Journal of Chromatography B, 910 (2012) 31–45

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

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

Review

Trends in data processing of comprehensive two-dimensional chromatography: State of the art☆

João T.V. Matos, Regina M.B.O. Duarte, Armando C. Duarte*

CESAM & Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal

ARTICLE INFO

Article history:
Received 26 October 2011
Accepted 29 June 2012
Available online 7 July 2012

Keywords:
Comprehensive two-dimensional chromatography
Multidimensional separations
Chemometrics
Peak detection
Resolution
Data treatment

ABSTRACT

The operation of advanced chromatographic systems, namely comprehensive two-dimensional (2D) chromatography coupled to multidimensional detectors, allows achieving a great deal of data that need special care to be processed in order to characterize and quantify as much as possible the analytes under study. The aim of this review is to identify the main trends, research needs and gaps on the techniques for data processing of multidimensional data sets obtained from comprehensive 2D chromatography. The following topics have been identified as the most promising for new developments in the near future: data acquisition and handling, peak detection and quantification, measurement of overlapping of 2D peaks, and data analysis software for 2D chromatography. The rational supporting most of the data processing techniques is based on the generalization of one-dimensional (1D) chromatography although algorithms, such as the inverted watershed algorithm, use the 2D chromatographic data as such. However, for processing more complex N-way data there is a need for using more sophisticated techniques. Apart from using other concepts from 1D chromatography, which have not been tested for 2D chromatography, there is still room for new improvements and developments in algorithms and software for dealing with 2D comprehensive chromatographic data.

Contents

☆ This paper belongs to the Special Issue Chemometrics in Chromatography, Edited by Pedro Araujo and Bjørn Grung.

* Corresponding author. Tel.: +351 234 370200; fax: +351 234 370084.
E-mail address: aduarte@ua.pt (A.C. Duarte).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.06.039

1. Introduction

The development of several one-dimensional (1D) separation techniques, such as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE), led to the generalized idea by the end of the twentieth century, that these techniques could be just finely tuned in order to solve all the practical problems in Analytical Chemistry [1]. However, the need for analysis of increasingly complex samples with a large number of compounds, highlighted the limitations of such techniques, and prompted the development of technologies with a much higher separation capacity in order they could take full advantage of coupling them to advanced detection systems, such as mass spectrometry and nuclear magnetic resonance spectroscopy.

The need for improving the analytical figures of merit associated to the research explosion in proteomics and metabolomics, and the ever increasing requirements for adequate identification and quantification of proteins, glycoproteins and metabolite products has prompted a need to push separation techniques to their limits. Furthermore, even when 1D chromatography could produce acceptable results, they do not have the separation power to deal with complex samples, and their use in such cases would mean spending a lot of time for analysis [1,2]. The obvious response to this lack of separation power of 1D techniques is the development of multidimensional chromatographic systems using two or more independent separation mechanisms.

Multidimensional separation can be understood as a separation system capable of discriminating the components from a mixture, using different separation mechanisms which are connected but do not interact among themselves, that is, they should be completely independent from each other. There are two modes of operation of multidimensional chromatography: heart-cutting and comprehensive. In the heart-cutting mode, only some selected fractions are transferred from the first into the second separation system, and the results become two separate 1D data sets. On the other hand, a separation is comprehensive when the whole sample is subjected to two different separation mechanisms, the separation (resolution) obtained in the first dimension is essentially maintained, and the chromatogram obtained is representative of the entire sample (after pre-treatment), which requires that either no sample goes to waste (everything passes through the detector) or a sufficient number of second-dimension chromatograms are recorded very frequently across the width of a first-dimension peak [3].

As reviewed by Phillips and Beens [4], the comprehensive multidimensional chromatography became more relevant after the development of comprehensive 2D gas chromatography (GC × GC), more than a decade after the development of a first comprehensive 2D liquid chromatography (LC × LC) by Erni and Frei [5]. Although the data sets resulting from GC × GC have received more attention than those resulting from any other comprehensive 2D chromatographic technique, they are formally equivalent and most of the work developed for GC × GC can be applied with small modifications to other chromatographic combinations, such as LC × LC, LC × GC, and LC × CE.

The data collected from advanced chromatographic systems designed for the analysis of complex samples contain huge amounts of information that need complex processing algorithms in order to take advantage of such powerful analytical systems. For instance, analysis of a sample with n replicates in a 2D chromatographic system coupled to a multichannel detector, such as a diode array detector (DAD) or a mass spectrometer (MS) can produce a so-called four-way data set. This terminology can be better understood through a schematic representation of different data sets derived from different types of analyses and orders of instruments, as shown in Fig. 1. The interpretation of these data sets is based on the order of the analytical signal, which was thoroughly discussed in 1994 by Booksh and Kowalski [6]. Fig. 1A represents both a first-order tensor (i.e., a vector) data that changes over the time of the first-dimension and a two-order tensor (i.e., a matrix) of data. While the former can be obtained by a first-order instrument, such as 1D chromatography system coupled to a single channel detector, the two-order tensor of data is derived from a second-order instrument, which is defined as an instrument capable of generating a data set that also changes over time. A 2D chromatographic system coupled to a single channel detector or 1D hyphenated chromatographic techniques (e.g. GC/MS or MS/MS) are good examples of such second-order instruments. It should be mentioned that the first-order tensor of data can also be produced by discarding information from a second-order tensor acquired in a second-order instrument. This is usually performed to build the first-order profile of the data set. The order of the data produced can be further increased if one combines an additional first-order instrument to the second-order instrument. This is depicted in Fig. 1B, which represents a third-order tensor data acquired in a 2D chromatography system coupled to a multichannel detector (third-order instrument). Finally, a four-way data, represented in Fig. 1C, is classified as a hypercube of data obtained by stacking the data from n replicates acquired in a third-order instrument.

The aim of this review is to discuss the state of the art in data processing for multidimensional data sets obtained in different types of 2D chromatography, from the pre-treatment until the quantification of the identified chromatographic peaks. The discussion will lead to the identification of the main trends in data processing of comprehensive 2D chromatography and it will pinpoint the gaps and research needs that should be tackled in this field.

2. Data acquisition and handling in comprehensive 2D chromatography

The massive amount of data generated from the current high-resolution analytical instrumentation requires the use of computerized assistance for data processing and transformation. The 2D chromatography is no exception, and the use of informatics tools has become essential for transforming the raw analytical data into fit for purpose information. The 2D chromatography produces a considerable amount of data in a relatively short time when applied to the separation of complex mixtures. Such an enhancement in performance provides an order-of-magnitude increase in peak capacity, when compared to 1D chromatography.

Handling of 2D chromatography data is a challenging task in Analytical Chemistry. The acquisition of data in real time from 2D chromatography coupled to detectors, such as DAD or MS, generates huge data files, that can reach more than 10 million data points which may lead to considerable problems in storage and processing [2,7]. The greatest challenges lies in producing automatic tools capable of processing and converting the data matrix under useful forms without losing control on the analysis of samples for obtaining raw data, and transformation of data into useful chemical

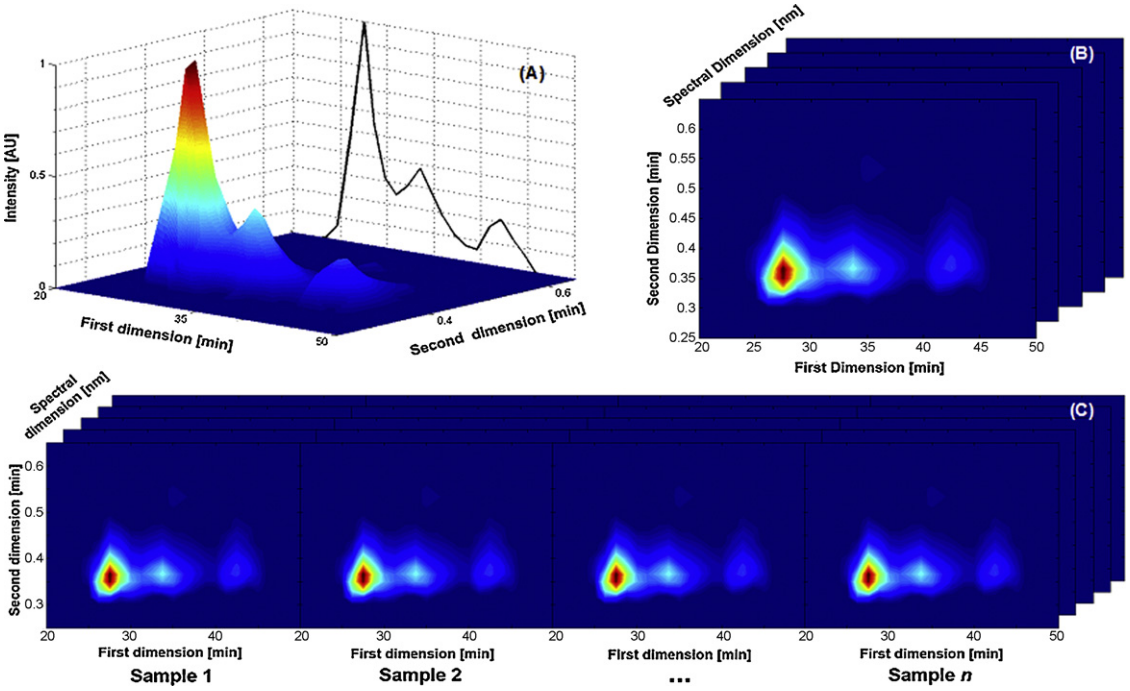

Fig. 1. Representation of (A) 1D chromatography (one-way data) and 2D chromatography (two-way data); (B) 2D chromatography coupled with a multichannel detector (three-way data); and, (C) 2D chromatography coupled with a multichannel detector with sample replicates (four-way data).

information. The scarcity of software available for data acquisition and handling in 2D chromatography, especially in LC × LC, is one of the most significant impediments for a generalized adoption of these separation techniques.

There are three general approaches that can be used to deal with the 2D chromatographic data. The first approach, and since the 2D chromatograms can be considered as a set of consecutive 1D chromatograms, is to deal with those chromatograms individually using all the data treatment tools already extensively developed for 1D chromatography; the second approach consists of dealing with the real nature of the data, applying the methods for second- (or third-) order tensor data, after previously having "folded" the vector of 1D data into a matrix (each column corresponding to a modulation period); a and finally, the third approach is to transform the data in an image file (also after a data modulation step) and follow further image treatment and processing. The first approach can take advantage of a full body of knowledge already designed and extensively studied for 1D chromatography and the availability of a large amount of software ready for use. On the other hand, when dealing with two or higher order data matrix, as it is the case of the second approach, it requires knowledge and expertise on complex chemometric algorithms. This is a developing field, and the evolving work is very promising, especially for dealing with three or higher order dimensional data. Finally, and despite the use of imaging tools to deal with 2D chromatograms may look like a misfit, this third approach has been in fact translated into commercially available software showing excellent results [8–11].

2.1. Data pre-treatment

In a typical experiment of 2D chromatography, the massive amounts of data acquired in both dimensions create large data files, which need to be processed in order to become useful. These data sets are extremely important in chemical analysis, and require the use of computational systems to process and extract the maximum chemical information possible. However, this processing requires different approaches and methodologies depending on the different ends or goals set by the analyst. Therefore, in order to accomplish this task, it becomes necessary a transformation of the data into a more suitable form, in accordance with the aim of the analytical work [2].

In general terms, the most common operations and important methodologies in data processing and pre-treatment in order to become possible the representation, detection and quantification of peaks are, the removal and correction of background, the attenuation of the signal noise, the correction of uncontrolled shifts in retention time, the identification and removal of signal artefacts, and the resolution of overlapping peaks [1,2,8]. The whole set of methodologies are not necessarily applied within each data processing and/or pre-treatment procedure: again, the reader should be aware that some operations are necessary, others are not, depending on the final goal of the analysis.

2.1.1. Modulation and interpolation of data

The first step in data processing of comprehensive 2D chromatography is to extract the experimental data from the detector response and build the corresponding 2D chromatogram. In the first applications of 2D separations, namely the development of LC × LC by Erni and Frei [5], two detectors were used for measuring the analytical signals at the end of each of two chromatographic columns. Nowadays, due to both the comprehensive methodology of the analysis and the available software, the experimental signal is only measured at the end of the second column. Consequently, in order to transform the output of the detector positioned at the end of the second column into a 2D chromatogram, it is necessary to slice the output according to the modulation time, and re-organize each sliced chromatogram along the time axis of the volume separated in the first column in this same modulation time. In order to perform this properly, the number of points acquired by the detector should be such that the number of points per modulation time is an integer number, otherwise interpolation is necessary.

Fig. 2A shows a simulated output from a detector positioned at the end of a 2D separation system, with a modulation time of 2 min. Since the modulation time is known, then the application of the slicing methodology is straightforward, and the resulting Fig. 2B shows the same data set but now under a form of a 2D chromatogram. After

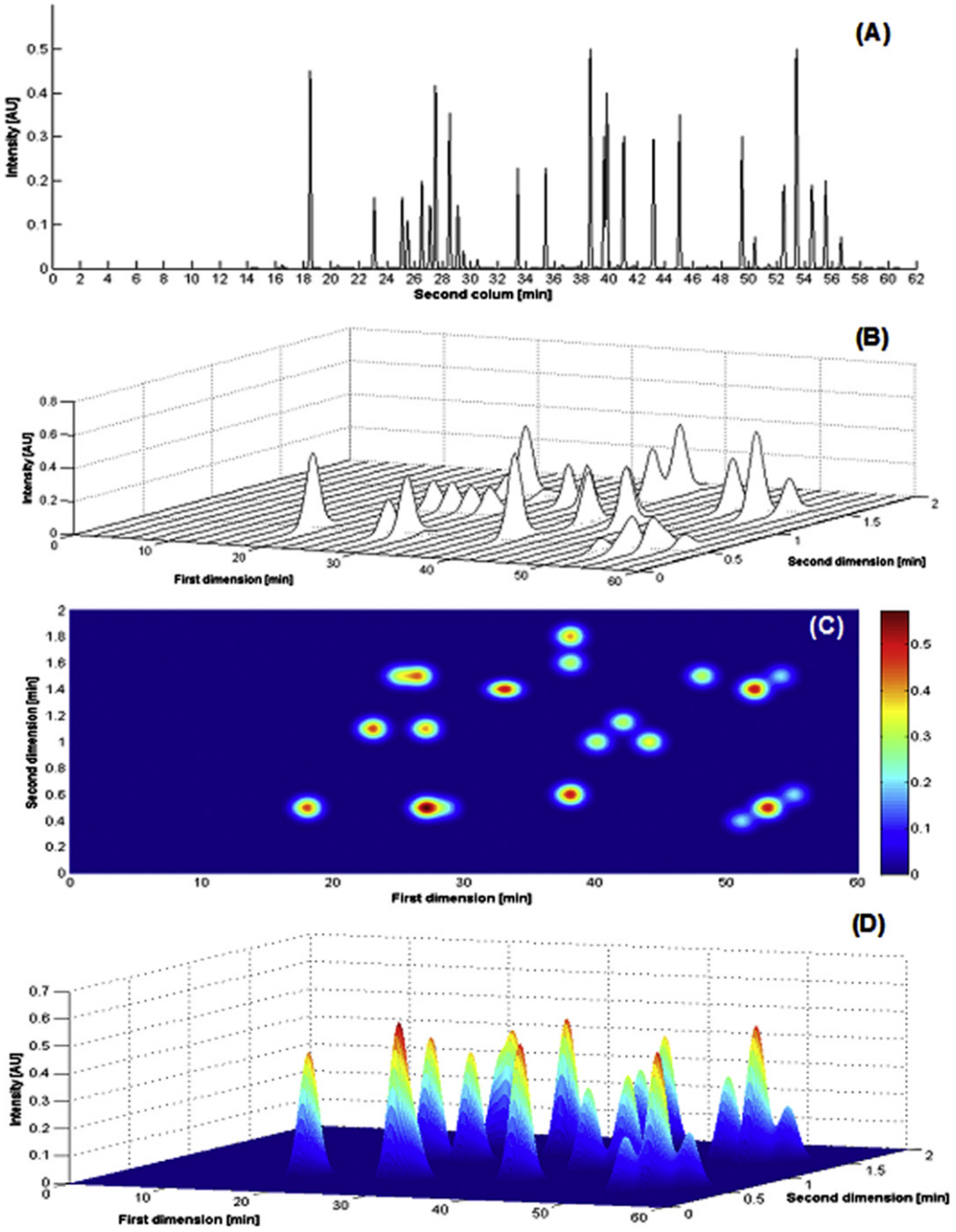

Fig. 2. Representation of a simulated data set for comprehensive 2D liquid chromatography: (A) raw data from the detector positioned at the end of the second column; (B) layout of the sliced 1D chromatograms defined by the modulation time; (C) contour plot of a 2D chromatogram after smoothing; and, (D) 3D plot of a 2D chromatogram after smoothing.

data modulation, it may be necessary to smooth the chromatogram by interpolation of data points between 1D chromatograms mainly along the first dimension in order to obtain a 2D chromatogram, as shown in Fig. 2C and D. Five different interpolation methods were recently compared by Allen and Rutan [12], for the case when producing a sufficient number of data points in the first dimension to allow the alignment of retention times from different injections. These methods, including linear interpolation followed by cross correlation, piecewise cubic Hermite interpolating polynomial, cubic spline, Fourier zero-filling, and Gaussian fitting, performed equally well.

Nowadays, the availability of multichannel detectors coupled to chromatographic systems allows obtaining a range of spectral data instead of a single intensity reading, and therefore the operator can have an additional order of measurement in the data array related to the spectral information. Prior to the slicing of the chromatograms according to the modulation time, the operator faces already a second-order data array that should be taken into account. Therefore, to convert the experimental data into usable chemical information it is necessary, for each wavelength signal, to divide the data according to the modulation time for building a 3D data matrix.

2.1.2. Data representation and visual features

For representation of a 2D data matrix either the contour colour coded plot or the 3D plot are the types of graphs that easily allow

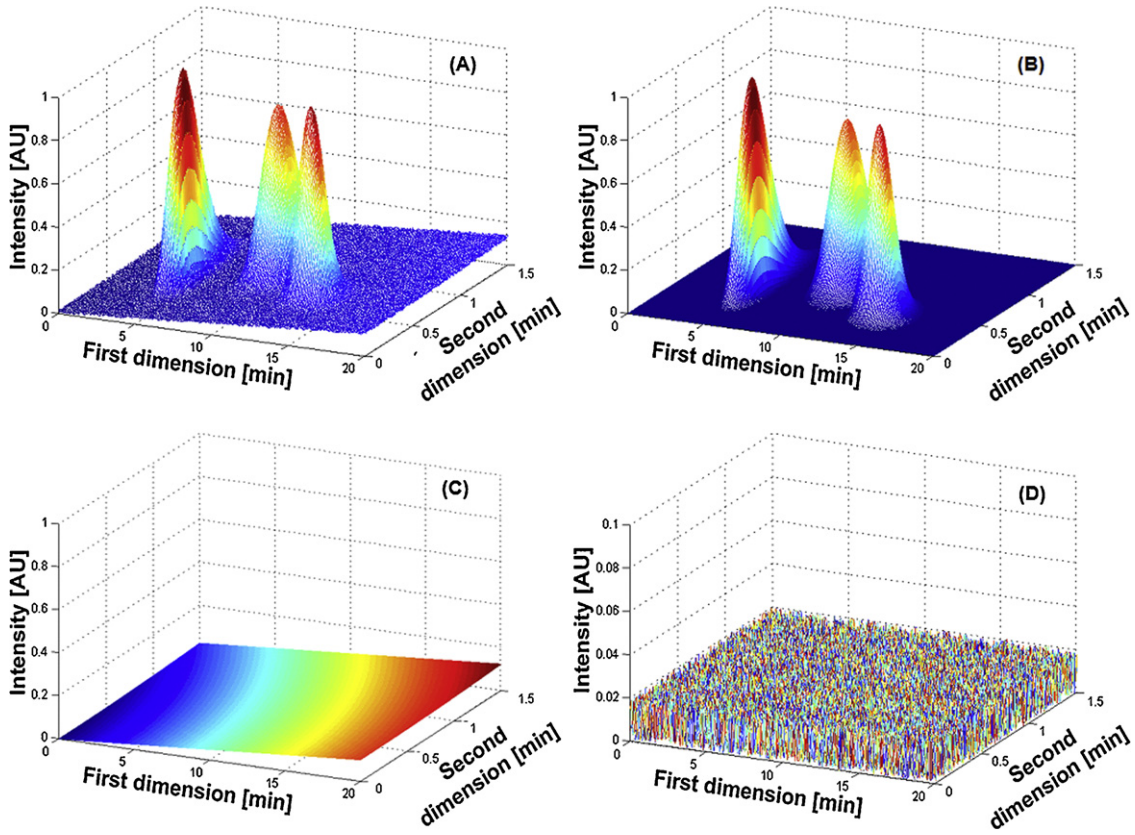

Fig. 3. Representation of a simulated 2D chromatogram and its component parts (A): analytical signal (B), background (C) and noise (D) (visualization inspired by Amigo et al. [14]).

visualizing and understanding the distribution of peaks in a chromatogram. As shown in Fig. 2C, such a graph does not provide quantitative information but it is an excellent visualization tool which becomes the first essential step for further processing of the data obtained. The use of a colour coded scale for peak intensity allows an immediate qualitative assessment of the different peaks and respective heights. Fig. 2D uses the same data as in Fig. 2C for simulating a typical 2D chromatogram, and the 3D plot provides already quantitative information, thus becoming the best approach to observe, analyse, and comment the visual features of 3D peaks shape and form [2,13]. When dealing with higher order data dimensions, it is not possible to produce conventional contours and 3D plots, since a 2D chromatographic system coupled to a multichannel detector will produce a three dimensional (3D) data array.

2.1.3. Background and noise signal

As in any other analytical methodology [14], it is possible to separate the 2D chromatographic experimental data signal (Fig. 3A) in to three major parts: the analytical signal, the background signal, and the noise. The analytical signal, as depicted in Fig. 3B, gives the chromatographic response of the analyte obtained for a particular set of experimental conditions and a particular detector free off any traces of noise and background effects. The background, represented in Fig. 3C, is associated with the systematic response from the chromatographic system not related to the analyte. Finally, the noise, represented in Fig. 3D, is associated with random variations and it is usually related to the sensitivity of the detector.

The background, particularly associated to noisy signal, may cause many problems, especially for quantification, since it can change both the shape and the elution time of peaks. This background drift can be caused by changes in the composition of mobile phase when in gradient mode, variation of pressure and temperature, and fluctuations caused by the injection valve. When the signal-to-noise ratio is low, it becomes difficult in practice to separate between noise and background, but in order to achieve a clear analytical signal with a flat baseline, it is necessary to identify and eliminate the interferences caused by noise and background signal. Although the noise interference can be usually reduced using smoothing algorithms, the strong background drifts are difficult to resolve. One of the easiest ways to deal with the background interference is performing a "mean centering" of the data, but only in the presence of a relatively stable background. Another possible way is to subtract a "blank" chromatographic run obtained in the same instrumental conditions, but even so this strategy may not be very accurate due to possible changes in eluent spectrum and differences in noise between chromatograms. For overcoming the drawbacks of using simple strategies in 2D chromatography, more complex algorithms have been proposed for removing the background signal and noise.

Zeng et al. [15] applied to each 1D peak of a 2D chromatogram (obtained in a GC × GC coupled with a flame ionization detector (FID) system) a baseline correction using a linear interpolation technique. The peak background was linearly simulated, subtracted from the 1D vector, and then a moving windows average was applied for smoothing the noise in the data. On the other hand, Zhang et al. [16] suggested the use of the trilinear decomposition method to remove the 3D background drift in 2D chromatography coupled with multichannel detectors (in this case a LC × LC coupled with a diode array detection (DAD) system). The authors [16] used the background drift and the analytical signal as factors for building a factor model by applying an alternating trilinear decomposition (ATLD) algorithm to the raw dataset in order to extract the background factor, and then subtract it from the raw dataset. The ATLD algorithm showed adequate proprieties of convergence and robustness to the excess of factors used, although others chemometric methods such as parallel factor analysis (PARAFAC),

self-weighted alternating trilinear decomposition (SWATLD) and alternating penalty trilinear decomposition (APTLD) algorithms can be used with similar results [16]. Finally, this methodology removes the background drift without the loss of peak information in entire spectral region and without the need for replicates and "blank" chromatograms.

An alternative technique is the use of image treatment software to estimate and remove the background from images files of 2D chromatograms, as suggested and applied by Reichenbach et al. [10] for GC × GC. This algorithm takes advantage of the following structural and statistical properties of the background from the images of 2D chromatograms: dead-bands that are regions without analytical signal; the mean of background level does not change much when compared to the characteristic peak widths; and, the noise present has the same statistical properties of the random white noise. This background-removal algorithm has been included in the GC Image and LC Image software packages, which has been reported in several works [8–11].

2.1.4. Correction of shifts in retention time of peaks

An important steps in any chromatography, either 1D or multidimensional, is to ensure the precision in the determination of retention time for each and every peak. These deviations are often observed in the chromatographic analysis, and can be easily identified by the comparison to the patterns of variation between replicates or standards. Due to the high sensitivity of the analytical methods, a poor precision of the retention times, when not corrected, can generate enormous deviations in the most of the chromatographic detection and quantification techniques actually used.

The fluctuations in retention time of peaks are always present in 2D chromatographic systems and can be originated from variations in temperature and pressure, degradation of the stationary phase, and matrix effects. For data processing to be successful there is a need to ensure that the retention times between replicates are repeatable and reproducible, and the axes are synchronized to prevent slice-to-slice misalignments and to ensure that peaks are properly aligned. Although there are data treatment methodologies and techniques that do not need to correct these deviations, this correction is a critical factor in data treatments and algorithms that relies upon the bilinearity or trilinearity of the data (such as, GRAM and PARAFAC).

The alignment between replicas/samples or warping can be accomplished by using different algorithms. Fraga et al. [17] suggested an alignment technique of the retention times, based on minimizing residuals in the generalized rank annihilation method (GRAM) which was addressed by Prazen et al. [18] for second-order hyphenated chromatography, namely GC–MS. On the other hand, van Mispelaar et al. [19] suggested a correlation-optimized shifting, based in an inner-product correlation associated with selected regions of the GC × GC data. This algorithm uses a 2D chromatogram reference to align all selections and as this alignment is performed, the inner-product correlation is calculated in order to identify the best-fit position. Johnson et al. [20] described an algorithm based on windowed rank minimization alignment with interpolative stretching between the windows. In this work [20] this algorithm was used to deal with shifted GC × GC retention times in quantification of naphthalene in jet fuel and produced better results than in the case where quantification of the chromatograms was not pre-aligned. Pierce et al. [21] reported the application of a comprehensive 2D retention time alignment algorithm that allows a warping in both chromatographic dimensions using a novel indexing scheme, and preserves the separation information in both dimensions. The algorithm was applied by Pierce et al. [21] to GC × GC but it can be applied to any 2D separation system with a gain on the retention time precision and also restoring the linearity of the data without losing quantitative information. Zhang et al. [22] developed a 2D Correlation Optimized Warping Algorithm, (2D COW) to align data obtained from 2D gas chromatography coupled with time-of-flight mass spectrometry (GC × GC–TOFMS). This powerful and flexible algorithm stretches and compresses a local sample segment of the 2D chromatogram to maximize the correlation from the sample relatively to the 2D chromatogram reference. Such calculations allow interpolating the warp non-grid points from the shifted grid points in order to align the chromatograms. When using image based software, the method for automatically aligning chromatograms developed by Hollingsworth et al. [23] can be applied, as in the case of Nelson et al. [24] and Wardlaw et al. [25] to study the weathering of an oil spill and oil seepage, and Cordero et al. [26] to compare coffee samples. Unfortunately, all of these methodologies and algorithms cannot deal with three- or higher-orders data structures, as those obtained from 2D chromatography coupled with multi-channel detectors. As the degree of hyphenation of several detectors to 2D chromatography increases, the more urgent becomes the search for more sophisticated techniques for correction of shifts in retention times of peaks. Recently, Allen and Rutan [12] developed an algorithm especially suited to LC × LC-DAD that allows dealing with four-way data with satisfactory results.

On the other hand, as already written slice-to-slice misalignments can occur. This time shifts occur between the subsequent second dimension chromatograms within a single 2D run preventing the linearity between the measuring orders. Ideally, the same peak in consecutive fractions should elute always at the same retention time, but sometimes such synchronization may not happen due to several operational variables, namely in GC × GC, such as control and timing of cooling and heating programmes, non-linearity of distribution of isotherms, and non-instantaneous reinjections into the second column [27]. These deviations between the transferred fractions can cause deformations in the shape of the 2D peaks or even give incorrect information about the number of peaks in case of more accentuated deviations. Skov et al. [28] developed a pre-treatment algorithm that uses cross-correlation methodologies to align shifted fractions. These authors handled retention time shifts in trilinear data structures from a GC × GC–TOFMS by comparing a standard retention time shift correction followed by PARAFAC with a PARAFAC2 algorithm, a "relaxed but powerful version of PARAFAC" as stated by the authors. Although the PARAFAC2 showed great potential, the shift correction followed by PARAFAC was found to be more robust at lower signal-to-noise ratio. Recently, Parastar et al. [29] proposed a new bilinear peak alignment (BPA) method, based on multivariate curve resolution, to correct this same within run retention time shifts in GC × GC coupled to a single channel detector (bilinear structure).

Another issue that it is necessary to take into account in 2D chromatography is the different time scale between the second dimension and the modulation period. Since the 2D chromatogram is generated from a 1D signal array, it is fundamental to make sure that there are not 2D peaks "wrapped-around" [30], i.e., the maximum range of a peak retention time in the second dimension is lower than the modulation period. One way to deal with this problem, which is quite common in chromatography of complex mixtures, is an algorithm capable of finding the absolute retention times in these cases as suggested by Micyus et al. [30]. In this algorithm, after the detection of "wrapped-around" 2D peaks, a series of chromatograms are reanalysed by an integer fraction of original modulation period and shifts of the retention of the second dimension are used to determine the absolute retention times. Another way, recently developed and presented by Weusten et al. [31], is to deal with the 2D chromatogram as a surface of a 3D cylinder. This cylindrical transformation, was tested in a set of 11 replicates of a human urine sample analysed by a GC × GC–MS, with good results.

3. Peak detection in comprehensive 2D chromatography

After the data pre-treatment, probably the most critical step in the analysis of the 2D chromatograms is the detection of the 2D peak. In the past decade, many authors dedicated to develop algorithms capable of detecting 2D peaks, especially for applications in GC × GC. A large part of these methodologies are direct consequence of the already developed algorithms for 1D chromatography, since in first instance the 2D dimensional separations can be understood as upgrading the classical column chromatography by just adding another column as a second dimension. However, with the evolution on the concept of 2D chromatography associated to the increasing use of multichannel detectors and the consequent production of huge amount of data, it is foreseeable an increase of methodologies and algorithms that can deal with these data more efficiently in a global manner rather than adding up 1D chromatograms.

Conceptually, the peak detection algorithms can have two main purposes: in non-targeted analysis, the algorithms should be capable of identifying 2D peaks in complex mixtures, without any pre-information from de sample (e.g., two-step peak detection algorithm [27], and the watershed algorithm [8]); in targeted analysis, the algorithms need a complete or a regional 2D chromatographic reference, and the peak detection is made by comparison with this standard (e.g., the GRAM method [32]). Several authors, who have developed algorithms for 2D peak detection, have also developed methods of quantification, since most of the analytical work involves not only the screening of peaks but also their quantification. Finally, it is important a word of caution regarding the choice of an algorithm because it may have to take into account several factors such as speed of analysis, accuracy of results, and components of interest, depending on the specific work involved.

3.1. Two-step peak detection algorithm

One of the most important contributions to the actual state of the art in 2D chromatography has been produced by Peters et al. [27,33], namely when suggesting new methodologies and important concepts for detection and resolution of peaks in GC × GC. Some of these methodologies, in special case for peak detection, have already been applied in LC × LC data sets [34] and the remaining can also be applied in principle to LC × LC with minor modifications, if any at all.

For the specific case of peak detection, Peters et al. [27] developed a detection algorithm that deals with this problem in two main steps: firstly, already known methods developed for peak detection in 1D chromatography are applied to detect the peaks one dimension at the time; secondly, some criteria are used to decide which peaks of the first dimension correspond to the same compound in the second dimension, that is, some criteria are set up for the process of merging the peaks from the first dimension with the peaks of the second dimension that are all produced by the same compound. It should be mentioned that this algorithm assumes non-bilinearity of the data, and consequently it is necessary to detect all 1D and merge them in 2D peaks. If the chromatographic data is bilinear, it is possible to use the second partial derivatives of the 2D matrix to identify the peaks, as suggested by Duarte et al. [35].

The raw data acquired in comprehensive 2D chromatography is constituted by a set of 1D chromatograms, each corresponding to a single injection in the second column. Once the obtained data has been organized in a 2D matrix, it becomes possible to analyse all chromatograms using 1D peak detection techniques. The two-step algorithm [27] detects 1D peaks based on the properties of the derivatives of the peaks computed by the Savitzky–Golay method [36]. From the original chromatogram and from the first- and second-order derivatives, it is possible to characterize the properties of chromatographic peaks, namely the height, and the peak starting-point and end-point (peak region). As shown in Fig. 4A, the starting-point and the end-point of a 1D peak correspond to the first and last point above zero, respectively, in the first-order derivative (Fig. 4B). Although, in practice the first derivative often does not reach zero, such problem can be overcome by defining a minimal value to quantify [27]. On the other hand, when the value of zero for the first-order derivative coincides with the minimum of the second-order derivative (Fig. 4C), then it means that the original chromatographic peak has reached the maximum value.

After defining the peaks in the previous step, it is possible to apply an algorithm following some pre-defined criteria, in order to merge the 1D peaks in the two dimensions. This algorithm produces a cluster with a collection of 1D peaks in the consecutive chromatograms that belong to the peak of the same compound [27]. The process of peak merging starts off with the first peak appearing in the first chromatogram obtained in the second-dimension and the attempt of finding all the merging combinations with all the peaks found in the subsequent chromatograms of the second-dimension, since these can be misaligned due to retention time shifts between the transferred fractions. Then, the overlap criterion and the unimodality criterion are applied to test whether it is possible to merge these peaks. Finally, the procedure repeats itself until all the merging combinations are tested [27].

For checking whether there is any overlap of the peak region (where the peak starts and ends) in the a chromatogram, the overlap criterion examines the degree of overlap of two peak regions, one from the existent 2D cluster (peak A in Fig. 5) and the other from the candidate to be merged (peak B in Fig. 5). The ratio of overlap, in percentage, is calculated according to the following equation, as suggested by Peters et al. [27]:

$$OV = \frac{b}{a} \times 100\% \quad (1)$$

where b is the length of the region where the two peaks overlap and a is the length of the peak region of peak A (Fig. 5).

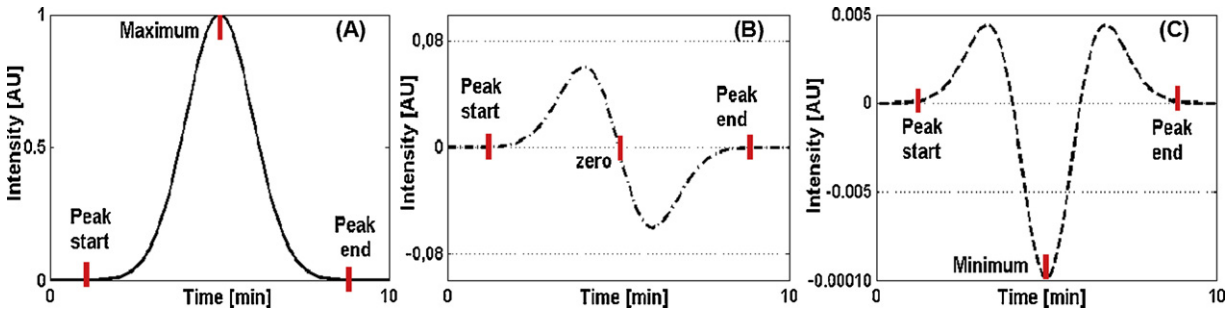

Fig. 4. Properties of a chromatographic peak assumed to be Gaussian (A), its first-order derivative (B), and its second-order derivative (C).

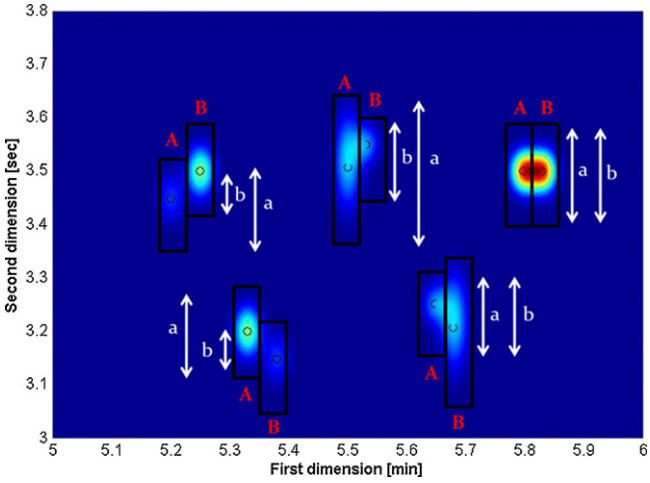

Fig. 5. Schematic representation of peak regions of adjacent 1D peaks for different scenarios in a chromatographic map (visualization inspired by Peters et al. [27]).

After that, it is necessary to define a threshold for the acceptability of the candidate peak. If the ratio of overlap is greater than this threshold than the peak can be subjected to the next criterion; otherwise, the peak cannot be merged and the algorithm continues to the next candidate [27].

The basis for the application of unimodality criterion, as suggested by Peters et al. [27], is the analysis of the peak-maxima profile in the first chromatographic dimension, between the peaks checked for merging in to 2D clusters. As an example, Fig. 6A shows the peak maxima profile for the 2D chromatogram shown in Fig. 6B. The observation of Fig. 6A, allows concluding that there are two maxima, peak A and peak D, since each 2D cluster can show only one maximum.

Finally, it is necessary to take into special consideration the handling of the closest maxima, that is, when more than one peak has been identified at the same retention time in first dimension. In such case, it becomes necessary to compare the retention time of the candidates of the second dimension, and the candidate peak with the retention time closest to the last peak of the 2D cluster is then the peak to be merged.

After the application of the two-step algorithm, Peters et al. [27] also describe an integration tool for the quantification of target 2D peaks: the 1D peaks that were used to build the target 2D peak, are integrated using a trapezoidal method and then summed in order to achieve the quantification of target analytes. The methodology based on summation of second dimension chromatograms has been also used by Pól et al. [37] and Kivilompolo and Hyötyläinen [38] for the quantification analysis of LC × LC data. A slightly different approach based on peak volume calculation for 2D contour plots has been suggested by Kivilompolo et al. [39].

In order to overcome the problem related to overlapped peaks, Peters et al. [27] also suggested two alternatives: (a) to integrate the area under the curve from the peak start to a perpendicular line that splits the overlapped peaks in the valley point; and (b) to subtract the area under the line connecting the peak start to the valley point from the total area under the curve, as above mentioned. This method is easy to implement, and do not require any user input, which makes it ideal for quantification of complex mixtures separated by 2D chromatography. However, for chemometric resolution and quantification of four-way data, Bailey and Rutan [40] showed the need for more sophisticated algorithms (Section 3.3.3). In this regard, Vivó-Truyols [41] have recently proposed an improvement of the two-steps algorithm which can set a path for the development of a methodology capable of handling four-way data. This algorithm, already applied and tested for LC × LC and GC × GC data, uses a Bayesian approach to perform the merging step, and it can be an extremely powerful tool since it follows the chromatographer's intuition by informing about the probability of other peak configurations, as the correct arrangement may not be the most probable one.

3.2. Inverted watershed algorithm

A completely different approach for peak detection in 2D chromatography is the use of software able to deal with the 2D data set using image analysis tools to extract the required information. To perform this task, Reichenbach et al. [8] suggests the use of an inverted version of the watershed algorithm, also known as the drain algorithm. This algorithm, which can also be used as a topographical tool, assumes that the surface of analysis has its highest point at a "mountain". When it "rains" in the "mountain", the water tends to go down throughout the cliff, and producing puddles of water surrounding the various "mountains". This algorithm is capable of foreseeing the movements of this rain and consequently delineates the peaks from an image [8]. Conceptually, the algorithm finds the highest peak and after that it will identify the neighbouring pixels until reaching the background [42]. After the detection is complete, each 2D peak identified is, in fact, a group of pixels. Unfortunately, the inverted watershed algorithm is very sensitive to noise and artefacts, which can cause a multi-peak detection in a single peak situation. However the application of smoothing tools to the image of the 2D chromatogram may reduce the effects of these artefacts.

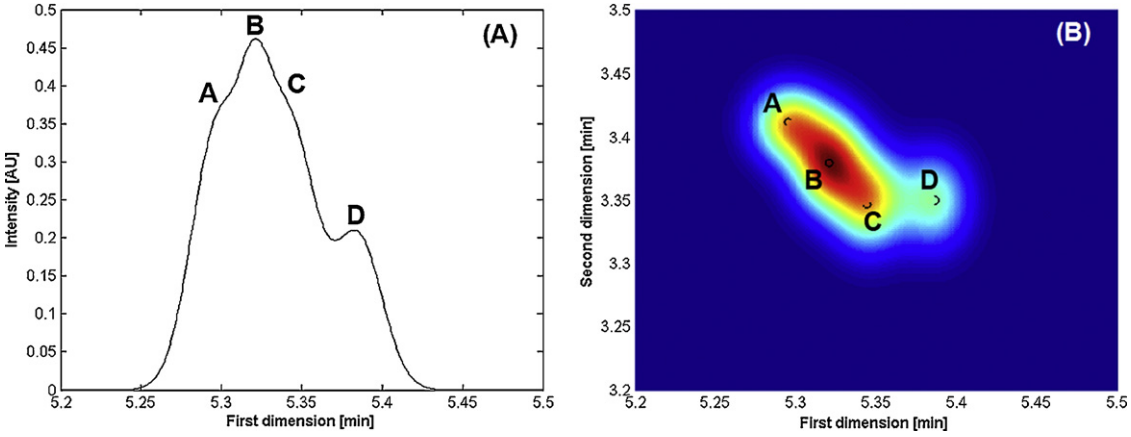

Fig. 6. Representation of a peak-maxima profile (A) and its corresponding 2D chromatogram after merging (B) (visualization inspired by Peters et al. [27]).

Once the peak detection is accomplished, then the statistical proprieties of the peak, such as the number of pixels (area of the peak) and the sum of pixels (volume of the peak), can be used to measure the characteristics of the 2D peaks, such as symmetry, orientation or eccentricity and also to obtain quantitative information. Finally, Reichenbach et al. [8,9] suggest the use of an algorithm for chemical identification by template matching. This algorithm uses the statistical proprieties of the peak to compare the sample peaks with statistical proprieties of several peak templates in order to find a match. While this algorithm can be used for peak detection and integration in 2D chromatography, a study carried out by Vivó-Truyols and Janssen [43] has shown that the probability of failure of this algorithm under normal conditions in GC × GC is around 15–20% when compared with the two-steps algorithm suggested by Peters et al. [27]. Such a high percentage value could be a discouraging factor for a wider use of such method in Analytical Chemistry. However, Latha et al. [42] tested 1000 cases for each set of parameter values, in order to compare the detection performance in both algorithms after a skew correction for various parameters, such as, different levels of noise, peak widths, and retention-time. The experiments allowed concluding that after a skew correction, the inverted watershed algorithm showed better results than the two-step algorithm. Furthermore, the accuracy of both algorithms decreases as the peak width and the noise increase, even after shift correction. Therefore, there is a need for improving the noise suppression techniques in order to achieve better results in both detection algorithms.

3.3. Multi-way chemometric methodologies

The increasing hyphenation of multi-array detectors such as DAD and MS detector to chromatographic systems has brought the need for developing and/or adapting more sophisticated techniques, even for 1D chromatographic systems, in order to identify and resolve the overlapped peaks in the resulting two or higher order data structures [14]. One of the legacies of the 1D chromatography to deal with those types of data structures is the adoption of chemometric methods for the discrimination of samples, depending on the patterns of both the chromatographic retention times and the spectral characteristics [7]. There is a general trend to use more and more complex chemometric algorithms to extract information from the experimental data sets that can attain very high order of instruments and contain huge amount of data points, namely when using multichannel detectors and sample replicates. Many of these methods and their variants have already been successfully applied to 2D chromatography (both LC × LC and GC × GC); as such are the cases of the generalized rank annihilation method (GRAM), parallel factor analysis-alternating least squares (PARAFAC-ALS), and multivariate curve resolution-alternating least squares (MCR-ALS).

This family of chemometric deconvolution techniques are already widely used to deal with overlapped signals from data acquired by 2D chromatography coupled to various detectors in quantitative analysis of complex samples. Their application may not be straightforward, because they are based on advanced mathematical concepts and, therefore, some caution should be exerted before direct application of such tools from the chemometric field. To ensure a good application of these models, it is necessary to guarantee that the detector gives always a linear response to the variation of concentration and there are no shifts in the retention time of the 2D peaks.

3.3.1. Parallel factor analysis model

The PARAllel FACtor analysis (PARAFAC) model has its origins in psychometrics sciences, but it has long been exploited in chemometrics and related areas to deal with the increased complexity of the multi-way data sets. This is an iterative and powerful method that has already been proven useful in deconvolution and quantification in 2D-chromatography [7,17,19,42,44–55]. The PARAFAC model applied to a three-way data array can be described as follows:

$$R = \sum_{n=1}^{N} x_n \otimes y_n \otimes z_n + E \quad (2)$$

where R, in 2D chromatography, is the instrumental response matrix, x_n is the second dimension chromatographic profile of each factor (N), y_n is the first dimension chromatographic profile of each factor (N), z_n is the detector response for each factor (N), and E is the error matrix with same size of R matrix [45].

Several algorithms have been described for finding the parameters x_n, y_n, and z_n and consequently fitting the PARAFAC model. These algorithms can be classified in three groups [56]: (a) non-iterative eigenvalue-based methods, such as Generalised Rank Annihilation method (GRAM) and the Direct Trilinear Decomposition method (DTLD); (b) alternating algorithms, such as the alternating least squares (PARAFAC-ALS) and the Self Weighted Alternating Trilinear Decomposition (SWATLD); (c) derivative-based methods such as Positive Matrix Factorisation for 3-way arrays (PMF3) and damped Gauss–Newton (dGN). GRAM and PARAFAC-ALS have been the most extensively discussed and applied to 2D chromatography.

3.3.1.1. The generalized rank annihilation method (GRAM). The GRAM has been developed by Sanchez and Kowalski [32] from others rank annihilation methods in order to deal with the analytical problem of detecting and quantifying one or a few components of interest from a complex mixture without the need for resolving the rest of the sample components. In practical terms, GRAM is a non-iterative eigenvalue-based method which needs two chromatograms: the sample 2D chromatogram and a 2D chromatogram obtained from one or more components of interest at well-defined concentrations. After that GRAM compares both 2D matrices in order to provide the pure elution profiles as well the relative concentration of the analyte in the sample [50,53].

The comparisons of chromatograms performed by GRAM require that both 2D matrices must be stacked to generate a three-way data [44]. Consequently, it is necessary that both matrices have the same size for each row from the sample matrix corresponding to the second dimension in the calibration matrix, and also for each column from the sample matrix corresponding to the first dimension in the calibration matrix [49]. For this reason, this method only supports the analysis of bilinear samples signals (signal represented by the product of two vectors) [44] and consequently does not allow the analysis of data structures of higher dimensions.

The GRAM was the first deconvolution method used in 2D separations and in spite of being widely used, both in GC × GC and LC × LC [17,44,49–55], it is necessary to ensure that the peaks associated with the components of interest have the same retention time and the same peak profile in both the sample and standard 2D chromatograms. Otherwise, the comparison of results through GRAM will be meaningless. Although the non-interactivity of the method justifies the speed of the calculation performed for reaching the results calculation, they tend to be worse than the results given by more interactive methods such as the PARAFAC-ALS.

3.3.1.2. The parallel factor analysis-alternating least squares (PARAFAC-ALS). The alternating least squares (ALS) was the first algorithm used to fit the PARAFAC model, since it is able to handle unresolved chemical components in three-way or higher-order data arrays [57]. PARAFAC-ALS, which has been already used in GC × GC and LC × LC [7,19,44–48], is not a completely automatic

method such as GRAM. Before fitting the model, it requires inputs from the user in terms of initial factor estimative and the definition of some constraints. However, selecting the appropriate number of factors in the PARAFAC model can be a real hard task to achieve. Usually, this number is given by the sum of interferences and the analytes present in the 2D chromatogram, but it is very difficult to forecast a priori this number for complex mixtures, particularly in the presence of either a low signal-to-noise ratio or overlapping peaks [45]. In order to deal with this limitation, Hoggard and Synovec [45] suggested an algorithm capable of automatically selecting the number of factors to be used in a PARAFAC model applied to comprehensive 2D separations using multichannel spectral detection, in this case a GC × GC–TOFMS.

Furthermore, there are non-negative constraints since all chromatographic signal are positive in relation to the mobile phase, and also the unimodality because chromatographic peak shapes are expected to be unimodal, that is, they usually are Gaussian-like with some fronting and/or tailing.

The ALS algorithm attempts, in each iteration, to improve the fitting of the PARAFAC model until reaching the global minimum which is the least-squares solution. This iteration process is one of the best qualities of the algorithm, but leads to one of its greatest drawbacks: the time spent on number crunching, especially with high number of variables [57]. One of the greatest advantages of PARAFAC-ALS method over the GRAM method is its capacity to resolve and quantify the chemical components of interest, taking into account only the sample information without the need for a standard chromatogram or multiple replicates [7]. The application of the PARAFAC-ALS method for resolution and quantification of targeted analytes from a four-way data array obtained by LC × LC-DAD has already been reported by Porter et al. [7]. Finally, the quantification can be accomplished by summing the outer product of the different dimensions of the chosen factor in the corresponding PARAFAC model across all points, thus yielding a scalar. This has already been shown by Hoggard and Synovec [45] in a PARAFAC of target analytes in GC × GC–TOFMS data.

3.3.2. Target finder algorithms

Since in most cases the algorithms based on the PARAFAC model consume a significant amount of computing resources, there have been several attempts to find algorithms capable of finding targeted components in the 2D chromatographic data set. These algorithms have, as the main feature, a process of screening a 2D peak to find the compound of interest in a short period of time. After the compound of interest has been found, then the chromatographic sub-region of this compound can be subjected to an algorithm, such as the PARAFAC-ALS, in order to deconvolute the pure component signals to confirm the result and produce quantitative information.

3.3.2.1. Window target testing factor analysis (WTTFA). The proliferation of 2D chromatographic techniques coupled with multi-channel detectors has led to the development of techniques of data analysis and algorithms robust enough to deal with very large multi-dimensional data arrays. Porter et al. [7] suggested the application of an algorithm capable of a fast qualitative screening of metabolites in LC × LC-DAD data sets using the window target testing factor analysis (WTTFA) to confirm the presence or absence of an analyte. The WTTFA algorithm used for data analysis was proposed by Lohnes et al. [58], and it is based on a search for a region in the sample's chromatogram that closely resembles known standard spectra of compounds, producing then the retention time if a match is found. The WTTFA algorithm starts to perform a singular value decomposition in a small retention time window of a 1D chromatogram. The resultant spectral matrix is truncated according to the user input value of the maximum number of components estimated in the window. After that, spectra of the standard compounds are projected into the subspace described by the analysed spectra and the correlation is calculated. Finally, after the previous steps have been accomplished, the window is incremented by one time unit, and the algorithm repeats this procedure for all windows until all possibilities are tested [7].

3.3.2.2. DotMap algorithm. The DotMap algorithm has been developed by Sinha et al. [59] for identifying spectra in GC × GC–TOFMS complex data matrices which are similar to the target spectra of interest. Firstly, there is a pre-treatment of the spectra data of the compound under study and also of the 2D chromatographic data: scaling, weighting and normalization of the data, besides correction of the baseline. The spectral information of the compound under study can be previously acquired from standards or from available libraries of spectra, such as NIST02. Afterwards, the algorithm computes the dot product "·" of the mass spectrum of the compound of interest with each mass spectrum point from the complete or partial 2D chromatogram, as follows:

$$\frac{m\sqrt{A_d}}{\sum m\sqrt{A_d}} \cdot \frac{m\sqrt{A_u}}{\sum m\sqrt{A_u}} \tag{3}$$

where A_d is the abundance of m/z signals at each point in the 2D chromatogram, A_u is the abundance of m/z signals of the compound of interest, and m is the vector containing m/z values used for weighting the signals.

The dot product result is then compared with a threshold value defined by 90% of the maximum dot product above the median of all dot products in the raw data, and a contour plot is generated with the location of the results above this threshold and the maximum value is extracted [59]. Finally, the data extracted is checked against a traditional mass spectra library in order to verify that the analysis has been well performed. This same algorithm has been evaluated by Hope et al. [60] for locating analytes of interest based on mass spectral similarity in data collected using GC × GC–TOFMS.

3.3.3. Multivariate curve resolution with alternating least squares (MCR-ALS)

Bailey and Rutan [40] have developed another methodology to deal with data from 2D chromatography associated to multichannel detection applied to complex mixtures. In this work, urine samples were analysed in a LC × LC-DAD system with replicates, producing a four-way data set. The aim of the work was to resolve and quantify the non-targeted overlapped compounds. In order to extract the maximum information from the complex data samples, Bailey and Rutan [40] developed a method that combines an iterative key set factor analysis (IKSFA) technique with an "in-house" MCR-ALS algorithm with a spectral selectivity constraint. Firstly, the 2D chromatograms are divided in sections due to the complexity of the whole sample and also due to regions of detector saturation. In the section to be analysed, the IKSFA starts by the determining the number of unique spectra and estimates the spectral initial guess for the next step. After that, the MCR-ALS is then applied in order to provide a resolution of the spectral different components, using non-negativity and spectral selectivity constraints. Finally, the relative concentrations are found using manual baseline integration and the %RSD values are determined by comparison with standard mixtures and control samples.

Bailey and Rutan [40] also highlighted that this algorithm does not assume multilinearity of data, which means that it can be applied when significant retention time deviations occur between samples, in both chromatographic dimensions. The algorithm is insensitive to shifts of retention time and distortions of peak shape, and consequently does not require pre-alignment of the data before application. The use of the non-negativity and selectivity constraints is enough to obtain results fit for purpose. According to

the authors [40] this lack of multilinearity even makes this algorithm more precise than the PARAFAC-ALS algorithm used by Porter et al. [7]. Finally, the main drawback of this method is its lack of full automation, since it requires some user intervention [12,40]. However, according to the authors [40], this intervention is easy to perform and very fast to accomplish.

Recently an application of a similar methodology has been proposed by Parastar et al. [61], where the MCR-ALS algorithm has been used to resolve and quantify a complex mixture of polycyclic aromatic hydrocarbons in Heavy Fuel Oil Sample by GC × GC–TOFMS. In this study, the results have been compared to those obtained by commercial software and to PARAFAC, showing an improvement of the results in terms of data fitting, elution process description, concentration relative errors, and relative standard deviations.

4. From 1D to 2D: an extension of the concept of resolution

Apart from the concepts already discussed for the pretreatment, detection and quantification of peaks, there are some other concepts from the 1D chromatography that could be extended to multidimensional chromatography, namely for the case of 2D chromatography, such as the measurement of peak overlap and resolution. The extension of these concepts to 2D chromatography implies the development of other concepts, such as the measurement of 2D retention time, and the concept of peak vicinity.

4.1. Retention time in 2D chromatography

The major obvious difference between 1D and 2D chromatography is the existence of a second dimension, and the first difficulty to overcome in 2D chromatography is how to deal simultaneously with the two dimensions. One way to solve this problem is to work with Euclidean distances. This metric, based in the Pythagoras theorem, finds the distance between two points considering that this distance is the hypotenuse of a right triangle whose sides are the X and Y coordinates. In practical terms, as shown in Fig. 7, the distance d_{0-1} between the origin and the maximum of peak 1 is given by:

$$d_{0-1} = \sqrt{(X_1)^2 + (Y_1)^2} \quad (4)$$

where X_1 is the retention time of the peak 1 in the first dimension and Y_1 is retention time of the peak 1 in the second dimension. With the use of Euclidean distance it is possible to replace the two chromatographic retention times of the 2D peak by one single metric, as in the case of 1D chromatography.

4.2. The concept of peak vicinity

In 1D chromatography, the distribution of peaks occurs only along one time axis and, consequently, it allows the existence of only two neighbouring peaks for each target peak: the peaks eluted just before and immediately after. In this sense, there are only two peaks that can have any degree of overlapping with the target peak. On the other hand, in 2D chromatography, the peaks are spread all over a surface defined by the two time axis, and there may be several neighbouring peaks surrounding a target peak [2,33].

For the purpose of calculating the resolution between peaks, Peters et al. [33] suggested that the only concern should be the study of the interaction between two consecutives peaks, thus concluding that resolution is meaningful only if no other interfering peak is in between the two peaks of interest. Therefore, prior to the resolution measurements, it becomes necessary to find the "peak vicinity", and for that it is necessary to define the 2D peak regions for all the peaks of the chromatogram. This region is composed of all regions of the 1D peaks, as defined by the peak detection algorithm of Peters et al. [27], and used for merging the 1D into 2D peaks. In order to advance in the definition of "peak vicinity", that is, to establish which peaks share the same neighbouring effect, Peters et al. [33] suggested that two peaks are neighbours if after plotting the non-interpolated trajectory profile line between two sets of 2D peak clusters, there are no other peak region crossing that trajectory. In this case, since the profile line does not cross any other peak region, the vicinity of the two peaks is validated and the resolution can be measured, as shown in Fig. 8A. On the other hand, as shown in Fig. 8B, if another peak region (peak 2) has been found crossing the trajectory line between the two peaks (peak 1 and peak 3) then the vicinity concept between these two peaks does not apply.

As highlighted by Peters et al. [33], a third case of interference can be found when between two peak clusters there is some profile trajectory line that crosses a peak region although the others do not. Fig. 8C is an example of this third case, where there are two profile lines (represented in white) between peaks 1 and 3 that cross the region of peak 2, although the other trajectory lines (represented in black) do not. In this third case, peaks 1 and 2 are considered as neighbours, and peak 2, in practical terms, is not considered as an interfering peak. The number of interfering peaks in 2D chromatography of complex mixtures can be extremely large, and Peters et al. [33] suggested the use of a threshold value below which such interfering peaks may be or should be neglected. Therefore, only the peaks of intensity higher than this threshold set above the background are considered for purpose of quantification, which represents a limitation for low-concentration compounds.

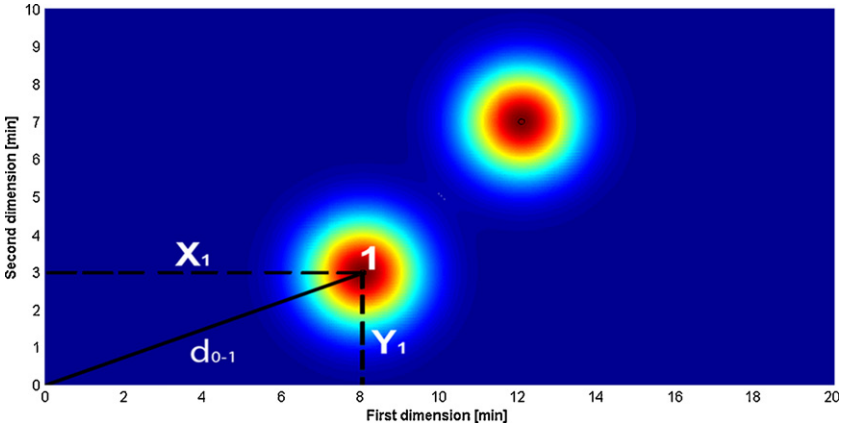

Fig. 7. Determination of Euclidean distance for peak 1 in a 2D chromatographic map.

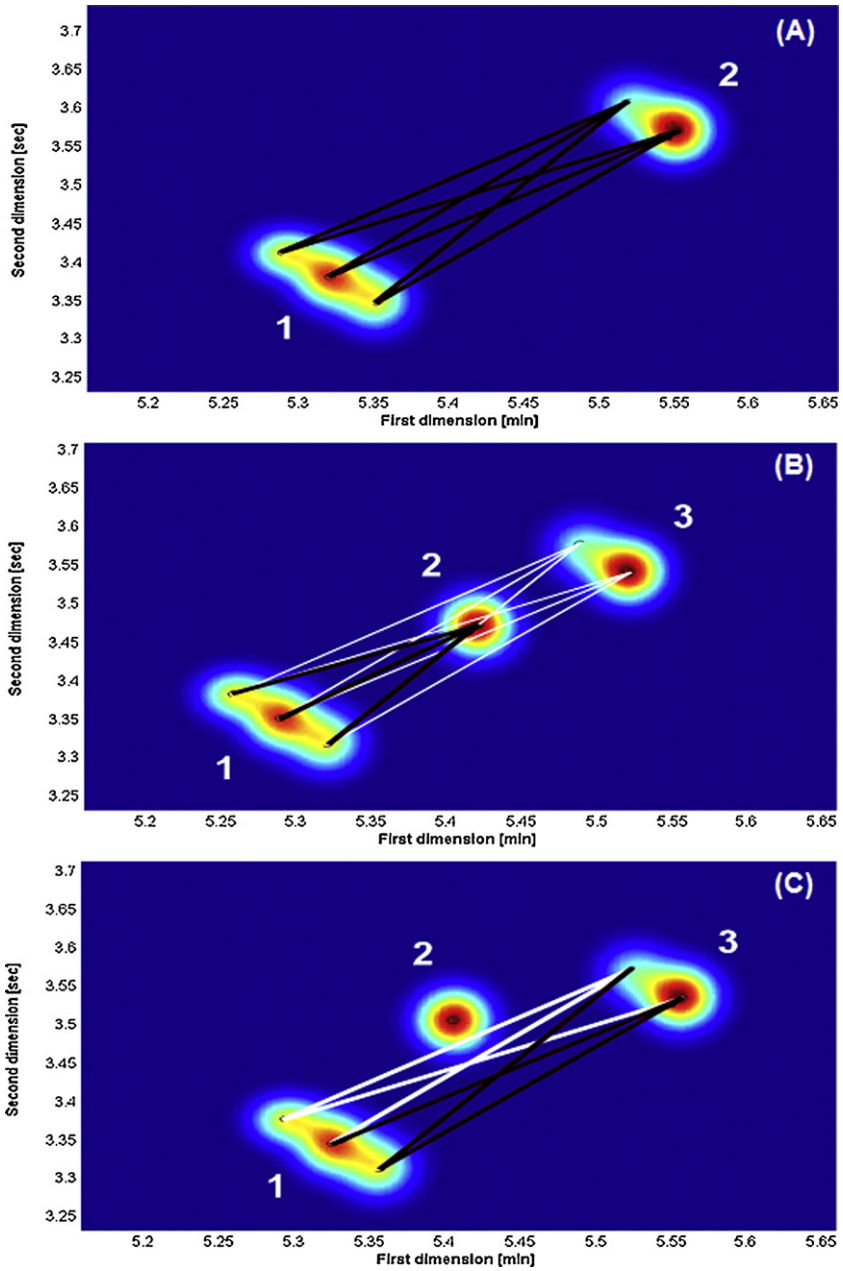

Fig. 8. Representation of peak vicinity in three scenarios: (A) without an interfering peak; (B) with an interfering peak; and, (C) with a partial interference of another peak. The white lines are the profile lines that do not cross the peak regions of other peaks, whereas the black lines are the profile lines that cross the peak regions of other peaks (visualization inspired by Peters et al. [33]).

4.3. Resolution of peaks in 2D chromatography

4.3.1. The saddle point as a measure of overlap

When two peaks in 2D chromatography are partially overlapped, the short trajectory between these two peaks shows the characteristics of a saddle point, i.e. it has a minimal point in one direction that is simultaneously a maximal point in another direction. As shown in Fig. 9A, if both peaks have a Gaussian shape, then the saddle point is the minimal point located in the shortest trajectory line between the two maximal points of these peaks. This point is considered a valley point in 1D chromatography (marked as a black dot). On the other hand, when the peaks do not show a Gaussian shape, it is necessary to test all possible profile trajectories between the 2D peak clusters, computing all the minimal points in these lines and verifying which of these minimal points have the highest value, becoming then the saddle point.

Fig. 9B shows the difference between considering the profile line between the two peak maxima (white line) and the profile line where the saddle point occurs (black line). The white dot in Fig. 9B represents the minimal value occurring at the profile line between the two peak maxima, while the black dot represents the value for the saddle point. That difference as described firstly by Peters et al. [33] can be quite significant in terms of resolution.

4.3.2. The valley-to-peak ratio in 2D chromatography

Once the saddle points and neighbouring peaks have been identified, it becomes possible to perform the calculations of resolution between the peaks [62]. This task is based on the concept of valley-to-peak ratio, V, between two 2D peaks, using the Kaiser's definition [63]:

$$V = \frac{f}{g} \tag{5}$$

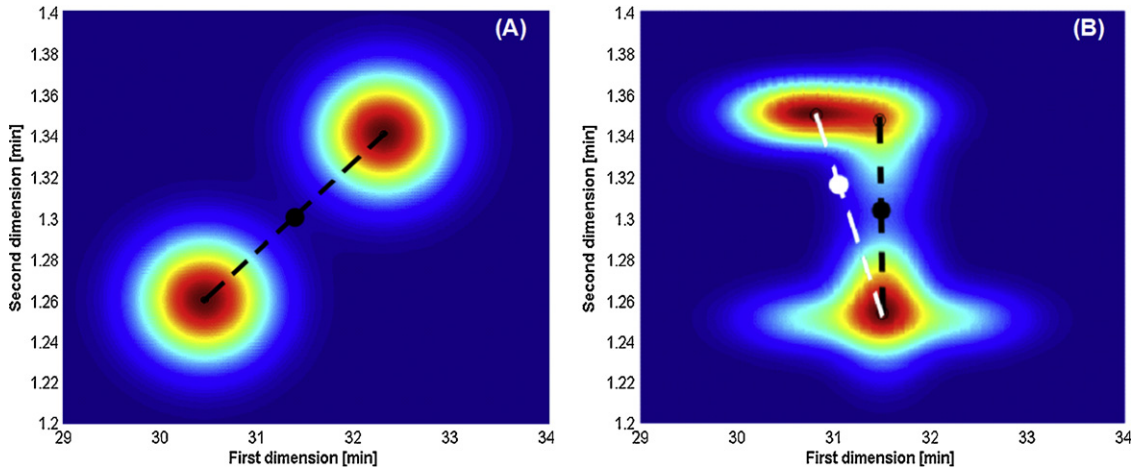

Fig. 9. Determination of the saddle point in Gaussian peaks (A), showing the shortest trajectory line between the two maximal points of the peaks (black line) and the saddle point location (black point), and differentiation between saddle point and minimal point in non-Gaussian peaks (B), showing the profile line between the two peak maxima (white line) and the minimal value (white dot) and the profile line (black line) where the saddle point occurs (black dot) (visualization inspired by Peters et al. [33]).

where g is distance from the baseline to the line linking the apexes of the two neighbouring peaks, and f is the distance between the height of the saddle point and this same line joining both peaks maximum, as shown in Fig. 10.

In an ideal chromatogram, with perfect Gaussian peaks, the saddle point is exactly in a midpoint between the peaks, thus becoming easy to define the g value as the average of the two peaks height [62]:

$$g = \frac{H_{\max_1} + H_{\max_2}}{2} \tag{6}$$

On the other hand, as shown in Fig. 10, the following relationship can be defined:

$$f = g - H_v \tag{7}$$

where H_v is the height of the saddle point [62].

However, in practical cases, perfect Gaussian peaks are not so common, and the saddle point is not positioned exactly at the midpoint between the peaks. Peters et al. [33] has overcome this problem by using only experimental values and geometric concepts to define the values of g and, consequently, using Eq. (7) to estimate the value of f. Therefore, using the geometrical proprieties represented in Fig. 10, the g value can be defined as [33]:

$$g = \frac{H_{\max_1} \times d_{s_2} + H_{\max_2} \times d_{s_1}}{d_{s_1} + d_{s_2}} \tag{8}$$

where d_{s_1} and d_{s_2} are the distances between the retention time of each peak and that of the saddle point, while $H_{\max_1}$ and $H_{\max_2}$ represent their respective heights. This procedure for the calculation of the valley-to-peak ratio was firstly suggested by Peters et al. [33] and it allows an easy way to measure the overlap of two peaks of any shape, using only the experimental raw data from the 2D chromatogram.

4.3.3. Measuring the resolution

Schure [62], based on the work of Giddings [64], proposed a metric for measuring the resolution (Rs) of Gaussian peaks from 2D chromatography using the valley-to-peak ratio (V):

$$Rs = \sqrt{-\frac{1}{2}\ln\left(\frac{1-V}{2}\right)} \tag{9}$$

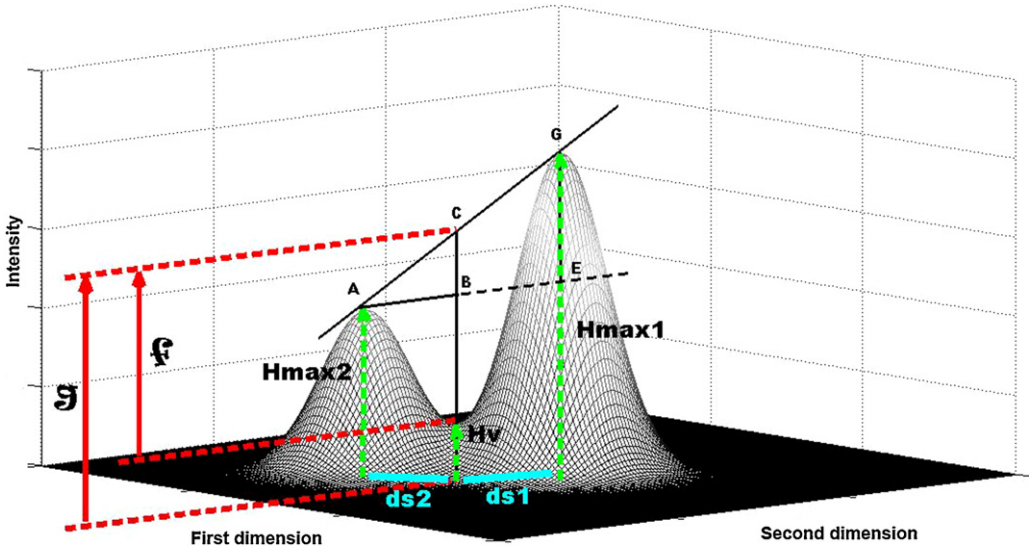

Fig. 10. Schematic diagram of an aid for calculation of the valley-to-peak ratio between two overlapping peaks in 2D chromatography. d_{s_1} and d_{s_2} are the distances from the retention time of peaks 1 and 2 and the saddle point, respectively; $H_{\max_1}$ and $H_{\max_2}$ represent the peaks height at maximum, H_v is the height of the saddle point, and g and f are the terms of the Kaiser's definition for valley-to-peak ratio (Eq. (5)).

Schure [62] also presented a compilation of resolution metrics proposed by other authors. The first formula was compiled from the work of Davis [65] and Shi and Davis [66]:

$$Rs(\theta) = \frac{\delta_t \sqrt{\gamma^2 \sin^2\theta + \cos^2\theta}}{4\sigma_x} \quad (10)$$

where δ_t are the deviations between peaks using the Euclidean distance, σ_x are the standard deviations of the x peak zone and γ is the ratio between the x and y standard deviations, and θ is the angle between the line that links both peaks and a parallel line to the x axis.

Another resolution metric presented by Schure [62] was proposed by Giddings [64], and it considers the total resolution following the Euclidean norm:

$$Rs = \sqrt{Rs_x^2 + Rs_y^2} = \sqrt{\left(\frac{\delta_x^2}{16\sigma_x^2}\right) + \left(\frac{\delta_y^2}{16\sigma_y^2}\right)} \quad (11)$$

where σ_x and σ_y are the standard deviations of peaks x and y, respectively; and δ_x and δ_y are calculated as $(\tilde{t}_{2,x} - \tilde{t}_{1,x})$ and $(\tilde{t}_{2,y} - \tilde{t}_{1,y})$, respectively.

Eq. (9) has been generalized by Peters et al. [33] for the determination of the valley-to-peak ratio for non-Gaussian 2D peaks, and it allows the description of the overall separation regardless of the peak shape. Such a generalization constitutes a huge advance for evaluating the chromatographic performance in the 2D chromatography. In order to avoid some problems related to multiple vicinities between neighbouring peaks in a 2D chromatogram, Duarte et al. [35] developed a new 2D chromatographic response function on basis of another concept from 1D chromatography: the peak purity. This concept, which assesses the volume peak free of interference, despite not being a measure of resolution, can be used to characterize the quality of the chromatographic separation [35]. The main disadvantage of this calculation is the need for simulating a 2D model of the chromatogram in order to identify the volume of each peak and the overlapping zone.

5. Data analysis software for 2D chromatography

The huge data sets produced by 2D separations make their analysis almost impossible without using some type of computer software in order to transform the 2D chromatographic data into usable information. Pierce et al. [67], in a recent review about advancements in comprehensive 2D separations with chemometrics, conclude that most of the available commercial and public domain software has been adapted from 1D chromatography. One of the few examples of commercial software developed and available for GC × GC is the software system developed as a spin off at the University of Nebraska-Lincoln [8–11], the GC Image, which incorporates the inverted watershed algorithm and allows a digital image processing for visualization, processing, analysis and reporting the GC × GC chromatographic data. Recently, a version of this software has been developed and released for dealing with LC × LC data [9].

Quite often, many researchers have developed their own algorithms, usually in MatLab (Mathworks, Natick, MA, USA), a well-known commercial software for numerical computation. Several of this algorithms and toolbox available in MATLAB language for analysis of multi-way data speeded all over the Internet. An example of this source code is the N-way toolbox available at http://www.models.kvl.dk/source/ [68]. There are no reports comparing the performance of different commercial software available but there are a few studies comparing different algorithms already incorporated in commercial software packages [42,43].

6. Conclusions and research needs

Data processing of comprehensive 2D chromatography is a rapid evolving subject since there is a general lack of commercial software associated to analytical instrumentation. Although the first algorithms developed for data processing in 2D chromatography were generalizations of concepts from 1D chromatography, nowadays there are already methods for non-targeted and targeted analyses fully developed to deal with bilinear chromatographic data or even higher orders. The emergence of multichannel detectors will easily lead to obtaining N-way data that need expertise drawn from chemometrics for proper data processing and attaining adequate information for analytical purposes. The first step will always be ensuring the linearity between the different orders of instruments. In the coming years, this step may pass through the improvement of the current 2D chromatographic systems, although the trend on this matter appears to be leading to the development and improvement of algorithms capable of ensuring synchronization between the orders of data. However, it is important to highlight that if this linearity is not guaranteed, than it is possible to sacrifice an order of measurement with adequate results, depending on the purpose of analysis. Within the approaches for calculating the quality of chromatographic separation, there is still room for improvement. This operation has only been developed for bilinear data. Nevertheless, the current state of progress within the chromatographic systems emphasizes more and more the need to develop methods of calculation for higher orders of data. Even for bilinear structures there is still some problems, namely the multiple vicinities between neighbouring peaks in the calculation of the resolution, or the need to produce a 2D chromatographic model for calculating the peak purity. Also the use of Euclidean distances should be the subject of study due to the different order of magnitude between dimensions. The use a of weighted distance measurement, such as the Mahalanobis distance, may be more appropriate. Finally, it should be stressed that the researchers often choose to develop their own algorithms and consequently it becomes hard to compare the results obtained. Besides, such an expert knowledge may take some time to be embedded in user-friendly software associated to the analytical instrumentation available in chemical laboratories.

Data processing of comprehensive two-dimensional chromatography is an area of great interest and expansion, thus becoming very difficult to define a clear cut trend, but rather allowing only establishing a set of methodologies and guidelines according to the type and order of magnitude of data, and purpose of analysis.

Acknowledgements

Centre for Environmental and Marine Studies (University of Aveiro, Portugal) and the Portuguese Science and Technology Foundation, through the European Social Fund (ESF) and “Programa Operacional Potencial Humano – POPH”, are acknowledged for financial support. One of the authors (João T. V. Matos) also thanks University of Aveiro for financial support through a research grant.

References

[1] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, J. Chromatogr. A 1168 (2007) 3.
[2] G. Guiochon, N. Marchetti, K. Mriziq, R.A. Shalliker, J. Chromatogr. A 1189 (2008) 109.
[3] P. Schoenmakers, P. Marriott, J. Beens, LC–GC Eur. 16 (2003) 1.
[4] J.B. Phillips, J. Beens, J. Chromatogr. A 856 (1999) 331.
[5] F. Erni, R.W. Frei, J. Chromatogr. 149 (1978) 561.
[6] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[7] S.E.G. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Anal. Chem. 78 (2006) 5559.

[8] S.E. Reichenbach, M. Ni, V. Kottapalli, A. Visvanathan, Chemom. Intell. Lab. Syst. 71 (2004) 107.
[9] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, J. Chromatogr. A 1216 (2009) 3458.
[10] S.E. Reichenbach, M. Ni, D. Zhang, E.B. Ledford, J. Chromatogr. A 985 (2003) 47.
[11] S.E. Reichenbach, V. Kottapalli, M. Ni, A. Visvanathan, J. Chromatogr. A 1071 (2005) 263.
[12] R.C. Allen, S.C. Rutan, Anal. Chim. Acta 705 (2011) 253.
[13] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, J. Chromatogr. A 1226 (2012) 140.
[14] J.M. Amigo, T. Skov, R. Bro, Chem. Rev. 110 (2010) 4582.
[15] Z.-D. Zeng, S.-T. Chin, H.M. Hugel, P.J. Marriott, J. Chromatogr. A 1218 (2011) 2301.
[16] Y. Zhang, H.-L. Wu, A.L. Xia, L.-H. Hu, H.-F. Zou, R.-Q. Yu, J. Chromatogr. A 1167 (2007) 178.
[17] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[18] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Anal. Chem. 70 (1998) 218.
[19] V.G. van Mispelaar, A.C. Tas, A.K. Smilde, P.J. Schoenmakers, A.C. van Asten, J. Chromatogr. A 1019 (2003) 15.
[20] K.J. Johnson, B.J. Prazen, D.C. Young, R.E. Synovec, J. Sep. Sci. 27 (2004) 410.
[21] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, Anal. Chem. 77 (2005) 7735.
[22] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Anal. Chem. 80 (2008) 2664.
[23] B.V. Hollingsworth, S.E. Reichenbach, Q. Tao, A. Visvanathan, J. Chromatogr. A 1105 (2006) 51.
[24] R.K. Nelson, B.M. Kile, D.L. Plata, S.P. Sylva, L. Xu, C.M. Reddy, R.B. Gaines, G.S. Frysinger, S.E. Reichenbach, Environ. Forensics 7 (2006) 33.
[25] G.D. Wardlaw, J.S. Arey, C.M. Reddy, R.K. Nelson, G.T. Ventura, D.L. Valentine, Environ. Sci. Technol. 42 (2008) 7166.
[26] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, S.E. Reichenbach, X. Tian, Q. Tao, J. Chromatogr. Sci. 48 (2010) 251.
[27] S. Peters, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, J. Chromatogr. A 1156 (2007) 14.
[28] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, J. Chromatogr. A 1216 (2009) 4020.
[29] H. Parastar, M. Jalali-Heravi, R. Tauler, Chemom. Intell. Lab. Syst. (2012), http://dx.doi.org/10.1016/j.chemolab.2012.02.003.
[30] N.J. Micyus, S.K. Seeley, J.V. Seeley, J. Chromatogr. A 1086 (2005) 171.
[31] J.J.A.M. Weusten, E.P.P.A. Derks, J.H.M. Mommers, S. van der Wal, Anal. Chim. Acta 726 (2012) 9.
[32] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[33] S. Peters, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, J. Chromatogr. A 1146 (2007) 232.
[34] E.J.C. van der Klift, G. Vivó-Truyols, F.W. Claassen, F.L. van Holthoon, T.A. van Beek, J. Chromatogr. A 1178 (2008) 43.
[35] R.M.B.O. Duarte, J.T.V. Matos, A.C. Duarte, J. Chromatogr. A 1225 (2012) 121.
[36] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.
[37] J. Pól, B. Hohnová, M. Jussila, T. Hyötyläinen, J. Chromatogr. A 1130 (2006) 64.
[38] M. Kivilompolo, T. Hyötyläinen, J. Chromatogr. A 1145 (2007) 155.
[39] M. Kivilompolo, V. Obůrka, T. Hyötyläinen, Anal. Bioanal. Chem. 391 (2008) 373.
[40] H.P. Bailey, S.C. Rutan, Chemom. Intell. Lab. Syst. 106 (2011) 131.
[41] G. Vivó-Truyols, Anal. Chem. 84 (2012) 2622.
[42] I. Latha, S.E. Reichenbach, Q. Tao, J. Chromatogr. A 1218 (2011) 6792.
[43] G. Vivó-Truyols, H.-G. Janssen, J. Chromatogr. A 1217 (2010) 1375.
[44] C.G. Fraga, C.A. Corley, J. Chromatogr. A 1096 (2005) 40.
[45] J.C. Hoggard, R.E. Synovec, Anal. Chem. 79 (2007) 1611.
[46] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, J. Chromatogr. A 1027 (2004) 269.
[47] A.E. Sinha, J.L. Hope, B.J. Prazen, C.G. Fraga, E.J. Nilsson, R.E. Synovec, J. Chromatogr. A 1056 (2004) 145.
[48] J.C. Hoggard, R.E. Synovec, Anal. Chem. 80 (2008) 6677.
[49] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Anal. Chem. 70 (1998) 2796.
[50] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 72 (2000) 4154.
[51] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, Anal. Chem. 71 (1999) 1093.
[52] C.G. Fraga, C.A. Bruckner, R.E. Synovec, Anal. Chem. 73 (2000) 675.
[53] L. Xie, P.J. Marriott, M. Adams, Anal. Chim. Acta 500 (2003) 211.
[54] A.E. Sinha, K.J. Johnson, B.J. Prazen, S.V. Lucas, C.G. Fraga, R.E. Synovec, J. Chromatogr. A 983 (2003) 195.
[55] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, J. Microcolumn Sep. 11 (1999) 97.
[56] G. Tomasi, R. Bro, Comput. Stat. Data Anal. 50 (2006) 1700.
[57] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[58] M.T. Lohnes, R.D. Guy, P.D. Wentzell, Anal. Chim. Acta 389 (1999) 95.
[59] A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, J. Chromatogr. A 1058 (2004) 209.
[60] J.L. Hope, A.E. Sinha, B.J. Prazen, R.E. Synovec, J. Chromatogr. A 1086 (2005) 185.
[61] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Anal. Chem. 83 (2011) 9289.
[62] M.R. Schure, J. Microcolumn Sep. 9 (1997) 169.
[63] G.C. Carle, Anal. Chem. 44 (1972) 1905.
[64] J.C. Giddings, Multidimensional Chromatography: Techniques and Applications, New York, NY, 1990.
[65] J.M. Davis, Anal. Chem. 63 (1991) 2141.
[66] W. Shi, J.M. Davis, Anal. Chem. 65 (1993) 482.
[67] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, J. Chromatogr. A 1184 (2008) 341.
[68] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst. 52 (2000) 1.