ChroMATHography: Solving Chromatographic Issues with Mathematical Models and Intuitive Graphics

José Manuel Amigo, Thomas Skov, and Rasmus Bro*

Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Received December 4, 2009

Contents

José Manuel Amigo (born 1978) studied Chemistry at the Faculty of Sciences, University of the Basque Country, Spain, and received his Bachelor degree in Sciences (Chemistry) in 2001. In 2007, he obtained his Ph.D. (Cum Laude) in Chemistry from the Autonomous University of Barcelona, Spain. Since 2007 he has been employed at the Department of Food Science, Quality and Technology of the University of Copenhagen, Denmark, as a postdoctoral researcher. His current research interests include hyperspectral analysis, pharmaceutical research, environmental modeling, process analytical technologies, chemometrics, multivariate calibration, multiway analysis, exploratory analysis, curve resolution, and MATLAB programming. He has authored more than 20 peer-reviewed scientific papers, 3 books on high school teaching, more than 10 conferences (3 of them plenary/invited conferences), and several posters for many diverse international meetings (receiving 2 poster awards).

1. Introduction

1.1. Hyphenated and Hypernated Chromatography

The evolution of hyphenated chromatographic systems for the detection, quantification, and/or identification of compounds has become one of the most important developments in various fields of chemistry in the last 30 years. The reliability of high resolution detectors coupled with the efficient separation power of modern chromatographic systems has changed many routines in analytical chemistry. Nowadays many reference methods for the determination and quantification of analytes in complex matrices are based on a chromatographic routine coupled with a highly sensitive detector.

The term *hyphenation*, proposed by Hirschfield[1] in 1980, refers to the online combination of a separation technique and a spectroscopic detection method, which provides information about the measured analytes. A number of spectroscopic detectors can be used, such as diode-array ultraviolet−visible absorbance (DAD-UV) for liquid chromatography (LC) or, most commonly, high performance liquid chromatography (HPLC), intensified linear diode array fluorescence (ILDA-FLU) and nuclear magnetic resonance (NMR) for LC, Fourier-transform infrared (FTIR) for LC and gas chromatography (GC), and, of course, mass spectrometry (MS) in LC and GC.[2−5] Another more recent development is *hypernation*, which according to Wilson and co-workers[2−4] is the coupling of two chromatographic techniques to enhance the resolution power (peak capacity). Examples could be LC × LC,[6,7] LC × GC,[8,9] or GC × GC.[10−12]

Hyphenation and hypernation create more complex data structures. This means that data holding more information are often obtained, but this is not necessarily information that can be easily extracted. More complex data structures

* Corresponding author. E-mail: rb@life.ku.dk.

10.1021/cr900394n
Published on Web 05/14/2010

Thomas Skov (born 1975) studied Food Science at the Faculty of Life Sciences, University of Copenhagen, Denmark, and received his Master degree in Food Science and Chemometrics in 2003. In 2008, he obtained his Ph.D. in Chemometrics, titled Mathematical Resolution of Complex Chromatographic Measurements, from the University of Copenhagen, Denmark. Since 2009 he has been employed at the Quality and Technology group, Department of Food Science, at the same university as an Assistant Professor, focusing mainly on Process Analytical Technology (PAT) and developing novel but user-friendly chemometric methods for chromatographic and metabolomics data. He has published more than 10 papers in peer-reviewed journals, supervised several B.Sc. and M.Sc. students, and taught chemometrics to students, staff, and industry during the last 5 years.

Rasmus Bro (born 1965) studied mathematics and analytical chemistry at the Technical University of Denmark and received his M.Sc. in 1994. In 1998 he obtained his Ph.D. (Cum Laude) in multiway analysis from the University of Amsterdam, The Netherlands. Since 1994 he has been employed at the Department of Food Science, Quality and Technology of the University of Copenhagen (former Royal Veterinary & Agricultural University), and in 2002 he was appointed Full Professor of chemometrics. He has had several stays abroad at research institutions in The Netherlands, Norway, France, and the United States. His current research interests include chemometrics, multivariate calibration, multiway analysis, exploratory analysis, experimental design, numerical analysis, blind source separation, curve resolution, MATLAB programming, and constrained regression. In 2000 he received the third Elsevier Chemometrics Award for noteworthy accomplishments in the field of chemometrics by younger scientists, and in 2004 he received the Eastern Analytical Symposium Award for Achievements in Chemometrics. He has authored more than 100 peer-reviewed scientific papers, 2 books on chemometrics, and more than 20 proceedings, book contributions, reviews, and patents.

make the data analysis more challenging, but they also allow for handling situations which could otherwise not be handled. This review aims at presenting these opportunities in light of *chromatographic data* and also highlighting that it is worthwhile to take advantage of them.

1.2. Dealing with Problems in Chromatographic Data

The ideal chromatographic signal (no matter the separation method(s) or detector(s) used) will have well-resolved peaks, adequate signal-to-noise ratios, no background contribution, and a large linear response range between analyte concentration and detector signal for individual samples/runs. When more than one sample is measured, the ideal situation also includes stable retention times and well-defined peak shapes (preferably Gaussian profiles) for all analytes. The success of the chromatographic separation as well as the robustness and stability, thus, depend on the appropriate selection of the chromatographic equipment (e.g., gas/liquid, pumps, column, mobile phase) and the experimental conditions, often based on experimental design (e.g., column type, temperature conditions, gradient of the mobile phase, etc.).

However, ideal chromatographic data are often not obtained and nontrivial problems can arise that need to be handled. There are a number of sources of variability in a chromatographic system (pumping systems, temperature gradients, stability of stationary phases, detection systems, etc.). These sources can be reflected in the signal and, therefore, may cause problems directly linked to two of the main objectives of chromatographic analysis:

(1) to achieve a perfect separation, detection, and quantification of the individual analytes

(2) to obtain sample specific fingerprints.

The problems can be related to the separation power of the analytical method and to the sample-to-sample stability. To increase the separation power, the chromatographer will most often reparameterize the chromatographic method (new column, changed temperature gradient, different solvents, etc). These operations are time-consuming and may generate other problems (e.g., changing the temperature gradient in GC may improve the separation of two analytes but result in worse separation for other analytes). The sample-to-sample stability relies more on the robustness of the individual instrumental parts. The data analytical community has provided tools to solve some of these problems postmeasurement from a data handling perspective. Mathematical modeling (chemometrics) has been applied in chromatography from the last 30 years on, offering robust and reliable data analytical alternatives to handle problems derived from the instability of the chromatographic system (baseline/background correction), the lack of sample-to-sample stability (alignment or normalization), or even the problems in quantifying overlapped peaks (coelution).

However, the synergy between chromatographers and chemometricians has been limited so far, and much closer collaboration is needed to really reap the combined benefits. Having a look at the literature (books, reviews, and papers), one will find that most of the research that involves chemometrics and chromatographic data has been written by chemometricians. And, despite the demonstrated usefulness, it is still rare to find chromatographers applying chemometrics. The problem may be caused by (1) the lack of implemented algorithms in instrument specific chromatographic software and, (2) which is more important, the different "languages" used by chemometricians and chromatographers.

1.3. Objectives of This Work

The objective of this review is to bring chemometricians and chromatographers closer together in order to encourage chromatographers to use chemometric tools to solve their problems and encourage chemometricians to take advantage of and learn from the chromatographic skills needed to provide accurate, reliable, and suitable data. The main focus of this review is novel data methods that are not already or only sparsely applied when handling chromatographic data. As we will expose throughout this review, there are many methods that can be employed, and we will introduce them, starting with the simplest ones in the first sections and then introducing more advanced methods along with more complex chromatographic problems. We will illustrate the most important methods with examples, as well as giving pointers to tutorials available in the literature.

The first part (section 2) offers an overview of the structure of the chromatographic signal and what challenges are present in chromatographic data. We describe how and to what extent chromatographic problems can be solved by data analysis. We start by presenting problems and solutions for individual chromatographic signals (section 3; the background contribution) and continue with aspects related to having more than one sample (correcting nonstable retention times), and finally in section 4 we describe how coelution problems can be handled. Pros and cons will be discussed, and also limitations and things to consider for the presented methods will be touched upon.

2. Chromatographic Data

Chromatographic data are not the same as informative data—properly preprocessed and analyzed chromatographic data are as close to informative data as we can get!

2.1. The Chromatographic Signal

The overall signal of a chromatogram can be divided into three constituting parts (Figure 1):[13]

(1) The analytical signal contains the signal of any analyte present. The analytical signal depends on the detector sensitivity and the capability of the chromatographic system, among others. For a given specific analysis, parts of the analyte signal may be considered an interferent signal, but this is not crucial for how the data will be treated in the following.

(2) The background signal is any signal that is not related to the analyte signal(s) and shows some sort of systematic behavior. The background often depends on the chromatographic conditions and is mostly referred to as the baseline.

(3) The noise is any unsystematic (random) variation in the signal. It basically depends on the detector sensitivity.

The sum of these three parts provides the overall chromatographic signal. In hyphenated chromatographic systems, the signal can be divided into the same parts (Figure 1b). The chromatographic signal can also be formulated as

$$\text{measured signal} = \text{analytical signal} + \text{baseline} + \text{noise} \quad (1)$$

An ideal situation would be to be able to split the chromatographic signal into the parts shown in Figure 1. Then, subsequent data analysis can be focused on the informative signal: a signal which can be complicated (overlapping peaks—Figure 1b) or very simple and straightforward (well-resolved peaks—Figure 1a).

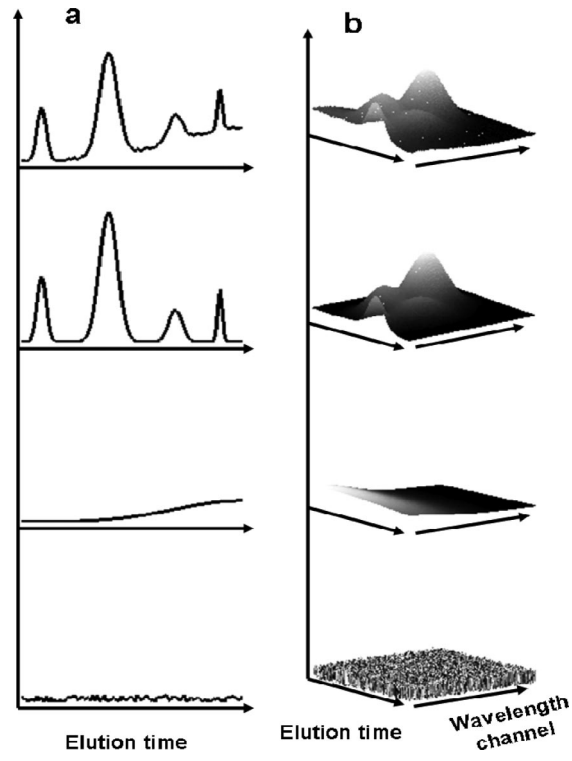

Figure 1. (a) Components of the chromatographic analytical signal obtained with a GC-flame ionization detector (GC-FID) system: upper, overall signal; middle, analytically relevant signal; bottom, background and noise, respectively (visualization inspired by Daszykowski and Walczak).[14] (b) Components of the chromatographic analytical signal obtained with a HPLC-DAD system: upper, overall signal; middle, analytically relevant signal; bottom, background and noise, respectively.

The structure of the data to be analyzed depends on the number of separation steps (e.g., GC × GC) and the detector(s) (e.g., MS-MS) used. A detector is *monochannel*, if only one number (e.g., intensity) is collected for each elution time. Typical examples are the flame ionization detector (FID), the UV monochannel detector, or even mass spectrometry when only the intensity of a single mass fragment (SIM —single ion monitoring) is collected. Nowadays, it is possible to use *multichannel detectors*, characterized by measuring a complete spectral range for each elution time. Two examples are the diode array ultraviolet−visible detector (DAD-UV) and the mass spectrometry detector (MS). Figure 2 summarizes the different kinds of data structures that can be collected depending on the instrumental setup.

2.2. Common Chromatographic Artifacts

Sometimes it is not possible to achieve perfect separation, either because of the complexity of the sample or because faster chromatographic runs are preferred. Also, problems with drifts in the baseline, changes in the shapes of the peaks, and shifts in the elution times (shifts that might be different for different peaks in the same sample) may decrease the quality of the final result of the analysis.[13] Throughout this review, we will use the term retention time for a specific elution time for a given analyte whereas elution time will be used as an overall description for something that elutes and provides a signal (peak, baseline, interferent, noise, etc.) An ideal situation would be achieved when

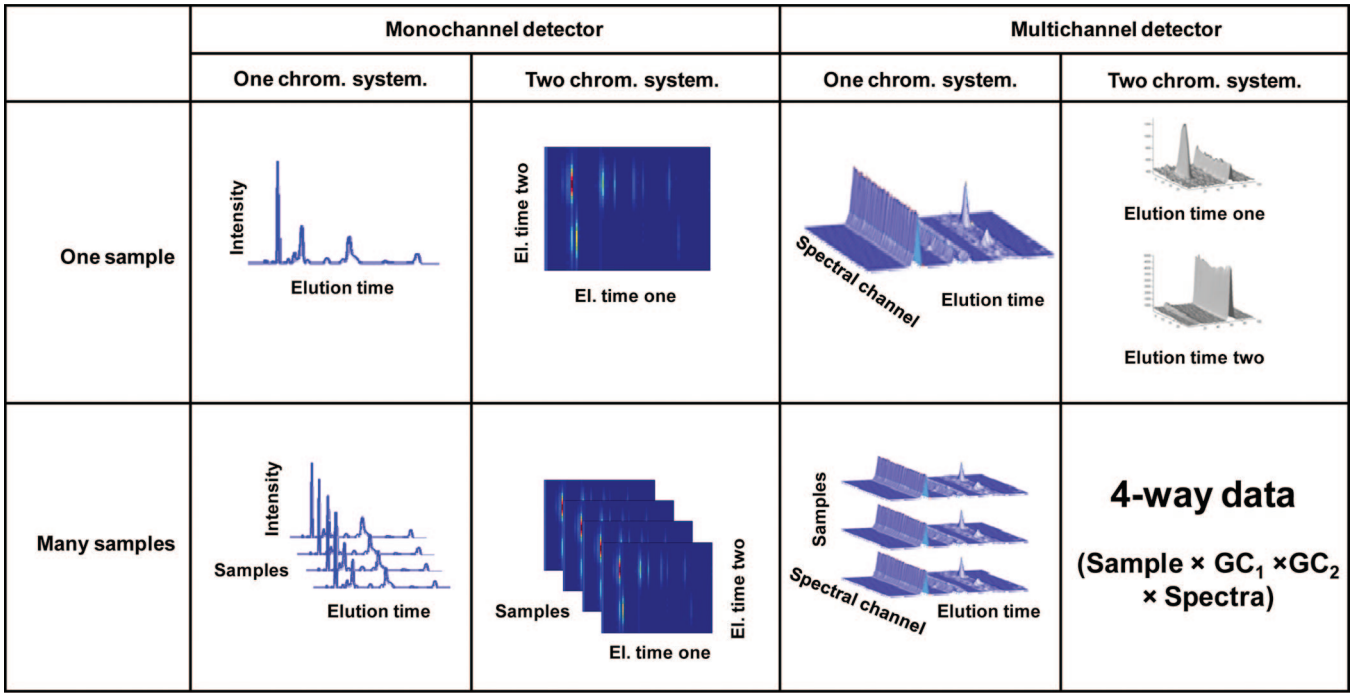

Figure 2. Number of directions (or modes) of chromatographic data using different combinations of chromatographic setup and detector for one or more samples (one mass channel per landscape for better visualization in multichannel detectors).

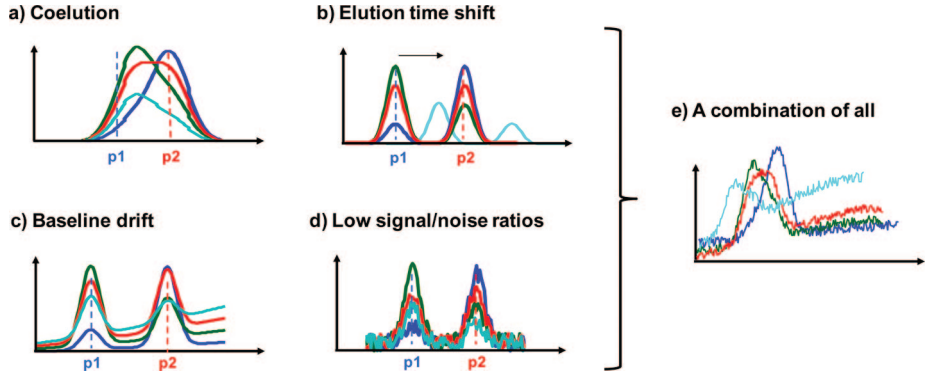

Figure 3. Common deviations from ideality in chromatography. p_1 and p_2 account for the retention times for the first and second peaks, respectively.

(a) all peaks maintain their retention time in different runs;
(b) if baseline is present, it should be rather stable in different runs;
(c) the signal-to-noise ratio is high;
(d) there is no (or little) peak tailing or fronting (peak shape changes);
(e) all peaks are resolved.

Unfortunately, nonideal situations, as shown in Figure 3, are often encountered. Typically, the problems in chromatographic data can be grouped into elution time shifts, baseline drifts, low signal-to-noise ratios, and peak shape changes or a combination of these.

2.3. A Basic Model of Chromatographic Data

Assume a single-channel chromatographic profile only containing the same single peak in several samples and no artifacts. The data can be efficiently described by a model using one so-called factor of the chemical/systematic part (Figure 4). All elution profiles can be described by the same common chromatographic profile (the peak position and shape) and a measure of how much each sample has of this profile. Thus, every sample can be described by the same elution profile multiplied by the respective concentration/peak area. Any remaining variation is then attributed to background (systematic) or noise (unsystematic) in the measurements.

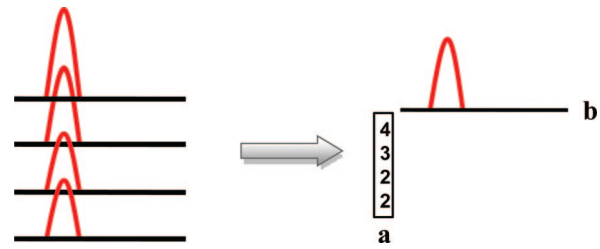

Figure 4. Illustration of a factor model applied to a chromatographic profile with one peak. If only one peak (one chemical component) is present, it can be described by a single common shape (loading, b) (estimated elution profile) and a measure of the amount of this profile (an individual score, a). Scores are thus an estimate of the relative concentration of this peak in each sample.

Figure 4 can also be written as an equation, and including a noise term, this becomes

$$\mathbf{X} = \mathbf{a}_1\mathbf{b}_1^{\mathrm{T}} + \mathbf{E} \qquad (2)$$

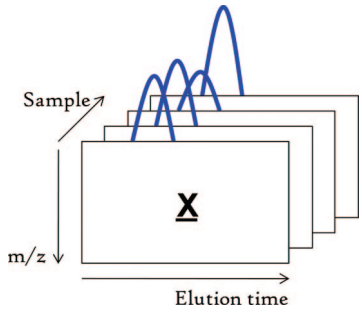

Figure 5. Data structure arranged as a three-way box of data, $\underline{\mathbf{X}}$, with GC-MS experiments as an example.

In this equation, $\mathbf{X}$ is a matrix where each row contains the elution profile of one sample, $\mathbf{a}_1$ is the score vector (see caption of Figure 4), and $\mathbf{b}_1$ is the loading vector. The subscript 1 denotes *one* part of the signal (that is, the analytical species in red in Figure 4). The matrix $\mathbf{E}$ is of the same size as the original data and contains the unexplained residual variation.

Equation 2 can easily be extended to describe the overall measured signal with several, I, analytes (several systematic signals) as presented in eq 1.

$$\mathbf{X} = \mathbf{a}_1\mathbf{b}_1^{\mathrm{T}} + \mathbf{a}_2\mathbf{b}_2^{\mathrm{T}}, ..., \mathbf{a}_I\mathbf{b}_I^{\mathrm{T}} + \mathbf{E} \qquad (3)$$

where $\mathbf{a}$ is a vector containing the concentrations of a certain analyte in the samples and $\mathbf{b}$ a vector that simply holds the chromatographic profile of that specific analyte. For analyte number i the contribution to the data is given by one factor $\mathbf{a}_i\mathbf{b}_i^{\mathrm{T}}$ and I is the total number of analytes whereas $\mathbf{E}$ is the residual part. As we will see further on, one method able to estimate the $\mathbf{a}$'s and $\mathbf{b}$'s given the actual profiles in $\mathbf{X}$ is called *multivariate curve resolution* (MCR),[15−18] which is a generic term that encompasses several different algorithms, but all with the aim of finding the $\mathbf{a}$'s and $\mathbf{b}$'s from the measured signal $\mathbf{X}$.

In Figure 4, a simple integration of the peak area or peak height would be enough to characterize the data, but for more complex problems, some factor models can be used to disentangle overlapping peaks, separate baseline from peaks, etc. Hence, the use of factor models will be crucial in the following section. The example in Figure 4 is a *two-mode* data set because there is a sample direction and a chromatographic direction. Extending the example in Figure 4 to more than two modes is straightforward; simply extend with an additional chromatographic direction (e.g., GC × GC-FID) or maintain the spectral direction (e.g., GC-MS). Data then becomes a three-mode data set and the factor models must be extended to be able to handle this (see eq 4 and Figure 5).

The natural extension of the simplified model shown in eq 2 for the data in Figure 5 can be written as

$$\underline{\mathbf{X}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \underline{\mathbf{E}} \qquad (4)$$

where $\mathbf{a}$ is the vector of concentrations of the analyte, $\mathbf{b}$ is the elution profile, $\mathbf{c}$ is the mass spectrum for the analyte, and $\underline{\mathbf{E}}$ is the residuals containing unexplained variation. Using the model in eq 4, it is possible to estimate $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ from the data as shown in Figure 6 and thus describe the underlying parameters of one chemical analyte.

The data in $\underline{\mathbf{X}}$ is now a box of data rather than a table, and mathematically, the product (∘) between $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ is a

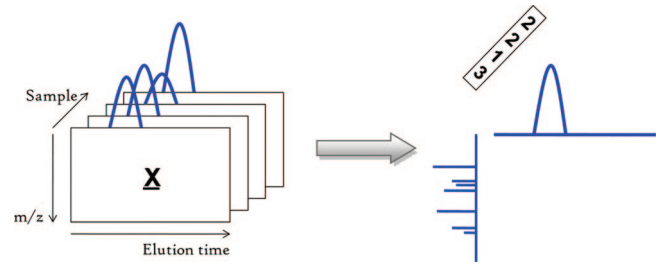

Figure 6. Data in three directions (samples, elution time, and mass spectrum). For simplicity, no baseline and noise are added to the data. The factor model extracts and describes chemical information for the analyte in one factor/chemical part.

bit more complex (a tensor product). However, the idea is exactly the same as above; the data can be modeled as a set of concentrations times an elution profile times a spectrum. A model that works on boxes of data is PARAllel FACtor analysis (PARAFAC),[19,20] which can even handle four or higher orders of data (hyper-) boxes. Just like the two-mode multivariate curve resolution (MCR) model, PARAFAC also extends to handling more chemical analytes.

The family of factor models techniques emerged from the need for solving the problems when separation cannot be improved by optimizing the chromatographic conditions.[15,16,21−23] The main feature of factor models is that they may recover the chromatographic profiles (elution profile, spectrum, and concentration), provided that each analyte has a distinct pattern.[13,24−28] Furthermore, we must at least approximately have the following:

(1) The spectrum collected for each elution time follows a linear relationship with the concentration. As an example, spectrophotometric patterns collected with a DAD-UV spectrophotometer follow the Beer−Lambert law.

(2) The intensity collected for each point of the spectral pattern can be assumed to be the sum of the absorbances of the analytes that form the mixture in each elution time.

(3) The elution time must be stable from sample to sample

It is important to note that one of the main advantages of factor models with respect to other techniques is that they do not assume any particular shape (e.g., Gaussian) of elution profiles.[29] The actual shape is determined from the data alone. We will return to a more complex use of factor models in the following.

When comparing different samples, normalization of chromatographic profiles is also an important issue and this step is crucial for obtaining data that contains the right quantitative difference between samples. Depending on the data structure, the normalization can be done in many ways. For quantitative measures (e.g., extracted peak areas), it is often recommended to normalize according to an internal standard peak. This has the advantage that each peak still contains a quantitative measure of the analyte concentration. The assumption behind this is that all peaks behave in the same way and can be corrected by the same internal standard. For many data sets, this is a valid assumption, but for other data sets, more internal standards can be needed to normalize more locally. For chromatographic profiles (e.g., TIC), it is common to normalize to a unit area of the profile. However, this means that the quantitative information is lost and only the pattern (which peaks are present) can be evaluated. This type of normalization is useful for characterizing different groups of samples and not so much for establishing a calibration model. If the latter is in focus, it is essential that

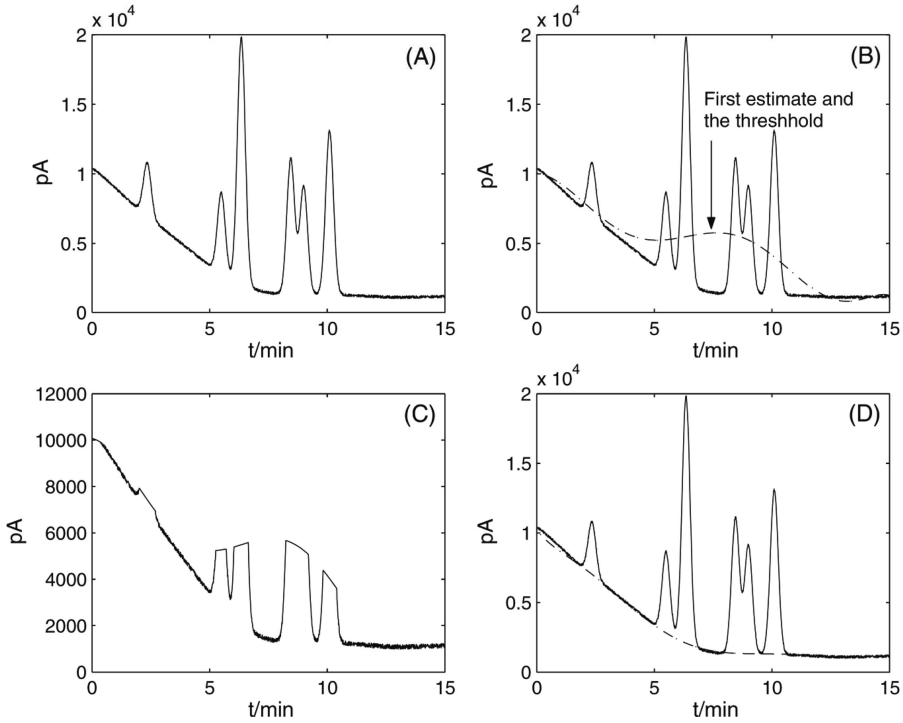

Figure 7. Illustration of baseline correction presented by Gan et al.:[34] (A) analytical signal; (B) first polynomial fitting; (C) signals above polynomial set equal to the polynomial and another polynomial is fitted; (D) final estimated baseline. Here an order of seven was used for the polynomial. Modified with permission from ref 34.

some other quantitative measure has also been obtained (e.g., the mass of the injected sample).

If the spectral mode is also included, normalization of the whole landscape is not recommended. Then it is more appropriate to find local peak regions and use factor models such as PARAFAC or PARAFAC2 to estimate the concentration of each analyte. After this, each analyte can be divided by the concentration of an internal standard and, thus, all peaks are normalized in the same way as indicated above for the quantitative measures. A further advantage is that the baseline (modeled and removed from the analyte signal) is not influencing the normalization, which would be a problem if the whole landscape of each sample was normalized. More can be said on normalization, but it is a subject that is very specific to specific areas of application and specific types of data. Hence, the reader is referred to the literature for details.[30,31]

In the following, we will start out by describing how baseline signals can be handled in different ways (section 3.1) and how retention time shifts can be approached (section 3.2). Subsequently, we will discuss ways to handle coeluting peaks (section 4) including handling coelution in combination with shifts and baseline variation.

3. Removing Artifacts by Preprocessing

3.1. Baseline Offset/Background Contribution

Baseline offset or background contribution (denoted baseline in the following) has always been an important issue in chromatographic analysis. Generally speaking, baseline correction can be divided into two approaches:

(1) Fit a certain curve (e.g., a polynomial) to be able to subtract this curve from the overall signal.

(2) Model the baseline as part of an overall (factor) model.

The first approach is often the simplest approach to implement, though not always the best from a data-quality point of view (section 3.1.1). The second approach is usually not an individual step but an added benefit of the use of a factor model (section 3.1.2).

3.1.1. Curve-Fitting

Almost all chromatographic software packages have some kind of baseline handling method. These methods are often based on a local baseline being constructed in a smaller region; often only a straight line (first order polynomial) is fitted. The fitted line can then be subtracted from the overall signal or the area below the fitted line subtracted when having found the overall peak area (from the analyte plus baseline).

Typically, the univariate response from a single channel sensor is corrected by fitting polynomial functions to the baseline in the vicinity of the peak to be corrected. This polynomial can be fitted through a predefined number of base points, and also polynomials of higher order can be applied if the baseline shows curvature.

These local methods work well if it is possible to find baseline points, if peaks are not coeluting (discussed further in section 4), and if the signal-to-noise ratio (S/N) is high. When this is not the case, the baseline is often better described using a global (throughout all or most of the elution profile) polynomial fit and typically using higher order polynomials to possibly account for a more complex, curved, baseline. In the following, the global methods will be discussed in more detail, as several methods are available.

Fitting (or estimating) the baseline can be achieved in different ways. When baseline points are not manually selected, they have to be selected by the algorithm. This is often done iteratively by first estimating the baseline from all points and then, e.g., removing the points furthest from the fitted polynomial in the positive direction (see Figure 7).[32−34]

Sometimes though, the curvature in the baseline is not readily described by higher order polynomials, and thus,

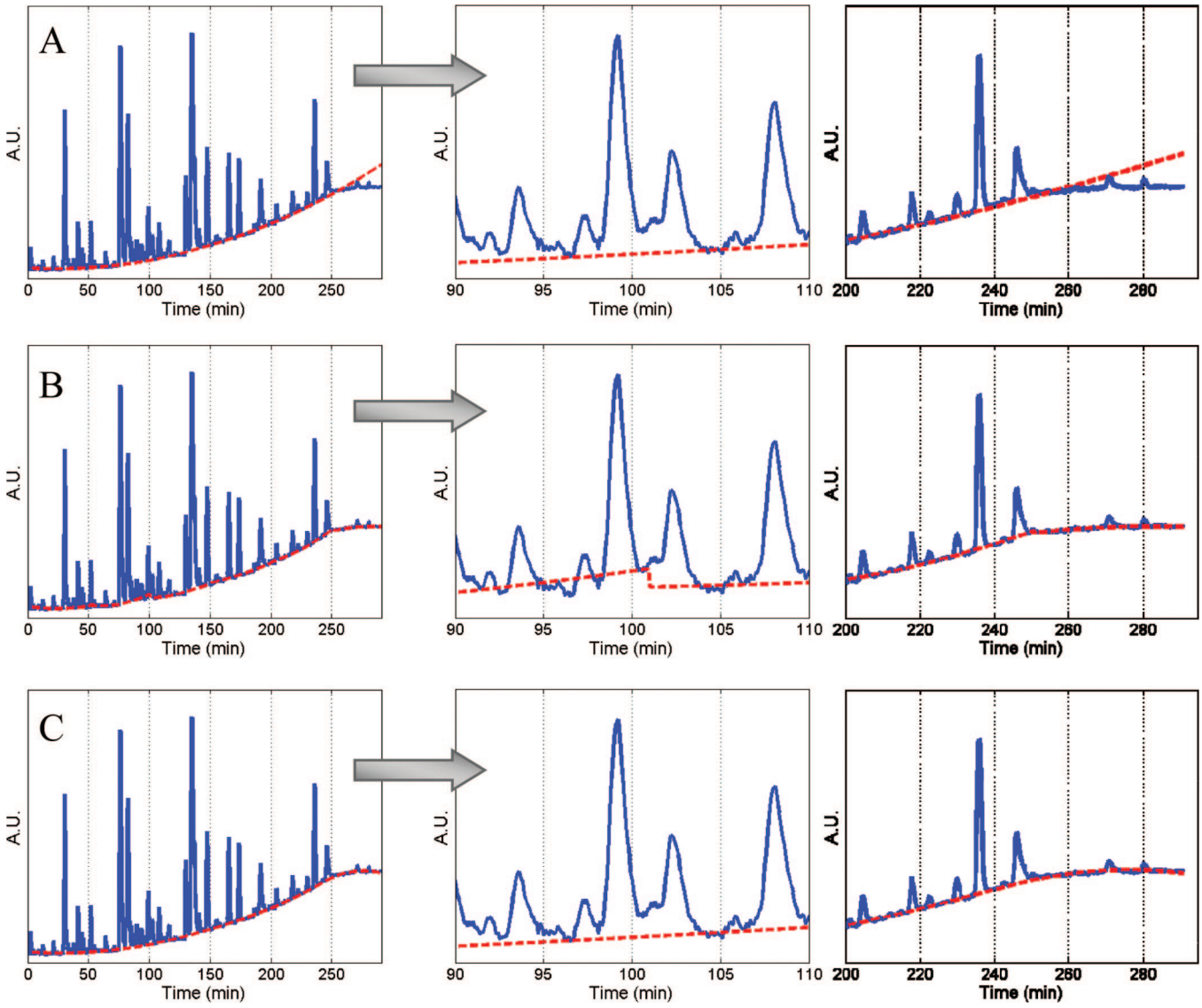

Figure 8. Examples of three baseline correction techniques with different degrees of flexibility: (A) global fit using a second order polynomial fit—smooth curve but improper correction; (B) local fit using several second order polynomials—nonsmooth curve but almost proper correction; (C) global fit using B-splines—smooth and proper baseline correction. LEFT figures show the whole chromatogram, whereas RIGHT figures show a zoom-in of the region between 90−110 min and 240−290 min.

artifacts are introduced if these methods are applied. In this case, e.g., B-splines can be used. Instead of fitting local polynomials, which are not necessarily continuous between regions, splines include many local polynomials but assembled into one smooth baseline which can attain a more flexible shape than a global polynomial. While B-splines offer high flexibility compared to polynomials, this flexibility can introduce artifacts in the baseline corrected signal. Often, visual inspection is used for guiding the selection of the baseline method.

When there are coeluting peaks, local methods are prone to errors due to the lack of baseline points between peaks. In these situations, it is often better to focus on a global estimation of the baseline or simply try to handle the baseline in a different way; the latter will be discussed in detail in the next section.

In general, a global polynomial fit requires higher order polynomials whereas a local fit (e.g., splines) requires lower order polynomials. Baseline correction methods for both single elution profiles[32−35] and data from hyphenated instruments[36−38] (e.g., GC-MS, LC-DAD) are available (Figure 8), but these will not be further discussed here.

3.1.2. Factor Model Approach

Instead of fitting a curve to the baseline of the data, it is also possible to take advantage of the baseline being a systematic part of the chromatographic signal, as indicated in eq 1. In fact, the baseline part can often be described in a similar way as is the case for the analytical relevant signal (the peak), i.e. by means of a factor model. For this to work, the baseline should have a similar shape for all samples; only the magnitude of it should vary.

When using factor models to estimate the baseline part, the best way is to work in intervals (local regions) for two reasons:

(1) The number of peaks present in the intervals should be fairly low (see section 5 for a further discussion of this aspect).

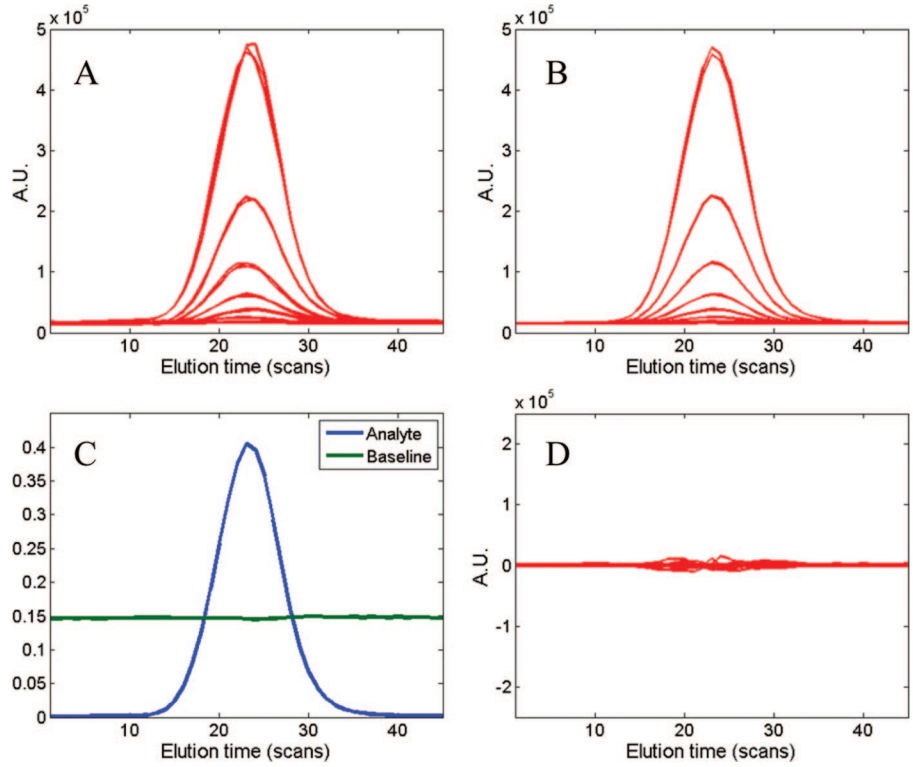

Figure 9. Use of PARAFAC to separate the chemical analytes and baseline in a small interval of a GC-MS experiment: (A) simple peak system with two systematic parts (one analyte and one baseline part) and some noise; (B) modeled data; (C) estimates from PARAFAC of the elution time mode; (D) residuals.

(2) The baseline part can change not only in the magnitude over the elution time but also in the shape in different runs.

Although the latter change might be relatively small, this will be better handled in smaller intervals, where differences are less pronounced. An illustration of finding the baseline part in a small region using factor models is shown in Figure 9, where the factor model, PARAFAC, has been used to separate baseline and chemical analyte signal. In this interval, there are two systematic parts (one analytically relevant peak and the baseline) as well as the unmodeled part (the residuals), and thus, two chemical factors can be derived from the factor model.

When applying factor models, it is important to have well aligned profiles and also important that there are no significant profile shape changes across samples, e.g. due to column overload or a degenerated column. Such differences will make the factor models shown so far less appropriate. In Figure 9, it is seen that there is still some systematic structure in the residuals due to slight misalignment of the elution profiles. For now, this is a minor detail, as the peaks have been described well, but in the following sections more advanced methods are introduced that can handle such shifts even when more profound.

3.1.3. Limitations and Things To Consider

When dealing with background/baseline contributions, two approaches are most often used: (1) fitting a curve or (2) modeling the baseline as an individual chemical part. When to use which method is not always an easy question to answer, but some guidelines can be given depending on the purpose of the baseline correction (Table 1).

3.2. Retention Time Shifts across Samples

In traditional chromatographic analysis, peaks are integrated and the concentration/area used in subsequent analyses. For well-resolved peaks with a simple baseline signal, this is a very feasible and simple approach. However, when more complicated data appear, methods such as factor models must be used. Because factor models work by finding the *underlying* elution profiles (among other) in several samples, it is important that the elution profile (and, e.g., mass spectral profile) of each compound remains the same across samples. If this is not the case (see Figure 10), alignment can be used to correct for differences in retention time.

For chromatographic data, alignment can be defined as "*a mathematical operation where similar chemical features are repositioned so that they appear at the same elution time in different runs*". The optimal alignment technique should require only a minimal or no input from a skilled technician or scientist, should be fast, should be robust, and should be applicable for a wide range of analytical situations without extensive customization needed.

Generally speaking, alignment can be divided into two approaches:

(1) Synchronize signals, based on a mathematical transformation of the time axis of each signal, so that all signals show similar phenomena at the same locations.

(2) Handle the shifted peaks implicitly within a factor model.

3.2.1. Synchronizing Signals

Alignment should preserve the relevant information in the data—i.e. keep peak areas and shape intact for subsequent quantification tasks. Different properties of the data and the alignment method must be considered: type and structure of

Table 1. Baseline Correction—Things To Consider

purpose of baseline correction	method	things to consider	limitations
look at profiles (e.g., individual UV spectra, MS spectra, TIC, SIC, etc.)	curve fitting	Baseline points are needed. Flexibility can be controlled by using window based curve fitting (e.g., with splines) or higher order polynomials. If changes in baseline between samples, parameters in curve fitting method must be changed from sample to sample!	Introduction of artifacts due to too simple a baseline estimate.
peak quantification	factor model	How many chemical components to have in the model? Does the baseline change (e.g., in shape) between samples?	If baseline changes from run to run, the baseline part found from a factor model will be incorrect (see Figure 14 for a way to solve this).
identify baseline components (e.g., *m*/*z*, UV spectrum) over time	factor model	If the baseline originates from few unique mass fragments or wavelengths with no overlap of the signal from chemical analytes, then identifying and removing these can remove the baseline.	
treat all samples at the same time	factor model	Are there shifts in profiles or baseline changes across samples?	Baseline must only vary in intensity and not in shape! If changes in shape, more advanced methods are needed; see section 3.2.
remove real baseline using blank sample	subtraction of curve	How to get the blank sample? Is one blank enough or is more needed—e.g. for every five samples?	Essential to have the same baseline profile for all samples.
special requirements	method	things to consider	limitations
Avoid introducing negative values	Factor model	Trying to extract too many chemical factors can provide model parameters that cross zero. Can be solved by non-negativity constraints—i.e. the profiles are not allowed to contain negative values.	
Noise must fluctuate around zero	Curve fitting	If proper baseline correction, then the peak area will be correct.	If baseline fluctuates above or below zero (improper baseline correction), the subsequently found peak areas can be biased.

data, shift correction needed, algorithmic parameters (quality measure, optimization criteria, reference selection, etc.), and, finally, what to do with the aligned data. These properties have been discussed in recent reviews by Tomasi et al.[41] and Vandenbogaert et al.[42] (for reviews, see refs 41−43, and for comparisons of alignment techniques, see refs 40, 41, and 44−50). Here the focus will not be on technical details regarding the individual alignment algorithms but merely on the practical and operational aspects as seen from a user-perspective.

Glancing at the literature, one alignment technique dominates for solving chromatographic shifts: correlation optimized warping (COW).[40,51,52] This is not surprising, as the method is rather simple. COW (illustrated in Figure 11) is a piecewise or segmented data alignment technique that aligns a sample chromatogram toward a *reference chromatogram* by stretching or compressing sample segments.

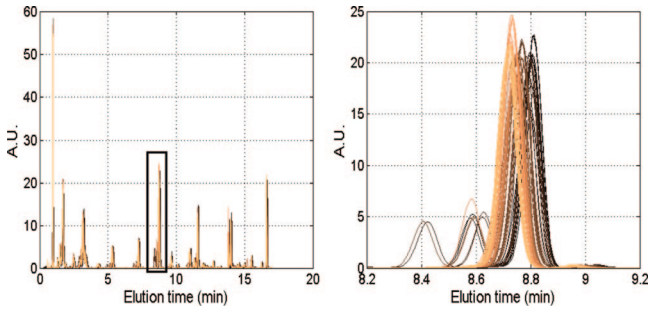

Figure 10. Illustration of unsynchronized chromatographic signals from GC-FID measurements on ground coffee samples.[39,40]

All alignment techniques work by moving certain parts of the chromatographic time axis either "left" or "right". With COW, this movement is done locally, as depicted in Figure 11.

Repositioning the boundaries of the local movements naturally moves the peaks within these boundaries, and thus, alignment is being performed. COW has been extensively

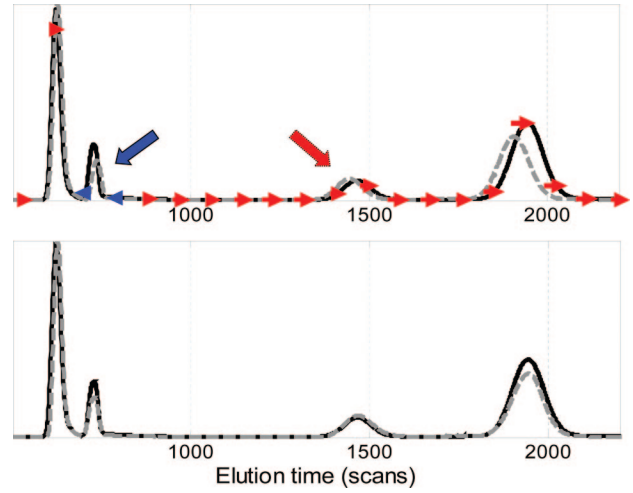

Figure 11. Illustration of alignment of local phenomena using COW. Some parts of the sample chromatogram (gray profile) go left (blue arrow), and some go right (red arrow), in accordance with the shifts to be corrected. Likewise, some parts are moved a lot, whereas other parts are only slightly moved or not moved at all. This makes the alignment method very flexible for unsystematic shifts.

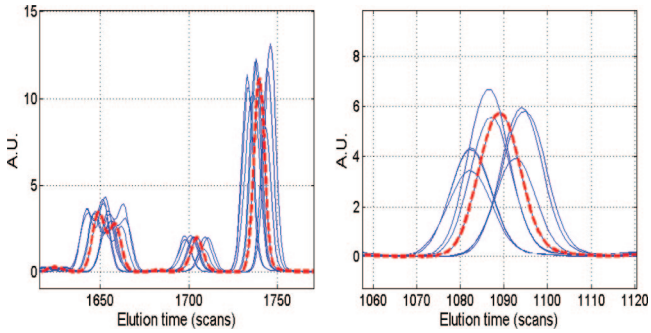

Figure 12. Reference chromatogram selection based on a given set of chromatographic profiles. Here the chromatogram most similar to the rest (red dashed profile) is selected.[39,59] Correcting the misaligned peaks means that some profiles must go left or right, but the flexibility needed is lower than if another chromatogram was selected as reference (e.g., the outmost right).

studied and improved since the first paper in 1998.[52] For gas chromatographic data, COW by Nielsen et al.[52] is by far the most tested alignment technique and is often the technique of choice when comparing new methods[40,45,46,48−50,53−57] or testing improvements.

Properly obtained chromatographic data often need only a constant or linear shift correction, e.g. stretching or shrinking of the whole elution time axis, or simply a movement of the whole chromatogram a certain integer sideways for proper alignment.[58] This is also known as a *systematic* shift. However, if the column is changed between runs, if different chromatographic columns are used, or if samples are measured over a long time, then a more complex shift correction might be needed. This, denoted an *unsystematic* shift, is characterized by a different degree of shift for multiple peaks across samples and can be seen as peaks shifting independently of one another in the same chromatogram (see Figure 11).

For the alignment method to work, it must synchronize the signals and preserve the peak information—i.e. the peak area. These two things go hand in hand. The more flexible the alignment method, the more adjustment is possible, but unfortunately also the higher the risk of introducing artifacts. On the other hand, less flexibility means that only smaller peak shifts can be corrected but then with reduced risk of changing the chromatographic profile. Using a method with low flexibility but still high enough flexibility to correct the shifts present is the goal for all alignment procedures. The best and easiest way to achieve proper alignment with a low flexibility from a given set of chromatographic profiles is by selecting a proper reference (or target) chromatogram (Figure 12). This can be done in several ways, but overall, the goal would be to use the chromatographic profile most similar (or in the middle if systematic shifts) to the other profiles.

When the reference chromatogram has been chosen, the settings of the alignment procedure must be chosen. In Figure 13 two examples of aligning using improper and proper conditions are shown. Obviously, selecting improper conditions is a problem, especially for the subsequent peak identification and quantification. A distinctly shaped profile (the black dotted profile in Figure 13) will always be difficult to align, but when artifacts are found in the normal profiles after alignment, this suggests that the conditions are not selected correctly.

To be able to obtain the best possible conditions for synchronizing signals, several aspects should be considered.

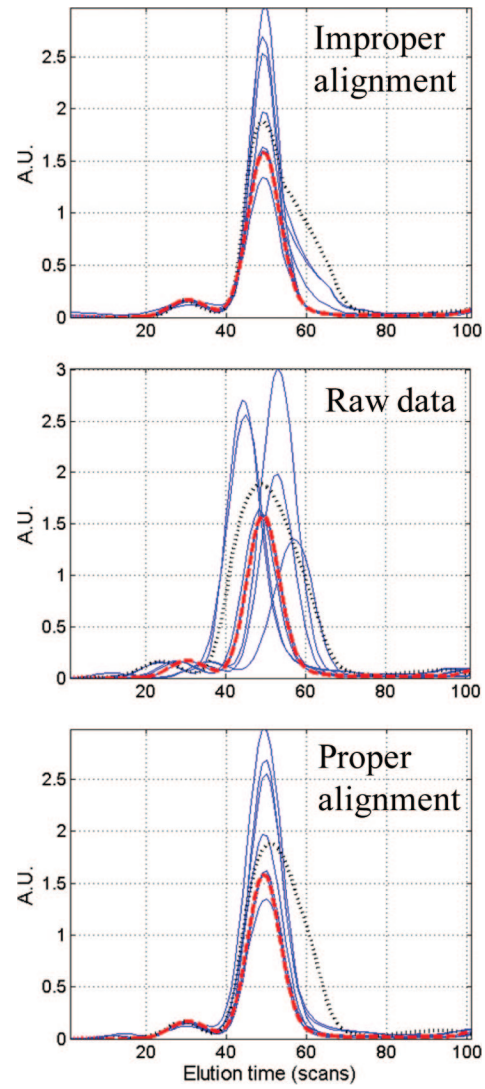

Figure 13. Example of alignment procedures using improper conditions and proper conditions. The raw data is shown in the MIDDLE plot with the improper alignment on TOP and the proper alignment on the BOTTOM. The red dashed profile is the reference chromatogram, and the black dotted profile is a different shaped profile.

(1) The data structure: (a) column characteristics (are all samples run on the same column); (b) deviating samples or peaks (tailing or fronting peaks); (c) sample classes with different peaks (a common reference chromatogram can be difficult to find if too few similar peaks are present across samples); (d) problematic regions in the chromatographic profile (e.g., step temperature profile, overlapping peaks, baseline).

(2) Shift behavior: (a) systematic/unsystematic; (b) large/small shift.

Combined with an appropriate alignment method, insight on the above issues can help define the right settings for alignment and help in checking visually in the right places of the chromatogram. Methods have been presented that guide the inexperienced user to find the proper conditions for alignment, and these principles can be used for any alignment technique.[39,52]

As a rule of thumb, the alignment methods, being rather simple and easy to apply, should always be tried out first. If these are not successful, the partly aligned chromatograms should be visualized to evaluate where problems are still

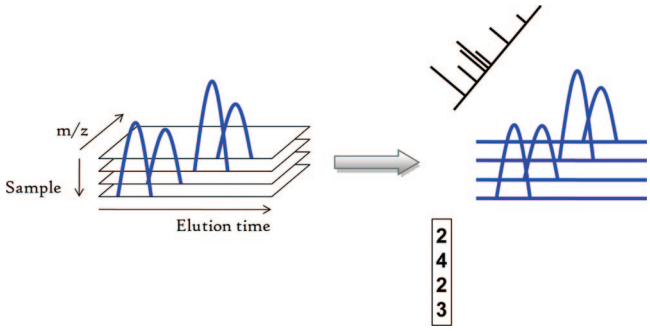

Figure 14. PARAFAC2 visualization of a single analyte found at different elution times across samples but with the same mass spectrum. This simple system can be decomposed into a chemical part consisting of a common mass spectrum, an individual elution profile, and a concentration measure.

observed and whether more flexible alignment methods or parameters are needed. It should also be noted that it might not be possible to align all peaks (at different elution times), even with the most flexible alignment method. This can be the case if a peak is eluting very close to the beginning or the end, where the alignment (reposition of the boundaries) has little chance of correcting larger shifts. Padding noise or random numbers to the ends can often help solve this problem.

If shifted peaks still occur, these might be handled implicitly within a factor model, as discussed in the next section.

3.2.2. Factor Model Approach

The simple factor models presented and used in sections 2 and 3 require that all identical phenomena are located at the same elution time and possess the same mass spectrum across samples. This explicitly implies that *one* elution profile shape and mass spectral fingerprint (the latter is very often the case, so this is not further discussed here) is typical for an analyte for *all* samples. A change in elution time across samples will invalidate this. However, a modified version of the simple factor model called PARAllel FACtor analysis 2 (PARAFAC2)[60−62] can be applied which allows for estimating the elution profile for a chemical analyte separately in each sample (Figure 14). This model allows for elution time shifts and even other complications that will be explained later (section 4).

In the example shown in Figure 9, some small deviations were seen between the peaks in the modeled samples (easily seen in systematic behavior of the residuals). Although small, it will be beneficial to handle these deviations and this can be done with PARAFAC2. Figure 15 shows the same data but with shifted profiles for the same peak in different chromatographic runs.

As mentioned around Figure 7, PARAFAC allows disentangling the chemical information in systems where the data is obtained as a three-way box of data. This can, for example, be data from several samples where each sample is measured chromatographically with spectral detection. For well-aligned peak systems, PARAFAC can describe the information from each chemical analyte even when they overlap. The information obtained from PARAFAC is the set of estimated concentrations/scores, a spectral profile, and an estimate of the elution profile. PARAFAC2, on the other hand, works in a slightly different way. Just as in the case of PARAFAC, each analyte is modeled by the concentrations and a spectral profile, but rather than one elution profile that has to be representative for all samples, *PARAFAC2 extracts as many*

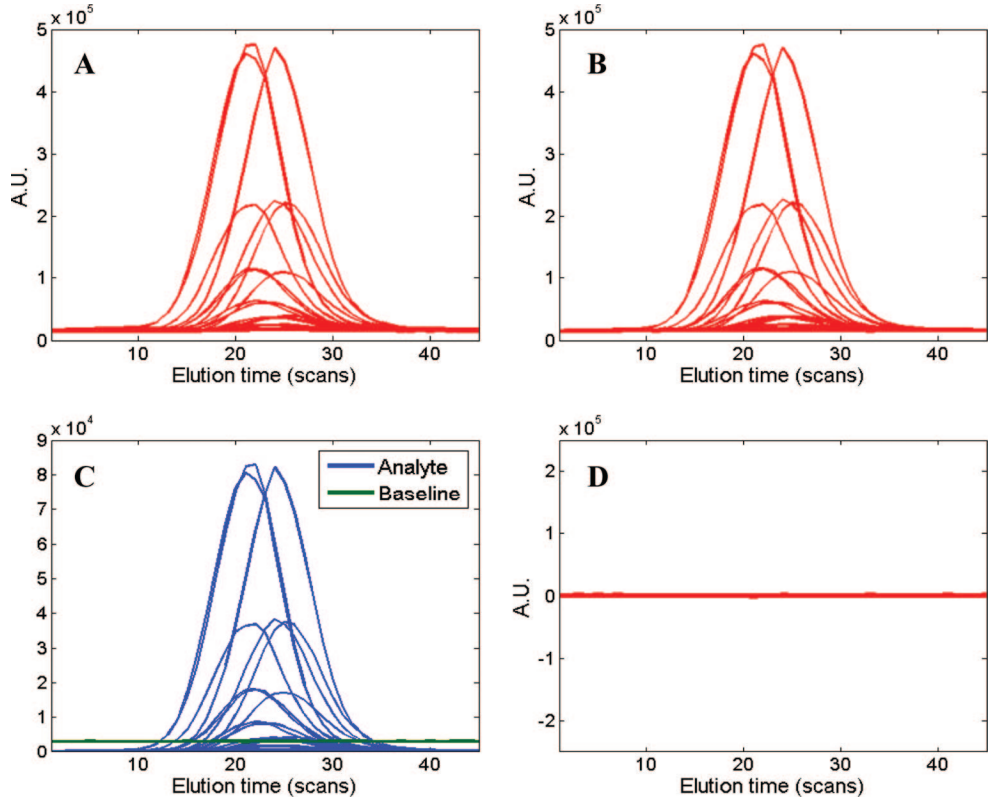

Figure 15. Use of PARAFAC2 to find the chemical parts from a small interval from a GC-MS experiment—illustration of shifted data: (A) raw data from one peak system with two chemical parts (one analyte and one baseline part) and some noise; (B) modeled data with two components; (C) model estimates of the elution time; (D) residuals.

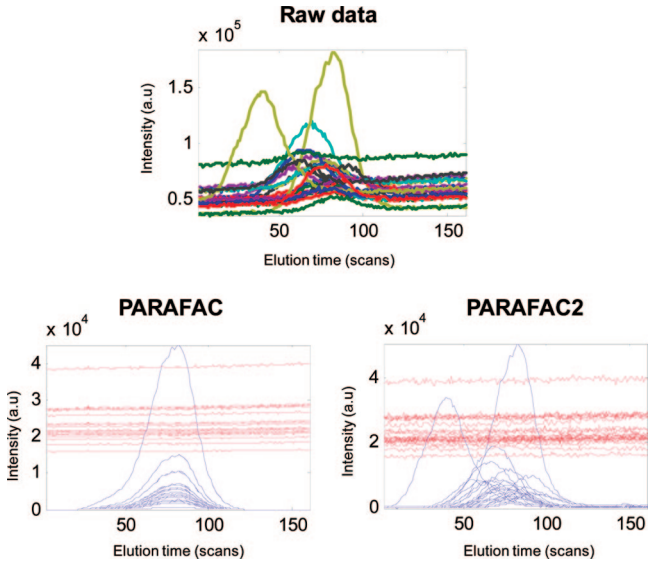

Figure 16. Comparison of PARAFAC and PARAFAC2 for data presented by Amigo et al.[29] The elution profile has been multiplied with the concentration measure for a better visualization of the curves. PARAFAC and PARAFAC2 both find the proper mass spectrum, but PARAFAC2 is by far better at estimating the correct elution profiles.

elution profiles for an analyte as there are samples. Compared to PARAFAC, PARAFAC2 offers a larger flexibility to an observed variation across samples. As only one elution profile is available to describe several shifted profiles, PARAFAC is not a good choice in a situation with severe peak shifts. Simply stated, the estimated profile can be considered as a weighted average of the shifted profiles and, thus, the profiles will often be broader and differently shaped compared to the original peaks (see Figure 16).

Despite being inferior at handling the shifted profiles, PARAFAC often finds a reasonable solution for the spectral mode (despite shifted peaks, the same analyte—i.e. the same mass spectrum—will be present in several scans in different runs if the shift is not too severe), but the main problem is an inferior estimation of the correct elution profile and, due to this, an incorrect concentration estimate.

3.2.3. Limitations and Things To Consider

Dealing with synchronization problems, two approaches are most often used: (1) alignment methods or (2) modeling the shifted behavior. Which method to use depends on the final purpose of the data analysis. These issues are highlighted in Table 2.

4. Co-eluting Peaks

Co-elution is one the most often observed chromatographic artifacts due to high sample complexity, inferior peak capacity, and/or the need for a fast chromatographic separation. There are two main ways of tackling the coelution problem:[64]

(1) *A priori* solution. By improving the classical chromatographic parameters (mobile phase composition and steepness of gradients, temperature, etc.) to separate overlapping peaks (Figure 17). Obviously, this approach is time-consuming and requires expert knowledge, and it is very common to find that other overlapping problems still remain after reprogramming of the methods.

Table 2. Alignment—Things To Consider

purpose after/with alignment	method	things to consider	limitations
Look at profiles (e.g., individual UV spectra, MS spectra, TIC, SIC, etc.)	Alignment methods	How much change in the signal can be accepted? If you look for patterns, then some changes (using, e.g., a very flexible alignment method) can be ok. Always start out with alignment methods no matter the shifted behavior. This will quickly indicate if the problem can be solved quickly or if more flexibility by means of other alignment parameters/methods or factor models is needed.	Be sure that peak shape changes are controlled—either visually or with metrics (ref 36).
Peak quantification	Factor model	Peak regions must be located for proper use of factor models. Try to find regions with five or fewer analytes to be able to use the factor models (see a discussion of this in section 5).	As with all factor models, the rank of the system should be low. As one chemical feature is described in one factor, this means that few chemical features (i.e., peaks) can be included in one model. However, combining model parameters (e.g., scores) afterward is straightforward (discussed in Figure 25).
Characterize peak shifts	Factor model Alignment methods	Look at model parameters or the alignment parameters (warping path[40]).	This is rather complex, and a visual characterization will be much simpler for the inexperienced user.
Local investigation	Alignment methods Factor model	Locate a region with the shifted peaks and either align these locally of apply factor models.	Including a global prealignment will make it easier to find local regions containing the peak(s) of interest.
Identify analytes	Factor model	Compare model parameters with libraries containing mass or UV spectra.	To confirm findings, always run a standard of the identified analyte on your own instrument.
Correct shifts smaller than the data acquisition rate (i.e., less than one scan)	Factor model	Alignment methods can be applied if each sample ($M \times N$) is interpolated to, e.g., five times M ($5M \times N$). This will make it easier to correct the shift.[63]	The increase in data points makes the alignment more time-consuming and, thus, should only be done if this minor shift correction is essential!

Table 3. General Comparison Dealing with Co-Elution Issues

method	main advantages	special requirements	things to consider	limitations
Curve fitting by empirical functions	Simple to use and explain in simple systems.	Selection of the proper shape. Selection of the fitting parameters.	Selection of the proper shape is not trivial.	Only very simple systems can be handled.
Factor models in general	Model artifacts such as baseline drift, sample shifts, etc., without forcing the model to any empirical function. Allows prediction when test sample has unknown interferents.	Works best on local intervals. The number of factors (chemical to use must be estimated).	Only necessary for complex part of the chromatogram. Other parts can often be handled by simpler approaches such as simple integration.	The success of the factor models is related to the simplicity of the local region. This means that too many analytes in the same region can make the model improper, depending on signal-to-noise ratio, similarity of analytes, etc.
PARAFAC	Model simple and easy to interpret. Less complicated than PARAFAC2 and MCR. Unique solution.	The presence of artifacts must be minimized (e.g., shifted peaks) or be consistent between samples (e.g., baseline offset).		
PARAFAC2	It can handle data sets with artifacts (baseline, shifted peaks). Unique solution.	Can be time-consuming due to the increased complexity of the model.	Are there embedded peaks? This may indicate the existence of isomers (difficult to detect).	Sometimes the detection of almost embedded analytes is difficult.
MCR	Faster than three-way models. It can handle only one sample.	Good initial estimates of chromatographs and spectra are needed.		

(2) *A posteriori*. By handling the artifacts after the chromatographic run. The overlapping peaks are resolved into the contributions of the different chemical components, by decomposing the original signal in a sum of different

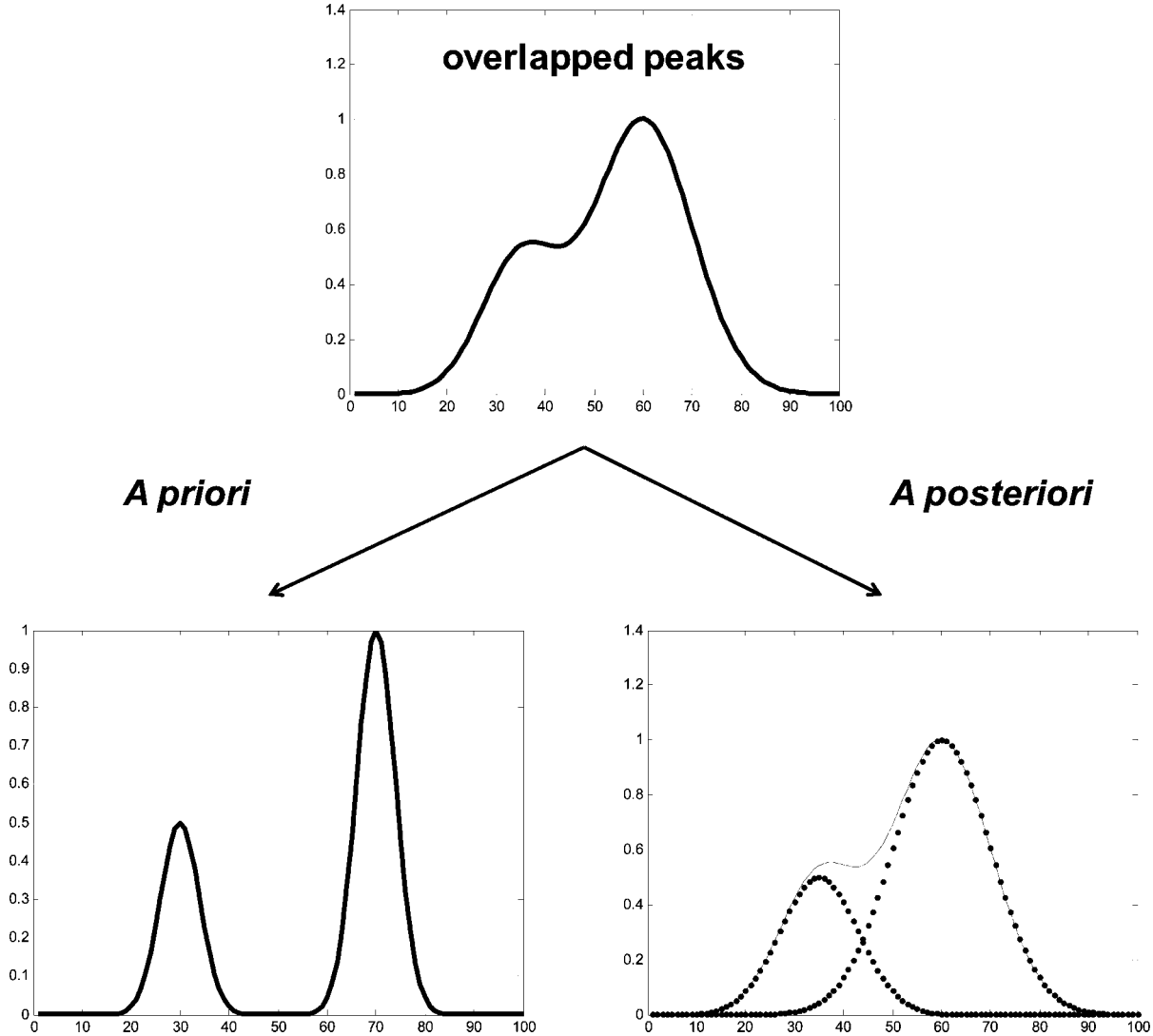

Figure 17. *A priori* and *a posteriori* solutions to solve the problem of two overlapping peaks. To the left, the chromatographic experiment is changed to provide better resolution, and to the right, mathematical modeling is used to resolve the overlapping peaks.

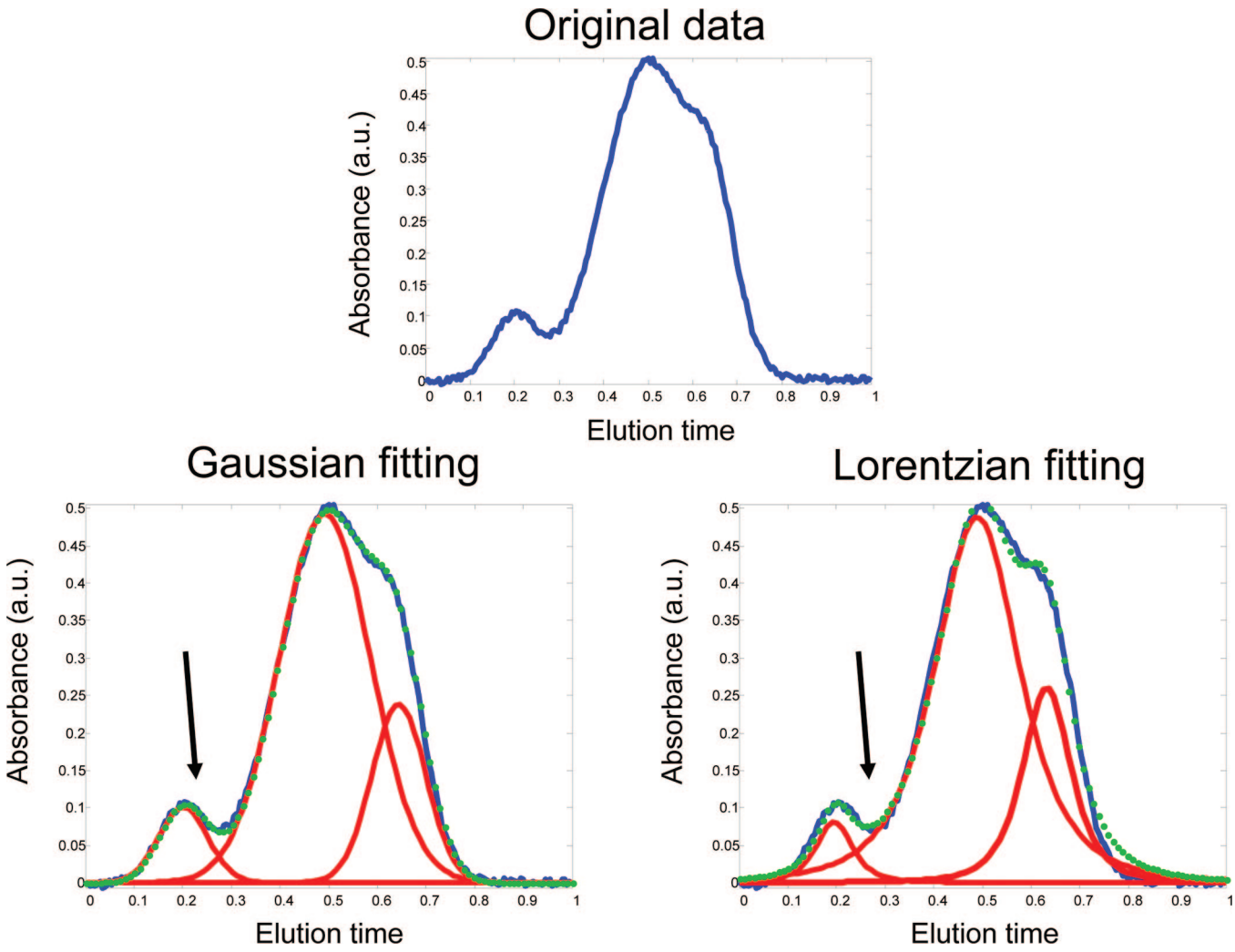

Figure 18. Resolution of a chromatographic profile (upper) where it can be appreciated three overlapped peaks. Bottom: Gaussian and Lorentzian fitting (blue, original data; green, calculated profile; red, the different profiles into which the original one has been divided).

profiles by means of modeling[65] Traditionally, this is done by fitting Gaussian or Lorentzians curves or combinations of these.[65] As an example, the problem in Figure 18 can apparently be easily solved by fitting the original data with Gaussian functions, in such a way that the sum of three Gaussian profiles (green) matches perfectly with the original data.

One of the main problems in this approach of fitting a functional shape is how to choose the proper function and the high number of parameters to be fitted.[66] Looking at the area indicated by the black arrow in Figure 18, it can be noticed that the fitting of the Gaussian and Lorentzian functions (green profile) is almost the same. However, the obtained profile for this analyte is quite different.

The success of the resolution of coeluted peaks is linked to the selective nature of the elution profiles. If one peak is totally embedded (i.e., located just below another peak), then the peak fitting and functional shape modeling become problematic.[65] Figure 19 represents a simulation similar to the one depicted in Figure 18, but now a small third peak has been included almost embedded in the second one. Hence, there are three overlapped peaks. The minor rightmost one is almost embedded. The immediate impression according to the overall chromatographic signal is that there are only two analytes. If three Gaussian are fitted, the minor additional peak is confounded with the major peak and, consequently, the solution obtained may lead to a wrong conclusion, as shown in Figure 19.

Other functions than Gaussian and Lorentzians can also be used based on *a priori* knowledge.[67]

4.1. The Importance of Multichannel Detectors To Solve Overlapping Issues

With multichannel detectors, it is sensible to profit from the spectral information. Factor models can be used to resolve coeluting peaks[21,68] more efficiently than in the monochannel scenario. Going back to the example in Figure 19, the third small analyte can be easily handled considering the whole spectral channel for each elution time and applying factor models. As there is only one sample, it is possible to apply MCR. As shown in Figure 20, the problem is correctly resolved by MCR, obtaining three peaks perfectly modeled as well as their respective spectra.

As already mentioned, multivariate curve resolution can be applied to resolve not only overlapping peaks but also several problems introduced before (low signal-to-noise ratio, baseline drifts, etc.).[17,69−77] Figure 21 illustrates the application of MCR in two situations. In the first one (Figure 21a), the signal-to-noise ratio is very low, which complicates the analysis. MCR, however, is able to estimate not only the two components but also the baseline effect (as well as

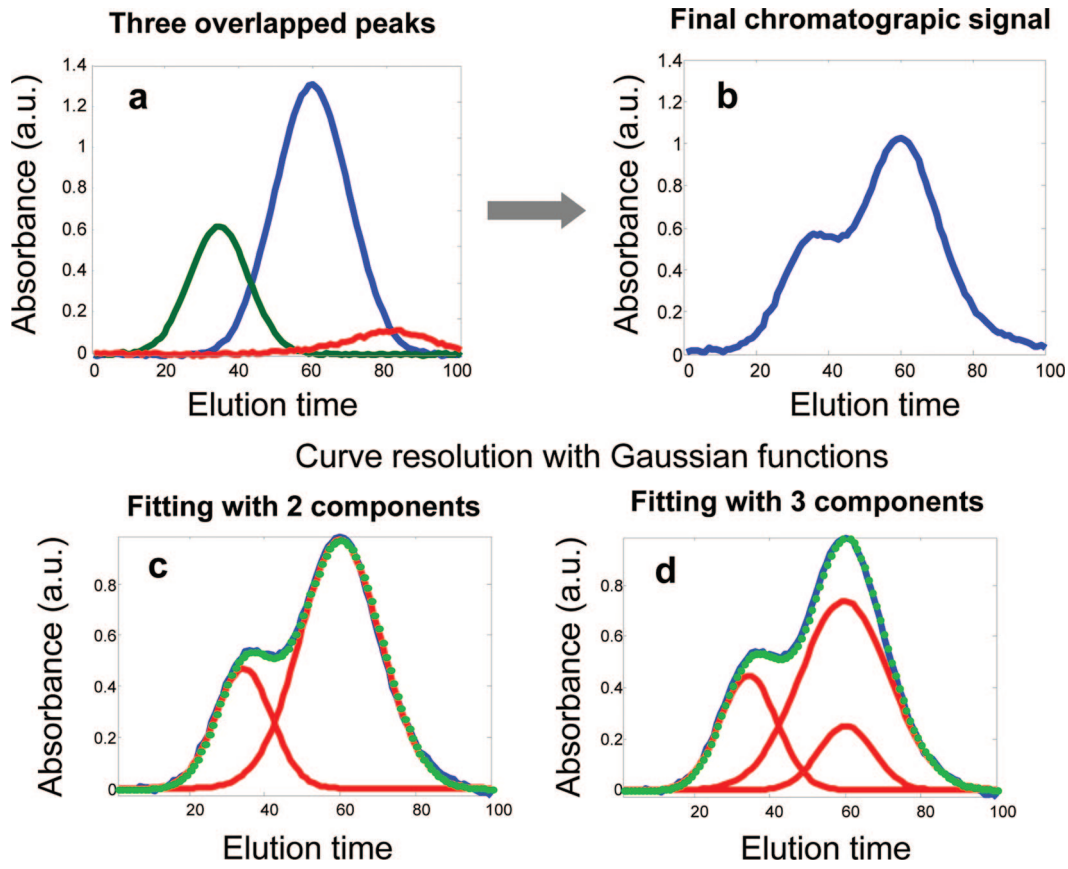

Figure 19. Curve resolution by using Gaussian fitting in a simulated chromatographic run using monochannel detectors: (a) simulated peaks; (b) overall signal. Lower part: Resolution by using Gaussian fitting with (c) two and (d) three Gaussian functions (blue, original data; green, calculated profile; red, peak profiles obtained for each peak).

the pure mass spectra).[29] The baseline effect is related to the mobile phase (carrier), and it can be observed that its elution profile (red) disappears when analyte appears. In the second case (Figure 21a), the problem is even more complicated, as there are two isomers (see Amigo et al.[29]) with practically the same mass spectrum. In this situation, it is even more complicated to find specific ions. Actually, it is even more difficult to detect that there is an overlapping problem. Still, MCR can solve the problem, and from the estimated mass spectra, the existence of two embedded isomers is easily detected.

4.2. Solving Problems in Hyphenated Chromatographic Sets with Many Samples

Resolving overlapping peaks from just one sample will not be possible if the chromatographic profile of a minor component is totally embedded in the peak of a major component or if their spectral patterns are highly similar.[26,69,78,79] Usually though, there is access to more than just one sample and then factor models such as PARAFAC and PARAFAC2 can also be used (see previous sections).

With several samples, it is possible to separate the data into estimated concentrations, spectra, and elution profiles (Figure 22). Unlike the two-way case handled by MCR, three-way models are generally easier to use, as the solution is mathematically more well-defined.[80,81] In MCR, there can be several solutions to the same mathematical problem, but the PARAFAC and PARAFAC2 models are unique, meaning that there is only one solution mathematically. The backside, though, is that the numerical implementation is more complicated in PARAFAC and PARAFAC2 and models can take a long time to estimate.

Figure 22 exemplifies schematically the differences between three-way models (PARAFAC and PARAFAC2) and the so-called augmented version of MCR.[17,26,76]

The augmented version of MCR makes it possible to handle three-way data by rearranging the data from several samples into a matrix. An example of the use of PARAFAC2 and MCR is shown in Figure 23. The visual appearance of the raw data of the chromatographic section selected makes it complicated to assess *a-priori* how many analytes are present and, hence, what number of factors to use in a factor model. It could seem that there is only one chemical analyte plus some baseline. However, both PARAFAC2 and MCR (Figure 23) find two chemical components (one major and one smaller interferent) and the baseline. To identify the components giving rise to the mass spectral profiles obtained for each model, a commercial library can be used (Database 1A, NIST/EPA/NIH Mass Spectral Library with Search Program. NIST 08, Software Version 2.0f, http://www.nist.gov/srd/nist1a.htm) to identify the analytes.

4.3. Limitations and Things To Consider

The power and usefulness of factor models to handle overlapping chromatographic peaks, separating baseline and analytical parts, handle peak shifts, and make it possible to identify analytes from estimated spectral profiles have been amply demonstrated. Nevertheless, for these methods to be successful, certain aspects must be considered.

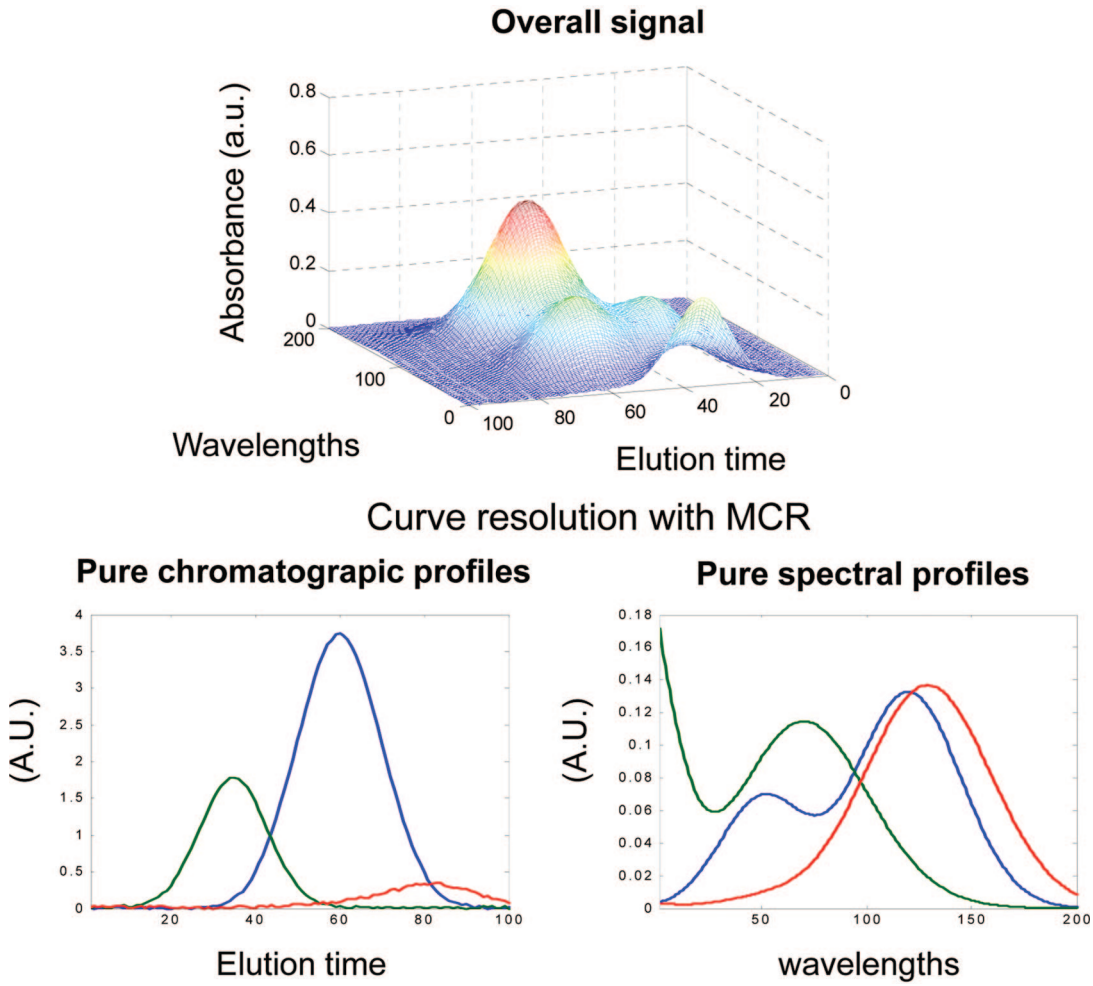

Figure 20. Curve resolution by using MCR. The upper part is the overall signal considering the whole spectral channel. The lower parts are the solutions obtained by using MCR: Pure chromatographic and spectral profiles.

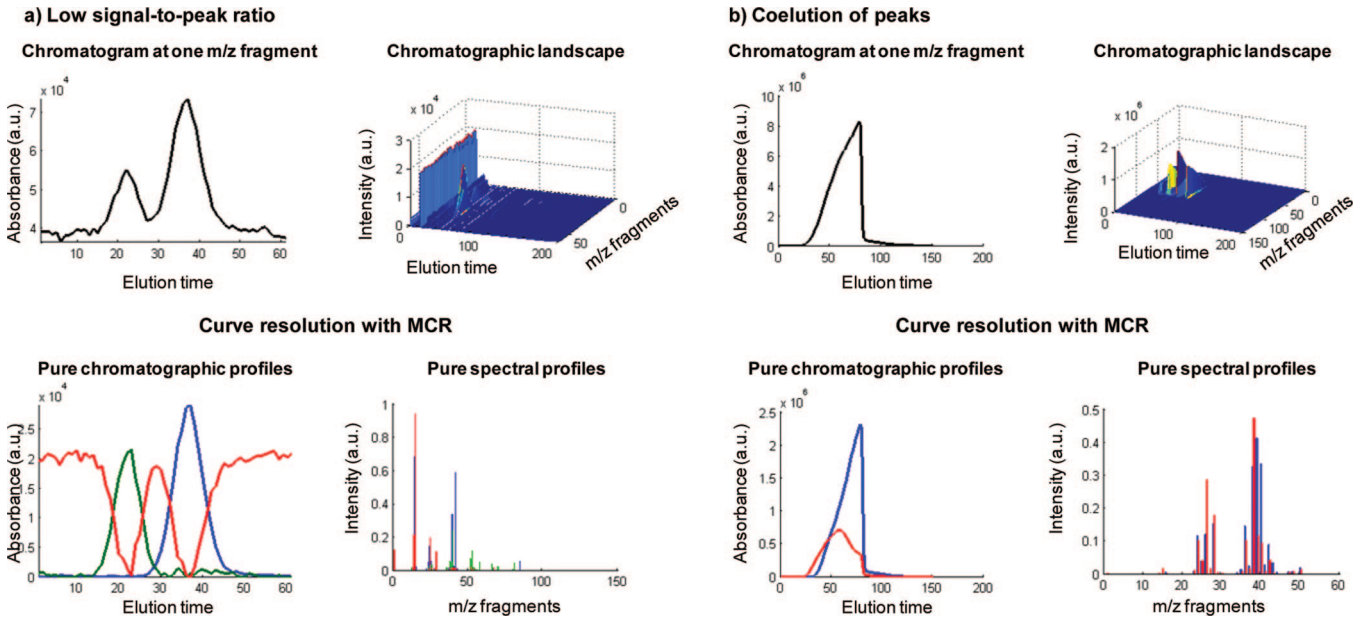

Figure 21. Two examples of the applicability of MCR in the analysis of one individual sample: (a) low signal-to-noise peaks with coeluting problems and (b) highly coeluted peaks (isomers). Upper part, left: Chromatogram at one *m*/*z* fragment illustrated. Upper part, right: Whole chromatographic landscape (one spectrum per elution time). Bottom: Pure chromatographic (left) and pure spectral (right) profiles obtained from MCR. More information on the data can be found in ref 29.

There is a limit to how many factors can be extracted from a given data system in a factor model. To be able to get the best description of the system, it is important and advisible to work locally, e.g., with baseline-resolved parts of the chromatogram. This implies being able to determine relevant peak regions containing a few peaks.[82−85] The complexity

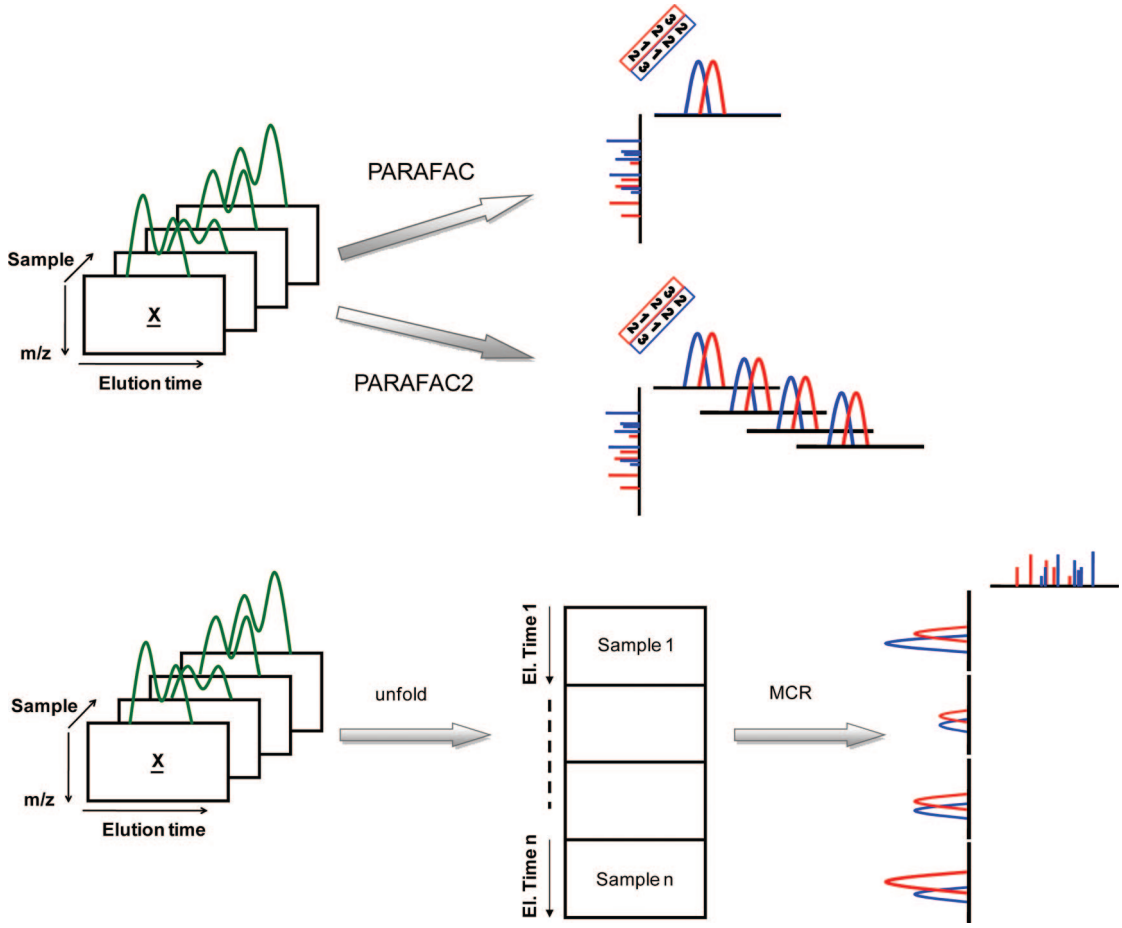

Figure 22. PARAFAC, PARAFAC2, and MCR solution of coeluting problems when more samples are analyzed at the same time.

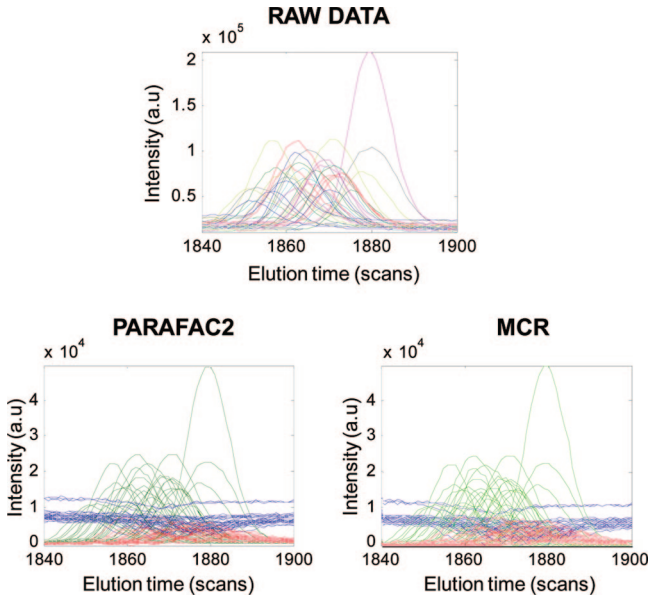

Figure 23. Comparison between PARAFAC2 and MCR results. The raw data is a segment of the GC-MS aroma profile of apples during ripening processes. The data set consisted of 36 samples. The MS spectrum was collected for each elution time.

of these local parts that can be successfully handled depends on the S/N ratio, the similarity of chemical components (both in elution and spectral profiles) in different runs, and the consistency of elution and spectral profiles (shifts) in different runs.

For the presented factor models, determining the proper number of factors is also important. These factors must behave according to the model characteristics—i.e. for PARAFAC, data must be trilinear[62,86] meaning that, for *each factor* estimated, the same elution profile and mass spectral profile multiplied by a certain number should give an adequate description of one chemical feature in each of the samples present. If a peak is absent in one sample, the same elution and mass spectral profile will be valid for this sample, but then the amount will be estimated to be zero.

The selection of a region to analyze can be made using both *a priori* knowledge and intuition by checking the original data and models with different numbers of factors (Figure 24). As a rule of thumb, when there is less noise, larger signals, less similarity, and higher consistency between systematic phenomena in different runs, more complexity (more chemical components) can be handled in local parts of the chromatograms. As an example, the PARAFAC2 model of the data set shown in Figure 24 can be perfectly explained and fitted by two factors (one peak and the baseline). The model with one factor is not defining the baseline correctly (so-called underfitting), whereas the model with three factors introduces a new factor (red one) that is not chemically meaningful. The negative shapes obtained are an indication that the model is using the noise in the data for a fitting variation that is not chemically relevant (overfitting).

It is a common and natural question to consider how many peaks can be resolved by PARAFAC or PARAFAC2. The question, though, is difficult to answer in a meaningful way.

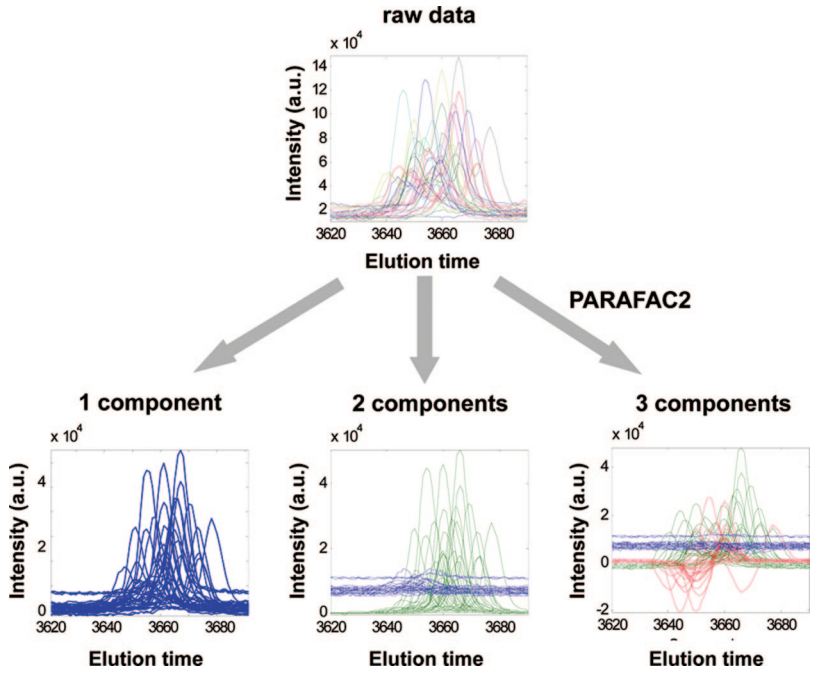

Figure 24. PARAFAC2 analysis of a chromatographic run varying the number of components.

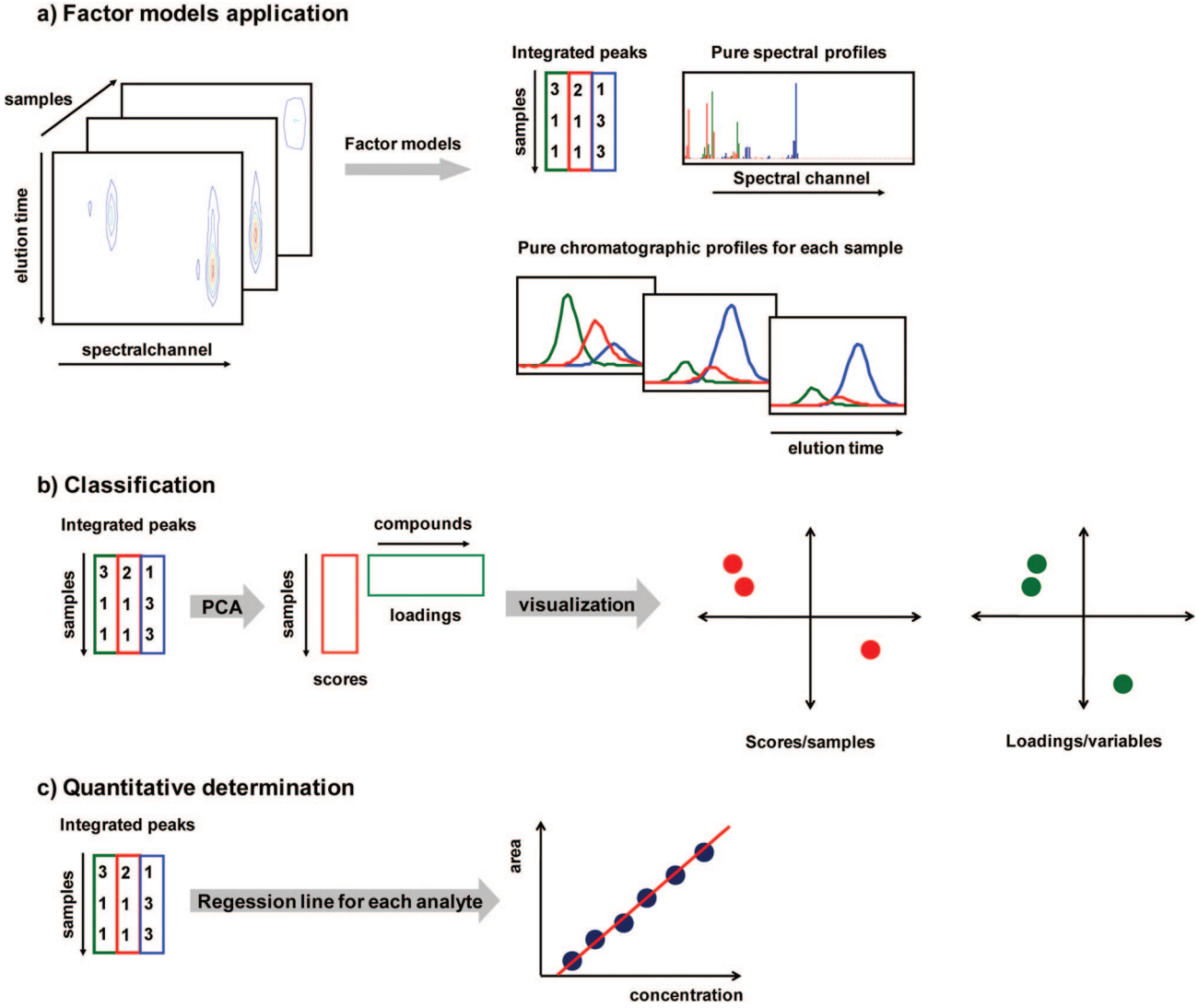

Figure 25. Overall data analysis procedure from raw data in multiple dimensions over the extracted peak areas given in a table to the pattern recognition to see how the samples are connected (b) and also providing quantitative information for any of the components (c).

Table 4. Chromatographic Artifacts[a]

action	one sample	more samples
Noise reduction	+ (filtering)	+ (factor models)
Baseline/background correction	+ (fitting polynomials)	+ (factor models)
Nonstable retention times	+[c]	+ (alignment or factor models)
Peak shape changes	+ (advanced curve fitting)	+ (advanced factor models)
Deconvolution[b]—resolved peaks	+ (curve fitting)	+ (factor models)
Deconvolution[b]—coeluted peaks	+ (curve fitting)	+ (factor models)
Deconvolution[b]—embedded peak	+ if visible (curve fitting)—(if nondetectable)	+ if not completely embedded peaks (factor models)

[a] + indicates that this issue can be handled by mathematical methods (typical method(s) are mentioned in parentheses). [b] Deconvolution means to find the peak area/height of and, if possible, identify the analytes (co)eluting. [c] Nonstable retention times can to some degree be solved by using marker peaks.

Table 5. Main Applications of Factor Models in Chromatographic Systems[a]

HPLC		GC-MS	
DAD		**MS**	
Resolution	Generalized rank annihilation method (GRAM)[137−139]	characterization and classification quantification	MCR[82,99,151]
			Comparison[131,132,152−154]
	PARAFAC[109,140−143]		PARAFAC and PARAFAC2[83,155,156]
	PARAFAC2[61]		*N*-way partial least squares regression (N-PLS)[156]
	MCR[119,144−146]		
	Comparison[86,102,113,128,147,148]		
	Residual bilinearization (RB)[149,150]		
kinetic monitoring	PARAFAC[157,158]		
modeling	MCR-ALS[158]	**GC × GC-MS**	
quantification	second order standard addition method (SOSAM)[159]	quantification and evaluation monitoring and modeling	several methods[173−181]
	GRAM[25,160−163]		PARAFAC[10,11,182]
	MCR[164−168]		GRAM[183]
	PARAFAC[130,133,169,170]		
	Bilinear least squares (BLLS)[171]		
	Comparison between some of them[172]		
	PARAFAC2[134]		
Fluorescence		**CE-DAD**	
Screening and peak purity	PARAFAC[93,184,185]	quantification	MCR[172,186−188]
MS		**LC × LC-DAD**	
Quantification	MCR[189−192]	Separation and quantification	PARAFAC[193,194]
Enzyme kinetics	MCR[195]	Metabonomics	Methods for 4-LC systems[196,197]
metabonomic fingerprinting	PARAFAC[198−200]	Evaluation of selectivity	Several methods[201]
MS and H-1-NMR			
discrimination, classification	Orthogonal projection approach (OPA)[200]		
Raman			
Reaction monitoring	MCR[202]		
ED			
resolution and quantification	MCR[203]		
FT-IR			
Biological samples	MCR[204]		
General applications	Comparison between some of them[205]		

[a] For an explanation of the abbreviations, see Table 6.

For noisefree and well-behaved data, there is almost no limit to how many peaks can be resolved.

In most cases, though, the practical limit is caused by minor artifacts in the specific data. Maybe one sample has an additional very minor compound that is indistinguishable from noise. Or maybe the shape of the elution profile of one compound is slightly different in different runs. Such artifacts are not necessarily detrimental for modeling the major part of the signal well, but they can put severe limits on how many compounds can be extracted. This is so because the more peaks there are to resolve, the smaller the unique part of the smallest peak will be and, hence, the more critical it will be whether the conditions (e.g., that residuals are identically and randomly distributed) are fulfilled. Smaller deviations such as the above-mentioned will therefore often be the limiting factor. In addition to artifacts disturbing the modeling, it is also important how many samples are available and how well the peaks to resolve vary independently in concentration—spectrally and with respect to retention time. In short, in theory, as many peaks can be resolved as needed if the data set supports it, but it is the chemical and physical specifics of the data that normally define the limit.

5. Dealing with Chromatographic Artifacts in Practice

As has been pointed out thus far, chromatographic problems can often be handled by mathematical approaches, some of which work on one sample at a time and some that take all samples into consideration (Table 4).

From a chromatographic point of view, there are four main benefits when applying factor models:

(a) Integration is automated as both interferents and baseline are automatically taken into account.

Table 6. Main References with Theoretical Background and Tutorials Available in the References

technique	acronym	refs
	Baseline Correction	
Asymmetric least-squares	ASL	Eilers[33]
Polynomial fit (vector based)	PFV	Gan et al.[34]
Polynomial fit (matrix based)	PFM	Boelens et al.[36]
		Gemperline et al.[206]
		Liang et al.[38]
Peak Alignment		
Correlation optimized warping	COW	Nielsen et al.[52]
		Tomasi et al.[40]
		Tomasi et al.[41]
		Van Nederkassel et al.[48]
Piecewise alignment	PWA	Forshed et al.[207]
		Lee et al.[208]
		Pierce et al.[53]
Parametric time warping	PTW	Eilers[33]
		Van Nederkassel et al.[48]
Semiparametric time warping	STW	Van Nederkassel et al.[48]
Dynamic time warping	DTW	Furlanello et al.[209]
		Keogh et al.[210]
		Pravdova et al.[46]
		Tomasi et al.[40]
		Tomasi et al.[41]
	Co-eluting Resolution/Peak Purity	
Self-modeling curve resolution	SMCR	Lawton et al.[211]
Principal component analysis	PCA	Davis et al.[212]
Theory of error and factor analysis in chemistry	FA	Malinowski et al.[21,68]
Target factor analysis	TFA	McCue et al.[213]
Evolving factor analysis	EFA	Maeder et al.[64,214]
Fixed size moving window-EFA	FSMW-EFA	Keller et al.[111]
Simple-to-use interactive self-modeling analysis	SIMPLISMA	Windig et al.[215]
Orthogonal projection approach	OPA	Sánchez et al.[106]
Heuristic evolving latent projections	HELP	Kvalheim et al.[96,216]
Iterative target transformation factor analysis	ITTFA	Gemperline et al.;[73] Vandeginste et al.[74]
Curve resolution	CR	Sharaf et al.[15,16]
Multivariate curve resolution	MCR	Osten et al.;[75] Tauler et al.[17,76]
Rank annihilation factor analysis	RAFA	Ho et al.[115]
		McCue et al.[217]
Generalized rank annihilation method	GRAM	Sánchez et al.[139,218,219]
Direct trilinear decomposition	DTLD	Sánchez et al.[220]
Parallel factor analysis	PARAFAC	Harshman;[19] Carrol et al.;[20] Bro et al.[62,80,81]
PARAFAC2	PARAFAC2	Bro et al.[60−62,81]
Shifted factor models		Harsman et al.[221]

(b) Detection and also modeling of interferents without the need for any complementary tool. It is not necessary to perform a time-consuming search for specific ions for each analyte.

(c) Handling of more complex situations.

(d) Run-time analysis can be shortened by helping with resolving overlapped peaks.

One disadvantage is that currently application of factor models requires some data analysis skills. Nevertheless, there are an increasing number of software packages giving easier access to the use of these models,[87−89] with some of them working under MATLAB software.[90]

So far, all factor models have been used to get well resolved chromatographic and spectral profiles that are directly related to the peak area (Figure 25a). These peak areas can then easily be set up in a table as visualized in Figure 25b. The use of this quantitative information depends on the aim of the analysis, but for example, analysis of variance (ANOVA) and pattern recognition techniques are now all valid as long as the individual columns contain the same information in all samples (i.e., the concentration of a specific analyte). Neither coelution nor baseline is affecting the data. Therefore, we end up with data that are free of chromatographic artifacts, allowing the visual interpretation and classification of the samples accordingly (Figure 25b) or, if a previous calibration was prepared, allowing us to obtain quantitative information for each analyte (Figure 25c).

Sometimes it is only needed to check if a peak is pure. Factor models can be applied for exploratory purposes to check the purity of a peak, determining the number of overlapped components and suggesting an elution profile of each component.[64,91−100] This is also named peak purity and is a very important issue in pharmacological research and in other areas.[101−109 110−114]

As we have mentioned in previous sections, one of the main benefits of factor models is its use for calibration purposes (Figure 25c). Since the first factor model reported in the literature that allows the quantitative analysis in the presence of several unknown interferents,[115−117] a number of authors have refined the information that can be obtained under various scenarios (precision and bias,[118−120] some common errors in calibration,[121−127] definition of limits of detection,[128] standardization of analytical methods[13,129]) to improve the quantitative results. As an example, Ortiz and co-workers focused on validating routines of analysis and calibration methods by using three-way factor models.[13] In their work, they studied the possibility of creating validation models and their prediction capability as well as other quality parameters according to legal prerequisites and different quality norms.[13,83,130−134]

The application fields of chemometrics for chromatographic analyses are very diverse, as evidenced in the nonexclusive list shown in Table 5. The table also highlights that there are several other methods than those specifically

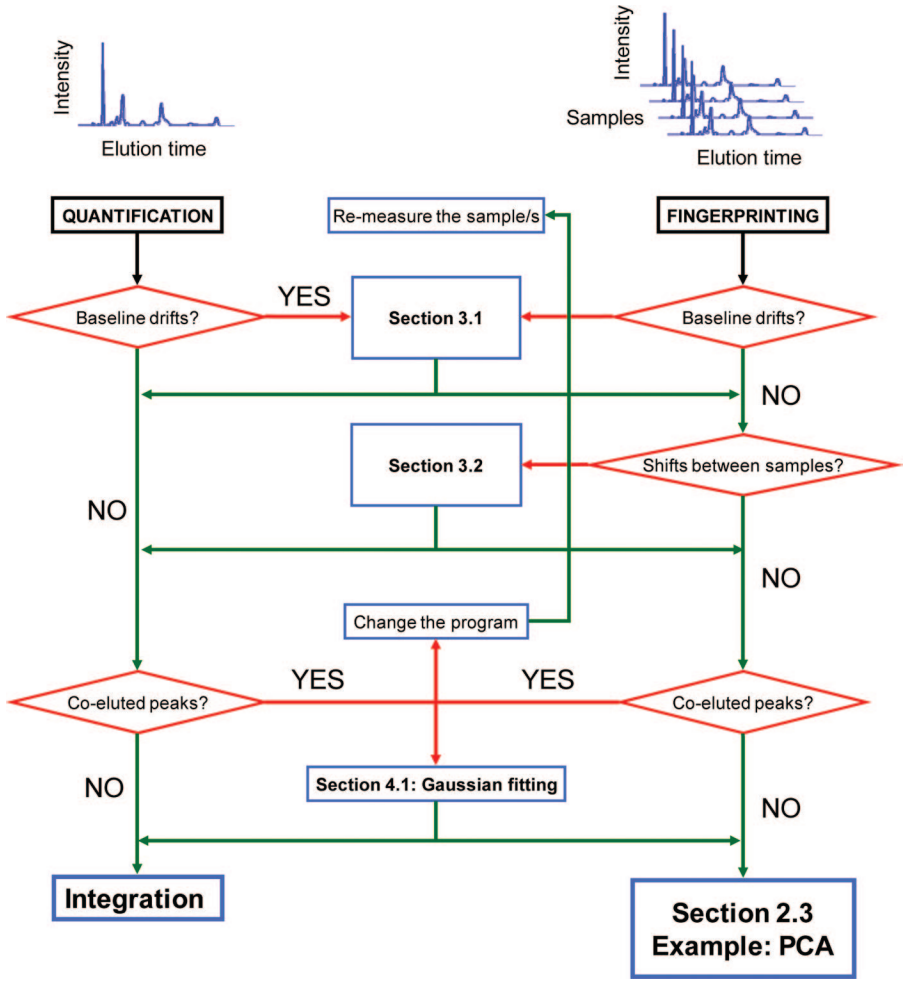

Figure 26. Actuation flow-chart when only monochannel detectors (e.g., FID, SIM) are available.

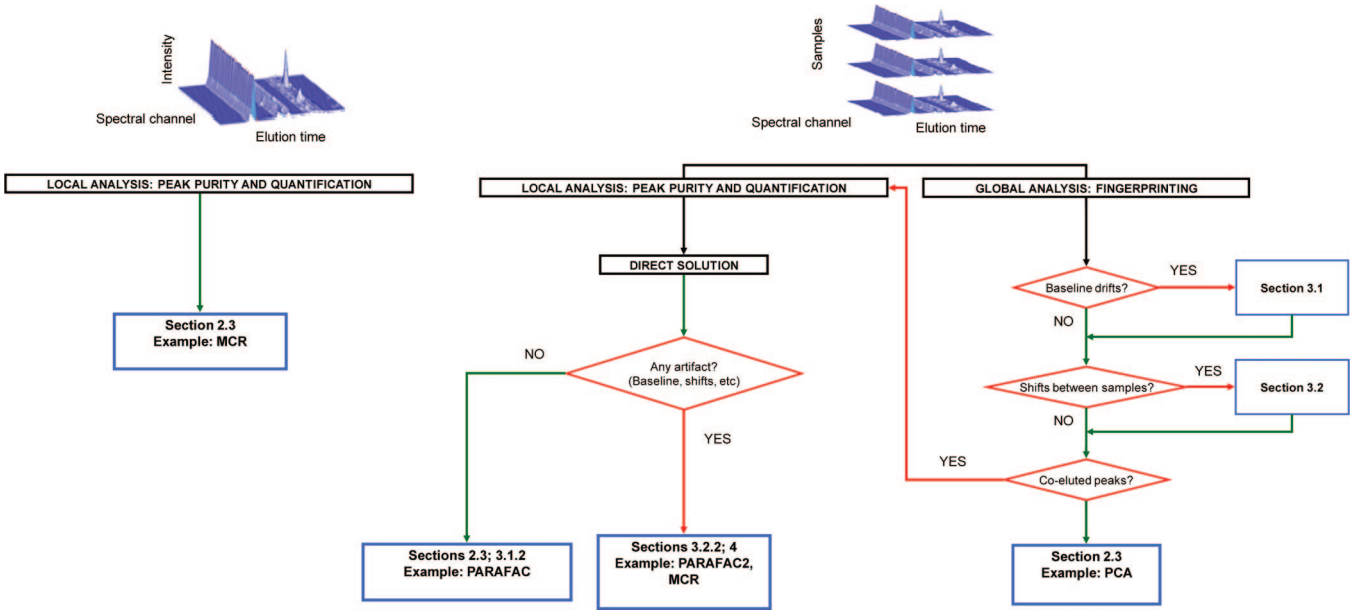

Figure 27. Actuation flow-chart when multichannel detectors (e.g., MS, florescence, UV) are available.

mentioned in this paper. However, most of them can be categorized into the types of methods discussed here.[135,136] Thus, the purpose of providing these many new methods here in Table 5 is not so much to impress with a long list of acronyms but to provide the reader with pointers to interesting applications. These papers may be using methods with unfamiliar names, but those methods are closely related or similar to the methods described in this paper.

To further detail the number of methods developed for and applied to chromatographic problems, Table 6 categorizes different algorithms and models and adds key references.

Several tools have been put forward to deal with chromatographic artifacts. Deciding which methods to apply for a certain data set requires some experience. Figures 26 and 27 show flow-charts of how such a selection of methods can be carried out. The use of each individual method depends on two issues: The final purpose of the analysis and the dimensionality of the detector (availability of monochannel and/or multichannel detectors). The charts include the most generic preprocessing and factor models that will be suitable for most chromatographic systems. If special needs arise, then more specific and dedicated methods can be applied, and some of these are briefly mentioned in Table 6.

Chemometrics in the form of preprocessing methods and factor models offers a multitude of solutions to common chromatographic problems/artifacts. This review has highlighted some of the more important and generic ones. What is needed now is that chromatographic users and software as well as hardware developers start collaborating more closely with chemometricians to bring these potential solutions into the hands of applied chromatographers.[222,223] This could be seen as state-of-the-art building blocks of chromatographic equipment (pumps, column, mobile phase, etc.) being made and optimized according to the needs prescribed by subsequently applied chemometric methods. Within this, the intermediate step, the chromatographers that use the equipment and transform a sample into a lot of numbers must be aware of where one can benefit the most when fine-tuning the individual method parameters (i.e., flow, temperature, etc.). This, with respect to both chromatographic equipment possibilities and what can be handled by subsequent chemometrics.

Soon, baseline and alignment methods and factor models should be included in commercial chromatographic software packages—we are not there yet, but we are closer than ever!

6. List of Symbols and Acronyms

Symbols

$\mathbf{X}$	bold uppercase denotes a matrix
$\mathbf{a}_i$	bold lowercase denotes vectors, and the subscript denotes the number of the analyte (that goes from 1 to i)
$\mathbf{b}^T$	superscript T denotes the transpose of the vector $\mathbf{b}$
$\circ$	tensor product
$\underline{\mathbf{X}}$	bold uppercase and underlined denotes a three way array

Acronyms

ALS	alternating least squares
ANOVA	analysis of variance
ASL	asymmetric least squares
BLLS	bilinear least squares
COW	correlation optimized warping
CR	curve resolution
DAD-UV	diode-array ultraviolet−visible
DTLD	direct trilinear decomposition
DTW	dynamic time warping
EFA	evolving factor analysis
FA	factor analysis
FSW-EFA	fixed size moving window-EFA
FID	flame ionization detector
FTIR	Fourier-transform infrared
GC	gas chromatography
GRAM	generalized rank annihilation method
HELP	heuristic evolving latent projections
HPLC	high performance liquid chromatography
ILDA-FLU	intensified linear diode array fluorescence
ITTFA	iterative target transformation factor analysis
LC	liquid chromatography
MS	mass spectrometry
MCR	multivariate curve resolution
NMR	nuclear magnetic resonance
N-PLS	*N*-way partial least squares regression
OPA	orthogonal projection approach
PARAFAC	parallel factor analysis
PARAFAC2	parallel factor analysis 2
PCA	principal component analysis
PFM	polynomial fit (matrix based)
PFV	polynomial fit (vector based)
PTW	parametric time warping
PWA	piecewise alignment
RAFA	rank annihilation factor analysis
RB	residual bilinearization
SIC	single ion (count) chromatogram
SIM	single ion monitoring
SIMPLISMA	simple-to-use interactive self-modeling analysis
SMCR	self-modeling curve resolution
S/N	signal-to-noise ratio
SOSAM	second order standard addition method
STW	semiparametric time warping
TFA	target factor analysis
TIC	total ion (count) chromatogram
TOF	time of flight
UV−vis	UV−visible monochannel detector

7. References

(1) Hirschfeld, T. *Anal. Chem.* **1980**, *52*, A297.
(2) Wilson, I. D. *Am. Lab.* **2002**, 13.
(3) Wilson, I. D.; Brinkman, U. A. T. *TrAC, Trends Anal. Chem.* **2007**, *26*, 847.
(4) Wilson, I. D.; Brinkman, U. A. T. *J. Chromatogr., A* **2003**, *1000*, 325.
(5) Brinkman, U. A. T. *J. Chromatogr., A* **1999**, *856*, 1.
(6) Stoll, D. R.; Cohen, J. D.; Carr, P. W. *J. Chromatogr., A* **2006**, *1122*, 123.
(7) Stoll, D. R.; Li, X. P.; Wang, X. O.; Carr, P. W.; Porter, S. E. G.; Rutan, S. C. *J. Chromatogr., A* **2007**, *1168*, 3.
(8) Hyotylainen, T.; Riekkola, M. L. *J. Chromatogr., A* **2003**, *1000*, 357.
(9) Tranchida, P.; Donato, P.; Dugo, P.; Dugo, G.; Mondello, L. *TrAC, Trends Anal. Chem.* **2007**, *26*, 191.
(10) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Pierce, K. M.; Young, E. T.; Synovec, R. E. *Analyst* **2007**, *132*, 756.
(11) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 2700.
(12) Xu, X.; van Stee, L. L. P.; Williams, J.; Beens, J.; Adahchour, M.; Vreuls, R. J. J.; Brinkman, U. A. T.; Lelieveld, J. *Atmos. Chem. Phys.* **2003**, *3*, 665.
(13) Ortiz, M. C.; Sarabia, L. *J. Chromatogr., A* **2007**, *1158*, 94.
(14) Daszykowski, M.; Walczak, B. *TrAC, Trends Anal. Chem.* **2006**, *25*, 1081.
(15) Sharaf, M. A.; Kowalski, B. R. *Anal. Chem.* **1982**, *54*, 1291.
(16) Sharaf, M. A.; Kowalski, B. R. *Anal. Chem.* **1981**, *53*, 518.
(17) Tauler, R.; Barcelo, D. *TrAC, Trends Anal. Chem.* **1993**, *12*, 319.
(18) Tauler, R.; Durand, G.; Barcelo, D. *Chromatographia* **1992**, *33*, 244.
(19) Harshman, R. A. *UCLA working papers in phonetics* **1970**, *16*, 1.
(20) Carrol, J. D.; Chang, J. *Psychometrika* **1970**, *35*, 283.
(21) Malinowski, E. R. *Factor Analysis in Chemistry*; John Wiley & Sons: New York, 2002.
(22) Vaidya, R. A.; Hester, R. D. *J. Chromatogr.* **1984**, *287*, 231.
(23) Lindberg, W.; Ohman, J.; Wold, S. *Anal. Chem.* **1986**, *58*, 299.
(24) Bro, R. *Anal. Chim. Acta* **2003**, *500*, 185.
(25) Comas, E.; Gimeno, R. A.; Ferre, J.; Marce, R. M.; Borrull, F.; Rius, F. X. *J. Chromatogr., A* **2004**, *1035*, 195.
(26) de Juan, A.; Tauler, R. *J. Chromatogr., A* **2007**, *1158*, 184.
(27) Arancibia, J. A.; Olivieri, A. C.; Escandar, G. M. *Anal. Bioanal. Chem.* **2002**, *374*, 451.
(28) Boque, R.; Ferre, J. *LC GC Eur.* **2004**, *17*, 402.
(29) Amigo, J. M.; Skov, T.; Coello, J.; Maspoch, S.; Bro, R. *TrAC, Trends Anal. Chem.* **2008**, *27*, 714.
(30) Bijlsma, S.; Bobeldijk, L.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2006**, *78*, 567.

(31) Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Oresic, M. *BMC Bioinf.* **2007**, 8.
(32) Eilers, P. H. C.; Marx, B. D. *Stat. Sci.* **1996**, *11*, 89.
(33) Eilers, P. H. C. *Anal. Chem.* **2004**, *76*, 404.
(34) Gan, F.; Ruan, G. H.; Mo, J. Y. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 59.
(35) Ruckstuhl, A. F.; Jacobson, M. P.; Field, R. W.; Dodd, J. A. *J. Quant. Spectrosc. Radiat. Transfer* **2001**, *68*, 179.
(36) Boelens, H. F. M.; Dijkstra, R. J.; Eilers, P. H. C.; Fitzpatrick, F.; Westerhuis, J. A. *J. Chromatogr., A* **2004**, *1057*, 21.
(37) Kaczmarek, K.; Walczak, B.; de Jong, S.; Vandeginste, B. G. M. *Acta Chromatogr.* **2005**, *15*, 82.
(38) Liang, Y. Z.; Kvalheim, O. M.; Rahmani, A.; Brereton, R. G. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 265.
(39) Skov, T.; van den Berg, F.; Tomasi, G.; Bro, R. *J. Chemom.* **2006**, *20*, 484.
(40) Tomasi, G.; van den Berg, F.; Andersson, C. *J. Chemom.* **2004**, *18*, 231.
(41) Tomasi, G. University of Copenhagen, 2006.
(42) Vandenbogaert, M.; Li-Thiao-Te, S.; Kaltenbach, H. M.; Zhang, R. X.; Aittokallio, T.; Schwikowski, B. *Proteomics* **2008**, *8*, 650.
(43) America, A. H. P.; Cordewener, J. H. G. *Proteomics* **2008**, *8*, 731.
(44) Christensen, J. H.; Mortensen, J.; Hansen, A. B.; Andersen, O. *J. Chromatogr., A* **2005**, *1062*, 113.
(45) Hendriks, M. M. W. B.; Cruz-Juarez, L.; De Bont, D.; Hall, R. D. *Anal. Chim. Acta* **2005**, *545*, 53.
(46) Pravdova, V.; Walczak, B.; Massart, D. L. *Anal. Chim. Acta* **2002**, *456*, 77.
(47) Szymanska, E.; Markuszewski, M. J.; Capron, X.; Van Nederkassel, A. M.; Heyden, Y. V.; Markuszewski, M.; Krajka, K.; Kaliszan, R. *Electrophoresis* **2007**, *28*, 2861.
(48) van Nederkassel, A. M.; Daszykowski, M.; Eilers, P. H. C.; Heyden, Y. V. *J. Chromatogr., A* **2006**, *1118*, 199.
(49) van Nederkassel, A. M.; Xu, C. J.; Lancelin, P.; Sarraf, M.; MacKenzie, D. A.; Walton, N. J.; Bensaid, F.; Lees, M.; Martin, G. J.; Desmurs, J. R.; Massart, D. L.; Smeyers-Verbeke, J.; Vander Heyden, Y. *J. Chromatogr., A* **2006**, *1120*, 291.
(50) Walczak, B.; Wu, W. *Chemom. Intell. Lab. Syst.* **2005**, *77*, 173.
(51) Bylund, D.; Danielsson, R.; Malmquist, G.; Markides, K. E. *J. Chromatogr., A* **2002**, *961*, 237.
(52) Nielsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr., A* **1998**, *805*, 17.
(53) Pierce, K. M.; Hope, J. L.; Johnson, K. J.; Wright, B. W.; Synovec, R. E. *J. Chromatogr., A* **2005**, *1096*, 101.
(54) Pierce, K. M.; Wright, B. W.; Synovec, R. E. *J. Chromatogr., A* **2007**, *1141*, 106.
(55) Torgrip, R. J. O.; Aberg, M.; Karlberg, B.; Jacobsson, S. P. *J. Chemom.* **2003**, *17*, 573.
(56) van Nederkassel, A. M.; Daszykowski, M.; Massart, D. L.; Vander Heyden, Y. *J. Chromatogr., A* **2005**, *1096*, 177.
(57) Yao, W. F.; Yin, X. Y.; Hu, Y. Z. *J. Chromatogr., A* **2007**, *1160*, 254.
(58) Van den Berg, F.; Tomasi, G.; Viereck, N. *Magn. Reson. Food Sci.: The Multivariate Challenge* **2005**, 131.
(59) Daszykowski, M.; Walczak, B. *J. Chromatogr., A* **2007**, *1176*, 1.
(60) Kiers, H. A. L.; Ten Berge, J. M. F.; Bro, R. *J. Chemom.* **1999**, *13*, 275.
(61) Bro, R.; Andersson, C. A.; Kiers, H. A. L. *J. Chemom.* **1999**, *13*, 295.
(62) Bro, R. University of Amsterdam, 1998.
(63) Johnson, K. J.; Prazen, B. J.; Young, D. C.; Synovec, R. E. *J. Sep. Sci.* **2004**, *27*, 410.
(64) Maeder, M. *Anal. Chem.* **1987**, *59*, 527.
(65) Grimalt, J.; Iturriaga, H.; Tomas, X. *Anal. Chim. Acta* **1982**, *139*, 155.
(66) King, M. D.; King, G. S. *Anal. Chem.* **1985**, *57*, 1049.
(67) *PeakFit*, Systat Software Inc.: 2010; http://www.sigmaplot.com/products/peakfit/peakfit.php, last accessed March.
(68) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 606.
(69) Manne, R.; Grande, B. V. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 35.
(70) Mason, C.; Maeder, M.; Whitson, A. *Anal. Chem.* **2001**, *73*, 1587.
(71) Karjalainen, E. J. *Chemom. Intell. Lab. Syst.* **1989**, *7*, 31.
(72) Karjalainen, E. J.; Karjalainen, U. P. *Chemom. Intell. Lab. Syst.* **1992**, *14*, 423.
(73) Gemperline, P. J. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 206.
(74) Vandeginste, B. G. M.; Derks, W.; Kateman, G. *Anal. Chim. Acta* **1985**, *173*, 253.
(75) Osten, D. W.; Kowalski, B. R. *Anal. Chem.* **1984**, *56*, 991.
(76) Tauler, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133.
(77) Rodriguez-Rodriguez, C.; Amigo, J. M.; Coello, J.; Maspoch, S. *J. Chem. Educ.* **2007**, *84*, 1190.
(78) Brereton, R. G.; Dunkerley, S. *Analyst* **1999**, *124*, 705.
(79) Manne, R. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 89.
(80) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149.
(81) Smilde, A. K.; Bro, R.; Geladi, P. *Multi-Way Analysis. Applications in the Chemical Sciences*; Wiley: Chichester, 2004.
(82) Ebrahimi, D.; Li, J. F.; Hibbert, D. B. *J. Chromatogr., A* **2007**, *1166*, 163.
(83) Arroyo, D.; Ortiz, M. C.; Sarabia, L. A. *J. Chromatogr., A* **2007**, *1157*, 358.
(84) Comas, E.; Ferre, J.; Rius, F. X. *Anal. Chim. Acta* **2004**, *515*, 23.
(85) Johnson, K. J.; Rose-Pehrsson, S. L.; Morris, R. E. *Energy Fuels* **2004**, *18*, 844.
(86) de Juan, A.; Tauler, R. *J. Chemom.* **2001**, *15*, 749.
(87) *PLS-Toolbox*, version 3.5; Eigenverctor Research: WA.
(88) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2005**, *76*, 101.
(89) http://www.models.life.ku.dk/source.
(90) *MatLab(R)*, version 7.0; The Matworks, Inc.: MA.
(91) Maeder, M.; Zilian, A. *Chemom. Intell. Lab. Syst.* **1988**, *3*, 205.
(92) Liang, Y. Z.; Kvalheim, O. M. *Chemom. Intell. Lab. Syst.* **1993**, *20*, 115.
(93) Poe, R. B.; Rutan, S. C. *Anal. Chim. Acta* **1993**, *283*, 845.
(94) Cren, H.; Lechevrel, C.; Roussel, G.; Lemoisson, E.; Goussard, J. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **1993**, *615*, 23.
(95) Li, H.; Hou, J. R.; Wang, K.; Zhang, F. *Talanta* **2006**, *70*, 336.
(96) Liang, Y. Z.; Kvalheim, O. M.; Keller, H. R.; Massart, D. L.; Kiechle, P.; Erni, F. *Anal. Chem.* **1992**, *64*, 946.
(97) Malinowski, E. R. *J. Chemom.* **1992**, *6*, 29.
(98) Schostack, K. J.; Malinowski, E. R. *Chemom. Intell. Lab. Syst.* **1993**, *20*, 173.
(99) Jalali-Heravi, M.; Zekavat, B.; Sereshti, H. *J. Chromatogr., A* **2007**, *1143*, 215.
(100) Jiang, J. H.; Sasic, S.; Yu, R. Q.; Ozaki, Y. *J. Chemom.* **2003**, *17*, 186.
(101) Sanchez, F. C.; Khots, M. S.; Massart, D. L.; Debeer, J. O. *Anal. Chim. Acta* **1994**, *285*, 181.
(102) Frenich, A. G.; Torres-Lapasio, J. R.; De Braekeleer, K.; Massart, D. L.; Vidal, J. L. M.; Galera, M. M. *J. Chromatogr., A* **1999**, *855*, 487.
(103) Grande, B. V.; Manne, R. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 19.
(104) Wasim, M.; Brereton, R. G. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 133.
(105) Kuczynska, A.; Walczak, B. *Acta Chromatogr.* **2002**, *12*, 11.
(106) Sanchez, F. C.; Toft, J.; van den Bogaert, B.; Massart, D. L. *Anal. Chem.* **1996**, *68*, 79.
(107) Sanchez, F. C.; vandenBogaert, V.; Rutan, S. C.; Massart, D. L. *Chemom. Intell. Lab. Syst.* **1996**, *34*, 139.
(108) Sanchez, F. C.; Khots, M. S.; Massart, D. L. *Anal. Chim. Acta* **1994**, *290*, 249.
(109) Wiberg, K. *J. Chromatogr., A* **2006**, *1108*, 50.
(110) Lincoln, D.; Fell, A. F.; Anderson, N. H.; England, D. *J. Pharm. Biomed. Anal.* **1992**, *10*, 837.
(111) Keller, H. R.; Massart, D. L. *Anal. Chim. Acta* **1991**, *246*, 379.
(112) Gilliard, J. A.; Cumps, J. L.; Tilquin, B. L. *Chemom. Intell. Lab. Syst.* **1993**, *21*, 235.
(113) De Braekeleer, K.; de Juan, A.; Massart, D. L. *J. Chromatogr., A* **1999**, *832*, 67.
(114) De Braekeleer, K.; Torres-Lapasio, J. R.; Massart, D. L. *Chemom. Intell. Lab. Syst.* **2000**, *52*, 45.
(115) Ho, C. N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1978**, *50*, 1108.
(116) Lorber, A. *Anal. Chem.* **1985**, *57*, 2395.
(117) Lorber, A. *Anal. Chim. Acta* **1984**, *164*, 293.
(118) Linder, M.; Sundberg, R. *J. Chemom.* **2002**, *16*, 12.
(119) Bezemer, E.; Rutan, S. *Anal. Chim. Acta* **2003**, *490*, 17.
(120) Smilde, A. K.; Doornbos, D. A. *J. Chemom.* **1991**, *5*, 345.
(121) Faber, N. M.; Meinders, M. J.; Geladi, P.; Sjostrom, M.; Buydens, L. M. C.; Kateman, G. *Anal. Chim. Acta* **1995**, *304*, 273.
(122) Faber, N. M.; Meinders, M. J.; Geladi, P.; Sjostrom, M.; Buydens, L. M. C.; Kateman, G. *Anal. Chim. Acta* **1995**, *304*, 257.
(123) Faber, K.; Lorber, A.; Kowalski, B. R. *J. Chemom.* **1997**, *11*, 95.
(124) Faber, N. M. *Anal. Lett.* **1999**, *32*, 2899.
(125) Faber, N. M.; Boque, R.; Ferre, J. *Chemom. Intell. Lab. Syst.* **2001**, *55*, 91.
(126) Faber, N. M. *Anal. Chim. Acta* **2001**, *439*, 193.
(127) Faber, N. M.; Ferre, J.; Boque, R. *Chemom. Intell. Lab. Syst.* **2001**, *55*, 67.
(128) Rodriguez-Cuesta, M.; Boque, R.; Rius, F. X. *Anal. Chim. Acta* **2003**, *476*, 111.
(129) Prazen, B. J.; Synovec, R. E.; Kowalski, B. R. *Anal. Chem.* **1998**, *70*, 218.
(130) Garcia, I.; Sarabia, L.; Ortiz, M. C.; Aldama, J. M. *Analyst* **2004**, *129*, 766.
(131) Garcia, I.; Sarabia, L.; Ortiz, M. C.; Aldama, J. M. *Anal. Chim. Acta* **2004**, *515*, 55.

(132) Garcia, I.; Sarabia, L.; Ortiz, M. C.; Aldama, J. M. *Anal. Chim. Acta* **2004**, *526*, 139.
(133) Ortiz, M. C.; Sarabia, L. A.; Garcia, I.; Gimenez, D.; Melendez, E. *Anal. Chim. Acta* **2006**, *559*, 124.
(134) Garcia, I.; Ortiz, M. C.; Sarabia, L.; Aldama, J. M. *Anal. Chim. Acta* **2007**, *587*, 222.
(135) Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B. *Appl. Spectrosc.* **2000**, *54*, 413.
(136) Wold, S.; Sjostrom, M. In *Chemometrics Theory and application*; Kowalski, B., Ed.; American Chemical Society: Washington, DC, 1977.
(137) Sakuma, I.; Takai, N.; Dohi, T.; Fukui, Y.; Ohkubo, A. *J. Chromatogr.* **1990**, *506*, 223.
(138) Josefson, M.; Tekenbergshjelte, L. *J. Pharm. Biomed. Anal.* **1991**, *9*, 643.
(139) Sanchez, E.; Ramos, L. S.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 151.
(140) Gaganis, V.; Pasadakis, N. *Anal. Chim. Acta* **2006**, *573*, 328.
(141) Wiberg, K.; Jacobsson, S. P. *Anal. Chim. Acta* **2004**, *514*, 203.
(142) Schmidt, B.; Jaroszewski, J. W.; Bro, R.; Witt, M.; Stærk, D. *Anal. Chem.* **2008**, *80*, 1978.
(143) Andersson, G. G.; Dable, B. K.; Booksh, K. S. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 195.
(144) Bezemer, E.; Rutan, S. C. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 239.
(145) Johnson, K.; De Juan, A.; Rutan, S. C. *J. Chemom.* **1999**, *13*, 331.
(146) Gomez, V.; Miro, M.; Callao, M. P.; Cerda, V. *Anal. Chem.* **2007**, *79*, 7767.
(147) Gerritsen, M. J. P.; Tanis, H.; Vandeginste, B. G. M.; Kateman, G. *Anal. Chem.* **1992**, *64*, 2042.
(148) Comas, E.; Gimeno, R. A.; Ferre, J.; Marce, R. M.; Borrull, F.; Rius, F. X. *Anal. Chim. Acta* **2002**, *470*, 163.
(149) Öhman, J.; Geladi, P.; Wold, S. *J. Chemom.* **2005**, *4*, 79.
(150) Öhman, J.; Geladi, P.; Wold, S. *J. Chemom.* **2005**, *4*, 135.
(151) Jalali-Heravi, M.; Vosough, M. *Anal. Chim. Acta* **2005**, *537*, 89.
(152) Fraga, C. G. *J. Chromatogr., A* **2003**, *1019*, 31.
(153) Pere-Trepat, E.; Ginebreda, A.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2007**, *88*, 69.
(154) Vosough, M.; Salemi, A. *Talanta* **2007**, *73*, 30.
(155) Mjos, S. A. *Anal. Chim. Acta* **2003**, *488*, 231.
(156) Durante, C.; Cocchi, M.; Grandi, M.; Marchetti, A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2006**, *83*, 54.
(157) Hindmarch, P.; Kavianpour, K.; Brereton, R. G. *Analyst* **1997**, *122*, 871.
(158) Lavine, B. K.; Ritter, J. P.; Voigtman, E. *Microchem. J.* **2002**, *72*, 163.
(159) Wu, H. L.; Yu, R. Q.; Shibukawa, M.; Oguma, K. *Anal. Sci.* **2000**, *16*, 217.
(160) Li, S. S.; Hamilton, J. C.; Gemperline, P. *J. Anal. Chem.* **1992**, *64*, 599.
(161) Li, S. S.; Gemperline, P. J.; Briley, K.; Kazmierczak, S. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **1994**, *655*, 213.
(162) Comas, E.; Gimeno, R. A.; Ferre, J.; Marce, R. M.; Borrull, F.; Rius, F. X. *J. Chromatogr., A* **2003**, *988*, 277.
(163) Gimeno, R. A.; Comas, E.; Marce, R. M.; Ferre, J.; Rius, F. X.; Borrull, F. *Anal. Chim. Acta* **2003**, *498*, 47.
(164) Gargallo, R.; Tauler, R.; CuestaSanchez, F.; Massart, D. L. *TrAc, Trends Anal. Chem.* **1996**, *15*, 279.
(165) Mas, S.; Fonrodona, G.; Tauler, R.; Barbosa, J. *Talanta* **2007**, *71*, 1455.
(166) Mas, S.; Jauregui, O.; Rubio, F.; de Juan, A.; Tauler, R.; Lacorte, S. *J. Mass Spectrom.* **2007**, *42*, 890.
(167) Tauler, R.; Lacorte, S.; Barcelo, D. *J. Chromatogr., A* **1996**, *730*, 177.
(168) Pere-Trepat, E.; Hildebrandt, A.; Barcelo, D.; Lacorte, S.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2004**, *74*, 293.
(169) Chen, Z. P.; Wu, H. L.; Li, Y.; Yu, R. Q. *Anal. Chim. Acta* **2000**, *423*, 187.
(170) Beltran, J. L.; Guiteras, J.; Ferrer, R. *J. Chromatogr., A* **1998**, *802*, 263.
(171) Braga, J. W. B.; Bottoli, C. B. G.; Jardim, I. C. S. F.; Goicoechea, H. C.; Olivieri, A. C.; Poppi, R. J. *J. Chromatogr., A* **2007**, *1148*, 200.
(172) Zhang, Y.; Wu, H. L.; Xia, A. L.; Han, Q. J.; Cui, H.; Yu, R. Q. *Talanta* **2007**, *72*, 926.
(173) Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **1998**, *70*, 2796.
(174) Prazen, B. J.; Bruckner, C. A.; Synovec, R. E.; Kowalski, B. R. *Anal. Chem.* **1999**, *71*, 1093.
(175) Prazen, B. J.; Bruckner, C. A.; Synovec, R. E.; Kowalski, B. R. *J. Microcolumn Sep.* **1999**, *11*, 97.
(176) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *Anal. Chem.* **2000**, *72*, 4154.
(177) Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *HRC, J. High Resolut. Chromatogr.* **2000**, *23*, 215.
(178) Gross, G. M.; Prazen, B. J.; Synovec, R. E. *Anal. Chim. Acta* **2003**, *490*, 197.
(179) van Mispelaar, V. G.; Tas, A. C.; Smilde, A. K.; Schoenmakers, P. J.; van Asten, A. C. *J. Chromatogr., A* **2003**, *1019*, 15.
(180) Sinha, A. E.; Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *J. Chromatogr., A* **2004**, *1027*, 269.
(181) Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. *J. Chromatogr., A* **2004**, *1056*, 145.
(182) Hoggard, J. C.; Synovec, R. E. *Anal. Chem.* **2007**, *79*, 1611.
(183) Xie, L. L.; Marriott, P. J.; Adams, M. *Anal. Chim. Acta* **2003**, *500*, 211.
(184) Baunsgaard, D.; Andersson, C. A.; Arndal, A.; Munck, L. *Food Chem.* **2000**, *70*, 113.
(185) Baunsgaard, D.; Norgaard, L.; Godshall, H. A. *J. Agric. Food Chem.* **2001**, *49*, 1687.
(186) Zhang, F.; Li, H. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 184.
(187) Zhang, F.; Li, H. *Electrophoresis* **2005**, *26*, 1692.
(188) Zhang, F.; Chen, Y.; Li, H. *Electrophoresis* **2007**, *28*, 3674.
(189) Salau, J. S.; Honing, M.; Tauler, R.; Barcelo, D. *J. Chromatogr., A* **1998**, *795*, 3.
(190) Pere-Trepat, E.; Lacorte, S.; Tauler, R. *J. Chromatogr., A* **2005**, *1096*, 111.
(191) Pere-Trepat, E.; Tauler, R. *J. Chromatogr., A* **2006**, *1131*, 85.
(192) Pere-Trepat, E.; Lacorte, S.; Tauler, R. *Anal. Chim. Acta* **2007**, *595*, 228.
(193) Fraga, C. G.; Bruckner, C. A.; Synovec, R. E. *Anal. Chem.* **2001**, *73*, 675.
(194) Fraga, C. G.; Corley, C. A. *J. Chromatogr., A* **2005**, *1096*, 40.
(195) Sanchez-Ponce, R.; Rutan, S. C. *Chemom. Intell. Lab. Syst.* **2005**, *77*, 50.
(196) Porter, S. E. G.; Stoll, D. R.; Rutan, S. C.; Carr, P. W.; Cohen, J. D. *Anal. Chem.* **2006**, *78*, 5559.
(197) Porter, S. E. G.; Stoll, D. R.; Paek, C.; Rutan, S. C.; Carr, P. W. *J. Chromatogr., A* **2006**, *1137*, 163.
(198) Idborg, H.; Edlund, P. O.; Jacobsson, S. P. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 944.
(199) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. P. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *828*, 14.
(200) Forshed, J.; Stolt, R.; Idborg, H.; Jacobsson, S. P. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 179.
(201) Cantwell, M. T.; Porter, S. E. G.; Rutan, S. C. *J. Chemom.* **2007**, *21*, 335.
(202) Dable, B. K.; Marquardt, B. J.; Booksh, K. S. *Anal. Chim. Acta* **2005**, *544*, 71.
(203) Berbel, F.; Kapoya, E.; Diaz-Cruz, J. M.; Arino, C.; Esteban, M.; Tauler, R. *Electroanalysis* **2003**, *15*, 499.
(204) Endelmann, A.; Diewok, J.; Baena, J. R.; Lendl, B. *Anal. Bioanal. Chem.* **2003**, *376*, 92.
(205) Istvan, K.; Rajko, R.; Keresztury, G. *J. Chromatogr., A* **2006**, *1104*, 154.
(206) Gemperline, P. J.; Cho, J. H.; Archer, B. *J. Chemom.* **1999**, *13*, 153.
(207) Forshed, J.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chim. Acta* **2003**, *487*, 189.
(208) Lee, G. C.; Woodruff, D. L. *Anal. Chim. Acta* **2004**, *513*, 413.
(209) Furlanello, C.; Merler, S.; Jurman, G. *IEEE Trans. Signal Process.* **2006**, *54*, 2436.
(210) Keogh, E.; Ratanamahatana, C. A. *Knowledge Information Syst.* **2005**, *7*, 358.
(211) Lawton, W. H.; Sylvestr, Ea. *Technometrics* **1971**, *13*, 617.
(212) Davis, J. E.; Shepard, A.; Stanford, N.; Rogers, L. B. *Anal. Chem.* **1974**, *46*, 821.
(213) Mccue, M.; Malinowski, E. R. *Appl. Spectrosc.* **1983**, *37*, 463.
(214) Maeder, M.; Zuberbuehler, A. D. *Anal. Chim. Acta* **1986**, *181*, 287.
(215) Windig, W.; Guilment, J. *Anal. Chem.* **1991**, *63*, 1425.
(216) Kvalheim, O. M.; Liang, Y. Z. *Anal. Chem.* **1992**, *64*, 936.
(217) Mccue, M.; Malinowski, E. R. *J. Chromatogr. Sci.* **1983**, *21*, 229.
(218) Sanchez, E.; Kowalski, B. R. *Anal. Chem.* **1986**, *58*, 496.
(219) Ramos, L. S.; Sanchez, E.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 165.
(220) Sanchez, E.; Kowalski, B. R. *J. Chemom.* **1990**, *4*, 29.
(221) Harshman, R. A.; Hong, S. J.; Lundy, M. E. *J. Chemom.* **2003**, *17*, 363.
(222) Concha-Herrera, V.; Vivo-Truyols, G.; Torres-Lapasio, J. R.; Garcia-Alvarez-Coque, M. C. *J. Chromatogr., A* **2005**, *1063*, 79.
(223) Schoenmakers, P. J.; Vivo-Truyols, G.; Decrop, W. M. C. *J. Chromatogr., A* **2006**, *1120*, 282.

CR900394N