Chemometric Strategies for Peak Detection and Profiling from Multidimensional Chromatography

*Meritxell Navarro-Reig, Carmen Bedia, Romà Tauler, and Joaquim Jaumot**

The increasing complexity of omics research has encouraged the development of new instrumental technologies able to deal with these challenging samples. In this way, the rise of multidimensional separations should be highlighted due to the massive amounts of information that provide with an enhanced analyte determination. Both proteomics and metabolomics benefit from this higher separation capacity achieved when different chromatographic dimensions are combined, either in LC or GC. However, this vast quantity of experimental information requires the application of chemometric data analysis strategies to retrieve this hidden knowledge, especially in the case of nontargeted studies. In this work, the most common chemometric tools and approaches for the analysis of this multidimensional chromatographic data are reviewed. First, different options for data preprocessing and enhancement of the instrumental signal are introduced. Next, the most used chemometric methods for the detection of chromatographic peaks and the resolution of chromatographic and spectral contributions (profiling) are presented. The description of these data analysis approaches is complemented with enlightening examples from omics fields that demonstrate the exceptional potential of the combination of multidimensional separation techniques and chemometric tools of data analysis.

1. Introduction

The technological revolution of the last decades has allowed the development of analytical instrumental techniques that enable a better understanding of biological systems. These advances have been possible by the combination of technical developments in the analytical instrumentation (i.e., resulting in equipment with higher analytical properties and faster acquisition rates) with the progress in the field of computer sciences, that allows the acquisition and storage of large amounts of information in extremely short times.[1,2]

Focusing on separation techniques and omics research fields, two different aspects should be discussed considering these analytical instrumentation advances. On the one hand, the increase in instrumental detection capabilities that allows obtaining much more information per unit of time (i.e., a high-resolution mass spectrum in every considered retention time).[3] This fact has permitted the evolution of the chromatographic separation monitoring from using a single or a reduced number of channels (i.e., few m/z values in MS measurements) to acquiring a virtually unlimited number of channels (i.e., thousands of m/z values in high-resolution MS). Regarding omics sciences, this revolution in the MS instrumentation performance has been crucial.[4] Currently, high-resolution mass spectrometers allow the acquisition of full spectra enabling the identification of compounds by comparison of their exact mass and, in some cases, their fragmentation patterns. On the other hand, extensive work has been done to improve chromatographic instrumentation (i.e., sub 2 μm particle columns in HPLC). Moreover, the option of coupling different separation techniques has to be considered.[2,5] In this work, we will fix our attention on these new hyphenation possibilities of combining various separation modes or techniques to increase the analytical resolution.

Therefore, the traditional single chromatographic separation (i.e., either in LC or GC) has evolved to allow multiple separation dimensions. This progress has been possible by the development of interfaces permitting the chromatographic analysis using two (or more in, for instance, 3D-LC) consecutive column separations. In these multidimensional separations, the comprehensive chromatography is of particular interest.[6] In this case, the effluent from the first column is automatically injected into the second chromatographic column by means of a modulator. The effluent of the first column is collected for a fixed period of time (known as modulation). For instance, two popular approaches are the comprehensive 2D gas chromatography (GC × GC) and the comprehensive 2D liquid chromatography (LC × LC). In both cases, selected columns attempt to use the most orthogonal separation modes possible to obtain optimal chromatographic resolution conditions.[7–9] For example, in multidimensional LC, this orthogonal separation can be achieved by using columns with stationary phases of almost orthogonal properties such as reversed-phase (RP) in one dimension and hydrophilic interaction liquid chromatography (HILIC) in the other dimension. However, this coupling between different chromatographic columns can also be performed offline (i.e., heart-cutting), but these most traditional approaches are being replaced by the more

M. Navarro-Reig, Dr. C. Bedia, Prof. R. Tauler, Dr. J. Jaumot
Department of Environmental Chemistry
Institute of Environmental Assessment and Water Research (IDAEA) -
Spanish National Research Council (CSIC)
Jordi Girona 18-34, E08034, Barcelona, Spain
E-mail: joaquim.jaumot@idaea.csic.es

DOI: 10.1002/pmic.201700327

automatic online comprehensive methods.[5] The main aim of multidimensional chromatography is to perform a rapid separation in the second dimension (i.e., usually few seconds in the case of GC × GC and few minutes in the case of LC × LC) to differentiate the overlapping compounds in the first dimension (much slower separation, from minutes to hours). Finally, other options should be considered linking entirely different separation modes such as, for instance, liquid chromatography and ion mobility spectrometry (i.e., LC-IMS).[10,11] In this case, LC is combined with IMS to separate the compounds according to both their chromatographic retention and ion mobility properties. Here, compounds in the second dimension are differentiated according to their drift time.

These new instrumental alternatives make it possible to obtain a large amount of information from the studied system. This progress has contributed to the arising of the omics studies that seek the complete measurement (if possible, in nontargeted studies) of the considered ome to understand its structure, function, or dynamics in an organism.[4] Regarding the primary focus of this work, two particular omes should be contemplated: the proteome that gives rise to the proteomics field and, especially, the metabolome that generates the metabolomics field. However, this amount of experimental information also implies the risk that the sought knowledge remains hidden within the massive experimental data.[12] This bottleneck motivates the use of advanced chemometric methods for the data analysis which can extract relevant quantitative and qualitative information.[13] This evolution in both the chromatographic and chemometric fields has also enabled the popularization of nontargeted omics. These nontargeted studies pursue the determination of the maximum possible number of compounds without an a priori limitation to a reduced set of known compounds within a family or pathway (as is usual in targeted omics). For this reason, this nontargeted omics approach is considered as a “discovery omics” because it allows the generation of new knowledge from the measured data but at the cost of an increase in the data processing complexity.[14,15]

In a similar manner to instrumental advances, chemometric methods have evolved in an attempt to adapt to these new challenging datasets.[16,17] Therefore, the same evolution can be observed from the analysis of simple datasets consisting of a single representative sample measurement (i.e., pH value or the absorbance at a single wavelength) to the evaluation of multidimensional datasets for every sample (i.e., LC × LC-MS). **Figure 1** shows a graphical summary of the evolution of this complexity of generated data and analyzed datasets.

The simplest case (zeroth-order data) corresponds to the description of a sample by a single measurement such as a pH value or the absorbance at a specific wavelength. The analysis of multiple samples results in a data vector (known as a one-way dataset of size *m* rows where *m* corresponds to the total number of samples) that can be analyzed using classical univariate statistical methods.

The next level (first-order data) corresponds to the description of a single sample by multiple measurements ordered, for example, in a vector. For example, first-order data is obtained when direct infusion MS is carried out or considering the total number of measurements provided by an array of sensors (e.g., pH, conductivity, the concentration of different ions). Here, when considering multiple samples, a data matrix is obtained generating a two-way dataset (with a size of *m* rows corresponding to the number of samples and *n* columns corresponding to the number of measurements per each sample, that is, the number of *m/z* values). In this case, the data analysis requires employing multivariate methods. The second-order data for a single sample has to be ordered in a data matrix. Each sample is represented by two dimensions as, for instance, monodimensional LC measurements using a mass spectrometer as a detector (LC-MS) or excitation–emission fluorescence spectra. When dealing with several samples, three-way datasets are obtained geometrically adopting the structure of an array or data cube. Finally, in the case of multidimensional separation techniques, third-order data is retrieved. Thus, each sample is represented by a data cube with two modes related to chromatographic dimensions and the other mode related to the detection technique. For example, in the case of LC × LC-MS, there are two chromatographic dimensions, and the third is related to the MS spectra. In this more sophisticated case, the analysis of multiple samples gives rise to higher-order structures: four-way datasets that are usually geometrically represented by hypercubes (see Figure 1 dataset order).

In this multidimensional chromatographic data, multilinearity must be taken into account.[18] Therefore, it is mandatory to check if sample profiles in the chromatographic dimensions are constant when considering different modulations within a sample and between different samples. These profiles are characterized by the retention time and the peak shape observed for the same compound in different modulations and samples.[19,20] The fulfillment of this same behavior condition (known as multilinearity condition, e.g., trilinearity for second-order data or quadrilinearity for third-order data) implies that a component (sample profile) of a higher-order data structure (second-order or more) can be described mathematically as a linear function of chromatographic and spectral profiles in the different data modes.[21,22] The accomplishment (or not) of this condition guides the selection of the most appropriate multivariate data analysis model for a particular dataset. Second-order datasets fulfilling this condition are known as trilinear whereas third-order datasets are quadrilinear.[17] In the case of datasets that do not obey this trilinearity condition, a bilinear behavior can generally be assumed.[23] In the case of multidimensional data, a GC × GC-MS measured sample commonly fulfills the trilinearity condition due to the high reproducibility in observed retention times and peak shapes.[24,25] In contrast, LC × LC-MS data usually does not accomplish this trilinear behavior, especially when an elution gradient is used in the second chromatographic dimension.[26,27] Therefore, LC × LC-MS is normally considered and analyzed as bilinear data. This data property has particular relevance since it deeply influences the selection of a chromatographic data profiling method. In the case of dealing with multilinear data, the use of methods taking advantage of this data structure is recommended. For example, the parallel factor analysis (PARAFAC) method is an example of a method extensively used in the analysis of GC× GC-MS data.[28] In the case of the analysis of nontrilinear data, a larger variety of methods are available.[13,29,30] For instance, the analysis of LC × LC-MS data has been performed using multivariate curve resolution by alternating least squares (MCR-ALS) or PARAFAC2 (a model similar to PARAFAC but allowing small deviations of this multilinear behavior) methods.

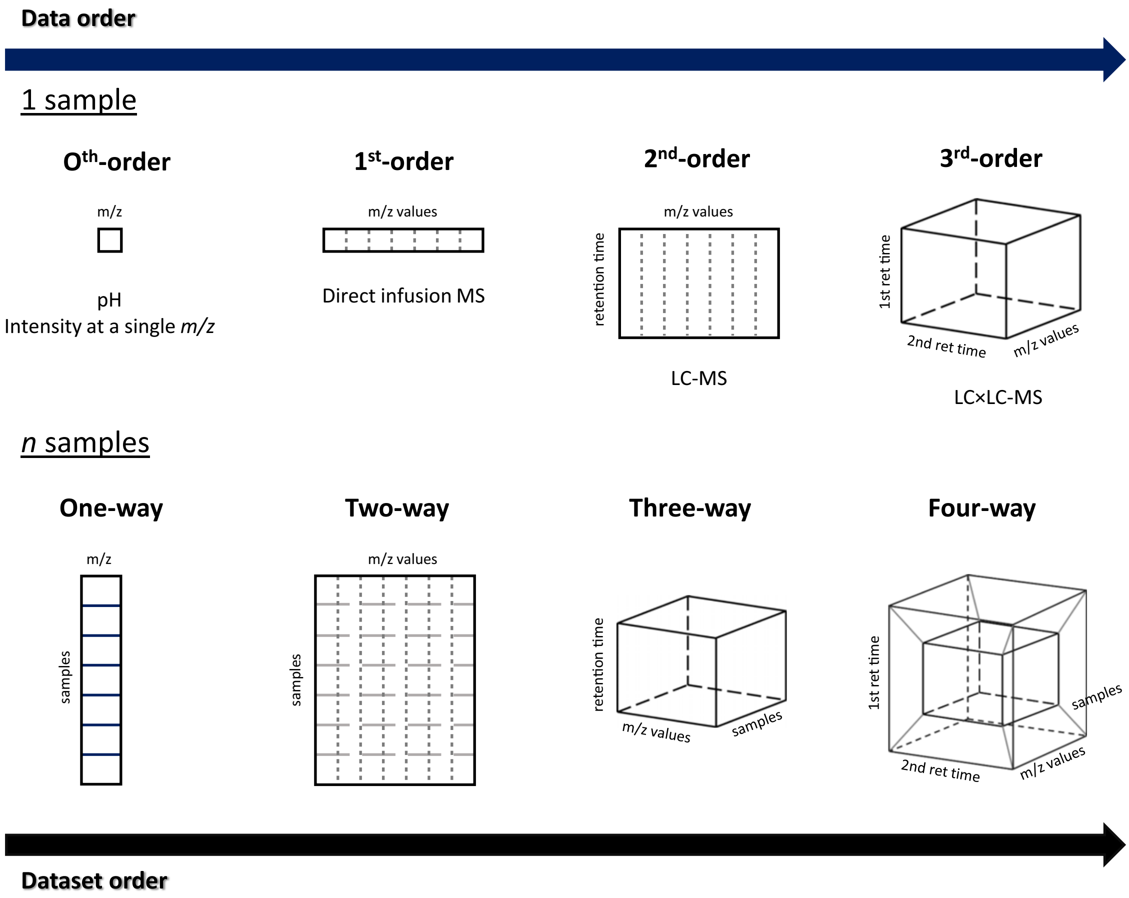

Figure 1. Graphical summary describing the increasing complexity of experimental data order and resulting datasets.

Alternatively, other approaches based on the chromatographic peak detection are also available and will be introduced below.

The following sections of this review will focus on the processing of these complex datatypes considering the experimental multidimensional raw data (in some cases known as pixel-level data in contrast with the processed peak tables). Firstly, different preprocessing methods that allow improvements in the data quality will be introduced. For instance, approaches useful for correcting small misalignments between chromatographic peaks in different samples will be presented. This correction permits the comparison among samples and the analysis of non-trilinear raw data using trilinear methods. Secondly, the most widely used chemometric approaches for the analysis of these multidimensional datasets will be depicted together with some guidelines related to their use. Real multidimensional examples within the omics field (e.g., GC × GC-MS and LC × LC-MS) will be used to illustrate the major advantages and drawbacks of each analysis method.

2. Multidimensional Data Preprocessing

The first step in the analysis of multidimensional data consists of the quality enhancement of the experimentally obtained chromatographic data.[20,31,32] This initial correction facilitates the subsequent detection or profiling of chromatographic peaks. In multidimensional chromatography, the most commonly encountered problems are similar to those found in monodimensional chromatography.[31] Therefore, in most of the cases, preprocessing methods used for multidimensional data are adaptations of approaches commonly employed for the correction of monodimensional chromatographic data. In this work, three major types of data preprocessing will be discussed: i) compression of the chromatographic and spectral information; ii) peak alignment within the same sample and between different samples; and iii) elimination of noisy contributions or baseline drifts from the experimental data. Finally, it is also important to mention these preprocessing methods used to remove sample-to-sample amount variation, which is commonly known as the size effect.[33,34] This data correction is of the utmost interest as this is an error source that could affect the validity of the biomarkers found by the subsequent sample comparison.[35] Different approaches aim to correct this undesired size effect using either chemical or mathematical normalizations.[36] As methods used for this normalization in multidimensional separations are analogous to those used in monodimensional separations, interested readers are referred to recent detailed reviews.[34–39]

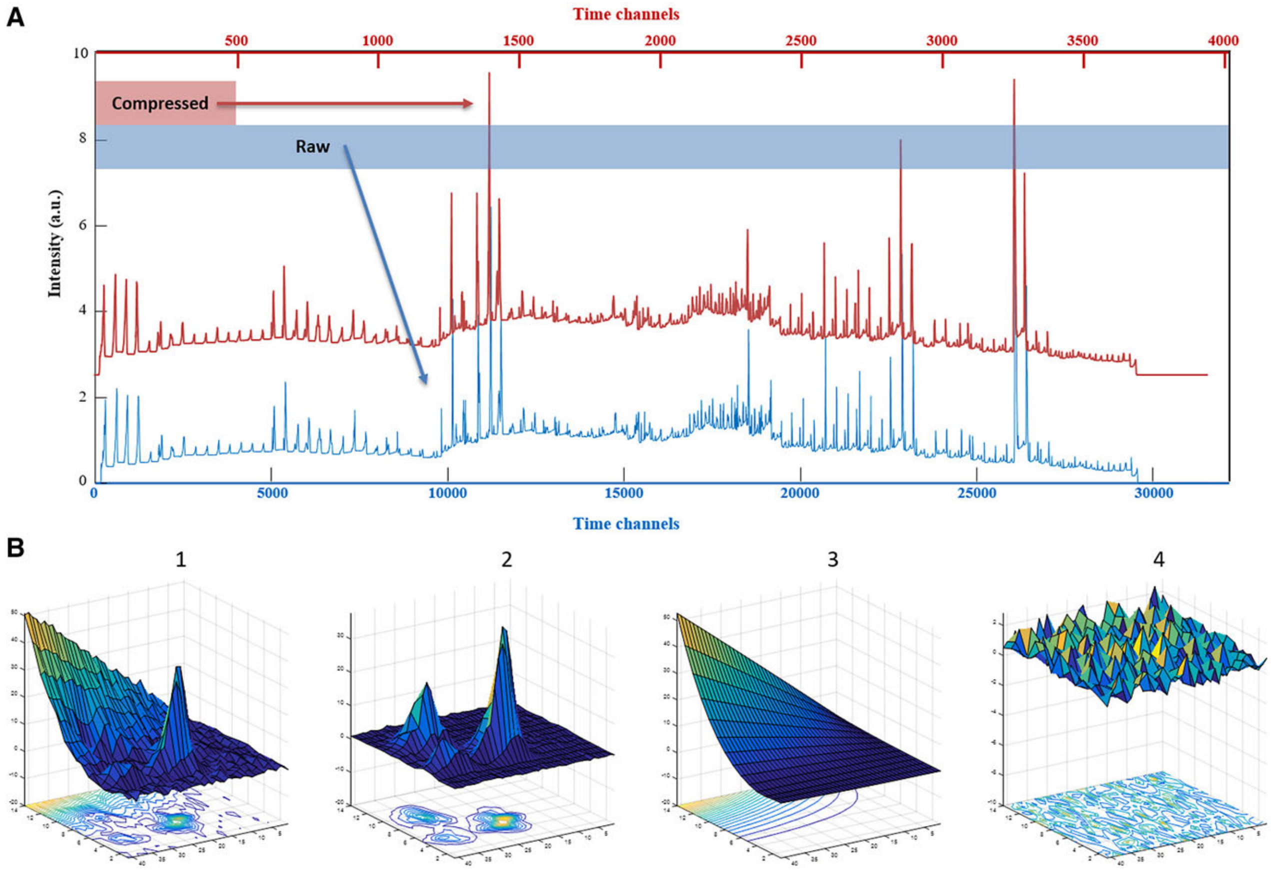

Figure 2. Multidimensional chromatographical data preprocessing. A) Example of wavelet compression of LC × LC-MS data. The blue line shows the TIC obtained from the raw data (32 000 time channels). The red line shows the compressed TIC (4000 channels). No relevant differences can be observed despite the eightfold compression depicted by the blue and red colored bars. B) Representation of the possible contributions in multidimensional data: 1) measured data, 2) chemical compounds, 3) baseline drifts, 4) random noise.

2.1. Multidimensional Chromatographic Data Compression

Multidimensional chromatographic techniques hyphenated to multivariate detectors (in particular, to MS) generate a vast amount of information for each sample. For example, in LC × LC-MS, a mass spectrum is acquired (if possible, at high resolution) for each of the considered retention times (combining the elution of the two chromatographic dimensions). Thus, the chromatogram of a single sample can contain hundreds of millions of elements (thousands of elution times containing each one thousands of m/z values) generating files in the order of gigabytes per sample. Consequently, this dimensionality causes problems for the data storage and processing.[2] Therefore, the first step in the multidimensional data preprocessing is the reduction of this amount of information keeping only that considered relevant. Most of these compression strategies take advantage of the sparse data characteristics because there is a majority of zero or close to zero elements (values below the instrumental noise threshold).[40,41] Moreover, the different compression approaches can be differentiated according to the compressed mode (chromatographic or spectral).

In the case of the chromatographic dimension reduction, the application of wavelet compression allows reducing the data size without losing information and with the additional advantage of filtering part of the experimental noise.[42,43] These wavelet transforms are based on the mathematical decomposition of elution data in different frequency components, keeping values of position, intensity, and shape of each chromatographic peak adapted to the new reduced scale. Wavelets can be applied in a 1D way (analogously to classical chromatography) or following a 2D approach (as in the case of image compression).[44] An example of the compression power of this approach is shown in **Figure 2**A. An eightfold reduction of the size of the chromatographic data can be obtained using a four-level wavelet without losing relevant information.

In the spectral dimension (particularly in the case of high-resolution MS), the primary goal is to maintain the acquired m/z values accuracy of signals of significant intensity. Therefore, those m/z values with an intensity signal below a particular threshold (i.e., experimental noise level) will be disregarded. Alternatively, there are compression approaches based on classical methods, such as binning, that allow a high compression rate but at the cost of losing data quality (i.e., MS spectral accuracy).[14] For this reason, alternative methods have emerged such as the *centWave* method which is based on the detection of regions of interest.[45] This approach achieves high compression rates while keeping the MS spectral resolution. Also, this approach performs preliminary filtering of the experimental noise by only maintaining those signals above a certain threshold value which define a relatively low number of mass traces. For these reasons, this method is widely used in the compression of metabolomics data since it is a fundamental part of the popular XCMS software.[46] However, the use of XCMS in multidimensional techniques is not straightforward. The optimization of algorithm parameters for chromatographic peak definition for each mass trace is complicated due to the presence of multiple peaks associated with a single compound in different modulations of the second chromatographic dimension. Nonetheless, the search for regions of interest in the mass spectral dimension (and the definition of mass traces) is still possible but the peak alignment and modelling steps present in the original algorithm should be prevented, as proposed in the work of Navarro-Reig.[47]

2.2. Chromatographic Peak Alignment

Chromatographic peak alignment within a single sample (due to the different modulations of the second dimension in which a single analyte can elute) or between different samples is of vital importance for the correct identification of compounds in omics studies.[48] Also, this peak alignment allows detecting the same compound in different samples to gain confidence in its quantification. If this correction is already crucial in monodimensional chromatography to correct small retention time drifts between different samples, it is even more critical in the case of multidimensional chromatography.[18,31] In the latter case, the retention time variability increases due to the second separation dimension which, as mentioned above, usually has very short separation times (seconds in the case of GC and minutes in the case of LC).

The correlation optimized warping (COW) algorithm is one of the most used methods in the monodimensional chromatography peak alignment.[49,50] This algorithm is based on taking a chromatogram as a reference and adjusting the target chromatograms to it from compressing or stretching segments. Zhang extended this algorithm to multidimensional chromatographies in the 2D COW method[51] and, more recently, Furbo proposed the peak alignment by fast Fourier transform (PAFFT)[52] method with a similar purpose. In both cases, there is the limitation of performing the peak alignment adjusting the total ion chromatograms. Therefore, the information coming from the multichannel detector is underused. In order to overcome this drawback, approaches using all acquired channels for the peak alignment have appeared. These methods consider the variations of all available channels as, for instance, the method proposed by Reichenbach[53] or the recent RT shift correction algorithm proposed by Zushi.[54] In addition, there are alternatives considering approximations based on carrying out a resampling by interpolation of the first chromatographic dimension.[55]

Despite all these different peak alignment approaches, this preprocessing is still a critical step that can lead to errors in the subsequent identification and quantification of proteins or metabolites. For this reason, methods that allow analyzing the data without the need for the peaks to be fully synchronized within and between the different samples should be considered as highly attractive.[17] Undoubtedly, this presupposes a significant advantage since it dramatically simplifies the data analysis.

2.3. Noise and Background Drift Correction

The next step in the multidimensional data quality enhancement is the evaluation of possible instrumental error sources such as experimental random noise or systematic drifts. Figure 2B represents the different contributions to the measured analytical signal (Figure 2B,1). The raw data can be decomposed in contributions due to the analytes of interest (Figure 2B,2) and, also, in contributions caused by unwanted baseline drift (Figure 2B,3) or random noise (Figure 2B,4). It is essential to perform these corrections since these undesired contributions could influence the detection of the compounds and their possible quantification since they can alter both the height and shape of the chromatographic peak.

There are different alternatives to minimize the contribution of random noise. Firstly, as mentioned above, during the compression of the data (for example, by wavelets or regions of interest approaches) a preliminary noise filtering can be carried out. Also, focused corrections can be performed using approaches based on data smoothing, such as the Savitzky–Golay or the CODA algorithms.[56,57] Finally, the experimental noise could be filtered in the subsequent analysis when using, for instance, peak resolution methods. These methods allow distinguishing the contributions due to the different chemical or biological compounds from contributions generated by random signals.

The second block of correction methods consists of eliminating experimental contributions that generate a drift in the chromatogram baseline.[18] Again, there are different options to perform this preprocessing. The simplest approach is based on the subtraction of the baseline from a chromatogram of a blank sample. Also, this correction can be carried out by determining the baseline of a particular chromatogram and, then, subtracting this calculated baseline from the other chromatograms. In this case, methods based on polynomial fitting can be employed but approaches based on least-squares baseline fitting are more advisable . For instance, the asymmetric least squares (AsLS) method provides excellent results.[58] Alternatively, there are methods based on penalized least-squares as the adaptive algorithm iteratively reweighted penalized least squares (airPLS) proposed by Zhang[59] which has the advantage of not requiring any user intervention for parameter optimization. More recently, other approaches getting a profit of the data sparsity have appeared as the baseline estimation and denoising with sparsity (BEADS) method.[60] This algorithm presents the advantage of allowing the baseline correction and, simultaneously, eliminating part of the random noise. However, this background drift correction is an active field of research with many recent proposals attempting to deal simultaneously with other drawbacks such as peak alignment.[61,62] Finally, as in the case of random noise, there is the option of modeling these baseline contributions as an additional component in the multidimensional data resolution methods that will be presented in the following section.

3. Peak Detection and Profiling

After the data preprocessing described in the previous step, the actual analysis of the multidimensional chromatographic data is carried out. This section is divided into two blocks corresponding to the two main types of chemometric approaches used for the analysis of multidimensional data: i) detection of 2D chromatographic peaks, and ii) resolution of elution and spectral profiles.

3.1. Peak Detection Methods

The simplest approach is the detection of 2D chromatographic peaks. However, it should be taken into account that the use of these methods only provides information relative to the two chromatographic dimensions so that the information present in the spectral mode is lost (unlike the profiling methods discussed below). In this peak detection approaches, the used procedures can be considered similar to those employed in the determination of compounds in 2-PAGE, a technique widely used in the proteomics field.

In a similar manner to omics studies, two approaches can be distinguished depending on the objective of the analysis. Targeted peak detection approaches are based on the comparison of the 2D chromatographic peaks in complex samples with those obtained in the analysis of reference standards. An example of these methods is the window target testing factor analysis (WTTFA) method.[63] More recently, Barcaru has presented an algorithm for comparing pairs of multidimensional chromatograms (demonstrated with a GC × GC-MS data example) using Bayesian statistics considering the Jensen–Shannon divergence.[64] This algorithm presents the advantage of eliminating other possible drawbacks such as small differences in retention times without the need for high computational requirements.

In the case of nontargeted peak detection approaches, the primary aim is to characterize the chromatographic peaks present in the multidimensional chromatogram without using any previous information (e.g., chromatograms obtained using reference standards). There are different methods to carry out these nontargeted multidimensional peaks detection. The first proposed methods were adaptations of algorithms previously described for the peak detection in monodimensional chromatography, such as, for example, the two-step peak detection algorithm presented by Peters.[65] This algorithm is based on the detection of peaks in the first chromatographic dimension (i.e., using the derivative properties of the Gaussian peaks) and, in a second step, clustering those peaks detected in the first dimension in the second dimension. Alternatively, methods based on image analysis for detecting the 2D peaks can be highlighted. For example, the watershed-based algorithms in which the image generated by the multidimensional response obtained in the analysis of a sample is considered. For each of the detected maxima in the image, the signal of neighboring pixels is evaluated to identify whether the measured signal corresponds to a chemical signal or just a background noise contribution.[66,67] In this way, multidimensional peaks caused by chemical or biological contributions are detected. However, these algorithms present problems both due to the splitting of a single compound peak as multiple peaks or due to the fading of minor chromatographic peaks into the background in cases of high levels of experimental noise. In these nonoptimal cases, the watershed algorithm parameters must be selected with special care.[68] For this reason, new methods for the detection of peaks in multidimensional chromatography are continuously being presented. Recent contributions based on the fulfillment of mathematical models can be highlighted. Kim et al. proposed the *msPeak* algorithm that focuses on the identification of regions with potential chromatographic peaks using the Normal–Exponential–Bernoulli (NEB) model and detecting the peaks in these regions from models considering first derivative tests.[69] Recently, adaptations of this method using other mathematical models such as Normal-Gamma (NG)[70] or Normal–Gamma–Bernoulli (NGB) have been proposed.[71] Finally, the contribution of Vivo-Truyols should be mentioned in which an approach based on Bayesian statistics is proposed for peak detection in multidimensional chromatography.[72]

A more in-depth explanation of these algorithms can be found in the references of each work and the revision done by van Stee.[73]

3.2. Peak Resolution Methods

The concept of the resolution of multidimensional data is based on the recovery of profiles (profiling) in all the measured modes (separation and spectral). In the case of multidimensional chromatography using MS as a detector, the elution profiles corresponding to each of the dimensions that provide the quantitative information will be solved and, also, the MS spectral profile that gives the qualitative information enabling the identification of compounds.[25] Additionally, in the particular case of LC-IMS-MS, the chromatographic elution and ion mobility profiles are obtained in addition to the MS spectra. Therefore, these strategies allow the analysis of chromatograms with highly overlapped peaks since the combination of chromatographic and spectral information permits the differentiation of the contributions corresponding to every single compound.

This double source of information is of particular interest in the framework of omics studies.[5,74] Here, the multidimensional data profiling allows the identification of the compounds and, also, the quantitative comparison between the different types of samples from the resulting elution profiles. On the one hand, this makes possible to differentiate between samples using established pattern recognition or classification methods, such as PCA or PLS-DA multivariate approaches. On the other hand, the quantitative information can be used to obtain possible markers related to the factor under study using methods such as statistical inference tests or the selected VIP variables obtained in the PLS-DA model.[19,75]

Several methods allow the profiling of these datasets. However, only the most widely used methods are presented in detail below: PARAFAC and its variant PARAFAC2 methods[22,76] and MCR-ALS method.[77] However, both classical and recent methods should also be mentioned as potential alternatives for peak resolution such as the generalized rank annihilation method (GRAM),[78–80] the trilinear decomposition (TLD),[81] the iterative key set factor analysis (IKSFA),[27,82–85] the alternative moving window factor analysis (AMWFA),[86] and the joint approximate diagonalization of eigenmatrices (JADE).[87]

3.2.1. PARAFAC and PARAFAC2

The PARAFAC method has been widely used for the analysis of multidimensional chromatographic data and, in particular, the work done by Synovec and collaborators in the GC × GC-MS data processing can be highlighted.[19,28,88,89] In this data approach, it is assumed that the data obtained from GC × GC-MS instruments behave in a multilinear way, which is not always the case. Therefore, the application of methods that impose this multilinear model is possible under some circumstances.[23,25,90] In these cases, the advantage of providing a unique solution without mathematical ambiguities is achieved. Variations of the PARAFAC method have been proposed for cases in which this multilinearity condition is not met as is the general case in LC × LC-MS (as well as some cases in GC × GC-MS, in particular, when a temperature gradient is used in the second dimension). These experimental issues generate small drifts in the elution time or changes in the shape of the chromatographic peaks between consecutive modulations and samples, especially when coelution among sample constituents exists.[26,89]

From a mathematical point of view, the PARAFAC method is based on the decomposition of the information of a second or higher-order data structure. In the case of considering a single sample measured by 2D chromatography and a multichannel detector, a data cube is generated (Figure 1) that can be decomposed according to the following equation:

$$\underline{\mathbf{X}} = \sum_{i=1}^{k} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{d}_i \ + \underline{\mathbf{E}} \qquad (1)$$

where $\underline{\mathbf{X}}$ represents the data cube obtained for this 2D sample, for example, of LC × LC-MS. This cube can be decomposed into three factor contributions using a reduced number of k components: $\mathbf{a}_i$ and $\mathbf{b}_i$ are vector profiles that describe the elution profiles in the two separation dimensions for the kth component, and $\mathbf{d}_i$ are profiles that contain the k mass spectra that can be associated with the resolved components. $\underline{\mathbf{E}}$ (residuals data cube) contains the variance not explained by the model and $\otimes$ represents the outer product. In this case, the trilinearity constraint forces that $\mathbf{a}_i$ and $\mathbf{d}_i$ profiles are equal in all the modulations, whereas the $\mathbf{b}_i$ profiles are free to vary between modulations. In the case of considering several samples, a four-way data structure is generated (i.e., hypercube) that can be decomposed in a similar manner allowing to retrieve a new set of profiles ($\mathbf{c}_i$) containing the quantitative information for each sample. Different algorithms allow the resolution of the previous equations, but the most used is the alternating least squares method under constraints (in particular, nonnegativity).[22]

PARAFAC 2 is a variant of PARAFAC specially developed for the analysis of multidimensional data that do not fulfill the trilinear behavior because of small deviations in any of the modes.[91,92] The PARAFAC2 algorithm similarly decomposes the data to PARAFAC but, in this case, the algorithm allows some freedom in the $\mathbf{b}_i$ elution profiles, allowing that every sample has its own set of profiles. However, the property of uniqueness is maintained by forcing the cross product of these profiles to be constant for all modulations. This property is useful in the analysis of multidimensional chromatographic data as it allows minor deviations on the retention time of the chromatographic peaks.[23,93] However, in these $\mathbf{b}_i$ profiles, it is not possible to apply additional constraints, such as nonnegativity, and therefore they may be not recovered and interpreted directly from a chromatographic viewpoint.

PARAFAC-based methods have been applied in the study of multidimensional data in several fields of research. In the case of omic studies, these examples are found mainly in the field of metabolomics using GC × GC-MS.[19,94–96] For example, the work of Snyder et al. shows how these methods can be used to separate, identify, and quantify compounds in targeted metabolomics studies.[97] Thus, the application of the PARAFAC method to GC × GC-MS data allows the analysis of L-β-methylamino-alanine (BMAA) in brain tissue extracts. Data of high complexity with multiple coeluting compounds was obtained as can be seen in **Figure 3**A. The described analysis protocol used the PARAFAC method to detect the presence of BMAA in brain tissue samples. As shown in Figure 3B,C, the use of PARAFAC allows obtaining the chromatographic profiles for each component in the two separation dimensions and the corresponding mass spectrum (Figure 3D).

3.2.2. Multivariate Curve Resolution by Alternating Least Squares

MCR-ALS is a widely used method for the analysis of large datasets variety from single or multiple chromatographic runs to hyperspectral images or environmental tables.[98] Focusing on multidimensional separation data, MCR-ALS adopts a different approach than PARAFAC based on the fulfillment of an underlying bilinear model.

MCR-ALS is a factor analysis method that gives much more freedom to small deviations in the behavior of the data (minor changes are allowed in the retention time between samples or in the shape of peaks). The model can be written in a matrix form as follows:

$$\mathbf{X}_{\mathrm{m}} = \sum_{i=1}^{k} \mathbf{a}_i \ \mathbf{d}_i{}^{T} + \mathbf{E} \qquad (2)$$

In the classical interpretation, $\mathbf{X}_{\mathrm{m}}$ is the matrix containing the chromatographic and spectral information (elution times in the rows and m/z values in the columns), $\mathbf{a}_i$ is related to the elution profiles of each component (quantitative information), and $\mathbf{d}_i$ is associated with the mass spectra of each component (qualitative information). One of the limitations of the MCR-ALS method is the possible existence of multiple solutions that adjust the experimental data in the same way (in contrast, to the uniqueness property of PARAFAC solutions).[77] For this reason, during the mathematical optimization, constraints such as the nonnegativity of the elution and spectral profiles or the mass spectra normalization are applied. These constraints provide a chemical or biological sense to the purely mathematical solution and

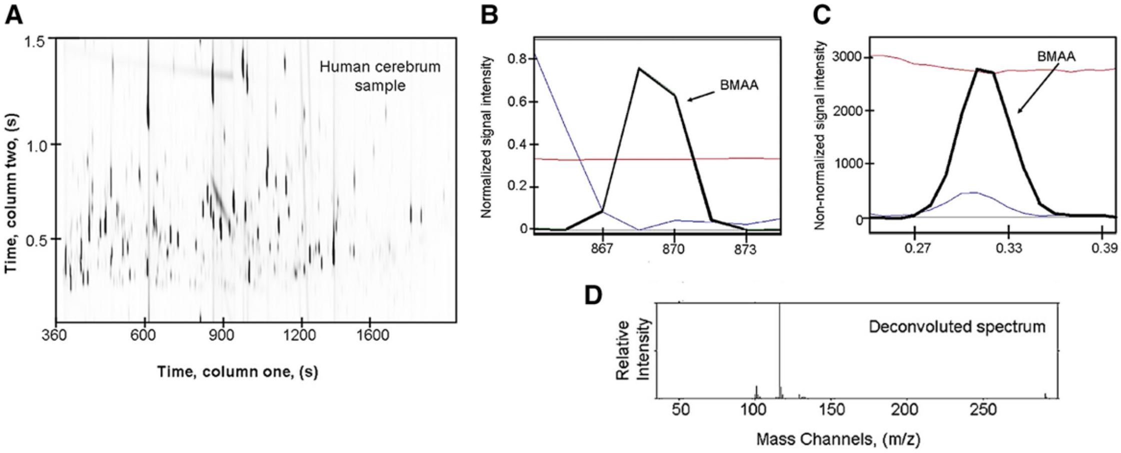

Figure 3. A) GC × GC-TOFMS example data of m/z 73 from extracted human cerebrum samples. PARAFAC resolved BMAA (bold line) and noisy contributions (thin lines) of (B) the first and (C) the second chromatographic dimensions. D) PARAFAC resolved mass spectrum. Adapted with permission.[97] Copyright 2010, Elsevier B.V.

minimize the existence of these ambiguities inherent to the factor analysis methods.[98] However, in case of working with multilinear data, the accomplishment of this multilinear behavior can be forced in the MCR-ALS solutions through additional constraints guaranteeing the uniqueness of the obtained solution.[90] When MS detection is used, the spectra of the resolved components are sparse and have a few number of highly selective or nearly specific signals, which reduce the possibility of these ambiguities.

Equation (2) also depicts how MCR-ALS deals with two-way matrices. However, as multidimensional separation data can be represented by a cube (i.e., LC × LC-MS of a single sample) or hypercubes (i.e., LC × LC-MS of multiple samples),[25] a preliminary data arrangement stage is required. So, the first step in the analysis is to unfold the data cube or hypercube in a two-way data matrix that can be analyzed by MCR-ALS (see **Figure 4**). This strategy of unfolding the data cube allows considering each modulation of the second dimension as an individual matrix. Then, each one of these single modulation matrices is set one on the top of each other. Next, this augmented data matrix is analyzed following the same model as in Equation (2). Multidimensional quantitative information is retrieved by refolding the $\mathbf{a}_i$ elution profiles by means of reversing the unfolding transformation whereas $\mathbf{d}_i$ contains the qualitative information.

Although MCR-ALS is initially only based on the fulfillment of the bilinear model explained above, it has been extended to model multiway data fulfilling multilinear models.[25,75] Thus, MCR-ALS can be used in most of the situations encountered in practice in multidimensional chromatography, when multilinear models are not fulfilled (only bilinear model is considered), when they are only partially fulfilled by some of the components, or when they are fulfilled by all the components of systems.

Recently, a paper by Cook et al.[99] discussed different strategies based on MCR-ALS to carry out the quantification of several analytes (furanocoumarins in apiaceous vegetables) in multidimensional data. In addition to the independent analysis of LC × LC chromatogram, hybrid approaches based on using the MCR-ALS resolved component spectra of the multidimensional analysis for a second analysis of only the first dimension chromatogram (2D assisted liquid chromatography [2DALC]) or using a matrix augmentation strategy to analyze simultaneously data coming from detectors at the end of the first and second columns (combined 2DALC [c2DALC]).[99,100]

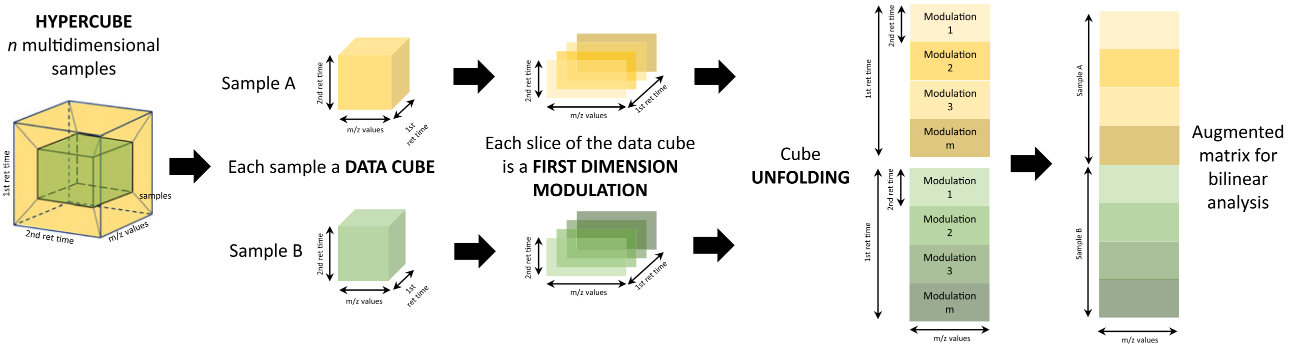

Figure 4. Data unfolding strategy allowing the bilinear analysis of hypercubes/data cubes structures.

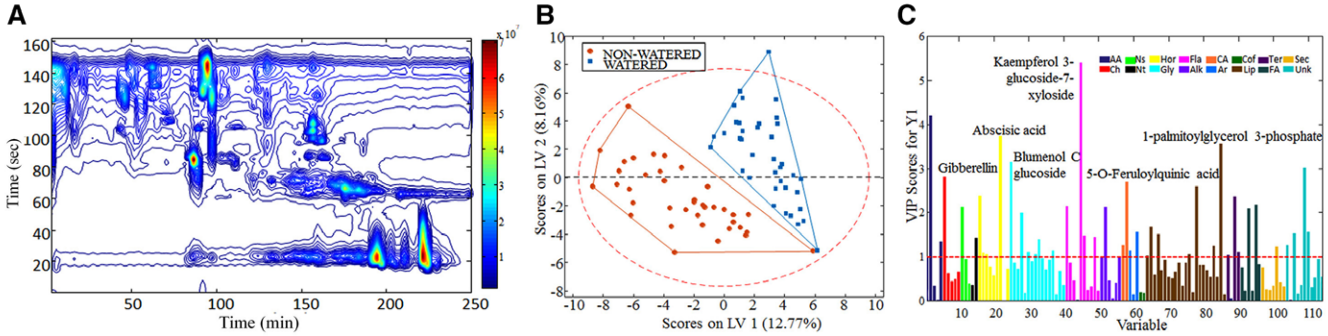

Figure 5. A) 2D LC × LC-MS chromatogram obtained in the analysis of a rice sample. Results of the classification analysis performed using MCR-ALS resolved peak areas of the stressed rice samples. B) PLS-DA scores plot showing drought stress effects, and (C) VIP scores plot for feature selection. Adapted with permission.[47] Copyright 2017, American Chemical Society.

The MCR-ALS method has been widely used in studies of different fields for both GC × GC and LC × LC data analysis. In the case of the omics studies, there are several recent examples in which its potential is shown for the analysis of nontargeted metabolomic studies. For instance, two environmental metabolomics examples are presented. In the first case, Yzadmanesh et al. described a protocol for the analysis of GC × GC-MS data taking as a case study the metabolome of *Daphnia Magna*.[75] In this case, the different steps of the analysis were described focusing on the evaluation of the data multilinearity and the data compression needed before the analysis. In the second case, Navarro-Reig analyzed LC × LC-HRMS data for the evaluation of the changes in rice metabolome caused by watering and harvesting time.[47] **Figure 5**A shows the complexity of the multidimensional data obtained in these metabolomic studies. The MCR-ALS analysis of LC × LC-MS data allowed the identification of approximately 150 metabolites that were affected by these environmental stressing factors. These metabolites were characterized by their elution profiles in each of the chromatographic dimensions and the mass spectrum from which their identification was achieved. This MCR-ALS resolution allowed establishing the variations with time of the different groups of metabolites and determining those metabolites most related to the absence of water using a PLS-DA analysis (see Figures 5B,C).

4. Guidelines and Prospects

Chemometric methods presented in this review allow the extraction of valuable information from highly complex data generated by multidimensional chromatographic techniques, such as those from nontargeted proteomic or metabolomic studies.

The different available options for multidimensional separations create data with unique properties that make necessary the development of tailored analysis procedures. Thus, the choice of the analysis method should take into account the data quality (i.e., the need for a baseline drift correction or peak alignment) as well as the information that is intended to retrieve. This reasoning will lead us to the use of detection (more straightforward but less informative) or profiling (more complicated but more informative) methods. In this last case, the structural properties of the multidimensional data must also be considered for the final selection of the analysis strategy. For instance, a clear difference exists between GC × GC-MS and LC × LC-MS data. In the first case, the behavior of the data approaches the multilinearity, so the use of trilinear methods such as PARAFAC may be possible and, in some cases, without the need to carry out complex data preprocessing. However, in the common case of a nonideal behavior, other more flexible methods such as MCR-ALS employing multilinear constraints appear as a more reliable option. In the case of LC × LC-MS, the data moves away from this ideal multilinear behavior so that more flexible methods such as bilinear MCR-ALS and, in some cases, PARAFAC2 should be considered. A visual summary of the main aspects to be considered when attempting to perform a multidimensional data analysis is shown in **Figure 6**.

All the presented methodologies allow the analysis of multidimensional chromatographic data. However, the remaining key challenge in these workflows is the development of algorithms that would enable the (total or partial) automation of the analysis, especially in the case of profiling approaches. For instance, there is a need of pipelines that allow carrying out the different stages of data compression, preprocessing, and detection or resolution of the chromatographic peaks with minimal user intervention. Here, it is also important to develop pipelines (in particular, for those automated) that assure the data integrity through all its lifecycle: from the raw data to the processed elution profiles and spectra. The maintaining of this data integrity will guarantee the quality of the obtained results and the reproducibility of not only the experimental protocols but also the data analysis approach used.

Also, the possibility of the development of new instrumental techniques or the coupling of more separation dimensions must be taken into account. An example of this is the popularization of LC-IMS-MS that combines different separation modes allowing the analysis of highly complex samples. This is a promising tool in fields such as lipidomics in which samples may contain multiple isobaric lipid species.[101,102] These new options will broaden the possibility of using these multidimensional techniques in the different omics sciences. In the case of metabolomics, several studies can be found in the literature demonstrating the benefits of the combination of multidimensional separation techniques and advanced data analysis tools. In contrast, few proteomics examples can be found. However, the development of

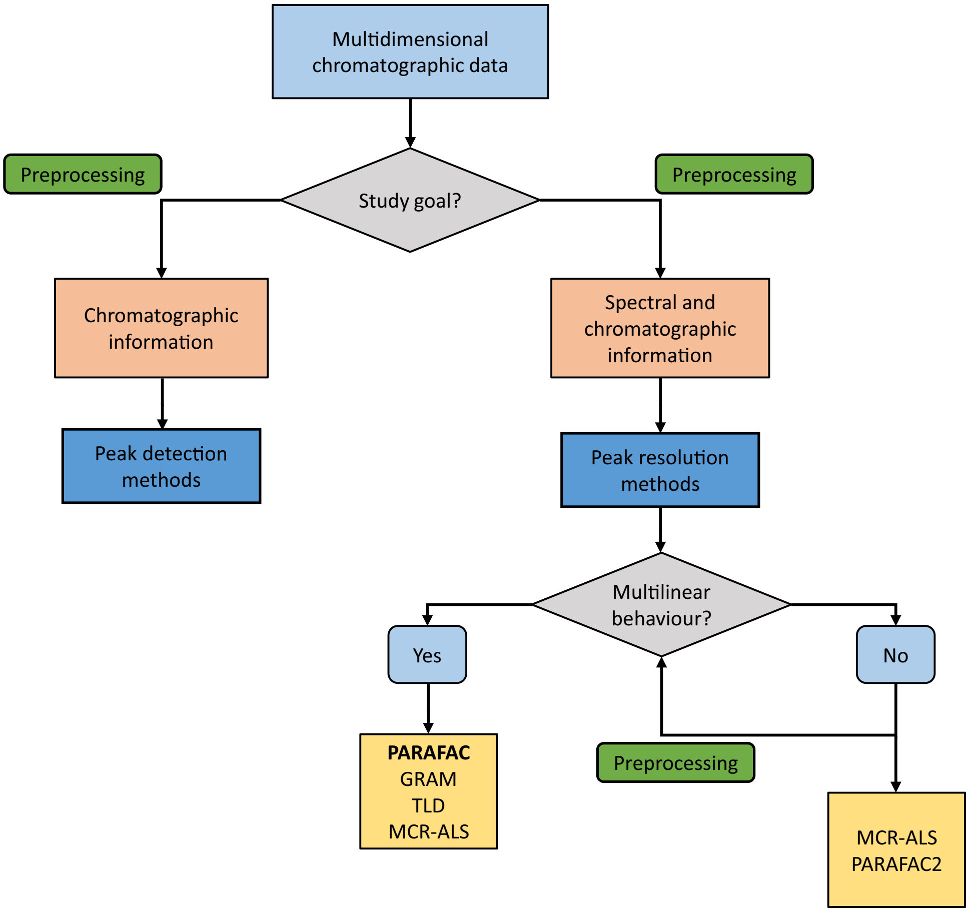

Figure 6. Flow-chart guiding the chemometrical method selection depending on the study goal and the data quality.

LC-based multidimensional techniques (e.g., LC × LC-MS and LC-IMS-MS) is an excellent starting point for the emergence of proteomics studies taking advantage of these new possibilities.

In conclusion, all these recent technological developments will allow obtaining more comprehensive and reliable information about the biological systems under study. However, the challenging issues related to the analysis of these complex datasets will grow in parallel.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 320737. Also, the authors acknowledge support from both the Ministry of Economy, Industry and Competitiveness (Grant CTQ2017-82598-P) and the Catalan Agency for Management of University and Research Grants (AGAUR) (2017SGR753).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

chemometrics, comprehensive chromatography, multidimensional chromatography, peak detection, peak resolution

Received: January 23, 2018
Revised: March 16, 2018
Published online: May 15, 2018

[1] E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, G. P. Nolan, *Nat. Rev. Genet.* **2010**, *11*, 647.
[2] J. C. May, J. A. McLean, *Annu. Rev. Anal. Chem.* **2016**, *9*, 387.
[3] A. G. Marshall, C. L. Hendrickson, *Annu. Rev. Anal. Chem.* **2008**, *1*, 579.
[4] G. J. Patti, O. Yanes, G. Siuzdak, *Nat. Rev. Mol. Cell Bio.* **2012**, *13*, 263.
[5] D. R. Stoll, P. W. Carr, *Anal. Chem.* **2017**, *89*, 519.
[6] P. W. Carr, J. M. Davis, S. C. Rutan, D. R. Stoll, *Adv. Chromatog.* **2015**, *50*, 139.
[7] P. Schoenmakers, P. Marriott, J. Beens, *LC GC Eur.* **2003**, *16*, 335.
[8] P. J. Schoenmakers, G. Vivó-Truyols, W. M. C. Decrop, *J. Chromatogr. A* **2006**, *1120*, 282.
[9] D. R. Stoll, X. Li, X. Wang, P. W. Carr, S. E. G. Porter, S. C. Rutan, *J. Chromatogr. A* **2007**, *1168*, 3.

[10] E. S. Baker, E. A. Livesay, D. J. Orton, R. J. Moore, W. F. Danielson III, D. C. Prior, Y. M. Ibrahim, B. L. LaMarche, A. M. Mayampurath, A. A. Schepmoes, D. F. Hopkins, K. Tang, R. D. Smith, M. E. Belov, *J. Proteome Res.* **2010**, *9*, 997.
[11] S. J. Valentine, X. Liu, M. D. Plasencia, A. E. Hilderbrand, R. T. Kurulugama, S. L. Koeniger, D. E. Clemmer, *Expert Rev. Proteomics* **2005**, *2*, 553.
[12] B. B. Misra, J. F. Fahrmann, D. Grapov, *Electrophoresis* **2017**, *38*, 2257.
[13] Z. D. Zeng, H. M. Hugel, P. J. Marriott, *Anal. Bioanal. Chem.* **2011**, *401*, 2373.
[14] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, *TrAC Trends Anal. Chem.* **2016**, *82*, 425.
[15] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, Y. Liang, *Anal. Chim. Acta* **2016**, *914*, 17.
[16] G. M. Escandar, H. C. Goicoechea, A. Muñoz de la Peña, A. C. Olivieri, *Anal. Chim. Acta* **2014**, *806*, 8.
[17] G. M. Escandar, A. C. Olivieri, *Analyst* **2017**, *142*, 2862.
[18] J. M. Amigo, T. Skov, R. Bro, *Chem. Rev.* **2010**, *110*, 4582.
[19] S. E. Prebihalo, K. L. Berrier, C. E. Freye, H. D. Bahaghighat, N. R. Moore, D. K. Pinkerton, R. E. Synovec, *Anal. Chem.* **2018**, *90*, 505.
[20] K. M. Pierce, B. Kehimkar, L. C. Marney, J. C. Hoggard, R. E. Synovec, *J. Chromatogr. A* **2012**, *1255*, 3.
[21] J. D. Carroll, J. J. Chang, *Psychometrika* **1970**, *35*, 283.
[22] R. Bro, *Chemometrics Intell. Lab. Syst.* **1997**, *38*, 149.
[23] S. A. Bortolato, A. C. Olivieri, *Anal. Chim. Acta* **2014**, *842*, 11.
[24] J. C. Hoggard, R. E. Synovec, *Anal. Chem.* **2008**, *80*, 6677.
[25] H. Parastar, R. Tauler, *Anal. Chem.* **2014**, *86*, 286.
[26] M. Navarro-Reig, J. Jaumot, T. A. van Beek, G. Vivó-Truyols, R. Tauler, *Talanta* **2016**, *160*, 624.
[27] H. P. Bailey, S. C. Rutan, *Chemometrics Intell. Lab. Syst.* **2011**, *106*, 131.
[28] D. K. Pinkerton, K. M. Pierce, R. E. Synovec, *Data Handling in Science and Technology*, Vol. *30*, Elsevier, Amsterdam, Netherlands **2016**, p. 333.
[29] K. M. Pierce, R. E. Mohler, *Sep. Purif. Rev.* **2012**, *41*, 143.
[30] Z. Zeng, J. Li, H. M. Hugel, G. Xu, P. J. Marriott, *TrAC Trends Anal. Chem.* **2014**, *53*, 150.
[31] J. T. V. Matos, R. M. B. O. Duarte, A. C. Duarte, *J. Chromatogr. B* **2012**, *910*, 31.
[32] E. Szymańska, A. N. Davies, L. M. C. Buydens, *Analyst* **2016**, *141*, 5689.
[33] M. Katajamaa, M. Orešič, *J. Chromatogr. A* **2007**, *1158*, 318.
[34] Y. Wu, L. Li, *J. Chromatogr. A* **2016**, *1430*, 80.
[35] P. Filzmoser, B. Walczak, *J. Chromatogr. A* **2014**, *1362*, 194.
[36] A. M. De Livera, M. Sysi-Aho, L. Jacob, J. A. Gagnon-Bartsch, S. Castillo, J. A. Simpson, T. P. Speed, *Anal. Chem.* **2015**, *87*, 3606.
[37] Y. Gagnebin, D. Tonoli, P. Lescuyer, B. Ponte, S. de Seigneux, P. Y. Martin, J. Schappler, J. Boccard, S. Rudaz, *Anal. Chim. Acta* **2017**, *955*, 27.
[38] A. Chawade, E. Alexandersson, F. Levander, *J. Proteome Res.* **2014**, *13*, 3114.
[39] A. Gardlo, A. K. Smilde, K. Hron, M. Hrdá, R. Karlíková, D. Friedecký, T. Adam, *Metabolomics* **2016**, *12*.
[40] D. W. Cook, S. C. Rutan, *Anal. Chem.* **2017**, *89*, 8405.
[41] J. J. De Rooi, C. Ruckebusch, P. H. C. Eilers, *Anal. Chem.* **2014**, *86*, 6291.
[42] M. Daszykowski, B. Walczak, *TrAC Trends Anal. Chem.* **2006**, *25*, 1081.
[43] C. R. Mittermayr, S. G. Nikolov, H. Hutter, M. Grasserbauer, *Chemometrics Intell. Lab. Syst.* **1996**, *34*, 187.
[44] A. W. Dowsey, J. A. English, F. Lisacek, J. S. Morris, G. Yang, M. J. Dunn, *Proteomics* **2010**, *10*, 4226.
[45] R. Stolt, R. J. O. Torgrip, J. Lindberg, L. Csenki, J. Kolmert, I. Schuppe-Koistinen, S. P. Jacobsson, *Anal. Chem.* **2006**, *78*, 975.
[46] R. Tautenhahn, C. Bottcher, S. Neumann, *BMC Bioinformatics* **2008**, *9*, 504.
[47] M. Navarro-Reig, J. Jaumot, A. Baglai, G. Vivó-Truyols, P. J. Schoenmakers, R. Tauler, *Anal. Chem.* **2017**, *89*, 7675.
[48] K. M. Pierce, L. F. Wood, B. W. Wright, R. E. Synovec, *Anal. Chem.* **2005**, *77*, 7735.
[49] N. P. V. Nielsen, J. M. Carstensen, J. Smedsgaard, *J. Chromatogr. A* **1998**, *805*, 17.
[50] G. Tomasi, F. Van Den Berg, C. Andersson, *J. Chemometr.* **2004**, *18*, 231.
[51] D. Zhang, X. Huang, F. E. Regnier, M. Zhang, *Anal. Chem.* **2008**, *80*, 2664.
[52] S. Furbo, A. B. Hansen, T. Skov, J. H. Christensen, *Anal. Chem.* **2014**, *86*, 7160.
[53] S. E. Reichenbach, D. W. Rempe, Q. Tao, D. Bressanello, E. Liberto, C. Bicchi, S. Balducci, C. Cordero, *Anal. Chem.* **2015**, *87*, 10056.
[54] Y. Zushi, J. Gros, Q. Tao, S. E. Reichenbach, S. Hashimoto, J. S. Arey, *J. Chromatogr. A* **2017**, *1508*, 121.
[55] R. C. Allen, S. C. Rutan, *Anal. Chim. Acta* **2012**, *723*, 7.
[56] W. Windig, J. M. Phalp, A. W. Payne, *Anal. Chem.* **1996**, *68*, 3602.
[57] A. Savitzky, M. J. E. Golay, *Anal. Chem.* **1964**, *36*, 1627.
[58] P. H. C. Eilers, *Anal. Chem.* **2004**, *76*, 404.
[59] Z. M. Zhang, S. Chen, Y. Z. Liang, *Analyst* **2010**, *135*, 1138.
[60] X. Ning, I. W. Selesnick, L. Duval, *Chemometrics Intell. Lab. Syst.* **2014**, *139*, 156.
[61] G. L. Erny, T. Acunha, C. Simó, A. Cifuentes, A. Alves, *J. Chromatogr. A* **2017**, *1492*, 98.
[62] F. Qian, Y. Wu, P. Hao, *Opt. Laser Technol.* **2017**, *96*, 202.
[63] S. E. G. Porter, D. R. Stoll, S. C. Rutan, P. W. Carr, J. D. Cohen, *Anal. Chem.* **2006**, *78*, 5559.
[64] A. Barcaru, G. Vivó-Truyols, *Anal. Chem.* **2016**, *88*, 2096.
[65] S. Peters, G. Vivó-Truyols, P. J. Marriott, P. J. Schoenmakers, *J. Chromatogr. A* **2007**, *1156*, 14.
[66] S. E. Reichenbach, M. Ni, V. Kottapalli, A. Visvanathan, *Chemometrics Intell. Lab. Syst.* **2004**, *71*, 107.
[67] S. E. Reichenbach, X. Tian, Q. Tao, D. R. Stoll, P. W. Carr, *J. Sep. Sci.* **2010**, *33*, 1365.
[68] G. Vivó-Truyols, H. G. Janssen, *J. Chromatogr. A* **2010**, *1217*, 1375.
[69] S. Kim, M. Ouyang, J. Jeong, C. Shen, X. Zhang, *Ann. Appl. Stat.* **2014**, *8*, 1209.
[70] S. Plancade, Y. Rozenholc, E. Lund, *BMC Bioinf.* **2012**, *13*, 329.
[71] S. Kim, H. Jang, I. Koo, J. Lee, X. Zhang, *Comput. Stat. Data Anal.* **2017**, *105*, 96.
[72] G. Vivó-Truyols, *Anal. Chem.* **2012**, *84*, 2622.
[73] L. L. P. van Stee, U. A. T. Brinkman, *TrAC Trends Anal. Chem.* **2016**, *83*, 1.
[74] D. W. Cook, S. C. Rutan, *J. Chemometr.* **2014**, *28*, 681.
[75] Y. Izadmanesh, E. Garreta-Lara, J. B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, *J. Chromatogr. A* **2017**, *1488*, 113.
[76] R. A. Harshman, *UCLA Working Papers in Phonetics* **1970**, *16*, 1.
[77] R. Tauler, A. Smilde, B. Kowalski, *J. Chemometr.* **1995**, *9*, 31.
[78] C. A. Bruckner, B. J. Prazen, R. E. Synovec, *Anal. Chem.* **1998**, *70*, 2796.
[79] C. G. Fraga, B. J. Prazen, R. E. Synovec, *High Resolut. Chromatogr.* **2000**, *23*, 215.
[80] C. G. Fraga, B. J. Prazen, R. E. Synovec, *Anal. Chem.* **2001**, *73*, 5833.
[81] A. E. Sinha, C. G. Fraga, B. J. Prazen, R. E. Synovec, *J. Chromatogr. A* **2004**, *1027*, 269.
[82] H. P. Bailey, S. C. Rutan, *Anal. Chim. Acta* **2013**, *770*, 18.
[83] H. P. Bailey, S. C. Rutan, P. W. Carr, *J. Chromatogr. A* **2011**, *1218*, 8411.
[84] H. P. Bailey, S. C. Rutan, D. R. Stoll, *J. Sep. Sci.* **2012**, *35*, 1837.

[85] E. D. Larson, S. R. Groskreutz, D. C. Harmes, I. C. Gibbs-Hall, S. P. Trudo, R. C. Allen, S. C. Rutan, D. R. Stoll, *Anal. Bioanal. Chem.* **2013**, *405*, 4639.
[86] M. He, Z. Y. Yang, T. B. Yang, Y. Ye, J. Nie, Y. Hu, P. Yan, *J. Chromatogr. B* **2017**, *1052*, 158.
[87] M. Zarghani, H. Parastar, *J. Chromatogr. A* **2017**, *1524*, 188.
[88] N. E. Watson, H. D. Bahaghighat, K. Cui, R. E. Synovec, *Anal. Chem.* **2017**, *89*, 1793.
[89] D. K. Pinkerton, B. A. Parsons, T. J. Anderson, R. E. Synovec, *Anal. Chim. Acta* **2015**, *871*, 66.
[90] A. De Juan, R. Tauler, *J. Chemometr.* **2001**, *15*, 749.
[91] R. Bro, C. A. Andersson, H. A. L. Kiers, *J. Chemometr.* **1999**, *13*, 295.
[92] H. A. L. Kiers, J. M. F. Ten Berge, R. Bro, *J. Chemometr.* **1999**, *13*, 275.
[93] J. M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, *TrAC Trends Anal. Chem.* **2008**, *27*, 714.
[94] A. C. Beckstrom, E. M. Humston, L. R. Snyder, R. E. Synovec, S. E. Juul, *J. Chromatogr. A* **2011**, *1218*, 1899.
[95] N. E. Watson, S. E. Prebihalo, R. E. Synovec, *Anal. Chim. Acta* **2017**, *983*, 67.
[96] M. F. Almstetter, P. J. Oefner, K. Dettmer, *Anal. Bioanal. Chem.* **2012**, *402*, 1993.
[97] L. R. Snyder, J. C. Hoggard, T. J. Montine, R. E. Synovec, *J. Chromatogr. A* **2010**, *1217*, 4639.
[98] A. De Juan, J. Jaumot, R. Tauler, *Anal. Methods* **2014**, *6*, 4964.
[99] D. W. Cook, M. L. Burnham, D. C. Harmes, D. R. Stoll, S. C. Rutan, *Anal. Chim. Acta* **2017**, *961*, 49.
[100] D. W. Cook, S. C. Rutan, D. R. Stoll, P. W. Carr, *Anal. Chim. Acta* **2015**, *859*, 87.
[101] C. Hinz, S. Liggi, J. L. Griffin, *Curr. Opin. Chem. Biol.* **2018**, *42*, 42.
[102] R. A. Harris, J. C. May, C. A. Stinson, Y. Xia, J. A. McLean, *Anal. Chem.* **2018**, *90*, 1915.