www.elsevier.com/locate/talanta

Analysis of multiple mass spectrometry images from different *Phaseolus vulgaris* samples by multivariate curve resolution

Carmen Bedia, Romà Tauler, Joaquim Jaumot

PII: S0039-9140(17)30811-1
DOI: http://dx.doi.org/10.1016/j.talanta.2017.07.087
Reference: TAL17786

To appear in: *Talanta*

Received date: 8 May 2017
Revised date: 26 July 2017
Accepted date: 28 July 2017

Cite this article as: Carmen Bedia, Romà Tauler and Joaquim Jaumot, Analysis of multiple mass spectrometry images from different *Phaseolus vulgaris* sample by multivariate curve resolution, *Talanta* http://dx.doi.org/10.1016/j.talanta.2017.07.087

Analysis of multiple mass spectrometry images from different *Phaseolus vulgaris* samples by multivariate curve resolution

Carmen Bedia*, Romà Tauler and Joaquim Jaumot

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain.

*Corresponding author: Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain. Telephone: +34934006100.

carmen.bedia@idaea.csic.es

Abstract

A new procedure based on the simultaneous analysis of multiple mass spectrometry images using multivariate curve resolution is presented in this work. Advantages of the application of the proposed approach are shown for three cases of plant studies demonstrating its potential usefullness in metabolomics studies, particularly in lipidomics. In the first dataset, a three stage germination time course process of green bean seeds is presented. The second example is a dose-response study where the stem bases of a non-exposed plant are compared to those of plants exposed to increasing concentrations of the pesticide chlorpyrifos. Finally, the third study is the simultaneous analysis of several sequential transversal and longitudinal cuts of the same green bean plant stem segment.

The analysis of these three examples required the comprehensive adaptation of different chemometric methodologies including data compression by selection of the regions of interest (ROI strategy), appropriate data normalization and baseline correction, all of them before MCR-ALS simultaneous image analysis of multiple samples and post processing of the achieved results. MCR-ALS resolved components provided spatial information about the changes in the spatial composition and distribution of the different lipids on the surface of the investigated samples. These results enabled the identification of single lipids and the clustering of those lipids that behaved similarly in the different images simultaneously analyzed. The proposed

strategy for MSI analysis represents a step forward in the simultaneous analysis of multiple sets of images providing an improved recovery of both spatial and structural information in environmental and biomedical studies.

Graphical abstract

Keywords

Mass spectrometry imaging, MALDI-TOF, multiple images, multiset, chemometrics, multivariate curve resolution, lipidomics

1. Introduction

There is an increasing need for obtaining both spatial morphological and structural information from samples coming from a broad range of research fields. Imaging techniques allow gathering this binary information related to chemical constituents located on the surface of the investigated samples [1]. Several examples of the usefulness of this group of techniques can be found in different research fields such as biomedicine, pharmaceutical process monitoring, food quality analysis or remote sensing [2-5]. Extensive studies in the imaging field have been done using vibrational spectroscopic techniques (such as IR or Raman) due to the possibility of multicomponent analysis at relatively low cost and time requirements. In recent years, the acquisition of images of samples using mass spectrometers as detectors has gained popularity due to the high quality of the information provided by them and due to the significant decrease in the cost of the analysis [6]. Therefore, mass spectrometry imaging (MSI) has emerged as a useful tool for the study of complex material surfaces providing simultaneously spatial and structural information. MSI image pixels are easily associated with the surface composition of the considered sample. On the other hand, MSI provides very

specific chemical information about a broad range of chemical compounds (from small metabolites to large proteins), allowing, in many cases, their unambiguous identification [7].

The huge amount of data obtained in one MSI experiment requires the application of different chemometric and data analysis tools to extract the hidden information in the samples, especially when untargeted approaches are used. In contrast, in targeted approaches where only a few number of a priori knew *m/z* values are evaluated, require lower data analysis effort but possibly omitting relevant information and knowledge. Development of new data handling tools are needed to aid in the different steps of the data analysis workflow such as in the data compression, pretreatment, exploration, and resolution [8-14]. The importance of these new tools is even more relevant when complex natural systems are investigated by means of MS imaging techniques. If several MS images are acquired, the unraveling of all the mixed information within and between the different investigated samples becomes more challenging. For instance, this is the case in the application of MSI to untargeted metabolomic studies that require the simultaneous analysis of several images to extract their spatial and structural information. MSI data arrays are usually arranged in data cubes defined by three data modes or directions assigned to x-pixels, y-pixels, and *m/z* values. However, in their analysis they are unfolded to give a data matrix for every image, where the spectra at all *x-y* pixels of every image are set in its rows one on top of the other. The whole set of pixel intensities measured at different *m/z* values give then a set of equally sized column vectors which are arranged in a data matrix for every sample analyzed. Multiple samples can be then arranged in a furtherly column-wise augmented data matrix, where the common *m/z* values are in the columns of this matrix. This augmentation is only feasible if all the considered samples have the same number of columns and, in particular, the same *m/z* values are considered for all the samples. This step facilitates the subsequent analysis using bilinear model data decomposition chemometric methods [15]. Multivariate curve resolution by alternating least squares (MCR-ALS) is a chemometric method that has previously

demonstrated to be suitable for the simultaneous analysis of multiple images from the same or different samples [16]. MCR-ALS has been applied successfully to multiset data structures built up by spectroscopic measurements of biological or pharmaceutical processes, to multiple chromatographic runs monitored by UV-vis molecular absorption or mass spectrometry, to multidimensional chromatography and spectroscopy, and to spectroscopic images obtained using vibrational techniques [16-23] and more recently also to mass spectrometric (MS) images [10, 21, 24]. However, the ability of chemometric tools (in particular, MCR-ALS) to evaluate simultaneously multiple MS images has been still little explored, especially in relation to the data compression and data pretreatment steps.

In this work, the capabilities of the combination of MSI with different chemometric tools for the analysis of multiple images are assessed with the goal to extract maximum information from them. In particular, MSI can be a powerful complementary tool in metabolomics studies to get spatial, morphological and structural information to be used for disease diagnosis in biomedical studies or for risk assessment in environmental and global change research (*e. g.* analyzing plant or animal tissues exposed to environmental stressors) [25, 26]. To illustrate these possibilities, in this work, MSI is applied to three different datasets focusing on the analysis of lipids in green bean plants. Lipids represent a major component in plant tissues and they are present in all cells as constituents of cell membranes, energy storage molecules or signal transducers. First, a time-course lipidomic evolution of green bean growing seeds is examined in detail, including three different seed MS images at various germination steps. The simultaneous analysis of the seed MS images obtained at different time points revealed useful information regarding the lipid changes underneath the germination process. In the second dataset, alterations in the lipid contents of the stem base of green bean plants induced by the presence of the organophosphate pesticide chlorpyrifos (CPF) in the irrigation water at different concentration levels are investigated. Information about the effects of this CPF on the composition of the plant stem base is shown. Finally, in the third dataset, a new set of MS

images obtained from the stem base of two green bean plants are compared. One stem was cut longitudinally, whereas the other stem was cut crosswise in seven different locations. In this case, the main aim of the study was the comparison of the information about the lipid composition provided by the diverse images obtained within different transversal cuts and between these transversal cuts and the longitudinal cut.

2. Materials and methods

2.1. Sample preparation

The seeds of green bean plant (*Phaseolus vulgaris*) were obtained from the local market.

Samples for the "germination process" study.

Selected plant seeds were germinated on wet cotton inside transparent plastic glasses at room temperature and sunlight. Samples were selected at different germination stages: seed deformation (2 days), seed with hypocotyl (3 days), and seed with young stem (4 days). Seeds were then flash-frozen in liquid nitrogen and stored at -80°C.

Samples for the "green beans exposed to chlorpyrifos" study

CPF analytical standard was obtained from Sigma. Selected green bean seeds were germinated as described above for 5 days. Then, seeds were transferred to plastic pots containing 15 g of soil and watered with 80 ml of Milli-Q water containing 0 (control), 0.02, 0.06 and 0.08 % of CPF. Three plant replicates were prepared for each condition. Plants were watered every 3 days and left to grown for 15 days. Then, plants were harvested, washed and stem bases (1.5 cm) were cut. Samples were flash-frozen and kept at -80°C.

Samples for the "stem base cuts" study

Two stem bases of control plants obtained as explained above were used.

Preparation of samples for MALDI imaging

All tissues were mounted in a cutting chock using Optimal Cutting Temperature (OTC, TissueTek) on the base of the tissue. The tissue slices were made at 15 μm thickness with a cryostat (Leica CM 3050) and placed directly onto ITO glass slides (Bruker). In the case of germinating seeds and the stem bases of plants exposed to CPF, they were cut longitudinally, whereas, in the third example, the stem bases were cut longitudinally and transversally at different levels. Before matrix application, ITO slides were scanned using an office scanner at 2400 ppi resolution in order to have optical images of the sample for further teaching inside the MALDI instrument. Then, 2-mercaptobenzothiazole (MBT, Sigma) a matrix that has been described to possess superior properties regarding the detection of lipids [27] was applied into the slides by sublimation [28]. Briefly, the sublimation device consisted in a flat bottomed condenser to which glass ITO slides were attached facing down using a double sided adhesive tape, and a beaker-shaped bottom piece to which the DHB matrix (100 mg) was added. The two pieces were assembled using an O-ring seal, put in a sand bath in a heating mantle and connected to a vacuum pump. Then, vacuum (100 mTorr) was applied and the mantle was heated to reach 121ºC. Set temperature was reached in 10 minutes and then the device was left in these conditions for 10 additional minutes. Then, heat and vacuum were removed and ITO plates were carefully removed from the sublimation condenser.

Spectra were acquired using an Autoflex III MALDI-TOF/TOF instrument (Bruker Daltonik GmbH) equipped with a Smartbeam laser operated at 200 Hz laser repetition rate at the “large focus” setting. Spectra were obtained in positive reflector ion mode in the 400 to 2000 m/z range. MALDI matrix and instrumental conditions used in this study favored the detection of lipids present in the analyzed tissues. Laser raster was set to 150 μm along both x- and y- axes.

2.2. MS image data compression and preparation

Each image raw data file was loaded into the SCiLS Lab software (version 2014b, SCiLS GmbH) and exported to an imzML file, the standard mass spectrometry data format. Each generated imzML file was then imported into the MATLAB® environment using the imzML converter tool

[29]. Most relevant *m/z* values were determined by using the regions of interest (ROI) approach [30, 31] that retains only those *m/z* values whose MS signal intensities are above a predetermined threshold value and are detected a minimum number of times. The usefulness of this approach has been demonstrated in previous works for different types of MS data [30], and in particular for the analysis of MS single image datasets [10]. In this work, this approach is extended to the analysis of multiple MS images preliminarily compressed consecutively in a pixel-by-pixel mode (see Figure 1a). So, this ROI approach allowed overcoming drawbacks as the difference in the number of detected *m/z* values in each pixel and, also, facilitated the column-wise augmentation strategy for the subsequent MCR-ALS analysis by forcing all the images in a dataset to be represented by the same columns (*m/z* values). Results obtained using this ROI approach allowed determining the most relevant *m/z* values having relevant MS intensity signals for a set of MS images simultaneously analyzed. In the first data example investigated in this work (different stages of the germination process), three images were simultaneously compressed (Figure 1b). Similarly, in the second data example (changes induced by CPF exposure), the four images related to control and exposed samples were jointly compressed (Figure 1c). Moreover, finally, in the third data example (differences between stem base cuts), the seven images containing transversal cuts and the image containing the longitudinal stem cut (number of pixels much larger than the images of transversal cuts) were compressed simultaneously too (Figure 1d). Parameters used for the ROI compression in the three cases were the following: *m/z* error tolerance of 0.25 amu (a multiple of the instrument mass accuracy), signal threshold of 5% of the maximum MS intensity of the image with the lowest intensity among all the analyzed ones, and forcing that the selected m/z value appeared in at least 100 pixels in the first and second examples, and 15 in the third one. This lower value for the minimum number of pixels in the third dataset is caused by the smaller number of pixels of the images containing transversal cuts. The selection of a larger value could cause that compound specific to a single small image could be neglected

from the analysis. Table 1 shows a description summary of images and final dimensions of the considered datasets.

2.3. Chemometric analysis

As a result of the MSI data compression procedure, a new MSROI augmented data matrix containing all the information of the mass spectrometry images was obtained for each one of the three data examples investigated in this work.

The first step of the chemometric analysis consisted in the normalization of the signal intensity values in every pixel to avoid excessive weight in the analysis from pixels with the highest intensity signals. This normalization also minimizes pixel variations caused by experimental issues during the acquisition of the MS image such as, for instance, ionization or matrix deposition efficiencies [1]. MS intensities of each pixel were normalized dividing them by the total ion count (TIC, sum of all intensities in the pixel) in the considered pixel. MSI data requires spectral normalization and, in many circumstances this TIC normalization has been recommended, especially if visual inspection of artifacts is performed before further analysis [8]. Comparative studies can be found in the literature using TIC normalization with good results [32, 33]. Moreover, a baseline correction using the asymmetric least squares (AsLs) method was applied to improve the quality of the experimental data and to minimize noise effects [34].

Next step is the chemometric resolution of the chemical constituents present in the investigated MS images of the plants section samples. From the diversity of methods able to deal with spectroscopic images, MCR-ALS was selected due to its ability to cope with this multiset type of data [35]. In this case, the joint analysis of different MS images was performed using a column-wise augmented data matrix ($\mathbf{D}_{aug}$), in which the ROI spectral dimension is common for all considered pixels (as a result of the previous compression step).

MCR-ALS decomposes a single MSROI image data matrix (**D**) according to a bilinear model following Equation 1:

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^{\mathbf{T}} + \mathbf{E} \qquad \text{Equation 1}$$

Where **D** is the matrix containing the MS information for a single image with a number of rows equals to the total number of pixels and a number of columns equals to the number of ROI m/z values. Then, **C** is the matrix of the concentrations of the different constituents of the image. Its dimensions are the total number of pixels in the image (rows) by the number of resolved image constituents (columns). The long column vectors of this matrix should be refolded in the *x-y* plane to give the distribution map of the concentrations of the different constituents over the investigated surface. $\mathbf{S}^{\mathbf{T}}$ is the matrix of the pure MS spectra of the different constituents of the image. Its dimensions are the number of constituents of the image (rows) by the number of finally selected ROI m/z values (columns). **E** is the residuals matrix having the variance not explained by the model (in the best case, only experimental noise). Its dimensions are the same as in the image data matrix **D**.

When multiple MS images are analyzed simultaneously, the column-wise augmented matrix, $\mathbf{D}_{\mathbf{aug}}$, containing the information of every i^{th} image, $\mathbf{D}_{\mathbf{image},i}$ is built up. The number of rows of this $\mathbf{D}_{\mathbf{aug}}$ matrix will be equal to the sum of the number of all pixels of every individual image data matrix, whereas the number of columns will be equal to the number of selected ROI *m/z* values, which is common for all the images (see above).

When MCR-ALS is applied to this new column-wise augmented matrix, $\mathbf{D}_{\mathbf{aug}}$, as shown in Figure 1E, the bilinear model can be extended to:

$$\begin{bmatrix}\mathbf{D}_{\mathbf{image},1}\\ \mathbf{D}_{\mathbf{image},2}\\ \mathbf{D}_{\mathbf{image},3}\end{bmatrix} = \mathbf{D}_{\mathbf{aug}} = \mathbf{C}_{\mathbf{aug}}\mathbf{S}^{\mathbf{T}} + \mathbf{E}_{\mathbf{aug}} = \begin{bmatrix}\mathbf{C}_{\mathbf{image},1}\\ \mathbf{C}_{\mathbf{image},2}\\ \mathbf{C}_{\mathbf{image},3}\end{bmatrix}\mathbf{S}^{\mathbf{T}} + \begin{bmatrix}\mathbf{E}_{\mathbf{image},1}\\ \mathbf{E}_{\mathbf{image},2}\\ \mathbf{E}_{\mathbf{image},3}\end{bmatrix} \qquad \text{Equation 2}$$

In Equation 2, $\mathbf{C}_{\mathbf{aug}}$ is now the augmented matrix of the concentrations of the different constituents in the different images simultaneously analyzed. Its dimensions are the total

number of pixels in all images simultaneously analyzed (rows) by the number of resolved image constituents (columns). The very long column vectors of this matrix should be refolded in the *x*-*y* plane for every image separately to give the distribution maps of the concentrations of the different constituents over the diverse images found in every i^{th} $\mathbf{C_{image,}}_{i}$. $\mathbf{S^T}$ matrix has the resolved MS spectra of all the resolved image constituents. From these MS spectra, identification of these constituents is feasible. $\mathbf{E_{aug}}$ is the residuals augmented matrix having the non-explained variance in the whole set of images simultaneously analyzed. It has the same dimensions as $\mathbf{D_{aug}}$

MCR-ALS workflow started with the initial estimation of the number of components from Singular Value Decomposition (SVD, See Supplementary Material Fig.1) [36]. In the case of MSI data, the selection of the optimal number of components is not simple and, different MCR-ALS analysis using a different number of components were performed. The number of components finally selected considered the more parsimonious (simpler) model explaining a reasonable large amount of data variance. Very low and noisy contributions were not considered. An initial estimation of spectra profiles for this number of components is also needed for ALS initialization, which was obtained by using a pure variable detection method such as the one proposed in the SIMPLISMA algorithm [37].

ALS optimization was carried out using non-negativity constraints for both **C** (distribution maps) and $\mathbf{S^T}$ (spectra) profiles. [38].Vector normalization (equal length) was imposed to each resolved spectra profile to minimize scale ambiguities of the resolved components during ALS. Rotational ambiguities can be disregarded in this case due to the sparseness of MS image data with a large number of zero values in both data directions, in the pixels and the mass spectra directions [39].

Explained variances R^2 (%) in the analysis of the different datasets (Equation 3) were used as a figure of merit to evaluate the goodness of the MCR-ALS resolution:

$$R^2(\%) = 100\frac{\sum_{ij} d_{ij}^2 - \sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2}$$ Equation 3

Where d_{ij}^2represents the signal of the $\mathbf{D_{aug}}$ matrix at a particular *i-th* pixel and *j-th m/z* value, and e_{ij}^2 corresponds to the residual of $\mathbf{E_{aug}}$ at the same pixel and *m/z* value.

A final step of the analysis of the MS images is the potential identification of the constituents (mostly lipids) detected in the images, by the comparison of their mass values with those obtained from public databases such as LipidMaps[40] and HMDB[41].

All calculations were performed using MATLAB® R2016a software (The Mathworks, Natick, MA, US) running on a Fujitsu Celsius R940n workstation equipped with two Intel® Xeon® CPU E5-2620 v3 processors and 128Gb RAM using Microsoft® Windows 7.

3. Results

Application of the proposed workflow to the simultaneous analysis of multiple MS images is demonstrated below for the lipidomics study of the three selected MSI data examples. The details about the image dimensions of each dataset and the number of ROIs values finally obtained are given in Table 1.

3.1. MCR analysis of MS images of bean slices during their germination

This dataset was composed of three MS images of bean slices obtained during its germination. The multiset augmented data matrix built up for the simultaneous analysis of these three images had 12332 rows (corresponding to all pixels in the three different images with 41x84, 44x94 and 48x99 pixels, respectively) and, after ROI compression, 398 columns (*m/z* values). Baseline correction using asymmetric least squares method and normalization considering the total ion count (TIC) obtained for each considered pixel were applied, prior MCR-ALS chemometric analysis. Visual inspection of the different images after TIC normalization allowed discarding those artifacts caused by undesirable size effects.

Different preliminary analyses using a different number of components were performed before deciding what could be the best model. Initial estimates of the pure MS spectra were

obtained by selecting the purest pixels by a procedure similar to the one used in the SIMPLISMA method [37]. A 28 components MCR-ALS model was finally fitted, explaining more than the 97% of the experimental data variance (R^2).

MCR-ALS resolved components defined by their relative contributions as distribution maps and by their mass spectra could be then linked to possible biological variations. In these components, trends related to the evolution of the germination process can be observed with appearance or disappearance of compounds contribution. A detailed explanation of two of these components is presented below. However, other MCR-ALS resolved components showed inconsistent changes in the distribution maps or an excessive number of peaks in their MS spectra that hindered their biological interpretation and forced their assignation to background and noise signal contributions.

Figure 2 shows an example of the distribution maps (concentration profiles mapped on two-dimensional surface images) and spectra of two of the components resolved by o MCR-ALS. For comparison purposes, the independent optical images of the same three bean seed slices acquired using an office scanner before matrix deposition and MSI analysis are shown in Figure 2a. Two of the MCR-ALS resolved components are shown in detail to highlight the advantages of the MCR-ALS results when compared with traditional MSI analysis methods where every *m/z* value is checked individually. Distribution maps of these two components show the spatial location of the detected compound within a particular bean slice and, and their evolution (appearance, disappearance or small variation) during the germination process at different stages. Figure 2b shows the mass spectra of these two components resolved by MCR-ALS. Component I (blue bars, left inset) showed a single relevant contribution at an m/z value equal to 785.3, which, taking into account that lipids detection is favored in the used experimental conditions, can be tentatively identified as a triacylglyceride (46:0). Distribution maps for this component showed the evolution of the contribution of this compound during the three considered germination stages (see Figure 2C). Triglycerides can account for more than 60% of

the weight of seeds. These lipids constitute the primary source of energy store whose breakdown is necessary to perform the transition from seed to seedling [42]. This is in agreement with the decreasing abundance observed for this component in the tissue at increasing germination stages. At the first germination time, the triacylglyceride showed a relevant contribution in the central part of the seed, which decreases significantly during the second stage to finally almost disappear at the last germination time.

On the other side, the mass spectrum of the second component shown in Figure 2b (component II, red bars, right inset) which has different mass peaks can be assigned to distinct adducts of the same compound or to different chemical compounds showing a rather similar spatial behavior. A tentative assignation of this component can be performed from its major peaks in Figure 2B. Table 2 shows the *m/z* values and a tentative identification of these peaks, which have been finally assigned to some of the main lipids families such as glycerolipids and phospholipids, including phosphatidylcholines, phosphatidylglycerols, phosphatidylserine and diacylglycerols. These are lipid species also present in the seeds which have been shown to be subjected to variations during the germination process [43]. The distribution maps of this component in the three slices showed that at the first germination stage, the presence of this component is dispersed among the whole seed at low concentration levels (Figure 2D). Then, at the second germination stage, these molecules are still distributed along the seed with higher intensities in some pixels. Finally, at the last germination stage, only minor contributions of this component were detected along the bean slice.

3.2. MCR analysis of MS images of stem base cuts exposed to chlorpyrifos

This dataset was composed of four MS images corresponding to four different stem base longitudinal cuts of green bean plants exposed to increasing concentrations of the

organophosphate pesticide CPF. This dataset allowed evaluating the variations in the concentration and distribution of lipids due to the CPF exposure. The size of the fused four images after column-wise augmentation was 4336 rows (pixels) and of only 143 m/z values after ROI compression. The resulting compressed data was subjected to TIC normalization and baseline correction using asymmetric least squares. As explained in the previous example, different trials were made to find the most suitable number of components in the MCR-ALS model to describe this multiset data adequately. Finally, 42 components were selected in this case, which explained more than a 98% of the experimental variance (R^2). As in the previous example, some of the MCR-ALS components were not biologically relevant (linked to background or noise contributions), but others showed increasing and decreasing tendencies influenced by the CPF exposure.

In this data example, the aim was to find what changes in the composition and distribution of the stem base compounds were due to the effect of increasing concentrations of CPF. As the CPF was added to the irrigation water, the roots and the base of the stem were in direct contact with the pesticide. Thus, these parts were more likely to present larger variations in their lipid composition. Figure 3 shows the mass spectra and corresponding distribution maps of two of the resolved MCR-ALS components illustrating more relevant CPF effects.

The optical image of the four stem bases from control to increasing doses of CPF, ordered from left to right, is shown in Fig. 3A in which the upper part of each image corresponds to the aerial part of the stem. From these images, it is difficult to obtain relevant information from the morphological analysis. MSI analysis provided a deeper insight of the system when the resolved MCR-ALS distribution maps were considered (Figures 3B and 3C). For instance, component I showed a mass spectrum with a single contribution of a compound with an m/z equal to 833.8 (in red in Figure 3D), which was tentatively assigned to a triacylglyceride (50:5) (see Table 2). This compound was homogeneously distributed in the inner part of the stem

base of the non-exposed plants (Figure 3B). The distribution of this triacylglyceride was not so homogeneous at 0.02% CPF exposure but, in general, the intensity of the pixels was higher than in controls. Then, a clear decrease was observed under exposure to CPF at 0.06% and 0.08%. A possible explanation for this could be the mobilization of triacylglycerides as a source of energy for the plant, in order to compensate the reduction of the photosynthetic performance experienced in the presence of the pesticide (unpublished data).

Other MCR components of this study, such as component II, probably represented simultaneously different species (in blue in Figure 3D), meaning that these compounds were similarly affected by the addition of CPF to the soil. The mass spectra of component II allowed the tentative identification of three lysophosphatidylcholine and three diacylglyceride species (see Table 2 for the recovered m/z values and tentative assignation of these compounds), which belong to phospholipid and glycerolipid classes, respectively. In the stem base of the control plant, the distribution map showed that these compounds accumulated in the inner and central part of the stem slice (Fig. 3C). The amount of these compounds at this localization decreased progressively at higher concentrations of CPF, being almost undetectable at CPF 0.08%. As diacylglycerides are part of the degradation pathway of triacylglycerides, a decrease in these species would reinforce the previous idea of triacylglyceride catabolism as an energy source for the plant. These results showed how CPF could induce changes in lipid composition and distribution in the stem base of green bean plants.

3.3. MCR analysis of MS images of stem base cuts

This dataset was composed of eight MS images of a green bean stem base: a longitudinal cut (L-cut) and seven transversal sections (T-cuts) at different levels from the aerial part connected to the rest of the stem to the ground part connected to roots (see Fig 4A, optical images from up to down). In this case, the column-wise augmented data matrix had 2701 rows (pixels) and

2529 columns (*m/z* values) after ROI compression. Again, same data pretreatments for baseline correction using asymmetric least squares and TIC normalization were applied.

The MCR-ALS model selected for describing this multiset data structure had 46 components (which includes interesting and non-interesting biological contributions). This MCR-ALS model explained more than a 98% of the experimental variance (R^2).

When considering all the resolved MCR-ALS components, in general, there was a good agreement between the information provided in the distribution maps resolved for the longitudinal and transversal cuts. When component contributions were located in the outer section of the longitudinal cut distribution map, then these contributions were also located in the outer parts of the transversal cuts whereas the inner parts showed negligible contributions. The same behavior was observed when assessing those regions of the longitudinal section closer to the roots or aerial parts that showed a more similar behavior to transversal cuts adjacent to roots (T1-T3) or aerial part (T6-T7). Two examples showing these trends are detailed below.

Figure 4 shows the optical images (Figure 4A), the MCR-ALS resolved spectra (Figure 4B) and distribution maps (Figures 4C and 4D) for two of the components resolved by MCR-ALS. In both cases, the resolved spectra showed a single relevant peak at an m/z value which was used for its identification. For the MCR-ALS component in blue in Figure 4B (component I in Table 2), the m/z value of the most intense peak was at 498.69. It was assigned to a lysophosphatidylethanolamine (18:3). Distribution maps showed a good agreement between L- and T- cuts (Figure 4C). The L- cut (L in Fig. 4C) showed strong contributions on the outer section of the stem, especially in the middle section of the tissue. This agreed with the distribution maps obtained for the T-cuts. The first two distribution maps (T1 and T2 in Fig.4C), corresponding to the stem narrow end, only showed really minor contributions, similar to the upper part of the L-cut. The other distribution maps presented more intense contributions. It is worth to mention that in the T-cuts corresponding to the stem broadest sections (T6 and T7 in

Fig.4C), intense contributions were detected in the outer region of the stem whereas, in the center, the resolved contributions were less intense.

In the second MCR-ALS resolved component, the situation observed in the resolved maps was the opposite. From the MCR-ALS resolved mass spectrum (Figure 4B red line), the m/z value detected was 565.41 which could be tentatively assigned to diacylglyceride (32:2). Here, the main contributions in the L cut were located at the aerial end of the stem (Fig. 4D) whereas in the broad region, corresponding to the ground end, only residual contributions could be found. Similar results were found by the study of the seven T-cuts. In this case, the most intense contributions were observed in the upper transversal cuts (T1 and T2 in Fig.4D). When considering wider transversal sections, the intensity of the contributions diminished and, at the widest T-cut (T7 in Fig. 4D), it was almost negligible (in agreement with the spatial information recovered from the L-cut).

4. Conclusions

In this work, the suitability of chemometric tools such as MCR-ALS to analyze series of images obtained by mass spectrometry in plant studies is demonstrated with potential applications in lipidomics and, by extension, metabolomics studies. Lipid constituents present in the investigated bean plant samples were detected and their tentative identification was performed. The proposed chemometrics strategy provides information related to the spatial distribution of these lipids in every analyzed image (sample) which were common for all images or specific for a particular image. Using this information, the chemical and biological interpretation of the studied set of images could be obtained allowing a deeper understanding of the considered systems.

The results obtained in the different examples of this work demonstrate that the combination of MSI with advanced data analysis tools, such as the MCR-ALS method, is a promising

approach to be used in untargeted -omics type of studies. For instance, the proposed MCR-ALS approach for the analysis of multiple images can be applied in lipidomics studies where changes induced in selected tissues are assessed by evaluating the changes in the concentrations of the compounds present in different samples (e.g. comparison of exposed and non-exposed samples). However, the unequivocal identification of these compounds (lipids or metabolites) will require the acquisition of high-resolutions mass spectrometry data or, at least, the confirmation of the proposed candidates by subsequent MS/MS independent to the proposed data analysis strategy. Spatial information provided by the proposed MSI chemometric analysis integrated strategy can be an excellent complementary approach to other approaches such as LC-MS, commonly used in metabolomics and lipidomics studies.

5. Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 320737.

Table 1. Description of the datasets analysed in this work.

Data set	No. of images	Image ID	No. of pixels[1]	Final no. of *m/z* values[2]
Germination process	3	Stage 1	3444 (41x84)	398
		Stage 2	4136 (44x94)	
		Stage 3	4752 (48x99)	
		Augmented	12332	
CPF exposure	4	Control	1170 (26x45)	184
		CPF 0.02%	1060 (50x53)	
		CPF 0.06%	874 (23x38)	
		CPF 0.08%	1232 (22x56)	
		Augmented	4336	
Stem base cuts	8	Longitudinal cut (L)	2214 (41x54)	2529
		Transversal cut 1 (T1)	30 (5x6)	
		Transversal cut 2 (T2)	30 (5x6)	
		Transversal cut 3 (T3)	25 (5x5)	
		Transversal cut 4 (T4)	36 (6x6)	
		Transversal cut 5 (T5)	49 (7x7)	
		Transversal cut 6 (T6)	121 (11x11)	
		Transversal cut 7 (T7)	196 (14x14)	

	Augmented	2701

[1] Total number of pixels (number of x-pixels and y-pixels)
[2] Number of m/z values of the augmented matrix after ROI compression

Table 2. *m/z* values and tentative identification of lipids discussed in the results section. DG: diacylglycerol; PC: phosphatidylcholine; PC-P: phosphatidylcholine plasmalogen; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; PS: phosphatidylserine; TG: triacylglycerol.

EXAMPLE 1- TIME COURSE GERMINATION PROCESS				
MCR-ALS component	*m/z found*	*Tentative id.*	*Adduct*	*Lipid class*
I	785.3	TG(46:0)	$[M+Na^+-2H]^-$	Glycerolipid
	749.6	DG(44:4)	$[M+Na^+-2H]^-$	Glycerolipid
	773.3	PG(36:2)	$[M-H^+]^-$	Phospholipid
	777.1	PG(36:0)	$[M-H^+]^-$	Phospholipid
II	779.4	DG(46:3)	$[M+Na^+-2H]^-$	Glycerolipid
	800.4	PS(36:6)	$[M+Na^+-2H]^-$	Phospholipid
	802.8	PC-P(38:0)	$[M-H^+]^-$	Phospholipid
	822.2	PC(38:8)	$[M+Na^+-2H]^-$	Phospholipid
EXAMPLE 2- CPF EXPOSURE				
MCR-ALS component	*m/z found*	*Tentative id.*	*Adduct*	*Lipid class*
I	833.8	TG (50:5)	$[M+Na^+]^+$	Glycerolipid
	522.4	LysoPC(18:1)	$[M+H^+]^+$	Phospholipid
	554.1	LysoPC(18:4)	$[M+K^+]^+$	Phospholipid
II	555.0	DG(31:0)	$[M+H^+]^+$	Glycerolipid
	556.1	LysoPC(18:3)	$[M+K^+]^+$	Phospholipid
	658.9	DG(38:8)	$[M+Na^+]^+$	Glycerolipid
	689.5	DG(38:1)	$[M+K^+]^+$	Glycerolipid
EXAMPLE 3 - STEM CUTS				
MCR-ALS component	*m/z found*	*Tentative id.*	*Adduct*	*Lipid class*
I	498.7	Lyso PE (18:3)	$[M+Na^+]^+$	Phospholipid
II	565.4	DG(32:2)	$[M+H^+]^+$	Glycerolipid

Figure 1. A) Schematic representation of the ROI compression approach in the case of multiple image analysis. Augmented data matrices for the B) "Germination process", C) “CPF exposure” and D) “Stem base cuts” datasets. E) Schematic representation of the MCR-ALS analysis of the augmented dataset showing the distribution maps obtained after refolding of matrix **C** columns and spectra contained in rows of matrix $\mathbf{S}^T$.

Figure 2. Results obtained in the MCR-ALS analysis of MS images of green bean germination process. A) Optical images of the seeds at the three considered germination stages. B) Resolved MCR-ALS MS spectra of two components, I (red) and II (blue). C) Resolved MCR-ALS distribution maps of component I. D) Resolved MCR-ALS distribution maps of component II.

Figure 3. Results obtained in the MCR-ALS analysis of MS images of stem bases from CPF exposed green bean plants. A) Optical images of the longitudinal cuts of the stem base. The upper part of each image corresponds to the aerial end of the stem bases. From left to right, control, 0.02, 0.06 and 0.08% CPF treated plants. B) Resolved MCR-ALS distribution maps of component I, identification of images as in A). C) Resolved MCR-ALS distribution maps of component II, identification of images as in A). C) Resolved MCR-ALS MS spectra of components I (red) and II (blue).

Figure 4. Results obtained in the MCR-ALS analysis of MS images of longitudinal (L) and transversal (T) stem base cuts. A) Optical images of the T cuts of the stem base (left), ordered from T1 to T7, from the aerial part to the ground end. Optical image of the L cut of the stem base (right).Scale bar 2mm. B) Resolved MCR-ALS MS spectra of components I (red) and II (blue). C) Resolved MCR-ALS distribution maps of component I, identification of images as in A). D) Resolved MCR-ALS distribution maps of component II, identification of images as in A).

References

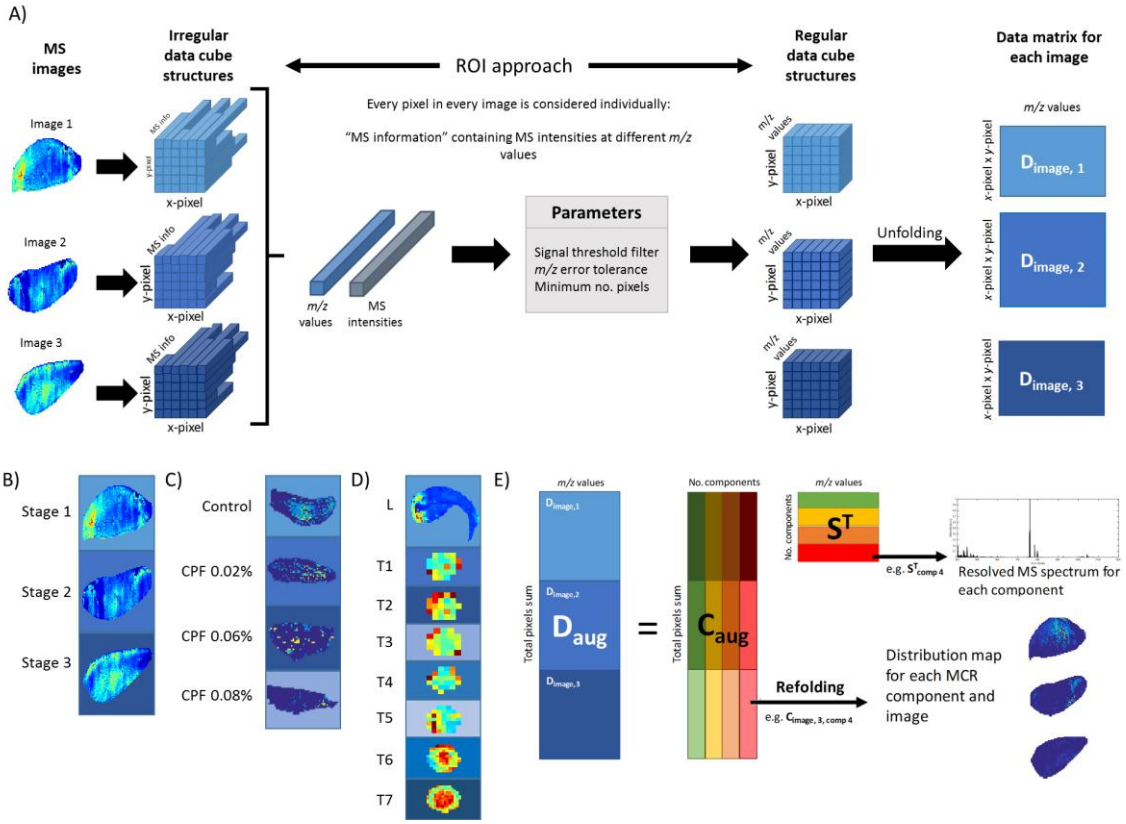
[1] J. Burger, A. Gowen, Data handling in hyperspectral image analysis, Chemometrics and Intelligent Laboratory Systems 108(1) (2011) 13-22.
[2] B. Boldrini, W. Kessler, K. Rebnera, R.W. Kessler, Hyperspectral imaging: A review of best practice, performance and pitfalls for in-line and on-line applications, Journal of Near Infrared Spectroscopy 20(5) (2012) 483-508.
[3] Y.Z. Feng, D.W. Sun, Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review, Critical Reviews in Food Science and Nutrition 52(11) (2012) 1039-1058.
[4] G. Lu, B. Fei, Medical hyperspectral imaging: A review, Journal of Biomedical Optics 19(1) (2014).
[5] S. Veraverbeke, E.N. Stavros, S.J. Hook, Assessing fire severity using imaging spectroscopy data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and comparison with multispectral capabilities, Remote Sensing of Environment 154 (2014) 153-163.
[6] M.M. Gessel, J.L. Norris, R.M. Caprioli, MALDI imaging mass spectrometry: Spatial molecular analysis to enable a new age of discovery, Journal of Proteomics 107 (2014) 71-82.
[7] B. Spengler, Mass spectrometry imaging of biomolecular information, Analytical Chemistry 87(1) (2015) 64-82.
[8] T. Alexandrov, MALDI imaging mass spectrometry: statistical data analysis and current computational challenges, BMC bioinformatics 13 Suppl 16 (2012).
[9] A. Bartels, P. Dülk, D. Trede, T. Alexandrov, P. Maaß, Compressed sensing in imaging mass spectrometry, Inverse Problems 29(12) (2013).
[10] C. Bedia, R. Tauler, J. Jaumot, Compression strategies for the chemometric analysis of mass spectrometry imaging data, Journal of Chemometrics 30(10) (2016) 575-588.
[11] P.J. Cumpson, I.W. Fletcher, N. Sano, A.J. Barlow, Rapid multivariate analysis of 3D ToF-SIMS data: graphical processor units (GPUs) and low-discrepancy subsampling for large-scale principal component analysis, Surface and Interface Analysis 48(12) (2016) 1328-1336.
[12] M. Galli, I. Zoppis, A. Smith, F. Magni, G. Mauri, Machine learning approaches in MALDI-MSI: clinical applications, Expert Review of Proteomics 13(7) (2016) 685-696.
[13] P. Källback, A. Nilsson, M. Shariatgorji, P.E. Andrén, MsIQuant - Quantitation Software for Mass Spectrometry Imaging Enabling Fast Access, Visualization, and Analysis of Large Data Sets, Analytical Chemistry 88(8) (2016) 4346-4353.
[14] A.M. Race, A.D. Palmer, A. Dexter, R.T. Steven, I.B. Styles, J. Bunch, SpectralAnalysis: Software for the Masses, Analytical Chemistry 88(19) (2016) 9451-9458.
[15] A. de Juan, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler, Chemometric Tools for Image Analysis, in: R. Salzer, H.W. Siesler (Eds.), Infrared and Raman Spectroscopic Imaging, Wiley-VCH2009, pp. 65-109.
[16] S. Piqueras, L. Duponchel, R. Tauler, A. De Juan, Monitoring polymorphic transformations by using in situ Raman hyperspectral imaging and image multiset analysis, Analytica Chimica Acta 819 (2014) 15-25.
[17] M. Abou Fadel, X. Zhang, A. De Juan, R. Tauler, H. Vezin, L. Duponchel, Extraction of Pure Spectral Signatures and Corresponding Chemical Maps from EPR Imaging Data Sets: Identifying Defects on a CaF<inf>2</inf> Surface Due to a Laser Beam Exposure, Analytical Chemistry 87(7) (2015) 3929-3935.
[18] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan, A. Gorzsás, Vibrational spectroscopic image analysis of biological material using multivariate curve resolution-alternating least squares (MCR-ALS), Nature Protocols 10(2) (2015) 217-240.
[19] G.L. Alexandrino, J.M. Amigo, M.R. Khorasani, J. Rantanen, A.V. Friderichsen, R.J. Poppi, Unveiling multiple solid-state transitions in pharmaceutical solid dosage forms using multi-series hyperspectral imaging and different curve resolution approaches, Chemometrics and Intelligent Laboratory Systems 161 (2017) 136-146.

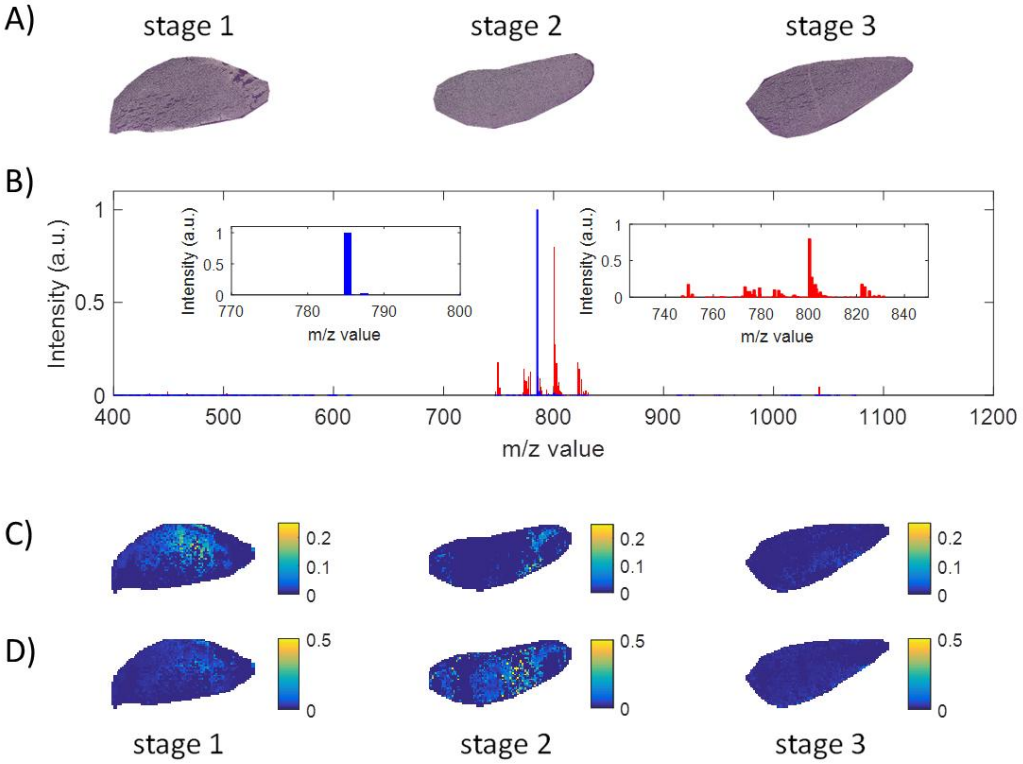
[20] R.L. Carneiro, R.J. Poppi, Homogeneity study of ointment dosage forms by infrared imaging spectroscopy, Journal of Pharmaceutical and Biomedical Analysis 58(1) (2012) 42-48.
[21] Y. Gut, M. Boiret, L. Bultel, T. Renaud, A. Chetouani, A. Hafiane, Y.M. Ginot, R. Jennane, Application of chemometric algorithms to MALDI mass spectrometry imaging of pharmaceutical tablets, Journal of Pharmaceutical and Biomedical Analysis 105 (2015) 91-100.
[22] M. Offroy, M. Moreau, S. Sobanska, P. Milanfar, L. Duponchel, Pushing back the limits of Raman imaging by coupling super-resolution and chemometrics for aerosols characterization, Scientific Reports 5 (2015).
[23] X. Zhang, R. Tauler, Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging, Analytica Chimica Acta 762 (2013) 25-38.
[24] J. Jaumot, R. Tauler, Potential use of multivariate curve resolution for the analysis of mass spectrometry images, Analyst 140(3) (2015) 837-846.
[25] B.A. Boughton, D. Thinagaran, D. Sarabia, A. Bacic, U. Roessner, Mass spectrometry imaging for plant biology: a review, Phytochemistry Reviews 15(3) (2016) 445-488.
[26] M. Lagarrigue, R.M. Caprioli, C. Pineau, Potential of MALDI imaging for the toxicological evaluation of environmental pollutants, Journal of Proteomics 144 (2016) 133-139.
[27] E. Astigarraga, G. Barreda-Gomez, L. Lombardero, O. Fresnedo, F. Castano, M.T. Giralt, B. Ochoa, R. Rodriguez-Puertas, J.A. Fernandez, Profiling and imaging of lipids on brain and liver tissue by matrix-assisted laser desorption/ ionization mass spectrometry using 2-mercaptobenzothiazole as a matrix, Anal Chem 80(23) (2008) 9105-14.
[28] A. Ly, C. Schöne, M. Becker, J. Rattke, S. Meding, M. Aichler, D. Suckau, A. Walch, S.M. Hauck, M. Ueffing, High-resolution MALDI mass spectrometric imaging of lipids in the mammalian retina, Histochemistry and Cell Biology 143(5) (2015) 453-462.
[29] A.M. Race, I.B. Styles, J. Bunch, Inclusive sharing of mass spectrometry imaging data requires a converter for all, Journal of Proteomics 75(16) (2012) 5111-5112.
[30] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, TrAC - Trends in Analytical Chemistry 82 (2016) 425-442.
[31] R. Tautenhahn, C. Bottcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinformatics 9 (2008).
[32] S.O. Deininger, D.S. Cornett, R. Paape, M. Becker, C. Pineau, S. Rauser, A. Walch, E. Wolski, Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations, Analytical and Bioanalytical Chemistry 401(1) (2011) 167-181.
[33] J.M. Fonville, C. Carter, O. Cloarec, J.K. Nicholson, J.C. Lindon, J. Bunch, E. Holmes, Robust data processing and normalization strategy for MALDI mass spectrometric imaging, Anal. Chem. 84(3) (2012) 1310-1319.
[34] P.H.C. Eilers, A perfect smoother, Analytical Chemistry 75(14) (2003) 3631-3636.
[35] A. De Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Analytical Methods 6(14) (2014) 4964-4976.
[36] G.H. Golub, C. Reinsch, Singular value decomposition and least squares solutions, Numerische Mathematik 14(5) (1970) 403-420.
[37] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Analytical Chemistry 63(14) (1991) 1425-1432.
[38] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemometrics and Intelligent Laboratory Systems 140 (2015) 1-12.
[39] P. Paatero, P.K. Hopke, X.H. Song, Z. Ramadan, Understanding and controlling rotations in factor analytic models, Chemometrics and Intelligent Laboratory Systems 60(1-2) (2002) 253-264.
[40] E. Fahy, M. Sud, D. Cotter, S. Subramaniam, LIPID MAPS online tools for lipid research, Nucleic Acids Research 35(SUPPL.2) (2007) W606-W612.
[41] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-

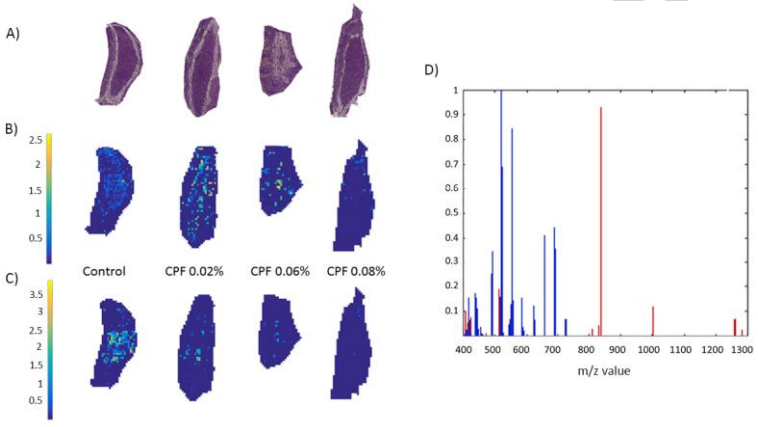
Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0-The Human Metabolome Database in 2013, Nucleic Acids Research 41(D1) (2013) D801-D807.
[42] I.A. Graham, Seed storage oil mobilization, Annual Review of Plant Biology, 2008, pp. 115-142.
[43] J.O. Offem, E.O. Egbe, A.I. Onen, Changes in lipid content and composition during germination of groundnuts, Journal of the Science of Food and Agriculture 62(2) (1993) 147-155.

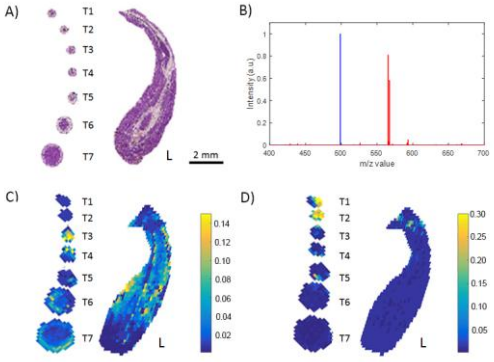
Highlights

- MCR-ALS was successfully applied on multiple mass spectrometry images of plants
- This strategy revealed the chemical composition and distribution within the systems
- The MCR simultaneous analysis allowed the biological interpretation of each case
- This method could be a useful procedure for both environmental and biomedical studies




A)
stage 1
stage 2
stage 3
B)
Intensity (a.u.)
m/z value
C)
D)
stage 1
stage 2
stage 3


A)
B)
C)
Control
CPF 0.02%
CPF 0.06%
CPF 0.08%
D)
m/z value


A)
B)
C)
D)
T1
T2
T3
T4
T5
T6
T7
L
2 mm