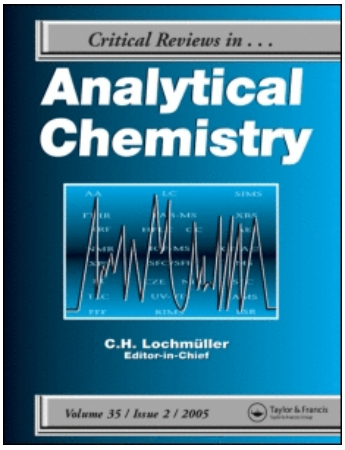

Critical Reviews in Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713400837

Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications

Anna de Juan[a]; Romà Tauler[b]
[a] Chemometrics Group Department of Analytical Chemistry, Universitat de Barcelona, Diagonal, 647. 08028, Barcelona, Spain [b] IIQAB-CSIC. Jordi Girona, 18, Barcelona, Spain

To cite this Article de Juan, Anna and Tauler, Romà(2006) 'Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications', Critical Reviews in Analytical Chemistry, 36: 3, 163 — 176
To link to this Article: DOI: 10.1080/10408340600970005
URL: http://dx.doi.org/10.1080/10408340600970005

Critical Reviews in Analytical Chemistry, 36:163–176, 2006

ISSN: 1040-8347 print / 1547-6510 online
DOI: 10.1080/10408340600970005

Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications

Anna de Juan
Chemometrics Group Department of Analytical Chemistry, Universitat de Barcelona, Diagonal, 647. 08028 Barcelona, Spain

Romà Tauler
IIQAB-CSIC. Jordi Girona, 18, Barcelona, Spain

This work is mainly oriented to give an overview of the progress of multivariate curve resolution methods in the last 5 years. Conceived as a review that combines theory and practice, it will present the basics needed to understand what is the use, prospects and limitations of this family of chemometric methods with the latest trends in theoretical contributions and in the field of analytical applications.

Keywords multivariate curve resolution, mixture analysis, soft-modeling, multicomponent systems

MULTIVARIATE CURVE RESOLUTION: EXTRACTING THE UNDERLYING MEASUREMENT MODEL

Analytical chemistry provides a vast number of examples of multicomponent systems. Nowadays, samples are far from simple and often contain many components to be simultaneously analysed (think of the generation of –omics data) or a few interesting analytes in the presence of many other chemical interferences (e.g., in environmental samples). More complex instrumentation is needed to cope with these systems in an efficient way and, equally needed are tools to handle and interpret the information obtained.

For the sake of simplicity, let us start with a clear example (a two-component chromatographic system with DAD detection). The HPLC-DAD system provides a two-way data set, a matrix **D**, with an elution and a spectral direction. Rows and columns of the data matrix will be spectra or elution profiles, respectively. Figure 1 and Eqs. [1–3] describe in detail the raw measurements and the underlying model of pure contributions.

$$\mathbf{D} = \mathbf{D_A} + \mathbf{D_B} \quad [1]$$

$$\mathbf{D} = \mathbf{c_A}\mathbf{s_A^T} + \mathbf{c_B}\mathbf{s_B^T} \quad [2]$$

$$\mathbf{D} = \mathbf{C}\mathbf{S^T} \quad [3]$$

Thus, the landscape of raw absorbances can be easily decomposed on the pure signals related to A and B, $\mathbf{D_A}$ and $\mathbf{D_B}$ (see Figure 2a and Eq. [1]). Each one of these landscapes, $\mathbf{D_i}$, is obtained from a dyad of profiles $\mathbf{D_i} = \mathbf{c_i}\mathbf{s_i^T}$, where $\mathbf{s_i^T}$ is a unit pure spectral profile and $\mathbf{c_i}$ is a concentration profile (elution profile in HPLC) that represents the weight (abundance) of that particular compound along the row direction in the data set (see Figure 2b and Eq. [2]). Finally, the additive model in Eq. [2] can be expressed in a more compact way, grouping together all the concentration profiles and all spectra in the **C** and $\mathbf{S^T}$ matrices, respectively (Figure 2c and Eq. [3]). This last expression is the most common way to express the Beer–Lambert law in matrix form and, by extension, the bilinear MCR model (1–7).

The goal of MCR is, thus, passing from the mixed non-selective information that comes from the instrument (**D**) to the real contributions of the pure components in our system (represented by the profiles in **C** ands $\mathbf{S^T}$) without using any behaviour model or a priori information about the system. How to do this in the most efficient and reliable way has been the task of MCR research from early days till present (1–5).

Exploration in MCR

MCR is, by definition, a model-free or a soft-modeling method that focuses on describing the evolution of the experimental multicomponent measurements through their pure component contributions. So, strictly speaking, only the matrix **D** of raw measurements is needed to perform the analysis. Nevertheless, exploring the data sets can provide information about the

Address correspondence to Anna de Juan, Chemometrics Group, Department of Analytical Chemistry, Universitat de Barcelona, Diagonal, 647, 08028 Barcelona, Spain. E-mail: anna.dejuan@ub.edu

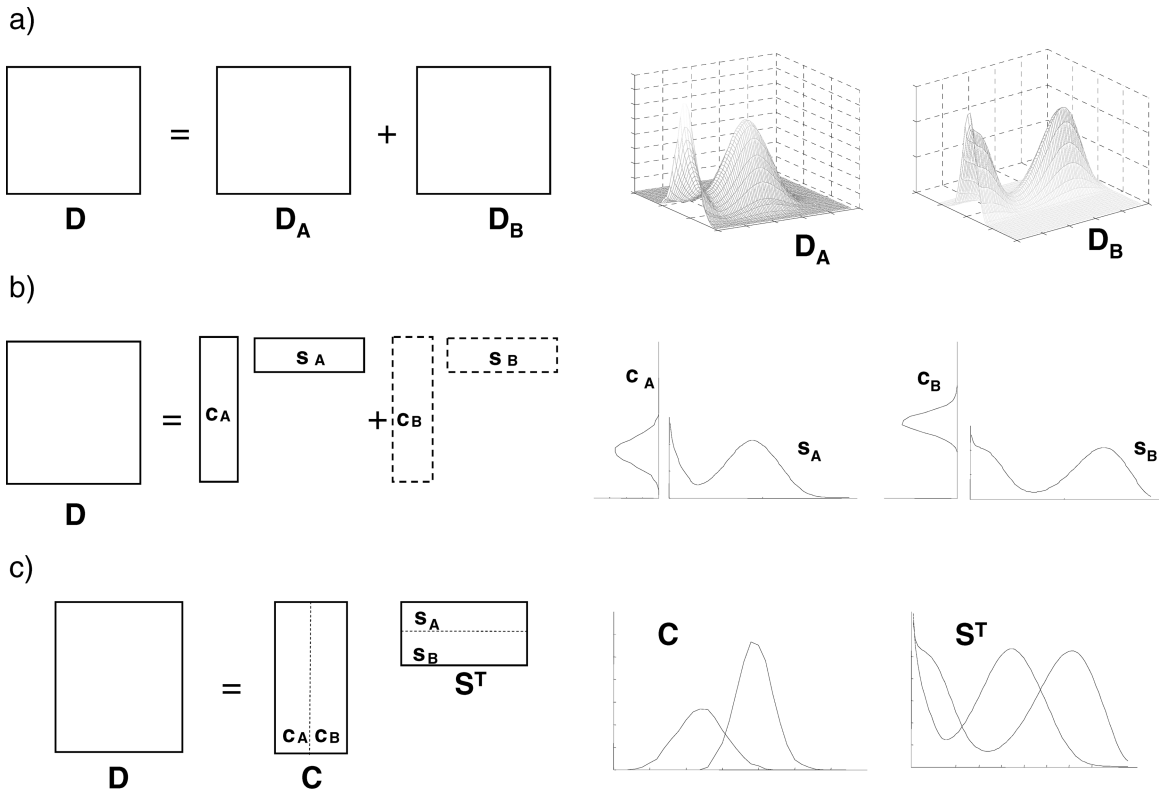

FIG. 1. The measurement model of a two-component HPLC-DAD system. (a) Described as an additive model of pure signal contributions, (b) described as a model of additive dyads of pure concentration profile and spectrum, (c) described as a bilinear model of concentration profiles and spectra.

number and evolution of components or in the form of initial estimates of concentration or response profiles. This previous information can orient the resolution process and improve significantly the final results obtained. The most usual resolution-oriented chemometric tools belong to the family of local rank analysis methods and to procedures for pure variable selection.

Local rank analysis methods perform repeated PCA analyses in small parts of the data set (the so called row or column windows) with the aim of knowing how the number and distribution of components evolve along the data set. The design of these windows, gradually growing in size or with a fixed size and moved along the data set, and the way to plot the results obtained in all the performed window PCA analyses determine the final information obtained.

Local rank analysis methods are particularly relevant in the study of processes, where the concentration profiles of the different components evolve smoothly and, often, following a sequential pattern, i.e., the first component emerging is the first decaying and so forth. Using gradually growing windows, mimicking the stepwise progress of a process, the evolution of the number of significant components indicates the process stage (e.g., time, pH value, temperature) at which the different components emerge and decay (8, 9). In sequential processes, the proper use of this information can provide concentration windows, i.e., the rows in the data set (process variable interval) where a particular component is present (10), and approximate concentration profiles for the different components (4, 11, 12). Evolving Factor Analysis was the parent method for local rank analysis of processes, from which most recent algorithms derive (8–10). Modifications are oriented to confirm the sequential evolution of components in unknown processes (13), to improve the setting of concentration windows (14) or to overcome problems linked to the analysis of rank-deficient processes (15).

Local rank analysis based on the use of fixed size windows covers the whole data set through the systematic study of small moving windows. The results inform about the local complexity of the data set, i.e., about how many components overlap in the different data set regions enclosed by the windows (16, 17). This information locates selective regions in the data set, i.e., windows with only one component, which are crucial in resolution analysis. Playing with the size of these windows, more capability to detect the presence of minor species (with larger windows) or more specificity in setting the boundaries of the different local rank regions (with smaller windows), can be achieved (18). Fixed Size Moving Window Evolving Factor Analysis has been the model for all recent algorithms (16). The fact that the whole data set is studied in small pieces makes these methods suitable for data sets that have a global smooth

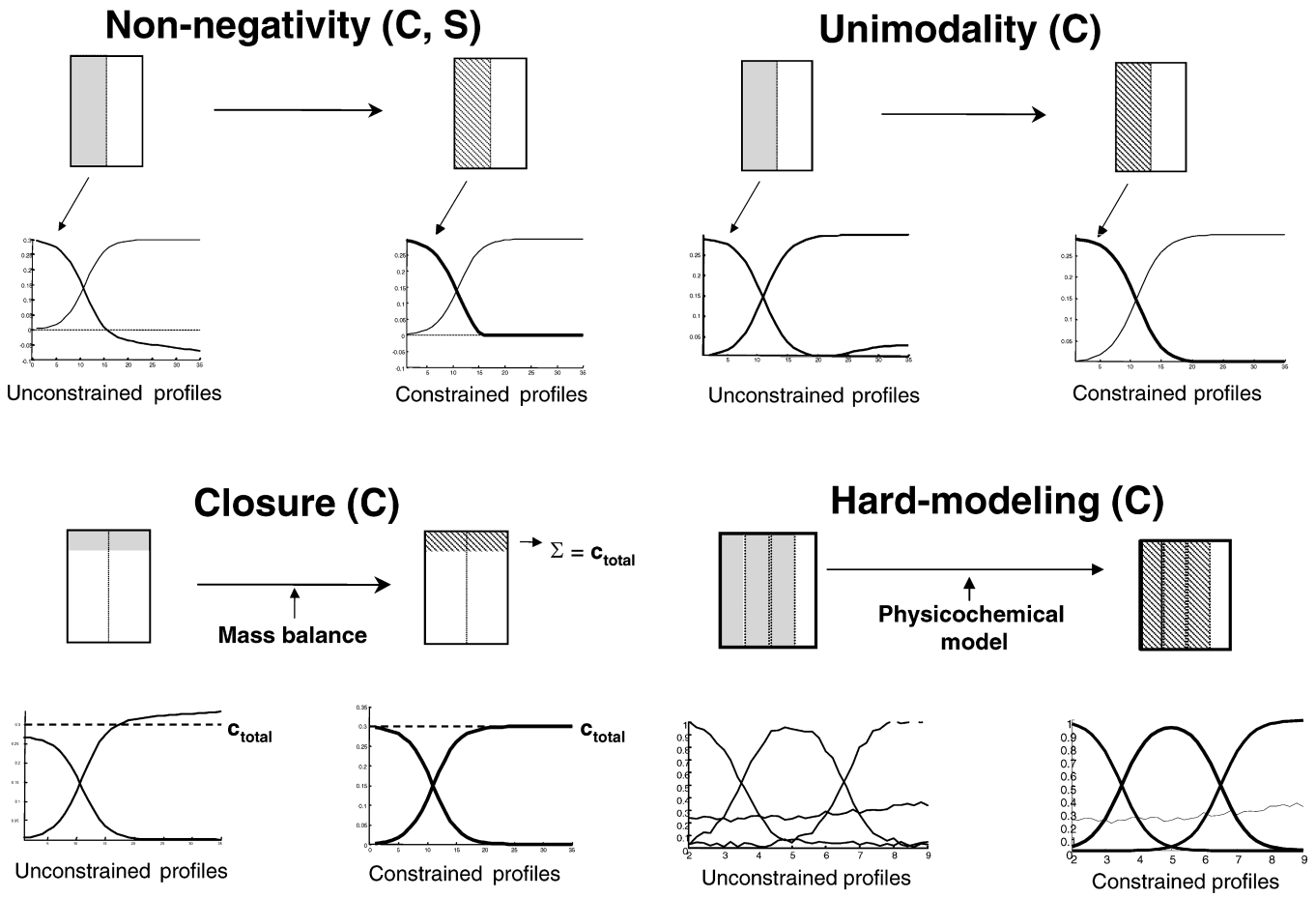

FIG. 2. Common constraints used in iterative MCR approaches.

variation in the concentration direction, such as processes (19, 20), or a local smooth pattern, such as spectroscopic images, where windows are designed so that the original 2D or 3D pixel neighbourhood is kept (21).

Methods of pure variable selection find the most representative row or column profiles for the different components in a data set. Most of these methods are indirectly helpful to determine the number of components in the data set and, when selectivity conditions are favourable, the row or column profiles obtained can be directly associated with pure concentration or response profiles and the resolution of the system can be achieved. Some of these procedures work on the abstract space of principal components (22, 23), whereas others use the space of the real measurements (24–27). The latter are most commonly used and, among them, SIMPLISMA, the pioneering approach, is still the most popular (24). Recent modifications of SIMPLISMA imply the combined use of raw and second derivative data for a better variable selection (28) and the use of the maximum angle among profiles as selection criterion in derived methodologies, such as SMAC (stepwise maximum angle calculations) (29). Pure variable selection methods play an important role in data sets where a smooth, sequential process direction is missing, because the performance of these approaches is not affected by the lack of ordering in the row or column directions. This fact explains that SIMPLISMA has encountered an extensive application in spectroscopic images (30–34) or environmental data (35, 36).

Non-Iterative Approaches

Most non-iterative resolution methods rely on combining information of small sections of a data set (subspaces) constructed from global and local rank information to obtain the pure component profiles. These subspaces can be concentration windows or regions of the data matrix that hold particular properties (presence or absence of particular components). In many non-iterative methods, the **C** or the **S** profiles are recovered one at a time and the paired matrix (**S** or **C**, respectively) is obtained afterwards through a single least-squares step, according to the model in Equation [3].

Window Factor Analysis (WFA) (37), Subwindow Factor Analysis (SFA) (38), or Heuristic Evolving Latent Projections (HELP) (39) are among the first and most significant approaches within this category. WFA recovers the concentration profile of each component using the original data set and the zero-concentration window of the component to be resolved, i.e., all

rows out of the concentration window. With this information, a vector representing the spectral variation of the component of interest uncorrelated to all other components is obtained. This vector, combined appropriately with the original data set, yields the sought concentration profile. SFA recovers the pure response profile of each component. The knowledge of the concentration windows is used in such a way that each pure spectrum is calculated as the intersection of two subspaces with only the compound to be resolved in common. HELP recovers one at a time the dyad of concentration profile and spectrum for each component. It starts finding a selective concentration region for the component to be resolved. This selective region provides directly the component spectrum and the concentration profile is recovered using appropriately the selective and the zero-concentration window. Deflating the matrix by subtraction of the signal contribution of the resolved component helps to find new 'selective' regions for other components in the deflated matrix that were not present in the original data set.

Most recent algorithms are usually evolutions of the parent approaches, such as Orthogonal Projection Resolution (OPR) from WFA (40) or Parallel Vector Analysis (PVA) from SFA (41). From the brief descriptions of the methods above, it can be easily noticed that a key point is the correct definition of the concentration windows and research has also gone in this direction (42). In terms of applicability, the main limitation of these methods stems from the fact that data sets need to have a sequential (or very ordered) concentration direction where concentration windows can be reasonably well set, i.e., the field of application is restricted to the analysis of evolving systems (processes, chromatography). Nevertheless, when a data set shows the suitable properties of component overlap in the concentration and spectral direction (43), these methods provide unique and correct solutions. An additional advantage is also linked to the one-at-a-time recovery of concentration profiles or spectra that can make these methods useful when only partial information of a system, such as the pure spectra or the pure concentration profiles of certain components, are of interest (44, 45).

Iterative Approaches

Iterative resolution approaches are currently considered the most popular MCR methods due to the flexibility to cope with many kinds of data structures and chemical problems and to the ability to accommodate external information in the resolution process. All of them share a common step of optimisation (of $\mathbf{C}$ and/or $\mathbf{S^T}$ matrices) that starts from initial estimates of $\mathbf{C}$ or $\mathbf{S^T}$ that evolve to yield profiles with chemically meaningful shapes, tailored according to chemical or mathematical information included in the optimisation process under the form of constraints (46–53). Iterative Target Transformation Factor Analysis, ITTFA, (46, 47) and Multivariate Curve Resolution-Alternating Least Squares, MCR-ALS (48–51) were the first approaches, although other methodologies with different principles, such as Gentle or Resolving Factor Analysis, RFA (52, 53) have appeared afterwards. General improvements applicable to most iterative methods above are linked to the explanations in the next paragraphs.

Constraints, defined as chemical or mathematical properties that should hold the resolved profiles, have been and still are an active field of research. Improvements have included the definition of new kinds of constraints and the progress in their implementation, trying to modify the profiles as smoothly as possible. Adding up to the classical constraints of non-negativity, unimodality, closure and the equality constraints linked to selectivity or to the use of known profiles (4, 49) the most outstanding contribution has been the introduction of the so-called hard-modeling constraints (see Figure 2). Until recently, systems were either hard- or soft-modelled. Hard-modeling fitted chemical data according to rigid models built from mathematical expressions that could define a physicochemical behaviour or, in a general sense, the shape of a signal or a profile. To do so, all the data set variation had to be described by that particular model (54, 55). Nowadays, hard models have been introduced in MCR as constraints to act in a partial or a total manner on concentration or response profiles (56–58). The main benefit is that models can be used that describe only part of the variation of the data set (e.g., the evolution of some components) whereas the rest of the system (the remaining components) can be soft-modelled. These constraints have allowed the hard-modeling of processes in the presence of inert interferences (soft-modelled) (59, 60) and the calculation of physicochemical parameters as additional outcomes of MCR. Because of the strength of this kind of constraint, the ambiguity in the resolved profiles is suppressed or significantly minimized (61, 62) and the chances to resolve severely overlapped concentration profiles and signals are substantially increased. Examples of hard-modeling constraints include incorporation of kinetic models (56–58, 63, 64), enzymatic models (60, 65, 66) or equilibria (59, 67) in concentration profiles and inclusion of signal shape models, such as peak-shaped functions for voltammetric measurements (68) or exponential decay curves in NMR DOSY data (69).

Profiles get their form modified by the action of constraints. The harsher the modification is, the further the profile goes from the smooth form that it has in reality. When corrections by constraints are too abrupt, divergences can appear in the optimisation process and the net effect of the constraint may not be as positive as expected. Constraint implementation should try to incorporate a smooth way to modify the profile shape and should permit a certain flexibility in the fulfillment of the constraint conditions to account for the experimental noise or for instrumental effects on the measurement. Smoothness has been often achieved by least-squares implementations of constraints (70–72) and different degrees of tolerance have been achieved by the use of penalty functions (73). These improvements have been particularly interesting in complex systems, where perturbations in the natural form of profiles can affect convergence more significantly and in the application of equality constraints (i.e., constraints that incorporate information on a partially or totally known profile shape), due to the unavoidable

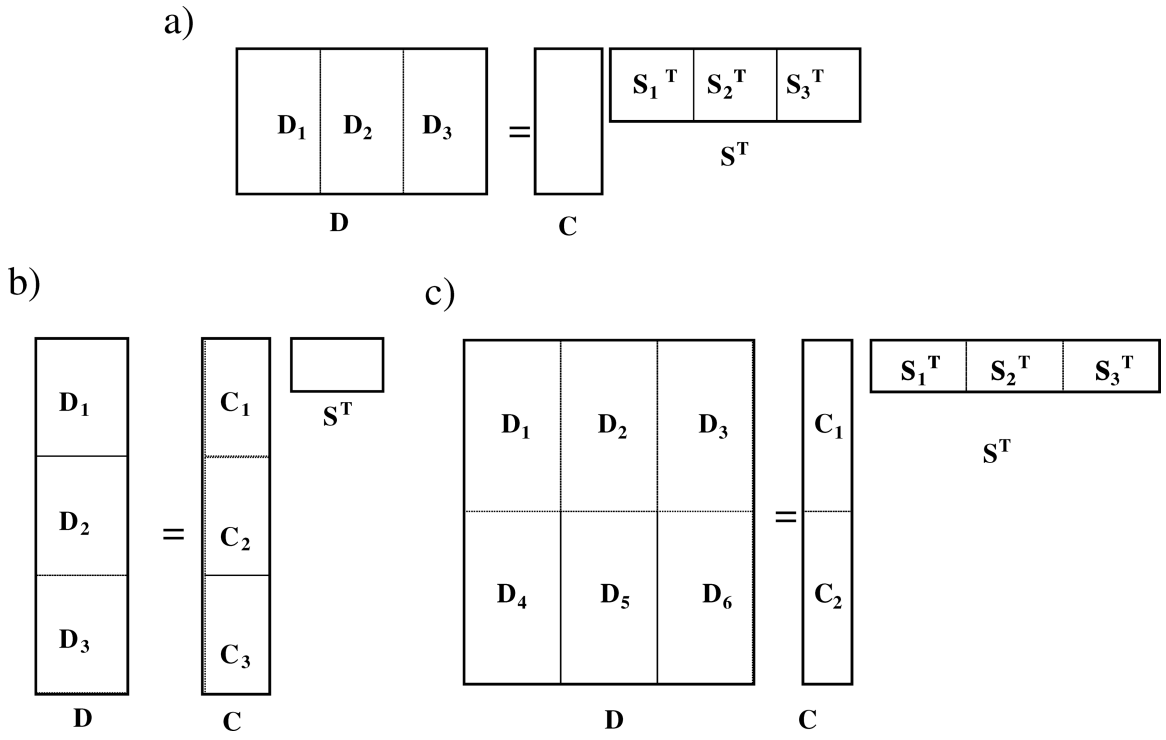

FIG. 3. Augmented MCR bilinear models. (a) model for a row-wise augmented matrix, (b) model for a column-wise augmented matrix, (c) model for a row- and column-wise augmented matrix.

small differences between the real profile shape and the known references. Nevertheless, classical substitution approaches that also allow for variable degrees of tolerance are interesting for they tend to allow a major flexibility in the combination of applied constraints and in the choice of profiles or directions to be constrained. Besides, their application in systems that naturally obey the imposed constraints leads also to the best fit when convergence is achieved.

A great MCR milestone was the extension of these methods beyond two-way data sets (49, 74–77). Classical MCR ambiguity problems get significantly reduced with the possibility to analyse structures with richer information (three-way data or row- and/or column-wise augmented matrices) (see Figure 3). Data fusion or multiset analysis are names given now to the merged measurements coming from one or more experiments monitored by different techniques but, well before these names were coined, MCR had already been applied to this and to other kinds of merged data arrangements. Data fusion responds often to the hyphenated or multi-way nature of modern instruments (coupling several detection systems or acquiring several responses at a time) (78–82) but, other augmented data arrangements, like multibatch or multiprocess data sets, are equally interesting (83–87). Bearing in mind the basic idea of resolution, i.e., the recovery of the pure component information from a data set in a set of pure profiles, we must not forget that many multiway methods, such as PARAFAC, PARAFAC2, TUCKER, have been often used for resolution purposes, although they do not fall into the classical MCR denomination (88). In this sense, it is interesting to note that MCR on augmented data can impose typical multi-way structures, such as trilinear (76, 89) and Tucker models (90), in the form of constraints and apply them to the complete set of profiles or to some of them. When extended MCR or multi-way methods are applied for resolution purposes, quantitative information linked to the data set can be obtained (59, 66, 91–95). Quantification by MCR presents advantages with respect to classical multivariate calibration approaches for it needs neither the knowledge nor the inclusion of interferences in the calibration model and the number of standards can be extremely low (even only one). The now extended use of MCR for quantitative purposes has promoted studies on the typical quality parameters of quantification (96, 97) and the emergence of new approaches that use the resolved concentration profiles by MCR as input in two- or multi-way calibration methods instead of the raw data sets (98).

Uncertainty in MCR

Although MCR gives astounding results from the sole use of raw experimental measurements, the profiles recovered may be affected by the so-called rotational and/or intensity ambiguities (6, 7, 49). Going back to the MCR model,

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}}$$

can be written as:

$$\mathbf{D} = \mathbf{C}\mathbf{T}\mathbf{T}^{-1}\mathbf{S}^{\mathbf{T}}$$

where **T** stands for any transformation matrix and this gives:

$$\mathbf{D} = \mathbf{C}'\mathbf{S}^{\mathbf{T}'} \quad \text{where } \mathbf{C}' = \mathbf{CT} \text{ and } \mathbf{S}^{\mathbf{T}'} = \mathbf{T}^{-1}\mathbf{S}^{\mathbf{T}}$$

This is the mathematical formulation of the rotational ambiguity, which means that we may obtain the same optimal lack of fit in the description of the data set **D** using sets of profiles (**C′** and **S'**) shaped differently from the true ones (**C** and **S**). Even in the absence of rotational ambiguity, Eq. [3] can still be rewritten as:

$$\mathbf{D} = \sum_{i=1}^{n} \left(\frac{1}{k_i}\,\mathbf{c_i}\right)\left(k_i\,\mathbf{s_i}^{T}\right) \qquad [4]$$

which means that the dyad of resolved profiles ($\mathbf{c_i}$, $\mathbf{s_i}$) for each pure component can present profiles with the sought shape but k_i times smaller, $(1/k_i)\mathbf{c_i}$, or bigger, $k_i\mathbf{s_i}^{T}$, than expected. The extent of ambiguity can be significantly decreased or even suppressed by the use of constraints. The more constrained a system is, the fewer the possible profile combinations that fulfill the required shape and intensity conditions and fit optimally the data set **D**.

Ambiguity has always been the Achille's heel of multivariate curve resolution. Only recently there have been solid attempts to evaluate the extent of this phenomenon in the MCR resolved profiles (99–103). Ambiguity is component and profile-dependent and, within the same data set, we may find components or profiles which lack ambiguity and others that have it in a large extent. Some of the latest tendencies opt for determining, one at a time for each component, the minimum and maximum boundaries of the feasible solution bands for the dyad of resolved profiles. The general idea is finding the 'boundary' dyads of profiles that provide the maximum and the minimum contribution of the component signal to the overall signal measured. The maximum and minimum boundaries should correspond to profiles that respect the constraints and should be part of a global set of profiles including all components and reproducing the data set with the optimal fit. All these conditions are set under different forms of optimisation problems and these approaches are valid for data sets with unlimited number of components (99, 100).

Other methodologies question whether the minimum and maximum boundaries really enclose all possible feasible profiles and, whether defined as maximum and minimum of signal contributions, they really represent the 'extreme' solutions. Instead of the boundaries, these alternative approaches display the band or the area enclosing all feasible solutions using visual tools or geometry concepts for this purpose (101, 102). However, their applicability restricts right now to 2- or 3-component systems. No matter the methodology used, it is very clear the dramatic decrease of ambiguity introduced by the effect of constraints and, above all, the major effect obtained when going from analysis of two-way data to the analysis of augmented data matrices (multi-way) (100, 101, 104, 105). Although there has been much progress in the knowledge and evaluation of ambiguity in MCR, works are developing and we shall see progress in years to come.

No matter the amount or the lack of ambiguity in the solutions, a common source of uncertainty in the results comes from the error in the experimental measurements. Finding analytical expressions that can help in the evaluation of error propagation in the MCR resolved profiles is extremely difficult, but there are strategies to obtain approximate estimates of the error linked to the resolution results (105, 106). Most of them obtain the error estimates as the average of a huge number of simulated or real replicate measurements. Simulations of replicates can include: repeated resolutions of data matrices with some rows or columns missing or with some complete matrices missing (in three-way data sets, jackknife methods) (107, 108), resolutions of a particular data set after having added some noise to the measurements (noise-added method) (109, 110) or construction of replicates adding noise to a noise-free simulated or reproduced data set (Monte Carlo approach) (106). Although the uncertainty induced in the resolved profiles by ambiguity and by noise effect arise from distinct sources and should be estimated separately, there is a thin border between the two when noise gets very large because of the ambiguity induced by the worse definition of concentration and spectral windows

Applications

MCR has been used since the origin for process analysis or, using a more general expression, for multicomponent evolving systems, but new areas of application appear that do not fit this definition. Typical MCR examples ranged from reaction monitoring at laboratory scale or at an industrial level to all kinds of hyphenated chromatographic or flow injection separations (75–98). These applications constitute still the most common fields of use of MCR. Advances in this context have gone in the direction of analysing data sets that are far beyond the classical two-way data table and that reflect the progress in analytical instrumentation and data analysis. Thus, processes tend to be monitored with multitechnique strategies (coming from multiresponse instruments or from separate measurements) (78–82) or taking multibatch or multiexperiment data sets (75, 83–98). These strategies have greatly improved the understanding of complex processes, like those involving biomolecules, that may include events happening at very different levels and that need specific multitechnique monitoring (68, 79, 81, 111, 112). In process analysis, Process Analytical Control has also benefited from multibatch analysis, setting methodologies to monitor and control the evolution of systems (75, 83–87). Process monitoring with instrumentally challenging 2-way measurements, such as 2D NMR (113), 2D DOSY NMR (69), with diverse kinds of electrochemical measurements (114) and with fast spectroscopies (115) have also been proposed.

Analytical characterisation and determination, often based on separation or flow injection techniques, has found clear improvement in resolution and quantification through the use of data arrangements combining standard and sample runs (91–95) and, eventually, multitechnique detection, such as LC-DAD-MS experiments (116). MCR allows for very flexible combinations

of information, e.g., chromatographic runs resolved well can be combined with routine fast overlapped runs coming from short columns (92), minor impurities can be more easily identified in augmented sample analysis (117), analytes in complex biological, food, industrial or environmental samples can be analysed (75, 91–95, 97, 118, 119).

In all the classical applications mentioned above, MCR is an obvious data analysis option, since the application of constraints (in the iterative methods) or the setting of concentration windows (in non-iterative methods) take advantage of the smooth continuous shape of the pure component profiles in both the response (**S**) and the concentration (**C**) direction. More challenging data sets are those where the concentration direction does not follow a global patterned shape, but just some local composition relationship. Measurements, such as spectroscopic images, environmental data or many of the –omics data sets, fall into this category.

Spectroscopic imaging has emerged in the last decade as a particularly powerful experimental measurement due to the presence of spatial-dependent information on the sample composition. Compounds spread and overlap on the scanned 2D or 3D image and the goal of data analysis is providing reliable distribution maps and characterization of the pure compounds in the image. Such a goal is well met by MCR, since the image data variation responds to the bilinear model that describes any pixel spectrum in the image as the linear combination of the signal contribution of its components. Making an image data cube ($x \times y \times \lambda$) suitable for MCR analysis requires only unfolding the cube into a data table that contains all pixel spectra. After the resolution analysis, the pure spectra of the constituents are recovered as well as their related pure distribution maps, once the profiles in **C** are folded back to recover the original spatial image structure (see Figure 4). MCR image resolution has gained relevance during the last years (30–34, 120, 121) and efforts are now focused on the use of spatial information, coming from exploratory methods based on the local rank analysis of image areas (21) or from the application of pixel classification tools (122), under the form of image-specific constraints. Many advances are still to come to improve the incorporation of information into the image resolution process and to extend the simultaneous analysis of several images (32) to the analysis of data sets scanned with different techniques, as a function of time, and depth.

Environmental information, coming from geographical, seasonal or compartmental monitoring can also be organized under the form of data tables, where the rows refer to geographical sites, seasonal times and the column variables are compound concentrations, physicochemical parameters, toxicological indexes or any other kind of data of environmental interest. Although the bilinear nature of this data is not as clear as for a spectroscopic measurement, it is well accepted that the total profile of a sample (the row that contains all the measured parameters in the table) can be the result of the contribution of different environmental sources, with a particular compositional

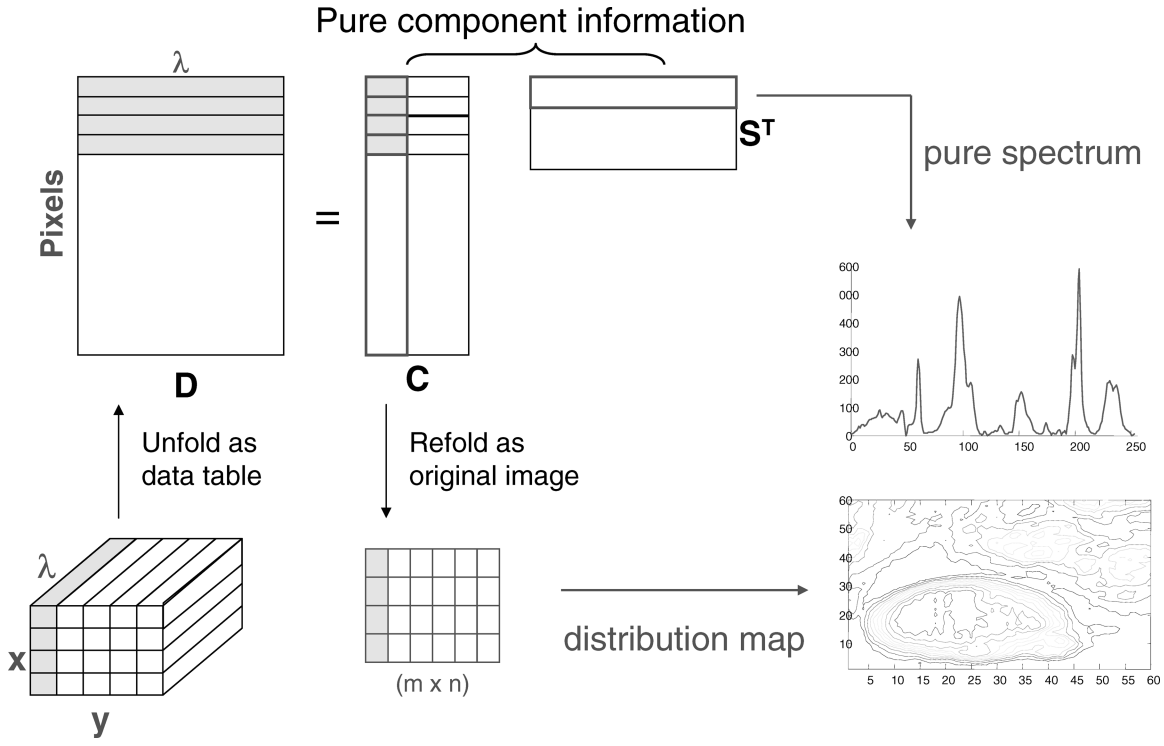

FIG. 4. Steps and results in the resolution of a spectroscopic image.

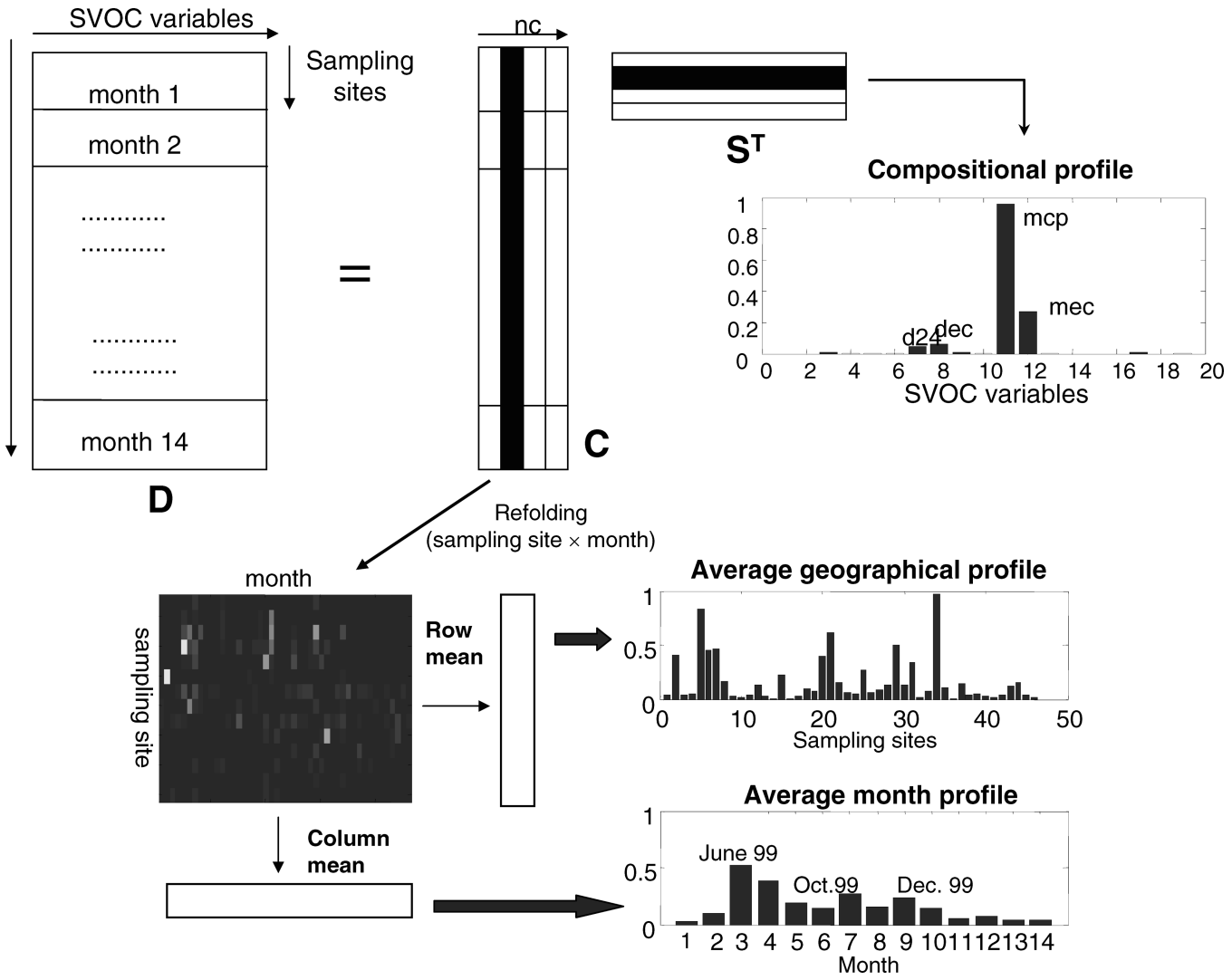

FIG. 5. Data arrangement and MCR decomposition of an environmental data set formed by monthly geographical monitoring of SVOC compounds (121). Compositional profile, average geographical and month profiles for a particular pollution source.

profile. This is the basis of the so-called receptor models (123, 124). $\mathbf{S}^{\mathbf{T}}$ and $\mathbf{C}$ would contain, in this context, the composition profiles of the different environmental sources and the geographical, seasonal, and source apportionment profiles, respectively. Figure 5 shows an example of environmental data analysis, where a data set containing monthly environmental tables of semivolatile organic compounds (SVOCs) concentrations collected at 48 sampling points in Portugal is analysed (125). The data arrangement and the profiles resolved for a particular pollution source, where the compositional profile, the geographical and the month profiles are obtained, is shown. MCR and other factorization approaches, such as Positive Matrix Factorization (PMF) (126, 127), multilinear engine (ME) (128) or unmix (129, 130), and most multi-way data analysis tools have been adapted to the analysis of 2-way or multi-way environmental data [131–133]. Specific contributions of MCR and related techniques to environmental analysis include the analysis of multidimensional data arrangements, incorporation of noise information in the factorisation process (126–130), incorporation of environmental parameters, such as weather parameters, wind speed and direction (134, 135) in the source modeling and the combination of the results obtained with typical geostatistical tools, such as Geographic Information Systems (GIS) (136).

Biological data are also a relevant field of application of MCR. Apart from the invaluable help of this method for the analysis and interpretation of biological processes, as essential as protein folding or DNA-drug interactions, for which no general physicochemical model is available [81, 110, 112, 137, 138], MCR finds a new challenging area in the new generation of -omics data (genomics, proteomics, metabolomics, ...). These are biological high-throughput data that contain

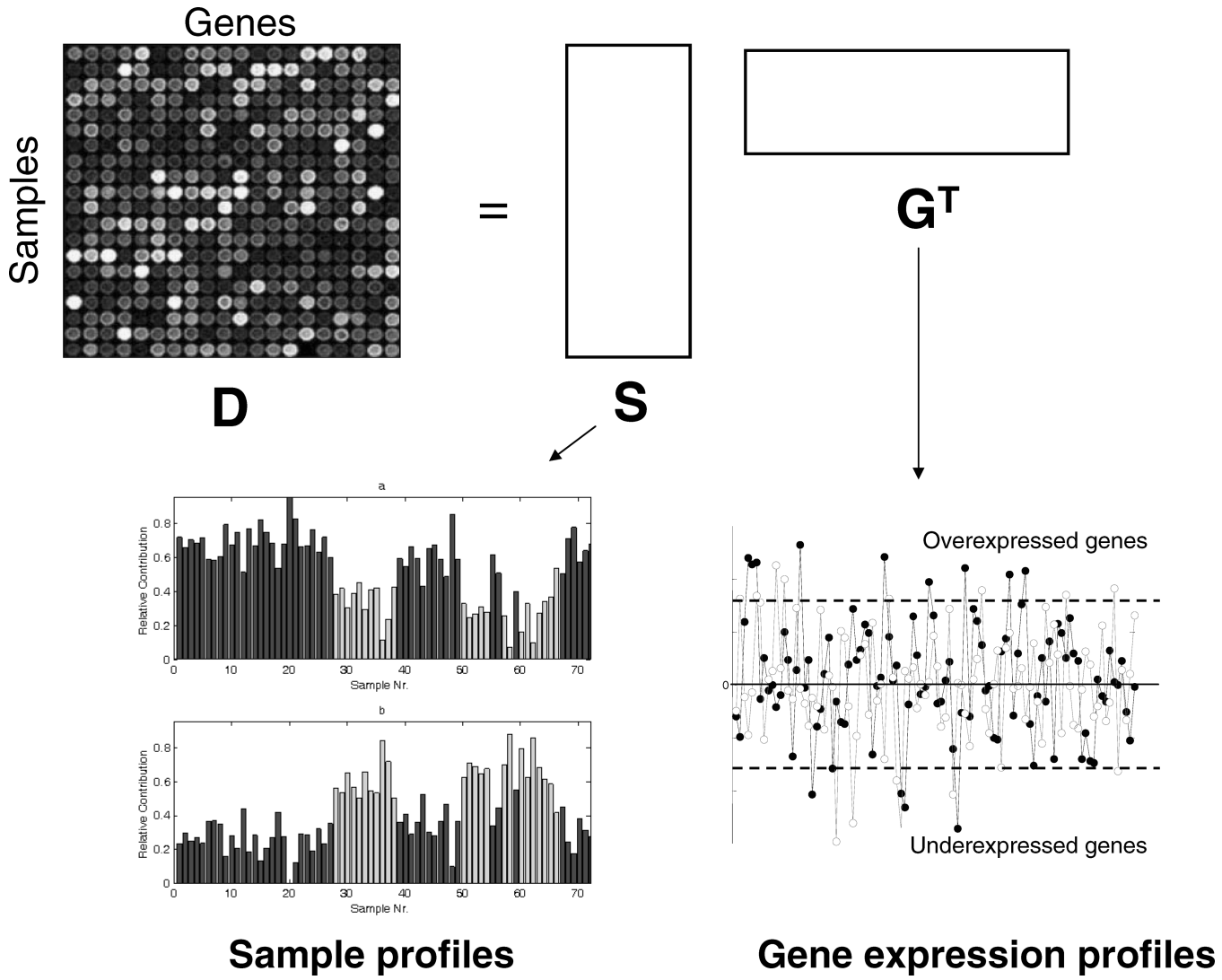

FIG. 6. Resolution of a DNA microarray. Sample profiles cluster samples according to their genic response. Over- and underexpressed genes (out of the area limited for the dashed lines) in the gene expression profiles characterise the nature of the clustered salient samples in the related sample profile. (Sample profiles are realistic and come from ref. (137); microarray picture and gene expression profiles are schematic representations of the real data set).

information about thousands of components. Metabolomic analysis is often performed with GC or LC/MS measurements. The chromatograms obtained contain a huge amount of components and classical global MCR of the complete data set is not feasible. Strategies are being introduced that use MCR in a semiautomated stepwise fashion on small time windows to recover all registered components (139). Microarrays are a typical tool used in genomics to identify genes and to know about their functions and expression levels under different conditions. In this case, a microarray data matrix $\mathbf{D}$ is sized (samples $\times$ genes). Samples can be cells under different conditions, e.g., normal vs. tumour cells, and a d_{ij} element in the data matrix is the expression of a certain gene in a particular cell. The bilinear model $\mathbf{D} = \mathbf{S}\mathbf{G}^{\mathbf{T}}$ contains gene expression profiles in $\mathbf{G}^{\mathbf{T}}$ and sample profiles in $\mathbf{S}$. For a dyad of a resolved microarray component, i.e., ($\mathbf{s_i}$, $\mathbf{g_i^T}$) the relationship between the most over- and underexpressed genes in $\mathbf{g_i^T}$ with the most salient samples in $\mathbf{s}_i$ can give indications on the genetic link of certain diseases. Figure 6 shows schematically the resolution of a microarray; in this case, the sample profiles are those obtained for a data set formed by samples showing two kinds of leukemia (140). The most salient samples in each resolved sample profile refer to one of the two kinds of leukemia. For details on the gene expression interpretation, see reference (141). Since microarray numerical measurements come from fluorescence imaging data, efforts are also focused on the improvement of resolution of the raw microarray image for a better translation between the spectroscopic value and the gene expression quantification (121). Other metabolomics data are

generated under very carefully designed conditions with the aim of identifying and relating metabolites with diseases or to understand the response to different treatments. Those data enclose very different levels (e.g., individuals, time, and experimental measurement) and give rise to the so-called multiset analysis. New methodologies are born to deal with these demanding data set structures (142, 143) and MCR can also have a future for the treatment of some of these problems.

CONCLUSIONS

MCR is a classical and, at the same time, a fully alive data analysis tool that is still in progress in terms of theoretical developments and new applications. Much development can still be foreseen in different areas, such as the incorporation of new kinds of information and models that respond to mathematical data structure and to properties of new measurements and processes, the increase in complexity of the data arrangements and the understanding and estimation of the uncertainty linked to the resolved profiles.

REFERENCES

1. J. C. Hamilton and P. J. Gemperline, An extension of the multivariate component-resolution method to three components. *Journal of Chemometrics* 4 (1990):1–13.
2. A. de Juan, E. Casassas, and R. Tauler, Soft modeling of analytical data in *Encyclopedia of Analytical Chemistry: Instrumentation and Applications* (Wiley, New York, 2000).
3. Y. Liang and O. M. Kvalheim, Resolution of two-way data: theoretical background and practical problem-solving—Part 1: Theoretical background and methodology. *Fresenius Journal of Analytical Chemistry* 370 (2001):694–704.
4. A. de Juan and R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta* 500 (2003):195–210.
5. J. H. Jiang, Y. Liang,, and Y. Ozaki, Principles and methodologies in self-modeling curve resolution. *Chemometrics and Intelligent Systems* 71 (2004):1–12.
6. W. H. Lawton and E. A. Sylvestre. Self modeling curve resolution. *Technometrics* 13 (1971):617–633.
7. E. A. Sylvestre, W. H. Lawton, and M. S. Maggio, Curve resolution using a postulated chemical reaction. *Technometrics* 16 (1974):353–368.
8. M. Maeder, and A. D. Zuberbühler, The resolution of overlapping chromatographic peaks by evolving factor-analysis. *Anaytica Chimica Acta* 181 (1986):287–291.
9. H. Gampp, M. Maeder, C. J. Meyer, and A. D. Zuberbühler, Calculation of equilibrium constants from multiwavelength spectroscopic data. III. Model-free analysis of spectrophotometric and ESR titrations. *Talanta* 32 (1985):1133–1139.
10. M. Maeder, Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Analytical Chemistry* 59 (1987):527–530.
11. H. Gampp, M. Maeder, C. J. Meyer, and A. Zuberbühler, Calculation of equilibrium constants from multiwavelength spectroscopic data IV. Model Free Least Squares Refinement by use of evolving factor analysis. *Talanta* 33 (1986):943–951.
12. R. Tauler and E. Casassas, Principal Component Analysis Applied to the Study of Successive Complex Formation Data in the Cu(II) ethanolamine Systems. *Journal of Chemometrics* 3 (1988):151–161.
13. A. C. Whitson and M. Maeder, Exhaustive evolving factor analysis (E-EFA). *Journal of Chemometrics* 15 (2001):475–481.
14. Z. P. Chen, J. Morris, E. Martin, R. Q. Yu, Y. Z. Liang, and F. Gong, Recursive evolving spectral projection for revealing the concentration windows of overlapping peaks in two-way chromatographic experiments. *Chemometrics and Intelligent Laboratory Systems* 72 (2004):9–19.
15. A. de Juan, S. Navea, J. Diewok, and R. Tauler, Local rank exploratory analysis of evolving rank-deficient systems. *Chemometrics and Intelligent Laboratory Systems* 70 (2004):11–21.
16. H. R. Keller and D. L. Massart, Peak purity control in liquid chromatography with photodiode array detection by fixed size moving window evolving factor analysis. *Analytica. Chimica Acta* 246 (1991):379–390.
17. H. R. Keller, D. L. Massart, Y. Z. Liang, and O. M. Kvalheim, Evolving factor analysis in the presence of heteroscedastic noise. *Analytica Chimica Acta* 263 (1992):29–36.
18. J. Toft and O. M. Kvalheim, Eigenstructure tracking analysis for revealing noise patterns and local rank in instrumental profiles: application to transmittance and absorbance IR spectroscopy. *Chemometrics and Intelligent Laboratory Systems* 19 (1993):65–73.
19. Z. D. Zeng, Y. Z. Liang, Y. L. Wang, X. R. Li, L. M. Liang, Q. S. Xu, C. X. Zhao, B. Y. Li, and F. T. Chau, Alternative moving window factor analysis for comparison analysis between complex chromatographic data. *Journal of Chromatography A* 1107 (2006):273–285.
20. Z. D. Zeng, C. J. Xu, Y. Z. Liang, and B. Y. Li, Sectional moving window factor analysis for diagnosing elution chromatographic patterns. *Chemometrics and Intelligent Laboratory Systems* 69 (2003):89–101.
21. A. de Juan, M. Maeder, T. Hancewicz, and R. Tauler, Local rank analysis for exploratory spectroscopic image analysis. Fixed Size Image Window-Evolving Factor Analysis. *Chemometrics and Intelligent Laboratory Systems* 77 (2005):64–74.
22. E. R. Malinowski, Obtaining the key set of typical vectors by factor analysis, and subsequent isolation of component spectra. *Analytica Chimica Acta* 134 (1982):129–137.
23. A. de Juan, B. van den Bogaert, F. Cuesta Sánchez, and D. L. Massart, Application of the needle algorithm for exploratory analysis and resolution of HPLC-DAD data. *Chemometrics and Intelligent Laboratory Systems* 33 (1996):133–145.
24. W. Windig and J. Guilment, Interactive self-modeling mixture analysis. *Analytical Chemistry* 63 (1991):1425–1432.
25. F. C. Sánchez, J. Toft, B. van den Bogaert, and D. L. Massart, Orthogonal projection approach applied to peak purity assessment. *Analytical Chemistry* 68 (1996):79–85.
26. B. V. Grande and R. Manne, Use of convexity for finding pure variables in two-way data from mixtures. *Chemometrics and Intelligent Laboratory Systems* 50 (2000):19–33.
27. J. H. Jiang, Y. Z. Liang, and Y. Ozaki, On simplex-based method for self-modeling curve resolution of two-way data. *Chemometrics and Intelligent Laboratory Systems* 65 (2003):51–65.
28. W. Windig, B. Antalek, J. L. Lippert, Y. Batonneau, and C. Bremard, Combined use of conventional and second-derivative

data in the SIMPLISMA self-modeling mixture analysis approach. *Analytical Chemistry* 74 (2002):1371–1379.
29. W. Windig, N. B. Gallagher, J. M. Shaver, and B. M. Wise, A new approach for interactive self-modeling mixture analysis. *Chemometrics and Intelligent Laboratory Systems* 77 (2005):85–96.
30. N. B. Gallagher, J. M. Shaver, E. B. Martin, J. Morris, B. M. Wise, and W. Windig, Curve resolution for multivariate images with applications to TOF-SIMS and Raman. *Chemometrics and Intelligent Laboratory Systems* 73 (2004):105–117.
31. L. Duponchel, W. Elmi-Rayaleh, C. Ruckebusch, and J. P. Huvenne, Multivariate curve resolution methods in imaging spectroscopy: Influence of extraction methods and instrumental perturbations.*Journal of Chemical Information and Computer Sciences* 43 (2003):2057–2067.
32. A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, and M. Maeder, Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. TrAC-*Trends in Analytical Chemistry* 23 (2004):70–79.
33. N. Dupuy and Y. Batonneau, Reliability of the contribution profiles obtained through the SIMPLISMA approach and used as reference in a calibration process—Application to Raman micro-analysis of dust particles. *Analytica Chimica Acta* 495 (2003):205–215.
34. Y. Batonneau, C. Bremard, J. Laureyns, J. C. Merlin, and W. Windig, Polarization effects of confocal Raman microspectrometry of crystal powders using interactive self-modeling analysis. *Journal of Physical Chemistry* 107 (2003):1502–1513.
35. R. Tauler, Interpretation of environmental data using chemometrics, in *Sample Handling and Trace Analysis of Pollutants: Techniques, Applications and Quality Assurance*. Ed. D. Barceló (Elsevier, Amsterdam, 2000), 689.
36. Y. Batonneau, C. Bremard, L. Gengembre, J. Laureyns, A. Le Maguer, D. Le Niaguer, E. Perdrix, and S. Sobanska, Speciation of PM10 sources of airborne nonferrous metals within the 3-km zone of lead/zinc smelters. *Environmental Science and Technology* 38 (2004):5281–5289.
37. E. R Malinowski, Window factor analysis: Theoretical derivation and application to flow injection analysis data. *Journal of Chemometrics* 6 (1992):29–40.
38. R. Manne, H. Shen, and Y. Liang, Subwindow factor analysis. *Chemometrics and Intelligent Laboratory Systems* 45 (1999):171–176.
39. O. M. Kvalheim and Y. Z. Liang, Heuristic evolving latent projections- resolving 2-way multicomponent data. 1. Selectivity, latent projective graph, datascope, local rank and unique resolution. *Analytical Chemistry* 64 (1992):936–946.
40. C. J. Xu, Y. Z. Liang, and J. H. Jiang, Resolution of the embedded chromatographic peaks by modified orthogonal projection resolution and entropy maximization method. *Analytical Letters* 33 (2000):2105–2128.
41. J. H. Jiang, S. ŠašiÊ, R. Yu, and Y. Ozaki, Resolution of two-way data from spectroscopic monitoring of reaction or process systems by parallel vector analysis (PVA):and window factor, analysis (WFA):inspection of the effect of mass balance. methods and simulations, *Journal of Chemometrics* 17 (2003):186–197.
42. E. R. Malinowski, Automatic window factor analysis. A more efficient method for determining concentration profiles from evolutionary spectra. *Journal of Chemometrics* 10 (1996):273–279.
43. R. Manne, On the Resolution Problem in Hyphenated Chromatography. *Chemometrics and Intelligent Laboratory Systems* 27 (1995):89–94.
44. Z. D. Zeng, Y. Z. Liang, and C. J. Xu, Comparing chemical fingerprints of herbal medicines using modified window target-testing factor analysis, *Analytical and Bioanalytical Chemistry* 381 (2005):913–924.
45. C. D. Brown and P. D. Wentzell, A modification to window target-testing factor analysis using a Gaussian window. *Chemometrics and Intelligent Laboratory Systems* 51 (2000):3–7.
46. P. J. Gemperline, A priori estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using target factor analysis. *Journal of Chemical Information and Computer Sciences* 24 (1984):206–212.
47. B. G. M. Vandeginste, W. Derks, and G. Kateman. Multicomponent self-modeling curve resolution in high performance liquid chromatography by iterative target transformation factor analysis. *Analytica Chimica Acta* 173 (1985):253–264.
48. R. Tauler and E. Casassas. Application of principal component analysis to the study of multiple equilibria systems. Study of copper (II) salicylate monoethanolamine, diethanolamine and triethanolamine systems. *Analytica Chimica Acta* 223 (1989):257–268.
49. R. Tauler, Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems* 30 (1995):133–146.
50. E. J. Karjalainen, The spectrum reconstruction problem. Use of Alternating Regression for unexpected spectral components in two-dimensional spectroscopies. *Chemometrics and Intelligent Laboratory Systems* 7 (1989):31–38.
51. J. Jaumot, R. Gargallo, A. de Juan, and R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems* 76 (2005):101–110.
52. R. Manne and B. V. Grande, Resolution of two-way data from hyphenated chromatography by means of elementary matrix transformations. *Chemometrics and Intelligent Laboratory Systems* 50 (2000):35–46.
53. C. Mason, M. Maeder, and A. Whitson, Resolving factor analysis. *Analytical Chemistry* 73 (2001):1587–1594.
54. M. Maeder and A. D. Zuberbühler, Nonlinear least-squares fitting of multivariate absorption data. *Analytical Chemistry* 62 (1990):2220–2224.
55. G. Puxty, M. Maeder, and K. Hungerbühler, Tutorial on the fitting of kinetic models to multivariate spectroscopic measurements with non-linear least-squares regression. *Chemometrics and Intelligent Laboratory Systems* 81 (2006):149–164.
56. A. de Juan, M. Maeder, M. Martínez, and R. Tauler. Combining hard- and soft-modelling to solve kinetic problems. *Chemometrics and Intelligent Laboratory Systems* 54 (2000):123–141.
57. S. Biljsma and A. K. Smilde. Application of curve resolution based methods to kinetic data. *Analytica Chimica Acta* 396 (1999):231–240.
58. E. Bezemer and S. C. Rutan. Multivariate curve resolution with non-linear fitting of kinetic profiles. *Chemometrics and Intelligent Laboratory Systems* 59 (2001):19–31.
59. J. Diewok, A. de Juan, M. Maeder, R. Tauler, and B. Lendl. Application of a combination of hard and soft modeling for equilibrium

systems to the quantitative analysis of pH-modulated mixture samples. *Analytical Chemistry* 75 (2003):641–647.
60. J. M. Amigo, A. de Juan, J. Coello, and S. Maspoch, Hard-soft modelling approaches to study and monitor enzymatic systems in biological fluids. *Analytica Chimica Acta* 567 (2006):245–254.
61. J. Jaumot, P. J. Gemperline, and A. Stang, Non-negativity constraints for elimination of multiple solutions in fitting of multivariate kinetic models to spectroscopic data. *Journal of Chemometrics* 19 (2005):97–106.
62. A. K. Smilde, H. C. J. Hoefsloot, H. A. L. Kiers, S. Bijlsma, and H. F. M. Boelens, Sufficient conditions for unique solutions within a certain class of curve resolution models. *Journal of Chemometrics* 15 (2001):405–411.
63. P. Jandanklang, M. Maeder, and A.C. Whitson, Target transform fitting: a new method for the non-linear fitting of multivariate data with separable parameters. *Journal of Chemometrics* 15 (2001):511–522.
64. A. R. Carvalho, J. Wattoom, L. F. Zhu, and R. G. Brereton, Combined kinetics and iterative target transformation factor analysis for spectroscopic monitoring of reactions. *Analyst* 131 (2006):90–97.
65. R. Sánchez-Ponce and S. C. Rutan, Steady state kinetic model constraint for Multivariate Curve Resolution-Alternating Least Squares analysis. *Chemometrics and Intelligent Laboratory Systems* 77 (2005):50–58.
66. J. M. Amigo, A. de Juan, J. Coello, and S. Maspoch, Hard-soft multivariate curve resolution strategies for the quantitative determination of oxipurines and uric acid in human urine. *Analytica Chimica Acta* 567 (2006):236–244.
67. J. M. Díaz-Cruz, J. Agulló, M. S. Díaz-Cruz, C. Ariño, M. Esteban, and R. Tauler. Implementation of a chemical equilibrium constraint in the multivariate curve resolution of voltammograms from systems with successive metal complexes. *Analyst* 126 (2001):371–377.
68. M. J. López, C. Ariño, S. Díaz-Cruz, J. M. Díaz-Cruz, R. Tauler and M. Esteban, Voltammetry assisted by multivariate analysis as a tool for speciation of metallothioneins: Competitive complexation of α- and β-metallothionein domains with cadmium and zinc. *Environmental Science and Technology* 37 (2003):5609–5616.
69. R. Huo, R. Wehrens, and L. M. C. Buydens, Improved DOSY NMR data processing by data enhancement and combination of multivariate curve resolution with non-linear least square fitting. *Journal of Magnetic Resonance* 169 (2004):257–269.
70. R. Bro and S. de Jong, A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics* 11 (1997):393–401.
71. R. Bro and N. D. Sidiropoulos, Least squares algorithms under unimodality and non-negativity constraints. *Journal of Chemometrics* 12 (1998):223–247.
72. M. H. Van Benthem, M. R. Keenan, and D. M. Haaland, Application of equality constraints on variables during alternating least squares procedures. *Journal of Chemometrics* 16 (2002):613–622.
73. P. J. Gemperline, and E. Cash, Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with penalty functions. *Analytical Chemistry* 75 (2003):4236–4243.
74. R. Tauler, A. Izquierdo-Ridorsa,, and E. Casassas, Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems* 18 (1993):293–300.
75. R. Tauler, B. Kowalski,, and S. Fleming, Multivariate curve resolution applied to spectral data from multiple runs of an industrial process, *Analytical Chemistry* 65 (1993):2040–2047.
76. R. Tauler and D. Barceló, Multivariate curve resolution applied to liquid chromatography-diode array detection. *TrAC-Trends in Analytical Chemistry* 12 (1993):319–327.
77. P. V. van Zomeren, A. Hoogvorst, P. M. J. Coenegracht, and G. J. de Jong, Optimisation of high-performance liquid chromatography with diode array detection using an automatic peak tracking procedure based on augmented iterative target transformation factor analysis. *Analyst* 129 (2004):241–248.
78. C. A. Holden, S. S. Hunnicutt, R. Sanchez-Ponce, J. M. Craig, and S. C. Rutan, Study of complexation in methanol/water mixtures by infrared and Raman spectroscopy and multivariate curve resolution – Alternating least-squares analysis. *Applied Spectroscopy* 57 (2003):483–490.
79. J. Jaumot, R. Eritja, R. Tauler, and R. Gargallo, Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand. *Nucleic Acids Research* 34 (2006):206–216.
80. B. Czarnik-Matusewicz, S. Pilorz, and J.P. Hawranek, Temperature-dependent water structural transitions examined by near-IR and mid-IR spectra analyzed by multivariate curve resolution and two-dimensional correlation spectroscopy. *Analytica Chimica Acta* 544 (2005):15–25.
81. S. Navea, A. de Juan, and R. Tauler, Detection and resolution of intermediate species in protein folding processes using fluorescence and circular dichroism spectroscopies and multivariate curve resolution, *Analytical Chemistry* 64 (2002):6031–6039.
82. M. S. Díaz-Cruz, J. Mendieta, R. Tauler, and M. Esteban, Multivariate curve resolution of cyclic voltammetric data. Application to the study of the cadmium-binding properties of glutathione. *Analytical Chemistry* 71 (1999):4629–4636.
83. A. Blanco, A. C. Peinado, and J. Mas, Elucidating the composition profiles of alcoholic fermentations by use of ALS methodology. *Analytica Chimica Acta* 544 (2005):199–205.
84. M. Garrido, I. Lázaro, M. S. Larrechi, and F. X. Rius, Multivariate resolution of rank-deficient near-infrared spectroscopy data from the reaction of curing epoxy resins using the rank augmentation strategy and multivariate curve resolution alternating least squares approach. *Analytica Chimica Acta* 5 (2004):65–73.
85. B. Ma, P. J. Gemperline, E. Cash, M. Bosserman, and E. Comas, Characterizing batch reactions with in situ spectroscopic measurements, calorimetry and dynamic modelling. *Journal of Chemometrics* 17 (2003):470–479.
86. E. N. M. Van Sprang, H. J. Ramaker, J. A. Westerhuis, A. K. Smilde, S. P. Gurden, and D. Wienke, Near-infrared spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model. *Applied Spectroscopy* 57 (2003):1007–1019.
87. C. B. Zachariassen, J. Larsen, F. van den Berg, R. Bro, A. de Juan, and R. Tauler, Comparison of PARAFAC2 and MCR-ALS for resolution of an analytical liquid dilution system. *Chemometrics and Intelligent Laboratory Systems* 83 (2006):13–25.

88. A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences* (John Wiley & Sons, New York, 2004).
89. R. Tauler, I. Marqués, and E. Casassas, Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid-base titration of salicylic scid at three excitation wavelengths. *Journal of Chemometrics* 12 (1998):55–75.
90. E. Peré-Trepat, A. Ginebreda, and R. Tauler, Main patterns of heavy metal ions distribution in fish, sediments and waters from Catalonia rivers using different multiway data analysis methods. *Chemometrics and Intelligent Laboratory Systems* (in press).
91. J. Saurina and S. Hernández-Cassou, Quantitative determinations in conventional flow injection analysis based on different chemometric calibration statregies: a review. *Analytica Chimica Acta* 438 (2001):335–352.
92. E. Peré-Trepat, A. Hildebrandt, D. Barceló, S. Lacorte, and R. Tauler, Fast chromatography of complex biocide mixtures using diode array detection and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems* 74 (2004):293–303.
93. J. Saurina and R. Tauler, Strategies for solving matrix effects in the analysis of triphenyltin in sea-water samples by three-way multivariate curve resolution. *Analyst* 125 (2000):2038–2043.
94. A. Pasamontes, and M.P. Callao, Sequential injection analysis for the simultaneous determination of clavulanic acid and amoxicillin in pharmaceuticals using second-order calibration. *Analytical Sciences* 22 (2006):131–135.
95. M. J. Rodríguez-Cuesta, R. Boqué, F. X. Rius, J. L. M. Vidal, and A. G. Frenich, Development and validation of a method for determining pesticides in groundwater from complex overlapped HPLC signals and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems* 77 (2005):251–260.
96. J. Saurina, C. Leal, R. Compañó, M. Granados, M.D. Prat, and R. Tauler, Estimation of figures of merit using univariate statistics for quantitative second-order multivariate curve resolution. *Analytica Chimica Acta* 432 (2001):241–251.
97. M. Rodríguez-Cuesta, R. Boqué, and F. X. Rius, Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data. *Analytica Chimica Acta* 476 (2003):111–122.
98. C. B. Zachariassen, J. Larsen, F. van den Berg, R. Bro A. de Juan, and R. Tauler, Multi-way analysis for investigation of industrial pectin using an analytical liquid dilution system. *Chemometrics and Intelligent Laboratory Systems* (2006):(in press).
99. P. J. Gemperline, Computation of the range of feasible solutions in self-modeling curve resolution algorithms. *Analytical Chemistry* 71 (1999):5398–5404.
100. 100. R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *Journal of Chemometrics* 15 (2001):627–646.
101. M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, and M. Maeder, On rotational ambiguity in model-free analyses of multivariate data. *Journal of Chemometrics* (2006):(in press).
102. R. Rajko and K. Istvan, Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR): Method based on computational geometry. *Journal of Chemometrics* 19 (2005):448–463.
103. M. N. Leger, and P. D. Wentzell, Dynamic Monte Carlo self-modeling curve resolution method for multicomponent mixtures. *Chemometrics and Intelligent Laboratory Systems* 62 (2002):171–188.
104. M. Garrido, M. S. Larrechi, F. X. Rius, and R. Tauler, Calculation of band boundaries of feasible solutions obtained by Multivariate Curve Resolution Alternating Least Squares of multiple runs of a reaction monitored by NIR spectroscopy. *Chemometrics and Intelligent Laboratory Systems* 76 (2005):111–120.
105. J. Jaumot, J. C. Menezes, and R. Tauler. Quality assessment of the results obtained by Multivariate Curve Resolution analysis of multiple runs of gasoline blending processes. *Journal of Chemometrics* (in press).
106. J. Jaumot, R. Gargallo, and R. Tauler, Estimation of error propagation and prediction intervals in MCR-ALS using resampling methods. *Journal of Chemometrics* 18 (2004):327–340.
107. J. Riu and R. Bro, Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemometrics and Intelligent Laboratory Systems* 65 (2003):35–49.
108. S. Bijlsma and A. K. Smilde, Estimating reaction rate constants from a two-step reaction: a comparison between two-way and three-way methods. *Journal of Chemometrics* 14 (2000):541–560.
109. K. Faber, Comment on a recently proposed resampling method. *Journal of Chemometrics* 15(2001):169–188.
110. J. H. Wang and P. K. Hopke, Estimation of the heteroscedastic noise in large data arrays. *Analytica Chimica Acta* 412 (2000):177–184.
111. M. Vives, R. Gargallo, and R. Tauler, Multivariate extension of the continuous variation and mole-ratio methods for the study of the interaction of intercalators with polynucleotides. *Analytica Chimica Acta* 424 (2000):105–114.
112. S. Navea, A. de Juan, and R. Tauler, Modeling temperature-dependent protein structural transitions by combined near-IR and mid-IR spectroscopies and multivariate curve resolution. *Analytical Chemistry* 75 (2003):5592–5601.
113. J. Jaumot, V. Marchan, R. Gargallo, A. Grandas, and R. Tauler, Multivariate curve resolution applied to the analysis and resolution of two-dimensional [H-1,N-15] NMR reaction spectra, *Analytical Chemistry* 76 (2004):7094–7101.
114. M. Esteban, C. Ariño, and J. M. Díaz-Cruz, Chemometrics for the analysis of voltammetric data. *TrAC- Trends in Analytical Chemistry* 25 (2006):86–92.
115. C. Ruckebusch, L. Duponchel, J. P. Huvenne, and J. Saurina, Multivariate curve resolution of step-scan FTIR spectral data. *Vibrational Spectroscopy* 35 (2004):21–26.
116. E. Peré-Trepat and R. Tauler, Analysis of environmental samples by application of multivariate curve resolution on fused high-performance liquid chromatography-diode array detection mass spectrometry data. *Journal of Chromatography A* 1131 (2006):85–96.
117. K. de Braekeleer, A. de Juan, and D. L. Massart, Purity assessment and resolution of tetracycline hydrochloride samples analysed using high-performance liquid chromatography with diode array detection. *Journal of Chromatography A* 832 (1999):67–86.
118. E. Peré-Trepat, S. Lacorte, and R. Tauler, Solving liquid chromatography mass spectrometry coelution problems in the analysis

of environmental samples by multivariate curve resolution. *Journal of Chromatography A* 1096 (2005):111–122.
119. N. R. Marsili, A. Lista, B. S. F. Band, H. C. Goicoechea, and A. C. Olivieri, New method for the determination of benzoic and sorbic acids in commercial orange juices based on second-order spectrophotometric data generated by a pH gradient flow injection technique. *Journal of Agricultural and Food Chemistry* 52 (2004):2479–2484.
120. J. H. Wang, P. K. Hopke, T. M. Hancewicz, and S. L. L. Zhang, Application of modified alternating least squares regression to spectroscopic image analysis. *Analytica Chimica Acta* 476 (2003):93–109.
121. J. A. Timlin, D. M. Haaland, M. B. Sinclair, A. D. Aragon, M. J. Martinez, and M. Werner-Washburne, Hyperspectral microarray scanning: impact on the accuracy and reliability of gene expression data. *BMC Genomics* 6 (2005):art. 72.
122. T. M. Hancewicz and J. H. Wang, Discriminant image resolution: a novel multivariate image analysis method utilizing a spatial classification constraint in addition to bilinear nonnegativity. *Chemometrics and Intelligent Laboratory Systems* 77 (2005):18–31.
123. P. K. Hopke, *Receptor Modeling in Environmental Chemistry* (John Wiley, New York, 1985).
124. P. K. Hopke, Recent developments in receptor modeling. *Journal of Chemometrics* 17 (2003):255–265.
125. R. Tauler, S. Lacorte, M. Guillamón, R. Céspedes, P. Viana, and D. Barceló, Chemometric modeling of main contamination sources in surface waters of Portugal. *Environmental Science & Technology* 23 (2004):565–575.
126. Y. L. Xie, P. K. Hopke, and P. Paatero, Positive matrix factorization applied to a curve resolution problem. *Journal of Chemometrics* 12 (1998):357–364.
127. P. Paatero and U. Tapper, Positive matrix factorization - a non-negative factor model with optimal utilization of error-estimates of data values. *Environmetrics* 5 (1994):111–126.
128. P. Paatero, The multilinear engine—A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model.*Journal of Computational and Graphical Statistics* 8 (1999):854–888.
129. R. C. Henry, Multivariate receptor modeling by N-dimensional edge detection. *Chemometrics and Intelligent Laboratory Systems* 65 (2003):179–189.
130. E. Kim, P. K. Hopke, T. V. Larson, and D. S. Covert, Analysis of ambient particle size distributions using unmix and positive matrix factorization. *Environmental Science and Technology* 38 (2004):202–209.
131. I. Stanimirova, B. Walczak, and D.L. Massart. Multiple factor analysis in environmental chemistry. *Analytica Chimica Acta* 545 (2005):1–12.
132. E. Peré-Trepat, M. Petrovic, D. Barceló, and R. Tauler, Application of chemometric methods to the investigation of main microcontaminant sources of endocrine disruptors in coastal and harbour waters and sediments. *Analytical and Bioanalytical Chemistry* 378 (2004):642–654.
133. R. Leardi, C. Armanino, S. Lanteri, and L. Alberotanza, Three-mode principal component analysis of monitoring data from Venice lagoon. *Journal of Chemometrics* 14 (2000):187–195.
134. E. Kim, P. K. Hopke, P. Paatero, and E. S. Edgerton, Incorporation of parametric factors into multilinear receptor model studies of Atlanta aerosol. *Atmospheric Environment* 37 (2003):5009–5021.
135. P. Paatero and P. K. Hopke, Utilizing wind direction and wind speed as independent variables in multilinear receptor modeling studies. *Chemometrics and Intelligent Laboratory Systems* 60 (2002):25–41.
136. M. Terrado, D. Barceló, and R. Tauler. Identification and distribution of contamination sources in the Ebro river basin by chemometrics modelling coupled to geographical information systems. *Talanta* 70 (2006):691–704.
137. M. Vives, R. Gargallo, and R. Tauler, Multivariate extension of the continuous variation and mole-ratio methods for the study of the interaction of intercalators with polynucleotides. *Analytica Chimica Acta* 424 (2000):105–114.
138. S. Navea, A. de Juan, and R. Tauler, Modeling temperature-dependent protein structural transitions by combined near-IR and mid-IR spectroscopies and multivariate curve resolution. *Analytical Chemistry* 75 (2003):5592–5601.
139. P. Johnson, A. I. Johansson, J. Gullberg, J. Trygg, J. A. B. Grung, S. Marklund, M. Sjöström, H. Antti, and T. Moritz, High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Analytical Chemistry* 77 (2005):5635–5642.
140. T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and S. E. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286 (1999):531–537.
141. J. Jaumot, R. Tauler, and R. Gargallo, Exploratory data analysis of DNA microarrays by Multivariate Curve Resolution. *Analytical Biochemistry* 358 (2006):76–89.
142. J. Van der Greef and A. K. Smilde, Symbiosis of chemometrics and metabolomics: past, present and future. *Journal of Chemometrics* 19 (2005):376–386.
143. A. K. Smilde, M. J. van der Werf, S. Bijlsma, B. J. C. van den Werff-van der Vat, and R. H. Jellema, Fusion of mass spectrometry-based metabolomics data. *Analytical Chemistry* 77 (2005):6729–6736.