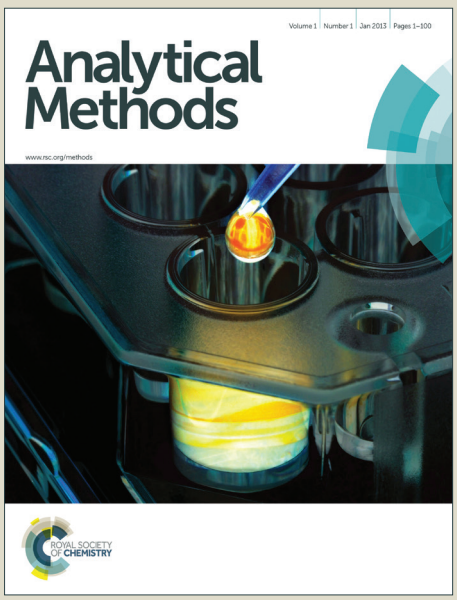
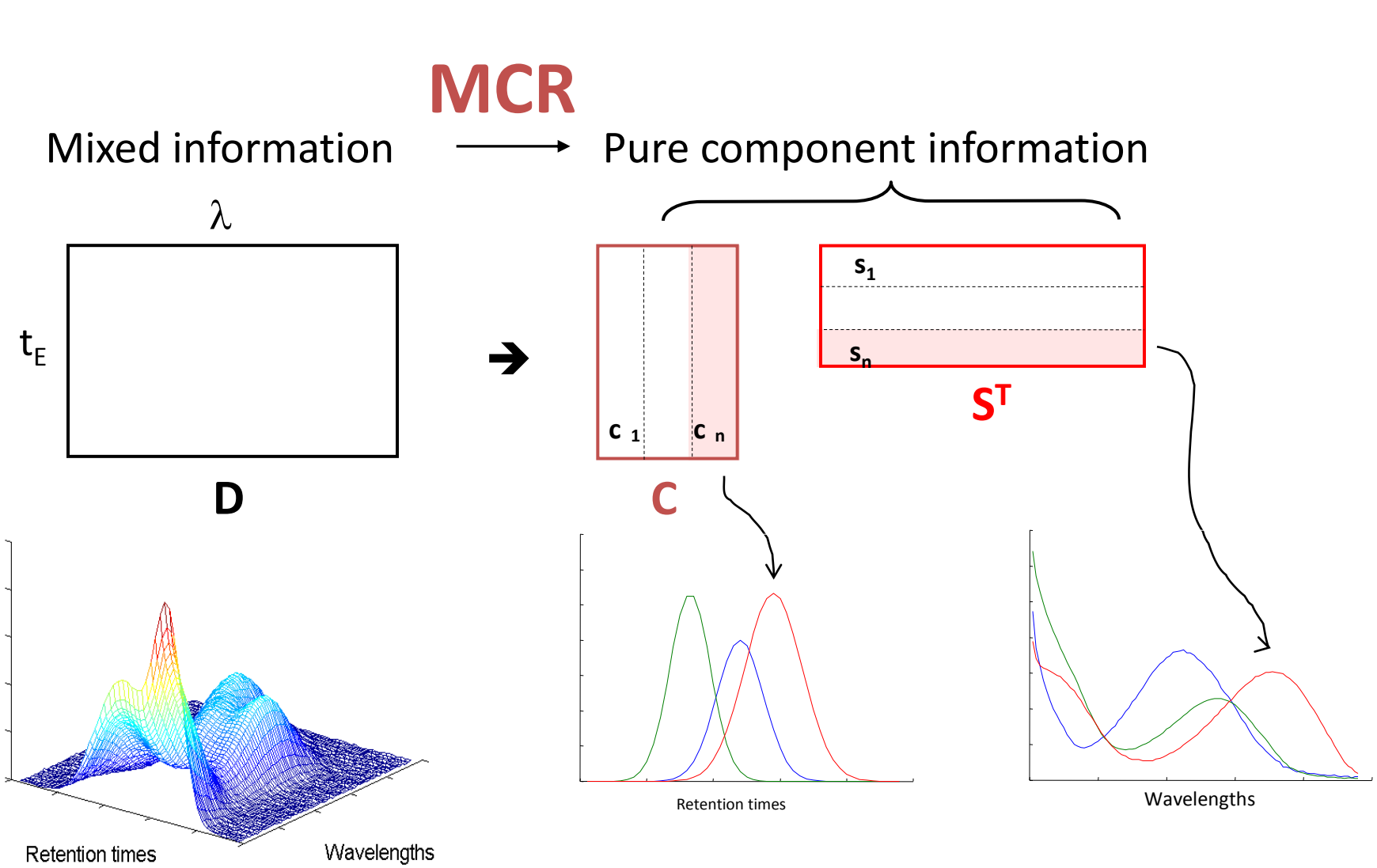
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods



Multivariate Curve Resolution (MCR). Solving the mixture analysis problem

Anna de Juan[1, (*)], Joaquim Jaumot[2], Romà Tauler[2].

1. Chemometrics group. Dept. Analytical Chemistry. Universitat de Barcelona. Diagonal, 645. 08028 Barcelona. anna.dejuan@ub.edu.
2. IDAEA-CSIC. Jordi Girona, 18. 08034, Barcelona.

Tutorial structure

1. Introduction. The concept and underlying model. Related methods.
2. When can MCR be used?
3. Data set configurations.
4. Algorithms.
5. Iterative MCR methods. The modus operandi.
    a. Initial estimates.
    b. Constraints.
6. Uncertainty in MCR solutions.
7. MCR. Case studies. A single image data set and an HPLC-DAD data set.

1. Introduction. The concept and underlying model. Related methods.

Multivariate Curve Resolution (MCR) is the generic denomination of a family of methods meant to solve the mixture analysis problem, i.e., able to provide a chemically (scientifically) meaningful additive bilinear model of pure contributions from the sole information of an original data matrix including a mixed measurement [1-6]. A paradigm for MCR analysis is the spectroscopic data sets. As an example, let us imagine a data set **D** related to an HPLC-DAD separation. The rows design the elution times and the columns the spectral channels. Each row in the matrix **D** contains a spectrum recorded at a particular elution time and each column a chromatogram related to a particular wavelength (see figure 1). MCR analysis describes this data set as the sum of the signal contributions coming from each of the eluted components (eq. 1). Each term of this additive model can be expressed by the product of a dyad of profiles, $\mathbf{c}_i\mathbf{s}_i^T$, where $\mathbf{s}_i^T$ is the pure spectrum of the component weighted by the related elution profile, $\mathbf{c}_i$.(eq. 2) Finally, the bilinear additive model can be expressed in a compact form with the equation 3:

$$\mathbf{D} = \sum_i \mathbf{D}_i + \mathbf{E} \quad \text{(eq. 1)}$$

$$\mathbf{D} = \sum_i \mathbf{c}_i\mathbf{s}_i^T + \mathbf{E} \quad \text{(eq. 2)}$$

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{(eq. 3)}$$

where **C** contains the elution profiles of all components and $\mathbf{S}^T$ the related pure spectra. **E** is the matrix expressing the error or variance unexplained by the bilinear model in all equations above. Although MCR is not restricted to the analysis of spectroscopic data, equation 3 is the form most often used to express the bilinear MCR model. In many examples of MCR analysis, the matrix $\mathbf{S}^T$ is related to the qualitative information of the components in the system analyzed (pure instrumental responses or basic signatures of different kinds), whereas the matrix **C** expresses the abundance or apportionment of the qualitative pure responses along the row direction of the data set.

As will be seen in next sections, the singularity of MCR methods is the fact that they provide meaningful models with component profiles that a chemist (or a scientist) can recognize as real instrumental responses or apportionment profiles because they present the natural properties expected from these chemical entities. Such a characteristic helps enormously in the interpretation of the results and makes that the results provided by MCR methods be easily understood by non-chemometricians. This is the most relevant difference with other data analysis methods that also describe the data through a bilinear model. Many of them are exploratory methods, such as Principal Component Analysis (PCA) [7] or Independent Component analysis (ICA) [8]. In these cases, the goal is finding a bilinear description of the data through orthogonal contributions sorted in decreasing variance order (PCA) or through statistically independent contributions. Although these methods have a clear exploratory value, they generally do not provide the true chemical or scientific model because the real mixed contributions do not hold the orthogonal or statistical independency as natural properties [9,10].

2. When can MCR be used?

Traditionally, MCR was conceived for evolutionary analytical data coming from a process or an analytical measurement [2-5]. Many analytical measurements, particularly all those based on spectroscopic methods are particularly suited to be analyzed by MCR, since the underlying analytical model (the known Beer-Lambert law) is formally a bilinear model of

pure signal contributions (see equation 3). Therefore, MCR fits naturally these measurements because it mimics exactly the structure of the analytical measurement.

It is logical that the first examples for which MCR found application were processes (reactions) monitored spectroscopically or chromatographic evolutions with a multivariate spectroscopic detection [11,12]. All spectroscopies and other analytical measurements, e.g., some electrochemical data [13], could also be well described by a bilinear model. MCR was also meant in the beginning to take advantage of the evolutionary character in the concentration direction (related to a process, to a chromatographic elution). This is why MCR is traditionally applied to second order data sets, where both directions in the bilinear model have a clear meaning.

Nowadays, the application of MCR has grown significantly and the fields of application have gained in complexity and diversity. Within the spectroscopic field, the structure in the concentration direction is no longer a requirement and this has allowed analyzing hyperspectral images, in which the image cube with two spatial directions (x- and y-) and a spectroscopic direction, can be unmixed previous unfolding of the cube into a data matrix with rows designing the pixel spectra and columns the spectral channels measured. After MCR, an operation of refolding of the concentration profiles is needed to recover the spatial structure of the distribution maps of the pure image components (see figure 2) [14-16].

Whether the pure contributions of a data set follow a natural bilinear model or not, MCR can always be used to furnish a simple bilinear description of the variation in the data that may include some natural properties related to the problem of interest. In these cases, the concept of component is redefined according to the new kind of problem studied. A clear example is found in the environmental field, where data tables of environmental measurements, e.g., tables with rows containing concentrations of contaminants in different locations, have been described as bilinear models of contamination sources, where geographic profiles and composition profiles form the dyads in the bilinear model [17,18]. One of the most recent examples that may be quite paradigmatic in this sense is constituted by genomic data, such as those coming from DNA microarrays. Here, each row of the data table is formed by gene expressions related to a particular cell line (sample) and the bilinear model is formed by dyads having basic gene expression signatures linked to sample profiles, where the abundance of a particular gene expression signature is encoded [19,20].

In this last kind of data sets, there is usually no physicochemical underlying bilinear model that can support the use of MCR but, still, the results obtained can be a good exploratory tool to understand the behaviour of the data and, often, instead of using mathematical assumptions in the data decomposition step, related basic knowledge (if available) can be implemented to drive MCR to the final solutions.

3. Data set configurations.

MCR was born to analyze single second-order data matrices [2-4]. This kind of data set provides the basic bilinear model in equation 3. In early 90's, with the birth of multi-way methods and the increase of complexity of analytical measurements, experiments and chemical problems, there was a clear need to analyze simultaneously several data tables together. In doing so, the amount and complementariness of information handled increased and so did the quality of the final solutions (profiles of pure contributions) retrieved by MCR [4,5,21- 25].

The data configurations including more than one data matrix are called augmented matrices or multisets and, hence, the denomination of multiset analysis for the data treatment of this kind of structures. Multiset structures can be extremely flexible and can be formed by data matrices of different size and meaning. Multisets can be structured as: a) a row-wise augmented matrix, when data tables are appended besides each other, e.g., a multitechnique monitoring of a process (eq. 4) [5,6,23,24,25] b) a column-wise augmented matrix, when data tables are one on top of each other, e.g., the different batches monitored in a production process (eq. 5) [23,26,27], and c) row- and column-wise augmented matrices, when the multiset extends in the row and column direction (eq. 6) [23,28,29]. Multiset structures look more complex, but they also obey the basic bilinear model and, therefore, MCR can be equally applied.

$$\begin{bmatrix}\mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & .... & \mathbf{D}_L\end{bmatrix} = \mathbf{C}\begin{bmatrix}\mathbf{S}_1^T & \mathbf{S}_2^T & \mathbf{S}_3^T & .... & \mathbf{S}_L^T\end{bmatrix} + \begin{bmatrix}\mathbf{E}_1 & \mathbf{E}_2 & \mathbf{E}_3 & .... & \mathbf{E}_L\end{bmatrix} = \mathbf{C}\mathbf{S}_{aug}^T + \mathbf{E}_{aug}$$

(eq. 4)

$$\begin{pmatrix}\mathbf{D}_1\\ \mathbf{D}_2\\ \mathbf{D}_3\\ ....\\ \mathbf{D}_K\end{pmatrix} = \begin{pmatrix}\mathbf{C}_1\\ \mathbf{C}_2\\ \mathbf{C}_3\\ ....\\ \mathbf{C}_K\end{pmatrix}\mathbf{S}^T + \begin{pmatrix}\mathbf{E}_1\\ \mathbf{E}_2\\ \mathbf{E}_3\\ ....\\ \mathbf{E}_K\end{pmatrix} = \mathbf{C}_{aug}\mathbf{S}^T + \mathbf{E}_{aug}$$

(eq. 5)

$$\begin{bmatrix}\mathbf{D}_{11} & \mathbf{D}_{12} & \mathbf{D}_{13} & .... & \mathbf{D}_{1L}\\ \mathbf{D}_{21} & \mathbf{D}_{22} & \mathbf{D}_{23} & .... & \mathbf{D}_{2L}\\ \mathbf{D}_{31} & \mathbf{D}_{32} & \mathbf{D}_{33} & .... & \mathbf{D}_{3L}\\ .... & .... & .... & .... & ....\\ \mathbf{D}_{K1} & \mathbf{D}_{K2} & \mathbf{D}_{K3} & .... & \mathbf{D}_{KL}\end{bmatrix} = \begin{pmatrix}\mathbf{C}_1\\ \mathbf{C}_2\\ \mathbf{C}_3\\ ....\\ \mathbf{C}_K\end{pmatrix}\begin{bmatrix}\mathbf{S}_1^T & \mathbf{S}_2^T & \mathbf{S}_3^T & .... & \mathbf{S}_L^T\end{bmatrix} + \begin{pmatrix}\mathbf{E}_{11} & \mathbf{E}_{12} & \mathbf{E}_{13} & .... & \mathbf{E}_{1L}\\ \mathbf{E}_{21} & \mathbf{E}_{22} & \mathbf{E}_{23} & .... & \mathbf{E}_{2L}\\ \mathbf{E}_{31} & \mathbf{E}_{32} & \mathbf{E}_{33} & .... & \mathbf{E}_{3L}\\ .... & .... & .... & .... & ....\\ \mathbf{E}_{K1} & \mathbf{E}_{K2} & \mathbf{E}_{K3} & .... & \mathbf{E}_{KL}\end{pmatrix} =$$

$$= \mathbf{C}_{aug}\mathbf{S}_{aug}^T + \mathbf{E}_{aug}$$

(eq. 6)

Building a meaningful multiset requires several conditions:

- The data tables analyzed together should have some kind of information (components) in common.
- The data tables appended should have a common mode (for simplification, the concentration or the spectral mode). Please note that having a 'common mode' is not only a question of size (dimensionality) but also of profile behavior. For instance, when a process is monitored simultaneously with different analytical techniques, not only there are as many measurements (rows) as process stages in all data matrices, but the concentration matrix for all data tables is common, because the process monitored is the same. A different example would be the one formed by several HPLC-DAD data tables with the same number of elution times and the same spectral range used for detection. Whereas from a dimensional point of view, a row-wise or a column-wise augmented matrix is possible, only the column-wise multiset is meaningful in this case. The reason

is that the only common mode is the spectral one, i.e., the spectral shape of a pure component is invariant in all chromatograms, whereas the elution profile for a certain component can vary in location and shape among chromatograms.

4. Algorithms.

There have been several algorithms since the beginning of MCR methods meant to provide the bilinear model of pure contributions from the sole mixed original data table. The main distinction was formerly among non-iterative methods and iterative methods [30,31]. Most non-iterative methods, such as Heuristic Evolving Latent Projections (HELP) [32], Window Factor Analysis (WFA) [33], Subwindow Factor Analysis (SFA) [34]... were born when MCR was circumscribed to the analysis of a single process-like data table. In these cases, the structured/sequential structure of the concentration profiles was conveniently used to define the windows of existence of the different components. Then, subspaces combining suitably windows with different conditions of component overlap (in terms of component identity and rank) were furnishing either the concentration profiles or the spectra and the counterpart of the bilinear model was obtained by a single least squares step. These algorithms are nowadays less in use because setting the concentration windows in systems with large number of components and non-sequential order or with non-structured concentration direction hinders the application of these methods. Moreover, they were not extended to deal with multiset structures, a common need in current data analysis.

Iterative methods work by optimizing a set of initial estimates (guesses) of concentration profiles or spectra under the action of constraints until a convergence criterion is achieved [30]. The most known are ITTFA (Iterative Target Transformation Factor Analysis) [35,36] and Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) [5,6,21-23,31]. ITTFA works optimizing the concentration profile direction under constraints and recovering the spectra matrix in the bilinear model by least-squares, whereas MCR-ALS performs an alternating optimization of both concentration profiles and pure spectra in each iterative cycle. In areas such as remote sensing analysis, MCR methods are called often unmixing methods or end-member analysis methods and also have an iterative nature [37] and in environmental sciences, these procedures are applied to find out the so-called receptor models [38] .

Iterative methods are usually the preferred option nowadays because it is not needed to have a structured concentration direction for them to be applied and much chemical and/or mathematical information can be introduced in the optimization process under the form of constraints [5,6,23,30]. Besides, algorithms such as MCR-ALS have been adapted to work with multiset structures and have allowed for the introduction of many different constraints related to chemical information [5,6,30,21], to mathematical properties (rank conditions or model structure) [23,39-41] or to data analysis tasks usually associated with other kind of algorithms (e.g., hard- modeling [42] or calibration [43]). These advantages justify that the rest of the manuscript be oriented to the details of application of MCR-ALS, since we think that it is an algorithm that can be used in many different instances and many of the comments performed can also apply to other iterative MCR methods.

5. Iterative MCR methods. The modus operandi.

As mentioned before, MCR-ALS works on a single data matrix or on a multiset by optimizing initial estimates (concentration profiles or spectra) in an alternating least-squares way within each iterative cycle under the action of suitable constraints until a convergence

criterion is fulfilled. The convergence criterion can be a preset number of iterations or a threshold value defining the difference in fit improvement between consecutive iterations. Once the optimization process has been finished, the MCR results are the set of concentration profiles and spectra and quality parameters related to the model fit, such as the variance explained or the lack of fit (LOF).

$$\%\, LOF = 100 \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \qquad \text{(eq. 7)}$$

Where d_{ij} is an element of the data matrix **D** and e_{ij} is the related residual. To obtain satisfactory results, it is relevant to pay attention to the steps of selection of initial estimates and constraints. Initial estimates are the starting point of the optimization process and one has to try to begin with sensible guesses. On the other hand, constraints are the properties that the profiles should obey. Therefore, selecting the appropriate ones and knowing how to apply them is the most crucial point to ensure obtaining meaningful and reliable solutions, affected by as less uncertainty as possible.

a. Initial estimates.

Initial estimates in MCR-ALS can be concentration profiles or spectra (or, in general, pure column or row profiles). The golden rule is starting with sensible initial guesses, which already obey the constraints that will be applied during the optimization process, i.e., initial random values are of no help if profiles with clearly definable properties are sought [5,6,30]. Initial estimates can come from previous knowledge if available (e.g., if the pure spectra of some of the components in the system are known, they can be used); however, there are methods that can help in the task of obtaining this initial information.

The main distinction is between methods designed to obtain initial estimates for process-like data and general methods that can work when no structure in the concentration direction is available. Among methods designed for process-like data, Evolving Factor Analysis (EFA) is the most known [44,45]. EFA is a local rank analysis method that detects the emergence and decay of the components in the data set and provides concentration profiles assuming a sequential order of emergence-decay for all components in the system. Whenever a sequential process is under analysis, the use of EFA is advisable, since it provides a set of initial estimates and relevant additional information, such as the windows of existence of the components in the system (used to set selectivity or local rank constraints, see afterwards). EFA is designed to act on a single data set. When multisets formed by different experiments are analyzed, EFA should be applied separately to each experiment. Afterwards, an augmented initial estimate appending the different individual EFA results is built, always keeping the correct correspondence among species in the different experiments.

Some other methods to build initial estimates can work irrespectively of the presence or absence of a structured concentration direction in the data set. A common denomination of these methods is pure variable selection methods [46]. Some of the most known are Simple-to-use self-modeling analysis (SIMPLISMA) [47], Orthogonal Projection Approach (OPA) [48], Key Set Factor Analysis (KSFA) [49], … All of them aim at selecting the most dissimilar rows or columns in a single data matrix or a multiset structure providing initial estimates of spectra or of concentration profiles, respectively. A general advice in using

these methods is applying them in the direction of least overlap among components, e.g., in an HPLC-DAD data set, the component overlap in the elution direction is lower than in the spectral direction. Therefore, the sensible option is trying to select the purest elution times and work with initial estimates formed by the spectra related to the selected times.

The methods of pure variable selection can be applied to a single data set or to a full multiset structure and provide in a single step initial estimates for single or augmented concentration or spectral matrices. The only drawback linked to many of these methods is that they are designed to work with positive data. When the data set contains negative values, e.g., in derivative spectra, some alternatives are offered to perform the selection of the purest rows or columns on 'positive modifications' of the data set, such as inverted second derivatives, data sets with an added offset,…[46] Initial estimates can be built afterwards with the measurements in the original data set related to the purest rows or columns selected.

b. Constraints.

Constraints are the properties that give shape to the row and column profiles in the bilinear model of the data set. Constraints are implemented as mathematical conditions and drive the MCR optimization process to the final solutions. There are constraints linked to natural properties of the measurement of the data set and others that express mathematical conditions or models [21,23,30].

Among the most known natural constraints are:

- **Non-negativity**. Forces the profiles to be formed by positive values. Can be implemented replacing negative values by zeros or with softer algorithms, such as non-negative least-squares or fast non-negative least-squares. It applies to all concentration profiles and to many instrumental responses.
- **Unimodality**. The presence of a single maximum per profile is allowed. It is the condition fulfilled by peak-shaped signals, e.g., elution profiles, and monotonic reaction profiles (always increasing or decreasing).
- **Closure**. It is the expression of the mass balance condition. Applies to some concentration profiles of reaction systems.
- **Known pure spectra /concentration profiles**. This is a kind of equality constraint, which makes the concentration profile and/or spectrum of a component to be equal to a certain known predefined shape.

Constraints related to mathematical conditions are:

- **Local rank/selectivity**. It relates to the zones of absence of components in some regions (windows) in the profiles. Usually it applies to concentration profiles. The zero-concentration regions in the different profiles are determined by local rank analysis methods, such as Evolving Factor Analysis (EFA) [45,46]. When only a component is present in a certain concentration window, we have selectivity. When only some components are absent, we talk about local rank information. Suitable local rank conditions ensure the recovery of the correct profiles by MCR [50].
- **Correspondence of species**. This constraint is only applied in column-wise augmented matrices (with an augmented concentration direction) and expresses the correspondence and presence/absence of components in the analyzed experiments [5,22,23].

- **Model constraints**. These are constraints applicable to multiset structures and only to the augmented matrices that may be present in the decomposition model. Although MCR gives by default a bilinear model, the conditions of trilinearity [21,39], multilinearity [41] or factor interaction (like in Tucker models) [40] can be implemented in a component-wise way. Therefore, completely bilinear, completely trilinear or hybrid models can be set to be obtained in the final MCR results.

Finally, a last group of constraints is linked to some tasks that can be done integrated in the MCR process.

- **Hard-modeling.** Forces the concentration profiles to be fitted by a parametric physicochemical model and the parameters of the model are obtained as an additional output. It can also be applied to pure analytical responses when the shape of the pure signal can be defined by a parametric equation. This constraint implies to perform a model fitting task during the iterative optimization process [23,42].

- **Correlation constraint**. In this case, internal univariate calibration models can be applied to particular concentration profiles of components in the system. The model is established between concentration values obtained with MCR (in arbitrary units) and real concentrations in calibration samples. The model is used to predict real concentrations in unknown samples [5, 27,43].

Some characteristics are common to all constraints: a) they are optionally applied, b) the application of constraints can affect differently the row and column profiles of the data set, the different components in the data set and the different subsets in a multiset structure, c) the application of constraints can be modulated according to tolerance criteria. The flexibility in the application of the constraints explains the versatility of MCR algorithms, which can adapt to very diverse scenarios through the proper selection of these restricting conditions.

To make a suitable selection of constraints, we should know first which kind of profiles we expect to obtain in the MCR decomposition model. Once this is known, constraints can be applied tuning the tolerance in the application according to some natural characteristics of the data set, such as the noise level. When in doubt about the application of a constraint, MCR models built with and without introducing this constraint can be compared. The introduction of a certain constraint should not produce a significant decrease in the variance explained in the data set. If so, the data set probably does not obey this constraint or it has been applied in a too strict way.

6. Uncertainty in MCR solutions.

Uncertainty in MCR is linked to the phenomenon of ambiguity and noise propagation. The concept of ambiguity means that different combinations of sets of concentration profiles and spectra can reproduce the original data set with the same fit quality [51]. Three different kinds of ambiguity can be observed:

- **Permutation ambiguity**. There is no sorting order on the MCR components. Therefore, they can be shuffled in the concentration and spectra matrix (always keeping the right correspondence of the dyads) and obtain identical results.

- **Intensity ambiguity**. Dyads of profiles having the same shape but different relative scales between concentration profile and spectrum reproduce equally well the original data set. This is why concentration values and pure response intensities in $\mathbf{c}_i$ and $\mathbf{s}_i^T$ profiles are always in arbitrary units unless reference information on real intensities is

available and actively used in the resolution process. Normalization of concentration profiles or resolved spectra or use of reference concentration values within the optimization help to suppress the intensity ambiguity.

$$\mathbf{D} = \sum_i \mathbf{c}_i \mathbf{s}_i^{\mathbf{T}} \quad \text{(eq. 2)}$$

$$\mathbf{D} = \sum_i (\mathbf{c}_i k_i) \left(s_i^T \frac{1}{k_i} \right) \quad \text{(eq.8)}$$

- **Rotational ambiguity**. Sets of concentration profiles and spectra with different shapes can reproduce the original data set with the same fit quality. This is the most relevant kind of ambiguity and can be expressed as follows:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \quad \text{(eq. 3)}$$

$$\mathbf{D} = (\mathbf{C}\mathbf{T}) \left(\mathbf{T}^{-1} \mathbf{S}^{\mathbf{T}} \right) + \mathbf{E} \quad \text{(eq. 9)}$$

where **T** is any transformation matrix, which used as in equation 9, keeps providing profiles that still obey the constraints set in the MCR optimization.

The main way to decrease/suppress ambiguities in MCR solutions is by means of the introduction of constraints [21,23]. Some of them can ensure the suppression of ambiguity in the profiles where they are applied, such as trilinearity and hard-modeling [21,52]. Other constraints, such as local rank can also suppress ambiguity, provided that some conditions of component overlap are fulfilled in the data set [50]. A relevant aspect to decrease ambiguity is working with multiset structures [21-23,51]. In this case, the diversity of the subsets appended makes that the possible combinations of profiles fulfilling constraints and describing adequately the variation in the multiset structure be smaller than those satisfying a single data set.

There are several approaches meant to assess the ambiguity in MCR solutions [51]. To find the full range of feasible solutions, only few approaches exist and they cannot go beyond systems with four components [53-55]. Other algorithms find the minimum and maximum boundaries of a certain objective function, i.e., amount of component signal over total signal with dyads of profiles obeying the constraints [56,57].

$$f_{i,min} = \min \frac{\|\mathbf{c}_i \mathbf{s}_i^{\mathbf{T}}\|}{\|\mathbf{C}\mathbf{S}^{\mathbf{T}}\|} \quad \text{(eq.10)} \qquad f_{i,max} = \max \frac{\|\mathbf{c}_i \mathbf{s}_i^{\mathbf{T}}\|}{\|\mathbf{C}\mathbf{S}^{\mathbf{T}}\|} \quad \text{(eq. 11)}$$

The results provided are the two extreme MCR dyads related to the minimum and maximum of the objective function for each component in the system. These methods can work with systems with unlimited of number of components and provide a good approximate description of:

a) **the extent of ambiguity**, by calculating $f_{i,max} - f_{i,min.}$ If the subtraction is zero, the component is resolved in a unique way. The higher the difference the larger the ambiguity associated with the resolution of a particular component.

b) **The location of ambiguity**. Plotting the dyads of profiles related to $f_{i,max}$ and $f_{i,min}$ for a particular component, it can be seen whether ambiguity affects concentration profiles and/or spectra for a particular component.

It is relevant to note that saying that an MCR solution is ambiguous is very vague and can transmit an unnecessarily negative impression, if no information on the extent and location of ambiguity is provided. Ambiguity is a concept related to each component in the system and to each profile within a dyad. An ambiguous solution of MCR can be very satisfactory if the components of interest are uniquely resolved or affected by a very low degree of ambiguity, even if the rest of the system is not perfectly defined.

7. MCR. Case studies. A single image data set and an HPLC-DAD multiset.

Nowadays, there exist freely available graphical user interfaces (GUIs) to perform MCR-ALS analysis in single and multiset analysis [58] and to assess the extent and location of ambiguity as described in the previous section [59]. Both GUIs mentioned do not have any limitation regarding the number of the components in the system or the data configuration and chemical meaning. The most common constraints can also be applied in an optional way and with varying mode-wise, component-wise and subset-wise combinations. Below and briefly, the steps to be followed for MCR analysis will be shown through two different examples: a single image data set and a multiset HPLC-DAD data set. Details on how to use the GUIs for MCR analyses are described in the references above. Soon a new release of the interface will appear with new and more powerful features in terms of diversity of constraints and use of noise structure in the MCR execution. The examples presented below can be found at: www.mcrals.info

Image data set

This data set is a hyperspectral Raman image of an emulsion, sized (60 x 60 x 253), where the first two figures refer to the pixels in the x- and y- direction and the third to the number of spectral channels [60]. In this case, the cube should be unfolded into a data matrix sized (number of pixels x spectral channels), i.e., 3600 x 253. For the sake of brevity, we will avoid talking extensively about preprocessing but, as before using any other data analysis method, pretreatments are recommended to remove intense backgrounds or other variations unrelated to the components of interest, e.g. scattering. If anything can be commented with respect to MCR analysis is that, whenever possible, pretreatments keeping the properties of the original data set will be favored. For instance, we will not use derivative treatments to correct backgrounds if other treatments keeping the positive character of the instrumental signal are available or scaling variables, e.g., in environmental data, instead of autoscaling will be preferred for the same reason.

The three necessary steps to perform MCR analysis are: a) the determination of the number of components, b) building initial estimates and c) selecting the constraints to be applied.

a) Determination of number of components. It can be known beforehand or found by auxiliary methods, such as Singular Value Decomposition (SVD). This step is typically considered a complex operation. However, when in doubt, MCR models with different sizes can be tried. Introducing a new necessary component should have a clear effect in the increase of the variance explained and should provide a model with interpretable profiles. If two MCR models with a different number of components provide equally plausible solutions, the one with least components should be chosen, unless sound chemical (scientific) knowledge points out towards the model with a larger size.

   In the case of the image data set, the number of components suggested by SVD is four, as the magnitude of the singular values indicates and the noisy pattern of the 5^{th} SVD score (see Figure 3a).

b) Building initial estimates. In the case of the image data set, the concentration profiles refer to the abundance of each component in the different pixels (unfolded from the original 2D structure into a single dimension). These profiles do not show any kind of continuous pattern, since the presence and amount of components from pixel-to-pixel does not follow any systematic pattern. The lack of structure in the concentration direction and the positive values of Raman signals suggest the use of a pure variable selection method. In this case, a method based on SIMPLISMA is applied to find the purest pixels because the overlap of components in the sample surface is lower than the overlap of spectral signatures. Once the purest pixels are selected, spectral initial estimates are built with the related spectra.

c) Selection of constraints. To decide which constraints can be applied, the first step is thinking about the meaning of the matrices in the bilinear model. In the case of an image, the matrices in the model $\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}}$ relate to the original image data set (**D**), the pure Raman spectra of the components in the image ($\mathbf{S}^{\mathbf{T}}$) and the related concentration profiles (**C**), which can be refolded to recover the original 2D image surface and be displayed as distribution maps. In this case, pure Raman spectra are positive and, therefore, the non-negativity constraint can be applied to the $\mathbf{S}^{\mathbf{T}}$ matrix and, likewise, concentration values are always positive and the non-negativity constraint applies to the matrix **C**. No typical process-like constraints, such as unimodality, closure or hard-modeling are suitable for this example because the concentration profiles lack this kind of systematic patterns. Instead, local rank constraints can be used and adaptations for image data sets can be used, although they will not be described here for brevity. The reader is sent to reference [61] for additional information.

Once all the information above is decided, the alternating least-squares optimization of **C** and $\mathbf{S}^{\mathbf{T}}$ starts. A convergence criterion should be proposed to stop the iterative process. In this case, a difference among lack of fit of 0.1% between consecutive iterations is proposed.

Figure 3 shows plots related to the MCR steps: a) determination of number of components, b) building initial estimates and c) concentration profiles and spectra (the concentration profiles are refolded to recover the (60 × 60) distribution maps of each component in the emulsion. Several things should be taken into account when analyzing MCR results:

- **the model fit quality**. In this case, the lack of fit is 7.8% (more than 99% variance explained), which is logical for a Raman hyperspectral image. This figure can be smaller, when instrumental measurements with a low noise level are analyzed, e.g., UV spectra from solutions, or much higher, e.g. in environmental data, when the amount of systematic variance in the data is very low. Therefore, the satisfactory lack of fit is a problem-dependent issue. Higher lacks of fits than expected in an MCR model can imply that more components are required to describe the variation of the data or that incorrect constraints (or applied in a too strict way) have been used
- **the interpretability of the profiles**. In this case, the pure spectra retrieved are recognizable as Raman spectra and the related distribution maps reflect the surface pattern of an emulsion, with an oily phase represented by two components (a big drop and the interface around), the outer aqueous phase and a small patch, which is an emulsion additive. Warnings of incorrect results in this context would have been the presence of completely random distribution maps or of pure spectra that had nothing to see with the spectral shapes seen in the original data.

The MCR analysis should be completed with an assessment of the ambiguity. Since this will be presented for the second example, it is skipped here. The reader is again referred to reference [61] to obtain a lot of detail in this issue. Qualitatively speaking, a certain ambiguity was still found in the definition of the components in the image when only non-negativity was applied. Introducing local rank constraints, the solutions obtained were unique.

HPLC-DAD example

This example is formed by three chromatograms equally sized (51 x 96), where the first figure refers to elution times and the second to spectral channels. Two of the chromatograms are standards of two different analytes ($\mathbf{D_A}$ and $\mathbf{D_B}$) and the third is a mixture sample containing the two analytes and an unknown interference ($\mathbf{D_m}$). A common practice in unknown problems involving several data sets is performing first the MCR analysis of each experiment to have a better idea on the species shared by the different experiments. In this case, since two pure standards are available, only the individual analysis of the mixture chromatogram is necessary.

Single data set analysis (chromatographic mixture)

The three basic steps to carry out MCR-ALS adapted to a single HPLC-DAD chromatogram set work as follows:

a) Determination of number of components. In this case, the number of components suggested by SVD is three.

b) Building initial estimates. In an HPLC-DAD data set, the concentration direction follows an evolutionary sequential pattern. Therefore, initial estimates can be built by Evolving Factor Analysis (EFA). Figure 4 shows the initial estimates obtained by this method, which clearly show the regions of emergence and decay of the different components in the data set.

c) Selection of constraints. The bilinear model of an HPLC-DAD data set $\mathbf{D} = \mathbf{C}\mathbf{S}^T$ relates to the original chromatogram with multivariate detection ($\mathbf{D}$), the pure UV spectra of the components in the system ($\mathbf{S}^T$) and the related elution profiles ($\mathbf{C}$), i.e., the chromatographic peaks. In this case, pure UV spectra are positive and, therefore, the non-negativity constraint can be applied to the $\mathbf{S}^T$ matrix. Elution profiles hold several properties that we can use as constraints, such as non-negativity and also unimodality, since a pure chromatographic peak should only have a maximum. This would be the simplest constraints that can be generally applied to resolve chromatographic data sets.

Figure 5 shows the results of the MCR analysis, with three overlapping peaks and the related spectra. The lack of fit related to the model is 3.1%, satisfactory for an HPLC-DAD data set with a low noise level.

To complete the analysis, the rotational ambiguity associated with these solutions can be estimated by the method described in section 7 based on the use of an objective function. To do so, the starting information needed is the concentration profiles and spectra obtained by MCR and the constraints used in the analysis, i.e., non-negativity and normalization in $\mathbf{S}^T$ and non-negativity and unimodality for profiles in $\mathbf{C}$. As can be seen, a certain amount of ambiguity is found in all components. The extent of ambiguity is estimated by the $f_{i,max} - f_{i,min}$ value, which varies from 0.22 to 0.30 depending on the component. Differences are also

obtained in the location of the ambiguity (concentration profiles or spectra) in the profiles recovered (see Figure 6).

Multiset analysis (several chromatographic runs)

To improve these results, multiset analysis combining the two standards and the mixture chromatogram can be carried out. The multiset built is a column-wise augmented data set because the common mode in this case is the spectral one (see explanation in section 3) and is structured as follows, [$\mathbf{D_A; D_B; D_m}$]. In the case of multiset analysis, the number of components of the system keeps being three and the steps of initial estimates and use of constraints are modified as follows:

b) Building initial estimates. For an HPLC-DAD multiset, initial estimates could be built by concatenating in the suitable way EFA analyses of the different experiments. However, a simpler option is applying a method based on the selection of purest variables to the full multiset in the elution direction and start with a single matrix of spectral estimates.

c) Selection of constraints. Non-negativity constraint is applied to the $\mathbf{S^T}$ matrix and $\mathbf{C}$ profiles and now, unimodality applies separately to elution profiles of each $\mathbf{C_i}$ subset, related to each of the chromatograms in the multiset. Since we work with a column-wise augmented data set, the constraint of correspondence of species can also be used. In this case, for the multiset [$\mathbf{D_A; D_B; D_m}$], information on the presence/absence of components can be set as follows:

$$\begin{array}{c} \\ \mathbf{D_A} \\ \mathbf{D_B} \\ \mathbf{D_m} \end{array} \begin{array}{c} \begin{array}{ccc} A & B & Int \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

where 1 implies presence and zero absence of components. Please note that the components of the chemical system should be sorted in the same way in the rows of the matrix of spectral estimates and in the columns of the matrix of correspondence among species. The introduction of this constraint implies in this case including selectivity in the elution profiles of components A and B in the standards of the multiset, which will lead necessarily to a significant improvement in the final results obtained by the MCR analysis for the full multiset.

In this example, the constraint of trilinearity will not be used since shifts and changes of peak shape and position among chromatograms can be expected. When in doubt, a mathematical way to check for the possibility of including this constraint is comparing the number of components needed to define the column-wise augmented matrix [$\mathbf{D_A; D_B; D_m}$], where the spectral mode is the common one, and the alternative row-wise augmented multiset, [$\mathbf{D_A\ D_B\ D_m}$], where the elution mode would be the common mode. In a trilinear situation, the number of components needed in both cases would be the same. If the elution mode behaves in a non-trilinear way, the number of components needed to define the row-wise augmented multiset will be larger than the one for the column-wise augmented multiset, because additional contributions will be needed to describe differences in elution pattern among chromatograms (peak shifts, peak shape changes,…). For this particular case, the singular values linked to multiset [$\mathbf{D_A; D_B; D_m}$] are 10^5 × (**7.7832, 1.9580, 0.6937**, 0.0829, 0.0452, 0.0430, 0.0420, 0.0404…) and to

multiset [$\mathbf{D_A}$ $\mathbf{D_B}$ $\mathbf{D_m}$] are 10^5 × (**7.4856, 2.7961, 0.9458, 0.3634, 0.1399**, 0.0535, 0.0445, 0.0434…) showing the need for more contributions in the row-wise augmented multiset and, hence, the lack of trilinearity in the multiset analyzed [23,62].

The final results of the multiset analysis will be a single matrix containing the pure spectra of the three components and an augmented concentration matrix with the elution profiles of each of the three chromatograms included in the multiset structure (the two standards and the mixture) (see Figure 7). The lack of fit of this analysis is equal to 3.6%, also satisfactory according to the noise level of the data set.

When working with analytical multiset structures, such as the one in this example, additional information can be eventually recovered, such as quantitative information, from the areas of the elution profiles recovered [12,23,63]. If absolute reference concentrations are available, quantitative concentration values will be obtained. Otherwise, the information per component can always be interpreted in relative concentration terms.

In this example, there is no ambiguity left in the final solutions of the multiset analysis. This comes from the inclusion of complementary/selective information in the full data set structure with the use of the standards and the fact that, in this case, the necessary conditions of component overlap to obtain unique solutions are fulfilled.

8. Conclusions

MCR techniques can be used in very diverse situations and scientific problems as long as an unmixing problem is present. In analysis of analytical responses, many times the bilinear model of MCR fits the measurement model but, even in other contexts (environmental, -omics data), the final results obtained show profiles scientifically meaningful. The reason is that, in difference with other bilinear decomposition methods, the driving force for the bilinear decomposition is the application of constraints that respect the chemical/scientific properties of the profiles to be modeled. Hence, the final results are not abstract mathematical solutions, but easily interpretable profiles.

Because of the bilinear basic scaffold, multisets formed by data matrices including very different information and different geometries can be also analyzed. Within the MCR-ALS algorithm, the range of constraints potentially applicable has increased enormously and, to the natural constraints linked to chemical (non-negativity, unimodality, closure,…) and mathematical properties (selectivity…) typically used, other conditions related to the definition of the multiset model (trilinear, multilinear, factor interaction…) have shown up and other constraints related traditionally to tasks carried out by other families of methods (hard-modelling, correlation) can also be applied.

Right now, there are freely downloadable graphical interfaces that allow applying MCR-ALS in multiset mode with the natural chemical and selectivity constraints and the option to use the trilinear condition. Estimating the uncertainty (ambiguity) of MCR solutions is also available in a user-friendly software. In the near future, a new version of the graphical interface will be released that will include the newest features, such as the hard-modelling and correlation constraints, more flexibility in the multiset mathematical model and the possibility to incorporate information on the data noise structure, allowing a weighted version of the algorithm to be run.

Acknowledgements

Financial support from the Spanish government through the project CTQ2012-38616 is acknowledged. This group belongs to the network of recognized research groups by the Catalan government (2014 SGR 1106).

References

1. S.C. Rutan, A. de Juan, R. Tauler. Introduction to Multivariate Curve Resolution, in *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009) **2**, 249-259.
2. W.H. Lawton, EA Sylvestre. *Technometrics*, 1971, **13**, 617.
3. J. Craig Hamilton, P. J. Gemperline. *J. Chemometr*., 1990, **4,** 1.
4. A. de Juan, R. Tauler. *Anal. Chim. Acta*, 2003, **500**, 195.
5. A. de Juan, R. Tauler. *Crit. Rev. Anal. Chem*., 2006, **36**, 163.
6. C. Ruckebusch, L. Blanchet. *Anal. Chim. Acta*, 2013, **765**, 28-36.
7. I. Jolliffe. Principal Component Analysis. Wiley, 2005.
8. A. Hyvärinen, J. Karhunen, E. Oja. Independent Component Analysis. J. Wiley & Sons, 2001.
9. H. Parastar. M. Jalali-Heravi, R. Tauler. *Trends in Anal. Chem*., 2012, **31**, 134.
10. J.M.P. Nascimento, J.M. Bioucas-Dias. *IEEE Trans. Geosc. Remote Sens*, 2005, **43**, 175.
11. J. Jaumot, R. Eritja, R. Gargallo. *Anal, Bioanal. Chem*., 2011, **399**, 1983.
12. A. de Juan, R. Tauler. *J. Chromatogr. A*., 2007, **1158**, 184.
13. M. Esteban, C. Ariño, J.M. Díaz-Cruz. Chemometrics in Electrochemistry. in *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009) **4**, 425.
14. A. de Juan, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler. 'Chemometric Tools for Image Analysis' in 'Infrared and Raman Spectroscopic Imaging'.(R.Salzer and H.W.Siesler; eds.) Wiley-VCH. Chapter **2**, 65-106 (2009).
15. A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder. *TrAC-Trends in Anal. Chem.*, 2004, **23,** 70.
16. J.M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco. *Trends Anal. Chem.*, 2008, **27**, 696.
17. R. Tauler. Interpretation of environmental data using chemometrics. *Sample Handling and Trace Analysis of Pollutants: Techniques, Applications and Quality Assurance*, Chapter 16. Edited by D.Barceló, Elsevier, 2000, 689.
18. M. Terrado, D. Barceló, R. Tauler. *Environ. Sci. Tech*., 2009, **43**, 5321.
19. J. Jaumot, R. Tauler, R. Gargallo. *Anal. Biochem*., 2006, **358**, 76.
20. P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martínez, C.P. Allen, M. Werner-Washburne. *BMC Bioinformatics*, 2006, **7,** art. 343
21. R. Tauler, A.K. Smilde, B.R. Kowalski. *J. Chemometr*., 1995, **9**, 31.
22. R. Tauler. *Chemom. Intell. Lab. Sys*., 1995, **30**, 133.
23. R. Tauler, M. Maeder, A. de Juan. 'Two-Way Extended Curve Resolution. *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009) **2**, 473-505
24. G. Muñoz, A. de Juan. *Anal. Chim. Acta*, 2007, **595**, 198.
25. S. Ruiz-Castelar, A. Checa, R. Gargallo, J. Jaumot. *Anal Chim Acta*, 2012, **722**, 34.
26. R. Tauler, B. Kowalski, B, S. Fleming. *Anal. Chem.*, 1993, **65**, 2040.

27. J. Jaumot, B. Igne, C.A. Anderson, J.K. Drennen, A. de Juan. *Talanta*. 2013, **117**, 492.
28. S. Navea, A. de Juan, R. Tauler. *Anal. Chem*., 2003, **75**, 5592.
29. C. Ruckebusch, M. Sliwa, P. Pernot, A. de Juan, R. Tauler. *J. Photochem. Photobiol. C: Photochem. Rev*., 2012, **13**, 1.
30. A. de Juan, S.C. Rutan, R. Tauler. 'Two-Way Data Analysis: Multivariate Curve Resolution–Iterative Resolution Methods, in *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009), **2**, 325-344
31. Y. Liang, Two-way data analysis: Multivariate Curve Resolution- Non-iterative methods in *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009),**2**, 309-324
32. O.M. Kvalheim, Y.Z. Liang. *Anal. Chem*., 1992, **64**, 936.
33. E.R Malinowski. *J. Chemometr*., 1992, **6**, 29.
34. R. Manne, H. Shen, Y. Liang. *Chemom. Intell. Lab. Sys*., 1999, **45**, 171.
35. B.G.M. Vandeginste, W. Derks, G. Kateman. *Anal. Chim. Acta*, 1985, **173**, 253.
36. P.J. Gemperline. *J. Chem. Infor. Comput. Sci*., 1984, **24**, 206.
37. J.M. Nascimento, J.M. Bioucas-Dias. (2005). *IEEE Trans. Geosci. Remote Sens*. 2005, **43**, 898-910.
38. P. Paatero, U. Tapper. *Environmetrics*, 1994, **5**, 111-126.
39. R. Tauler, I. Marqués, E. Casassas. *J. Chemometr*., 1998, **12**, 55.
40. M. Alier, M. Felipe, I. Hernández, R. Tauler. *Anal. Bioanal. Chem*., 2011, **399**, 2015.
41. A. Malik, R. Tauler. *Anal. Chim. Acta*, 2013, **794**, 20.
42. A de Juan, M Maeder, M Martínez, R Tauler. *Chemom. Intell. Lab. Sys*., 2000, **54**, 123.
43. M. C. Antunes, J. E. J. Simão, A. C. Duarte and R. Tauler. *Analyst*, 2002, **127**, 809.
44. M. Maeder, A.D. Zuberbühler. *Anal. Chim. Acta*, 1986, **181**, 287.
45. M. Maeder, A. de Juan. Two-Way Data Analysis: Evolving Factor Analysis, *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009), **2**, 261-274.
46. W. Windig. Two-Way Data Analysis: Detection of Purest Variables, in *Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009), **2**, 275-307
47. W Windig, J Guilment. *Anal. Chem*., 1991, **63**, 1425.
48. F. Cuesta Sánchez, B. van den Bogaert, S.C. Rutan and D.L. Massart. *Chemom. Intell. Lab. Sys.*, 1996, **34**, 139.
49. E.R. Malinowski. *Anal. Chim. Acta*, 1982, **134**, 129.
50. R. Manne. *Chemom. Intell. Lab. Sys*., 1995, **27**, 89.
51. R. Tauler, M. Maeder. Two-Way Data Analysis: Multivariate Curve Resolution – Error in Curve Resolution, in ,*Comprehensive Chemometrics*, S. Brown, R. Tauler, and R. Walczak, Eds.(Elsevier, Oxford, 2009), **2**, 345-363.
52. A. Golshan, H. Abdollahi, M. Maeder. *Anal. Chim. Acta*, 2012, **709**, 32.
53. H. Abdollahi, M. Maeder, R. Tauler. *Anal. Chem*., 2009, **81**, 2115.
54. R. Rajkó, K. István. *J. Chemometr*., 2005, **19**, 448.

55. M. Sawall, C. Kubis, D. Selent, A. Börner, K. Neymeyr. *J. Chemometr*., 2013, **27**, 106.
56. R. Tauler. *J. Chemometr*. 2001, **15**, 627.
57. P.J. Gemperline. *Anal. Chem*., 1999, **71**, 5398.
58. J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, *Chemom. Intell. Lab. Sys*. 2005, **76,** 101.
59. Jaumot J, Tauler R. *Chemom. Intell. Lab. Sys*. 2010, **103**, 96.
60. .J. Andrew and T.M. Hancewicz. *Appl. Spectrosc*.**.** 1998, **52,** 797.
61. A. de Juan, M. Maeder, T. Hancewicz, R. Tauler. *J. Chemometr*., 2008, **22**, 291.
62. A. de Juan, R. Tauler. *J. Chemometr*., 2001, **10**, 749.
63. E. Peré-Trepat, S. Lacorte, R. Tauler. *Anal. Chim. Acta*, 2007, **595**, 228.

Figure captions

Figure 1. Bilinear model obtained from MCR for an HPLC-DAD data set. Expressed as: a) sum of pure signal contributions, b) sum of the dyads of pure concentration profile and spectra, c) product of matrices of pure concentration profiles and spectra.

Figure 2. MCR decomposition of a hyperspectral image data set.

Figure 3. Steps of MCR application in the analysis of a hyperspectral image. a) Determination of number of components by SVD, b) selection of spectral initial estimates by a method based on SIMPLISMA (right plot), global intensity map showing the location of the pixels related to the initial estimates (pixels numbered have the same color code than related spectra found by SIMPLISMA), c) resolved distribution maps and spectra (the GUI windows in the figure are described in [57]).

Figure 4. Raw HPLC-DAD data set (top plot) and related EFA initial estimates of elution profiles (in colour, bottom plot).

Figure 5. Top plots: raw HPLC-DAD data set. Columns (chromatograms, left plot) and rows (spectra, right plot). Bottom plots: MCR resolved elution profiles (left plot) and spectra (right plot).

Figure 6. Estimation of ambiguity in the resolved elution profiles and spectra applying non-negativity (in elution profiles and spectra) and unimodality (in elution profiles). Dotted lines: MCR resolved profiles, solid lines: minimum and maximum boundaries per each resolved profile (the GUI window in the figure is described in [58]).

Figure 7. MCR resolved elution profiles (top plot) and spectra of an HPLC-DAD multiset [$\mathbf{D_A; D_B; D_m}$] formed by two standards of A and B components and a sample mixture of A and B and an interference.

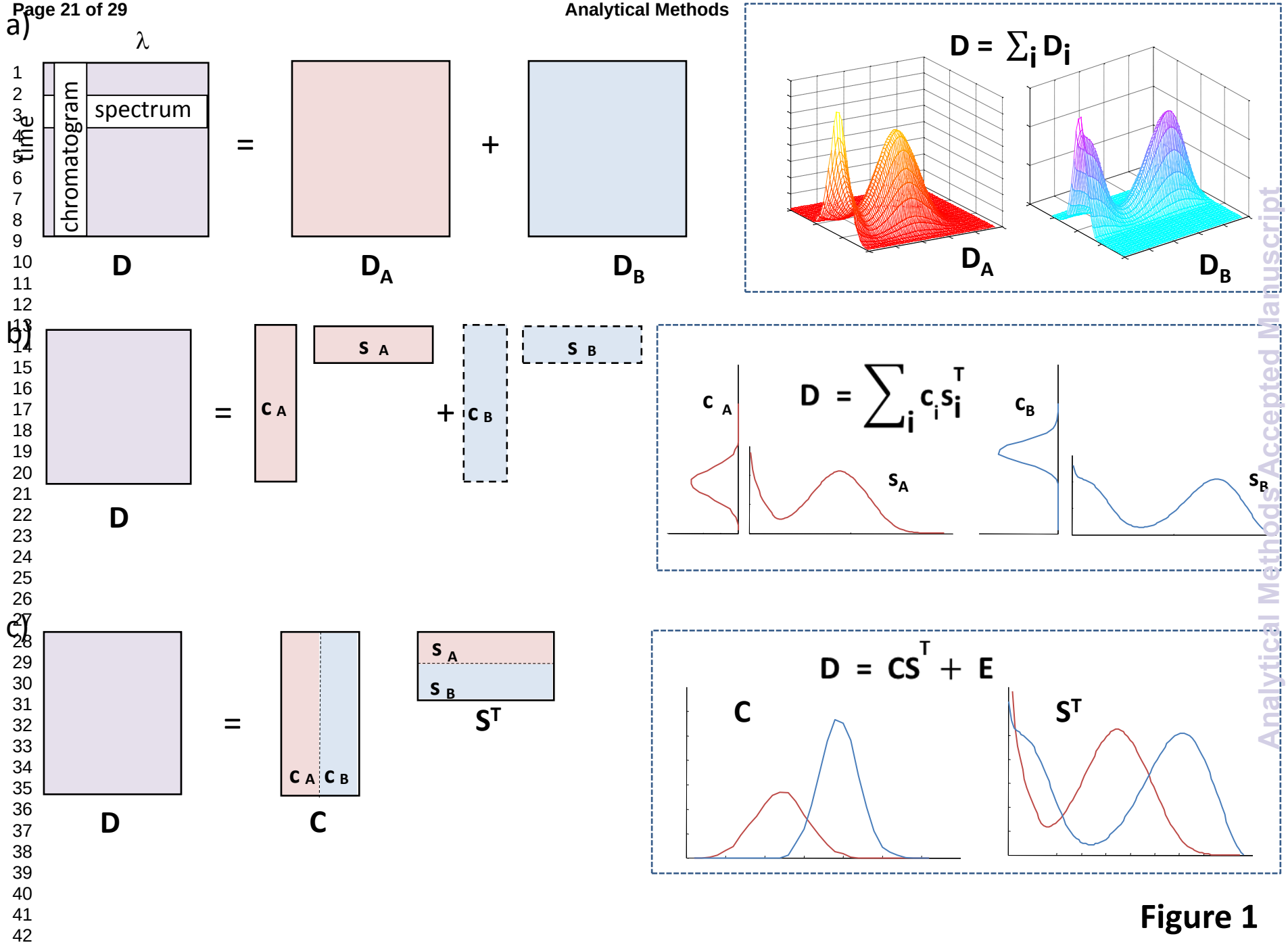


Figure 1

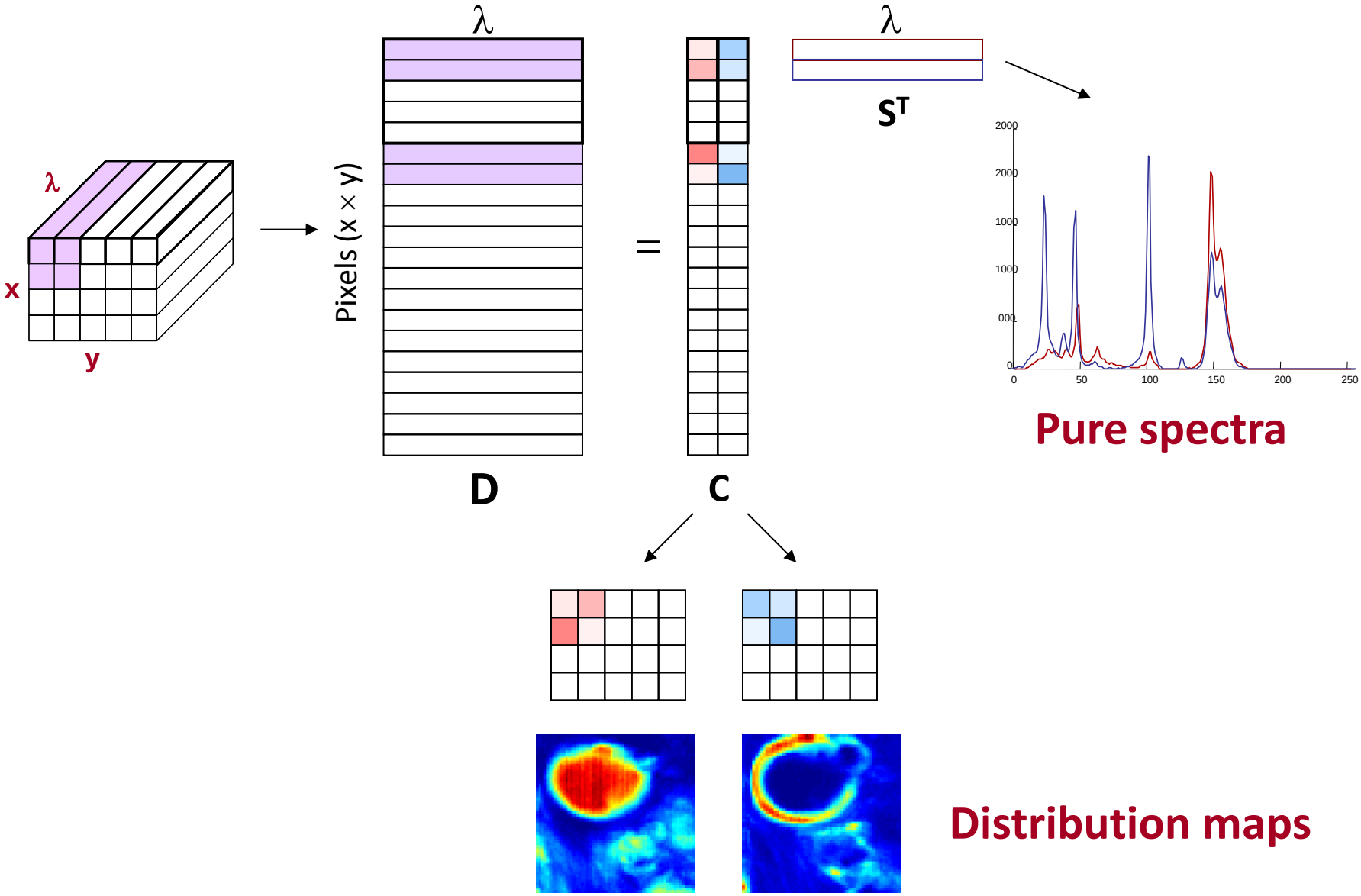
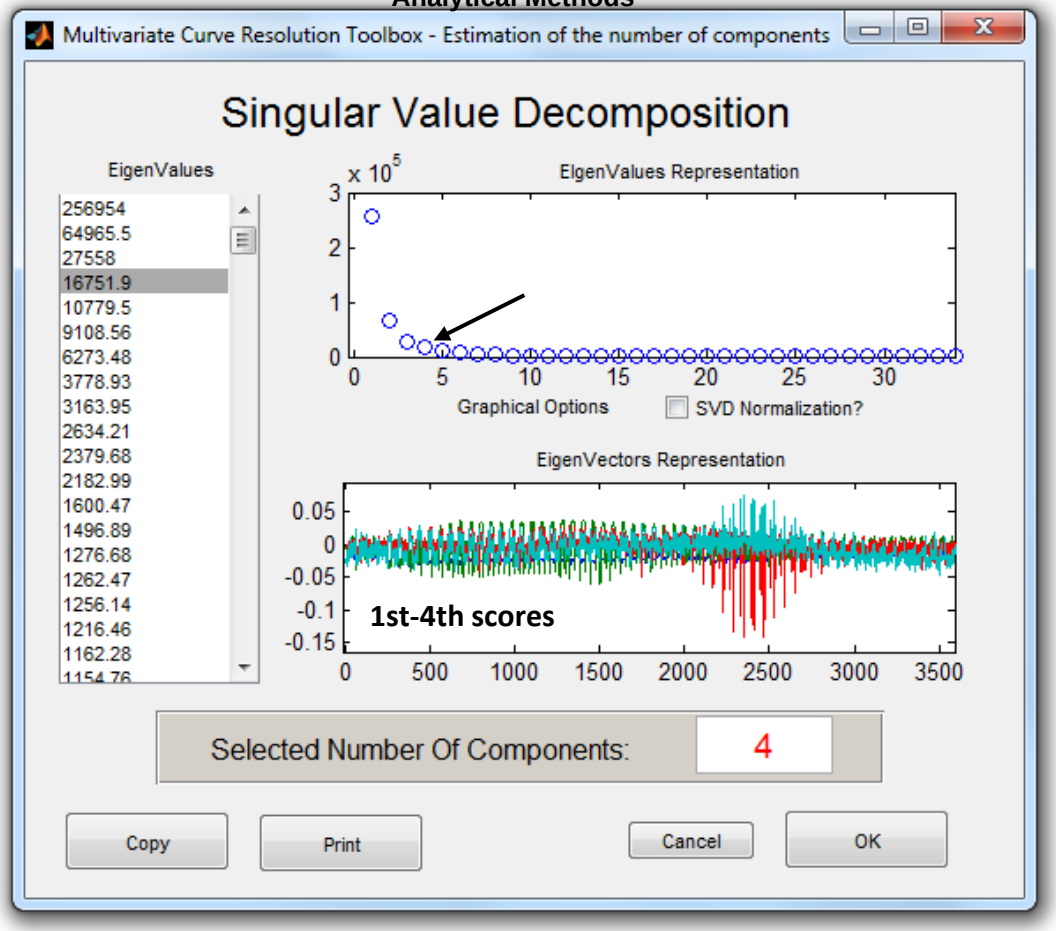


Figure 2



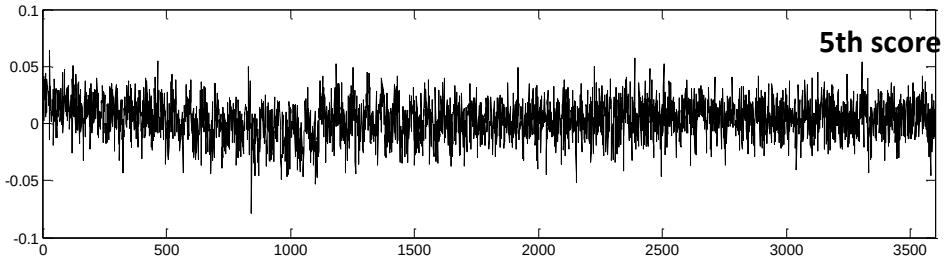
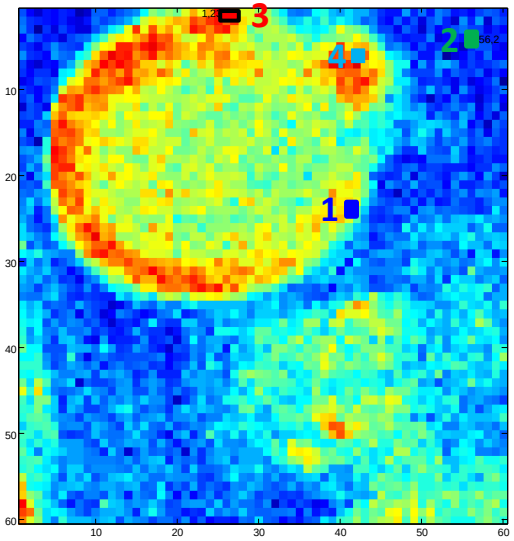


Figure 3a



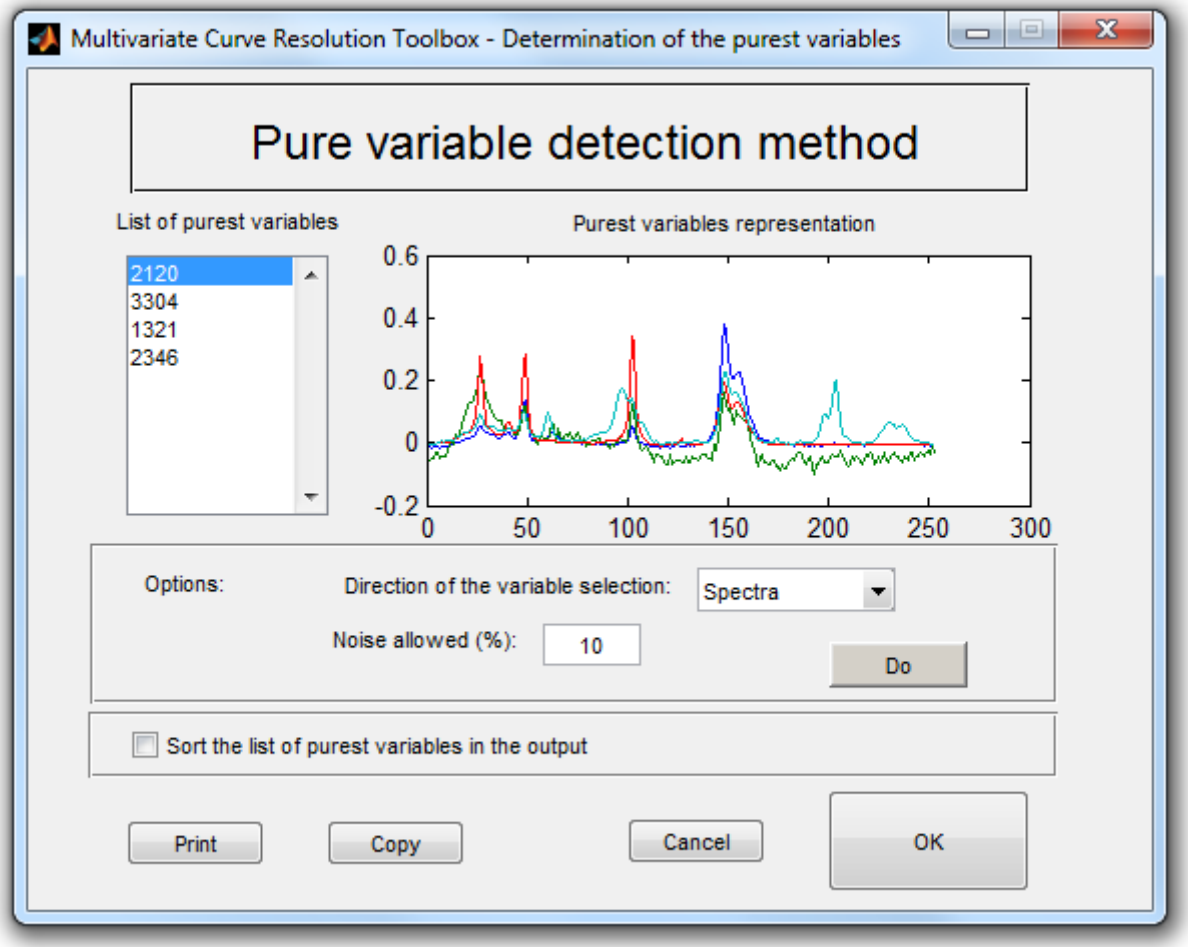
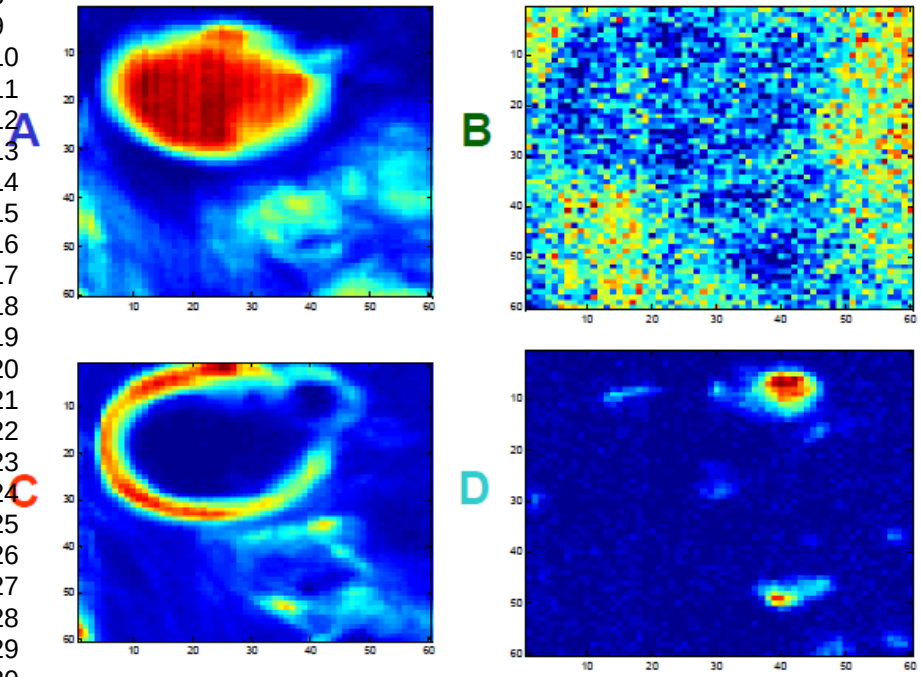
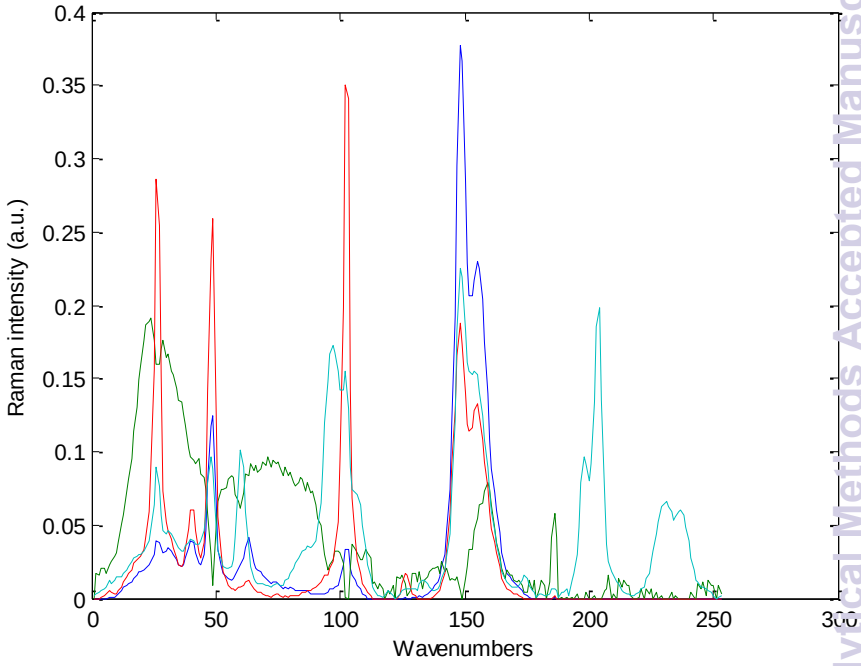


Figure 3b



Distribution maps (C)



Pure spectra (S^T)

Figure 3c

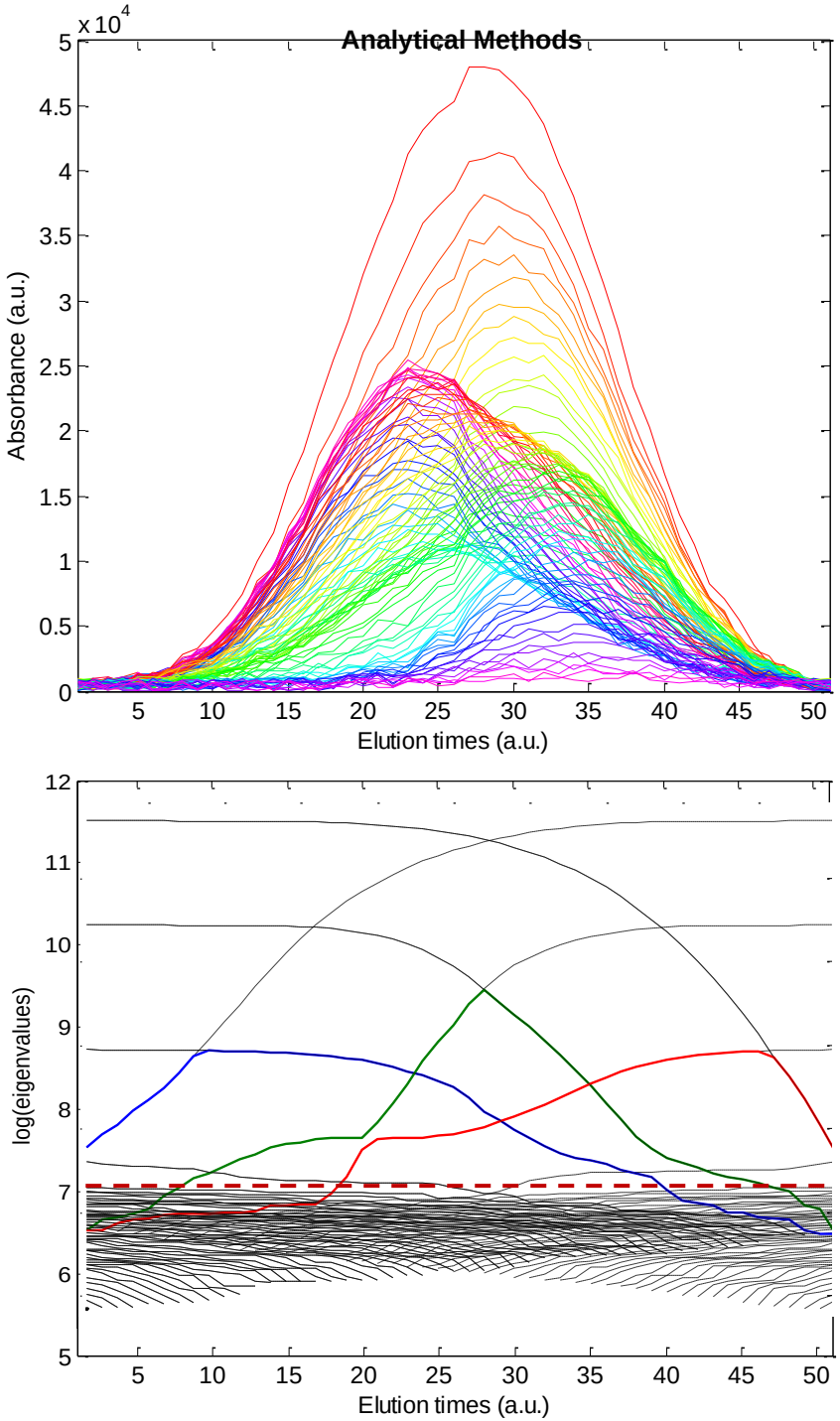


Figure 4

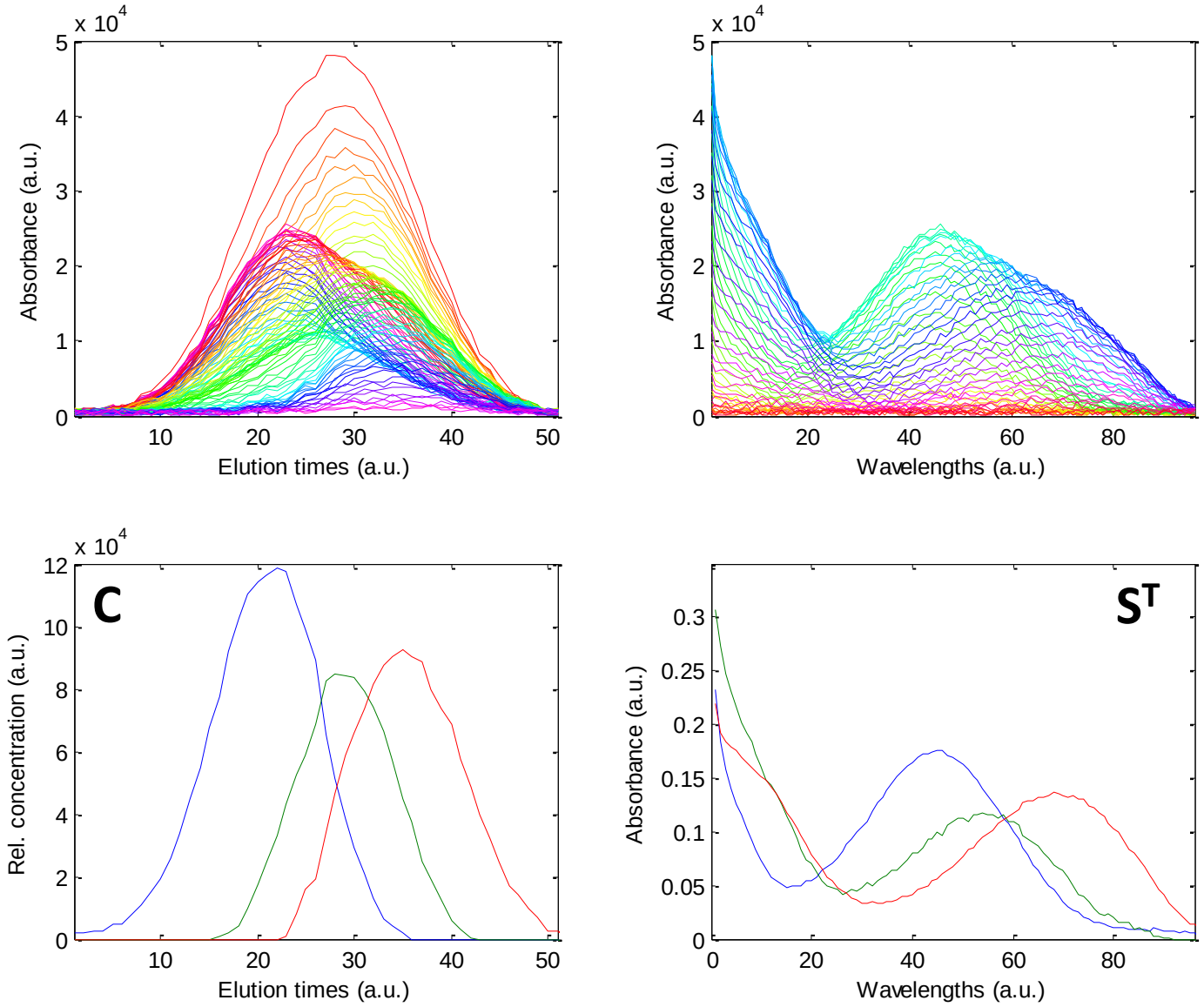


Figure 5

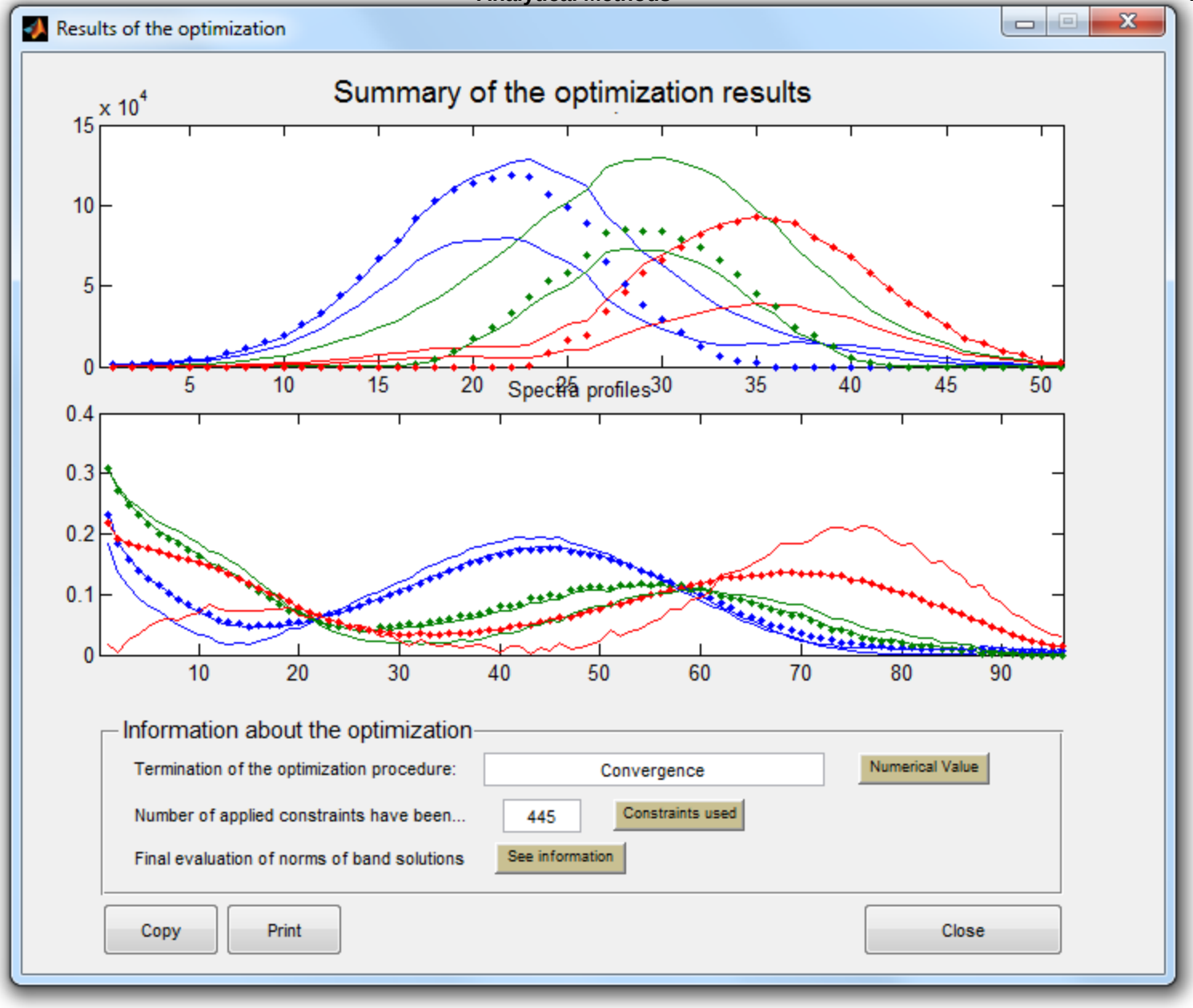


Figure 6

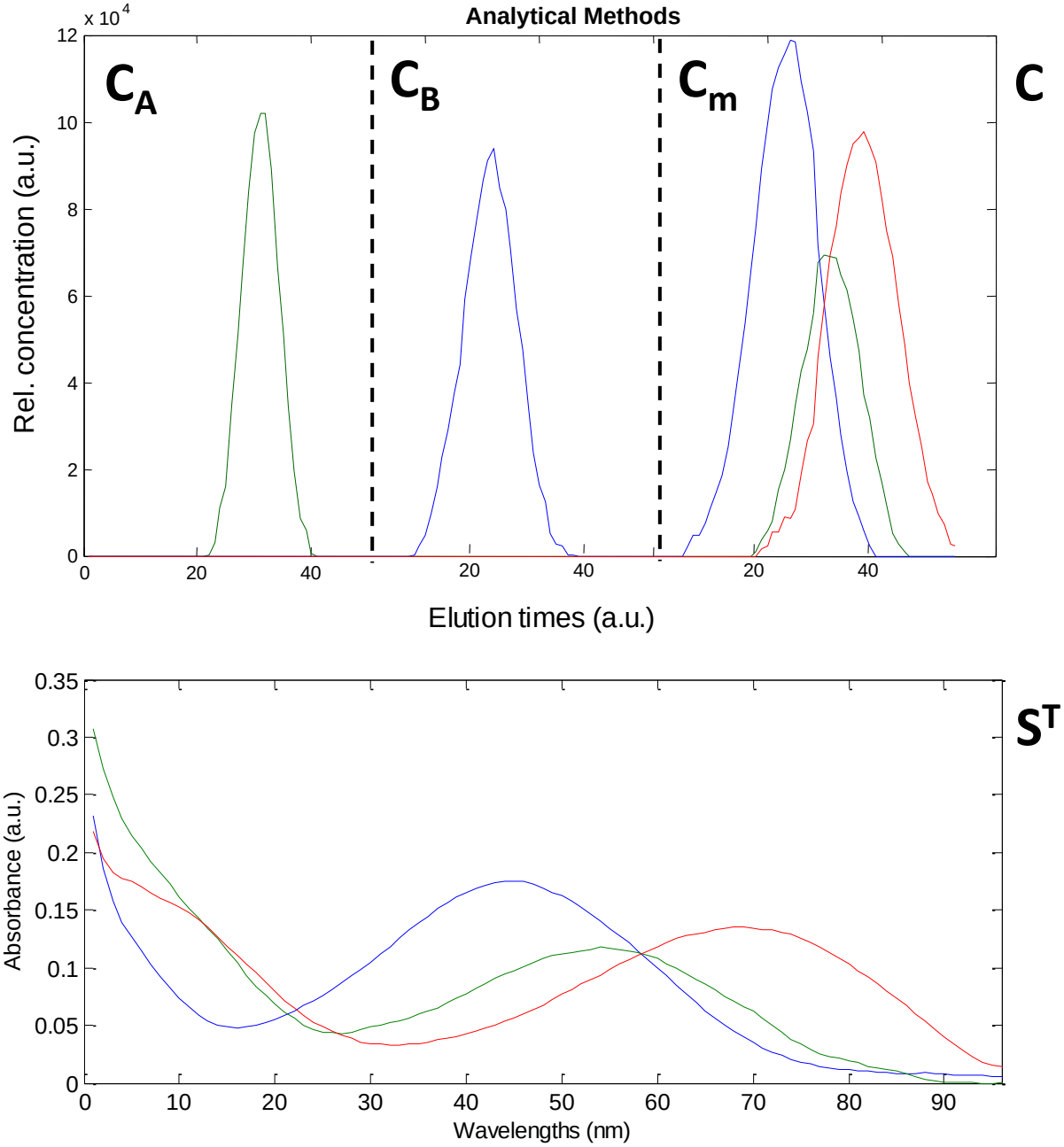


Figure 7