properly? J. Chromatogr. A 1488 (2017) 45–56. https://doi.org/10.1016/j.chroma.2017.01.066.
[103] M.R. Euerby, P. Petersson, W. Campbell, W. Roe, Chromatographic classification and comparison of commercially available reversed-phase liquid chromatographic columns containing phenyl moieties using principal component analysis, J. Chromatogr. A 1154 (2007) 138–151. https://doi.org/10.1016/j.chroma.2007.03.119.
[104] R. Græsbøll, N.J. Nielsen, J.H. Christensen, Using the hydrophobic subtraction model to choose orthogonal columns for online comprehensive two-dimensional liquid chromatography, J. Chromatogr. A 1326 (2014) 39–46. https://doi.org/10.1016/j.chroma.2013.12.034.
[105] S. Brown, R. Tauler, Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Elsevier Science, 2015. https://books.google.es/books?id=heXEDwAAQBAJ.
[106] D. Bressanello, E. Liberto, M. Collino, F. Chiazza, R. Mastrocola, S.E. Reichenbach, C. Bicchi, C. Cordero, Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography: revealing fructose-induced changes in mice urinary metabolic signatures, Anal. Bioanal. Chem. 410 (2018) 2723–2737. https://doi.org/10.1007/s00216-018-0950-9.
[107] X. Sun, Y. Lv, J. Wang, H.Q. Cheng, J. Huang, Y. Du, J. Dong, Differential protein expression profiling by iTRAQ-2D-LC-MS/MS of rats treated with oxaliplatin, J. Cell. Biochem. 120 (2019) 18128–18141. https://doi.org/10.1002/jcb.29116.
[108] J.G. Scheiner, S.M. Manova, multiple response variables and multispecies interactions. Design and Analysis of Ecological Experiments, in: C. Press (Editor), Des. Anal. Ecol. Exp., 1993, pp. 94–112.
[109] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, M.E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (2005) 3043–3048. https://doi.org/10.1093/bioinformatics/bti476.
[110] E. Saccenti, A.K. Smilde, J. Camacho, Group-wise ANOVA simultaneous component analysis for designed omics experiments, Metabolomics 14 (2018) 1–18. https://doi.org/10.1007/s11306-018-1369-1.
[111] P.D.B. Harrington, N.E. Vieira, J. Espinoza, J.K. Nien, R. Romero, A.L. Yergey, Analysis of variance-principal component analysis: a soft tool for proteomic discovery, Anal. Chim. Acta 544 (2005) 118–127. https://doi.org/10.1016/j.aca.2005.02.042.
[112] J. Engel, L. Blanchet, B. Bloemen, L.P. Van den Heuvel, U.H.F. Engelke, R.A. Wevers, L.M.C. Buydens, Regularized MANOVA (rMANOVA) in untargeted metabolomics, Anal. Chim. Acta 899 (2015) 1–12. https://doi.org/10.1016/j.aca.2015.06.042.
[113] F. Marini, D. de Beer, E. Joubert, B. Walczak, Analysis of variance of designed chromatographic data sets: the analysis of variance-target projection approach, J. Chromatogr. A 1405 (2015) 94–102. https://doi.org/10.1016/j.chroma.2015.05.060.
[114] B. Govaerts, B. Francq, R. Marion, M. Martin, M. Thiel, The Essentials on Linear Regression, ANOVA, General Linear and Linear Mixed Models for the Chemist, second ed., Elsevier, 2020. https://doi.org/10.1016/b978-0-12-409547-2.14579-2.
[115] M. Martin, B. Govaerts, LiMM-PCA: combining ASCA+ and linear mixed models to analyse high-dimensional designed data, J. Chemom. 34 (2020) 1–20. https://doi.org/10.1002/cem.3232.
[116] A. Rácz, A. Gere, D. Bajusz, K. Héberger, Is soft independent modeling of class analogies a reasonable choice for supervised pattern recognition? RSC Adv. 8 (2018) 10–21. https://doi.org/10.1039/c7ra08901e.
[117] L. Stahle, Cross-validation for the two-class Problem: a monte Carl0 study, Analysis 1 (1987) 185–196.
[118] Y. Sato, T. Nakamura, K. Aoshima, Y. Oda, Quantitative and wide-ranging profiling of phospholipids in human plasma by two-dimensional liquid chromatography/mass spectrometry, Anal. Chem. 82 (2010) 9858–9864. https://doi.org/10.1021/ac102211r.
[119] M. Navarro-Reig, J. Jaumot, R. Tauler, An untargeted lipidomic strategy combining comprehensive two-dimensional liquid chromatography and chemometric analysis, J. Chromatogr. A 1568 (2018) 80–90. https://doi.org/10.1016/j.chroma.2018.07.017.
[120] M. Bylesjö, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, J. Chemom. 20 (2006) 341–351. https://doi.org/10.1002/cem.1006.

[121] E.D. Kantz, S. Tiwari, J.D. Watrous, S. Cheng, M. Jain, Deep neural networks for classification of LC-MS spectral peaks, Anal. Chem. 91 (2019) 12407–12413. https://doi.org/10.1021/acs.analchem.9b02983.
[122] Y. Xu, S. Zomer, R.G. Brereton, Support vector machines: a recent method for classification in chemometrics, Crit. Rev. Anal. Chem. 36 (2006) 177–188. https://doi.org/10.1080/10408340600969486.
[123] A. Checa, C. Bedia, J. Jaumot, Lipidomic data analysis: tutorial, practical guidelines and applications, Anal. Chim. Acta 885 (2015) 1–16. https://doi.org/10.1016/j.aca.2015.02.068.
[124] U.W. Liebal, A.N.T. Phan, M. Sudhakar, K. Raman, L.M. Blank, Machine learning applications for mass spectrometry-based metabolomics, Metabolites 10 (2020) 1–23. https://doi.org/10.3390/metabo10060243.
[125] M. Marcinkiewicz-Siemion, M. Kaminski, M. Ciborowski, K. Ptaszynska-Kopczynska, A. Szpakowicz, A. Lisowska, M. Jasiewicz, E. Tarasiuk, A. Kretowski, B. Sobkowicz, K.A. Kaminski, Machine-learning facilitates selection of a novel diagnostic panel of metabolites for the detection of heart failure, Sci. Rep. 10 (2020) 1–11. https://doi.org/10.1038/s41598-019-56889-8.
[126] S.I. Abba, A.G. Usman, S. Işik, Simulation for response surface in the HPLC optimization method development using artificial intelligence models: a data-driven approach, Chemometr. Intell. Lab. Syst. 201 (2020). https://doi.org/10.1016/j.chemolab.2020.104007.
[127] P. Bonini, T. Kind, H. Tsugawa, D.K. Barupal, O. Fiehn, Retip: retention time prediction for compound annotation in untargeted metabolomics, Anal. Chem. 92 (2020) 7515–7522. https://doi.org/10.1021/acs.analchem.9b05765.
[128] D. Panagopoulos Abrahamsson, J.S. Park, R.R. Singh, M. Sirota, T.J. Woodruff, Applications of machine learning to in silico quantification of chemicals without analytical standards, J. Chem. Inf. Model. 60 (2020) 2718–2727. https://doi.org/10.1021/acs.jcim.9b01096.
[129] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425–1432. https://doi.org/10.1021/ac00014a016.
[130] H.P. Bailey, S.C. Rutan, Comparison of chemometric methods for the screening of comprehensive two-dimensional liquid chromatographic analysis of wine, Anal. Chim. Acta 770 (2013) 18–28. https://doi.org/10.1016/j.aca.2013.01.062.