Multi-excitation hyperspectral autofluorescence imaging for the exploration of biological samples

Madiyeh Ghaffari, Anne Laure Chateigner-Boutin, Fabienne Guillon, Marie Francoise Devaux, Hamid Abdollahi, Ludovic Duponchel

Multi-excitation hyperspectral autofluorescence imaging for the exploration of biological samples

Mahdiyeh Ghaffari[1,2], Anne-Laure Chateigner-Boutin[3], Fabienne Guillon[3], Marie-Françoise Devaux[3], Hamid Abdollahi[1], Ludovic Duponchel[2*]

[1]*Department of Chemistry, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan 45137-66731, Iran.*

[2] *Université de Lille, CNRS, UMR 8516 - LASIR - Laboratoire de Spectrochimie Infrarouge et Raman, F-59655 Villeneuve d'Ascq, France.*

[3] *INRA – Centre de recherche Angers-Nantes, UR BIA ''Biopolymères Interactions Assemblages" 1268, Nantes, France.*

(*)corresponding author : ludovic.duponchel@univ-lille.fr

Abstract

Many plant tissues can be observed thanks to autofluorescence of their cell wall components. Hyperspectral autofluorescence imaging using confocal microscopy is a fast and efficient way of mapping fluorescent compounds in samples with a high spatial resolution. However a huge spectral overlap is observed between molecular species. As a consequence, a new data analysis approach is needed in order to fully exploit the potential of this spectroscopic technique and extract unbiased chemical information about complex biological samples. The objective of this work is to evaluate multi-excitation hyperspectral autofluorescence imaging to identify biological components in wheat grains during their development through their spectral profiles and corresponding contribution maps using Multivariate Curve Resolution - Alternating Least-Squares (MCR-ALS), a signal unmixing algorithm under proper constraints. For this purpose two different scenarios are used: 1) analyzing the total spectral domain of data sets using MCR-ALS under non negativity constraint in both spectral and spatial modes; 2) analyzing a reduced spectral domain of data sets using MCR-ALS under non negativity in both modes and trilinearity constraint in spectral mode. Considering the original instrumental setup and our data analysis approach, we will demonstrate that extracted contribution maps and spectral profiles of constituents can provide complementary information used to identify molecules in complex biological samples.

Key words

Multivariate Curve Resolution - Alternating Least-Squares; multi-excitation hyperspectral images; Autofluorescence; Trilinearity constraint; Wheat grain.

1. Introduction

Plants are complex organisms with a specialized body plan that requires multilevel organization (organs, tissues, cell types, subcellular compartments) and highly specialized tissues with distinct properties. Plant organs are therefore a mosaic of different cell types, with variation in their structure and composition under genetic and environmental controls. Plant organs are widely used for food, feed and industrial applications. Studying the tissue composition of organs is therefore of tremendous interest to understand plant biological functions and to evaluate the quality of plant tissues for nutritional and industrial applications.

Due to the heterogeneous nature of plant tissues and to compare multiple plant samples, there is a need for imaging techniques that rapidly provide information concerning the chemical composition of tissues with good spatial resolution. Spectroscopic methods such as Raman, Mid-infrared or UV-Visible fluorescence are classically used to analyze plant tissue sections.[1–6] However, to be applicable to plant comparison, spectral imaging techniques must deal with minimal sample preparation (e.g. avoid tissue dehydration and resin embedding), high spatial resolution (< 3 μm per pixel), high acquisition speed and a high field of view while keeping enough spectral information to ease molecular identification. As a consequence full field fluorescence imaging systems such as confocal microscopes equipped with spectral detectors is certainly the only technique meeting such requirements. Taking advantage of the autofluorescence properties of many plant compounds[2,3,6–9] confocal or multiphoton imaging can be performed with little tissue preparation and, more important, without labeling.

Hyperspectral autofluorescence imaging was considered in this work to follow the evolution of wheat grain tissues during grain development. For this purpose, we collected confocal hyperspectral images of wheat grains sections at different stages of development. The wheat grain comprises several tissues: embryo, endosperm and outer layers. Autofluorescence in cereal grain is due to several compounds, including pigments such as chlorophyll in green tissues, carotenoids, anthocyanidin and proanthocyanidin in colored grains.[10–12] In the grain outer layers, autofluorescence is particularly important in cell walls where it has been linked to the presence of phenolic compounds such as lignin and hydroxycinnamic acids (ferulic acid, para-coumaric acid).[13–15] However, the fluorescence properties of major compounds largely overlap in the spectral domain. For instance, lignin and hydroxycinnamic acids both fluoresce after UV

excitation while only lignin is fluorescent after visible excitation.[2,16,17] Chemometric methods have already been used to overcome this problem and identify tissues in dry mature grain and in particles from autofluorescence multispectral images.[11, 14–16] However we need to get beyond that because ambiguity has not been totally lifted. Indeed combining autofluorescence hyperspectral images obtained by confocal microscopy after different excitations may help to further identify tissues with overlapping autofluorescence responses or in cases of the co-localization of components. For each excitation wavelength, one hyperspectral image is acquired. The set of hyperspectral images forms a multi-excitation hyperspectral image. It can also be considered that for each pixel in an image, an excitation emission fluorescence matrix is acquired. Thus we can see that such experiments will impose constraints. First, as emission wavelengths are always longer than excitation ones, it results in a different spectral domain depending on the excitation wavelength and a partial excitation–emission matrix. Second, the final multi-excitation hyperspectral image cannot be strictly considered as a full 3-way data set. Finally, with the purpose of comparing samples during grain development, several images have to be analyzed together in a consistent way. As a consequence data sets have to be reorganized and chemometric methods have to be adapted for this specific analysis.

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is one of the signal unmixing techniques that can provide pure spectra and contribution maps of different components in the samples.[21] It is an iterative algorithm that solves the bilinear model with no prior knowledge. This algorithm extracts from the data set pure contributions (concentration and spectral profiles) of all compounds present in the sample from suitable alternating least-squares optimization subjected to different constraints. The latter is imposed based on the chemical knowledge of the studied system.[22–25] When data sets are analyzed by MCR-ALS, the result might be challenging due to lack of unique solutions, which is an intrinsic characteristic of bilinear matrix decompositions if incomplete information is available about the system.[26] The resulting uncertainty may be dramatically large in certain cases, e.g., when extensive profile overlap occurs in one of the data modes. However, imposing different constraints such as non-negativity, unimodality, selectivity and trilinearity may drastically decrease the extent of rotational ambiguity by adding more information to the MCR analysis of the system under study.[27–31] A Trilinearity constraint can guarantee accurate unique profiles for constituents with trilinear structure.[32,33] To fulfill the trilinearity condition, the augmented data matrix should

contain some similar factors in its sub-matrices, factors which should not be affected by experimental conditions so that they share a common spectral shape. Only their areas (and vertical heights) should proportionately change according to the constituent concentration. It is a very strong constraint that forces the decompositions to give unique solutions under mild conditions.

The objective of this work is to evaluate the method of multi-excitation hyperspectral imaging by confocal microscopy combined with the exploration of second order data sets with the MCR-ALS approach to identify different components in the developing wheat grain based on their autofluorescence properties. Contribution maps and spectral profiles of different compounds are complementary information which will be extracted from the exploration of this kind of data sets. For this purpose two different scenarios are used: 1) Analyzing the whole spectral domain - obtained by merging the data sets of each excitation wavelength - using MCR-ALS under non negativity constraint in both modes; 2) Analyzing the reduced spectral domain - common to all excitation wavelength - using MCR-ALS under non negativity in both modes and trilinearity constraint in spectral mode in order to get second order advantage. The two approaches will be compared and discussed on the basis of extracted contribution maps and pure spectral profiles of all components. These results will be also compared with images generated from the standard method based on emission signal integration.

2. Material and methods

2.1. Plant materials and growth conditions

Wheat plants (*Triticum aestivum* L. cv. Recital) were grown in containers filled with plain soil at INRA Clermont-Ferrand (France) under conditions of natural day length and temperature. Upon flowering plants were transferred to a covered structure allowing to control and monitor the temperature. Temperature was set at 21°C from 6 a.m. to 9.30 p.m. and 14°C at night. Wheat grain development is influenced by the time elapsed since flowering, the temperature and the position of the grain within the spike. Thus its development was monitored using the thermal time method[34] which is more robust than only considering time when temperature is varying. The thermal time unit is Celsius degrees days after flowering (°DAF). This measure is the

cumulated daily average temperature since flowering. Individual spikes were tagged when the first flowering event was observed. Grains were harvested at different desired developmental stages as already described in a previous work.[35] More precisely samples were collected at 150, 270, 350, 460, 560, 630, 790 °DAF and at mature stage.

2.2. Multi-excitation hyperspectral autofluorescence imaging

Freshly harvested grain samples were frozen and cut in the equatorial region of the grain using a cryotome (HM 500 OM, Microm) into 20 µm cross sections. In the case of dry grains, the embryo was removed and the grains were placed onto moist paper for 24 h at 4°C to facilitate sectioning. For each grain, serial sections were placed on a slide. This was then mounted in water and analyzed using a confocal laser-scanning system (A1, Nikon) equipped with a x40 objective for confocal imaging. The microscope was equipped with a spectral detector unit and provided three excitation wavelengths: 375 nm (UV), 488 nm (blue) and 561 nm (green). At the chosen magnification, the field of view was 317 × 317 µm and images were digitized as matrix of 512 × 512 pixels with a pixel size of 0.62 × 0.62 µm. As a consequence each hyperspectral data cube contained 262,144 emission spectra. Moreover for each field of view, three hyperspectral images were recorded by collecting emitted light from 404 to 714 nm for UV excitation, 504 to 744 nm for blue excitation and 574 to 744 nm for green excitation with a 10 nm step between spectral variables. A dichroic mirror was used to filter the three excitation wavelengths. For each development stage, two or three serial sections were imaged. Two regions per section were analyzed: (1) opposite to the crease and (2) on the side of the grain, called dorsal and side respectively, with the objective to visualize all the tissues of the wheat grain (Figure 1). In the case of the youngest grains, one field of view was not always sufficient. In this case, two images were acquired to observe the whole region. Finally the whole data set contained 40 multi-excitation hyperspectral autofluorescence images considering different parts of the grain and development stages; this corresponds to more than 10 million emission spectra in total.

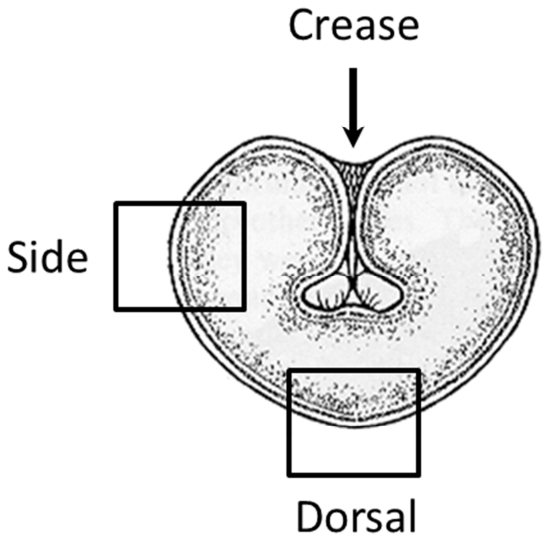


Figure 1. A simple scheme of wheat grain cross section with spectral acquisition areas.

2.3.Spectral data arrangement

In this work, multiple excitation wavelengths were used to generate autofluorescence hyperspectral images of the wheat grain; which means one 3D autofluorescence hyperspectral image (two spatial modes *x* and *y* and one emission mode λ_{em}) per excitation wavelength denoted $\lambda_{ex,1}$, $\lambda_{ex,2}$ and $\lambda_{ex,3}$, respectively (Figure 2a). $\lambda_{em,1}$, $\lambda_{em,2}$ and $\lambda_{em,3}$ are the three emission wavelength ranges corresponding to the three excitation wavelength. So for these three excitation wavelengths, there are three 3D cubes thus forming together a 4D data set with two spatial modes *x* and *y*, one excitation mode and one emission mode. Each multi-excitation hyperspectral autofluorescence image corresponds to a 4D image as shown in figure 2a by considering together the purple, blue and green cubes. As a consequence, an excitation emission autofluorescence matrix is acquired for each pixel of the sample surface instead of only one emission spectra as it is usually the case for conventional fluorescence analysis. However, this matrix is only partial because the spectral domains corresponding to each excitation wavelength differs. In order to analyze the acquired data sets by using a curve resolution method like MCR-ALS, a 2D matrix of data is needed. Consequently each 3D autofluorescence hyperspectral image matrix corresponding to each excitation is unfolded to a matrix with $x \times y$ rows and $\lambda_{em,i}$ columns (i=1, 2 and 3) as shown in figure 2b.

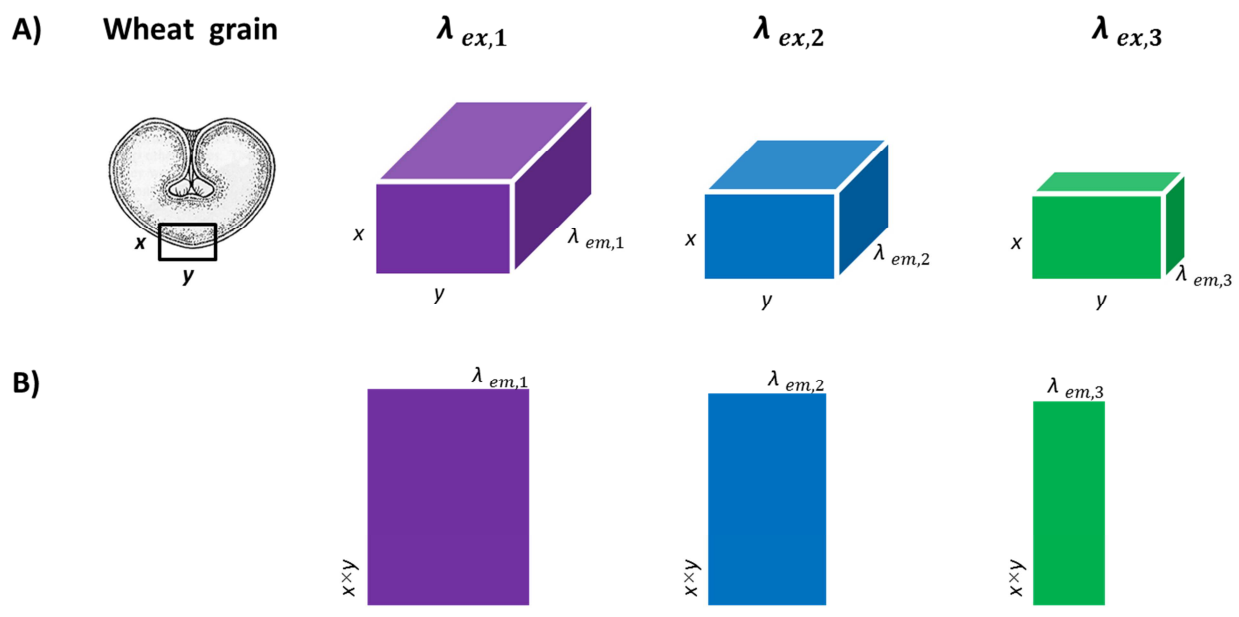


Figure 2. Data arrangements: a) the three 3D blocks (denoted $\lambda_{ex,1}$ $\lambda_{ex,2}$ and $\lambda_{ex,3}$) are three autofluorescence hyperspectral images collected using the three excitation wavelengths. b) Unfolded version of the three cubes.

2.4. Data analysis – the MCR-ALS approach

The MCR-ALS method is used to decompose simultaneously a hyperspectral image into the pure spectra of the image constituents and corresponding contribution maps. MCR-ALS is based on a bilinear model which assumes that each observed spectrum is a linear combination of pure components spectra present in the system.[36,37] This model can be written in matrix form as:

$$\mathbf{D}=\mathbf{C}\mathbf{S}^{T}+\mathbf{E} \quad (1)$$

Where, **D** is the unfolded hyperspectral image data set, **C** is the matrix of the relative amounts or contributions, $\mathbf{S}^{T}$ is the pure spectra matrix and **E** is the matrix associated to non-modeled part of data and potentially only noise. Naturally signal intensity in every pixel of the image should not be negative and neither should the contributions of the different constituents. It is therefore logical to use non-negativity constraint during alternating least squares optimization. The spectral matrix $\mathbf{S}^{T}$ is also normalized to avoid scale ambiguity. At the end of the ALS optimization procedure, the optimized **C** matrix can be refolded, to recover a 2D image

(distribution map) of the contributions of every component. To evaluate the quality of extracted profiles obtained from the MCR-ALS procedure, the percentage of lack of fit (*lof*) and the percentage of explained variance (R^2) are calculated according to the two following equations:

$$lof\ (\%) = 100 \times \sqrt{\frac{\sum_{ij} d_{ij} - \hat{d}_{ij}}{\sum_{ij} d_{ij}^2}} \qquad (2)$$

where d_{ij} is the element of the hyperspectral image data set **D**, and $\hat{d}_{ij}$ is the corresponding element of this data matrix recalculated by MCR-ALS. This lack of fit value gives a measure of the fit quality in relative terms with the same units as the measured data, and comparable to the experimental relative error estimation. For the explained variance, R^2 is calculated as:

$$R^2 = 100 \times \left(1 - \frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2}\right) \qquad (3)$$

where e_{ij} are the elements of the **E** matrix. The main advantage of this algorithm is the amount of information that can be included in the optimization process and the ability of working with either a single data matrix or multiset data structures i.e. using simultaneously several data cubes. However it is known that such an approach does not always extract unique solutions if not well managed. Therefore additional chemical information and well-selected constraints used in the decomposition process can significantly reduce or even eliminate this uncertainty. One of the most well-known constraints to ensure uniqueness of extracted solutions in MCR-ALS is trilinearity. Due to the trilinearity constraint, data arrays should contain some similar factors in their modes while factors should be independent from experimental conditions.[37] Trilinearity is a very strong constraint that forces trilinear decompositions to have unique solutions under mild conditions.[38,39] An important achievement in the analysis of complex data matrices resulted when Multivariate Curve Resolution methods were applied to several data matrices simultaneously to the so-called Matrix Augmented - Multivariate Curve Resolution - Alternating Least Squares, MA-MCR-ALS.[40] Generally speaking, resolution ambiguities and rank deficiency problems in the analysis of two-way data sets can be reduced significantly if it is possible to analyze data structures with more information. MA-MCR-ALS, as an extension of MCR-ALS, can be easily adapted to the trilinear analysis of the three-way data sets generated from data matrices with the same row or column, taking advantage of their structure. When the trilinearity constraint is used

in the MA-MCR-ALS framework, profiles of a component in different data matrices are forced to have a common shape during each ALS optimization but may differ by a scaling factor. When this constraint is inserted into the ALS iterative optimization procedure, it forces the shape of the loading vectors to be the same in all sub-matrices. An interesting aspect of the trilinearity constraint in the ALS optimization is that its use is optional for each component of the system.

MCR-ALS optimization always starts with the generation of initial estimates of either concentration or spectral profiles. In general, the use of chemically meaningful estimates is an essential factor that can lead to not only a rapid convergence of the extractions but also a decrease in ambiguity of solutions in some cases. Different methods can be used to find suitable initial estimates to start the MCR-ALS calculation. The so-called SIMPLISMA based on the concept of purest variables is usually used to calculate them.[41] However it was not possible to use it in this work due to the very high number of spectra in the considered multiset analysis. We have, therefore, decided to use the Kennard–Stone (KS) algorithm in order to generate initial estimates. [42] This algorithm allows for the selection of spectra from all parts of the data space even in the case of nonhomogeneous distribution. It starts by finding the two most distant spectra in the data set using Euclidean distance, and then continues with other points until the selected rank is reached.

The rank of the data set i.e. its total number of independent signals is usually obtained from Principal Component Analysis (PCA) considering the magnitude of singular values. Singular values related to chemical contributions are typically large, whereas singular values related to noise are smaller and similar among themselves. This diagnostic can be complemented by looking at the emergence of noisy patterns in scores and loadings profiles, typical in noise-related components. When in doubt, few MCR models with different number of components can be calculated. The final model is selected as the smallest one providing an optimal model fit and chemically meaningful resolved profiles.[43] In this work for most of the data sets several MCR models were calculated and the selected one had minimum residuals and optimal fit. In order to extract spectral profiles and contribution maps of all active components in the sample, MCR-ALS approach has been applied following two different scenarios.

In a first scenario, an augmented data set **D** was obtained by considering the whole spectral domain of each data cube and merging the corresponding three 2D unfolded matrices as shown

in figure 3b. In this case, MCR-ALS was applied under non-negativity constraint only in order to extract spectral autofluorescence emission profiles and contribution maps of active components. As explained in the data collection section, several regions were observed for each grain section and several serial sections were also observed for each development stage, resulting in several multisets (in red and orange in figure 3a). In the figure 3b, purple, blue and green cubes correspond to the three different excitation wavelengths used while light and dark colors correspond to two sample regions of interest. Then we use column- and row-wise data augmentation to fuse all matrices and generate a global matrix **D**. Lastly, the augmented data set is used to extract pure concentration and spectral matrices using MCR-ALS (Figure 3c). To tackle the large volume of data from the 40 multi-excitation hyperspectral images (i.e. more than 10 million of emission spectra), augmented data sets were built and analyzed separately for each development stage that still correspond to the multivariate analysis of 1.3 million spectra each.

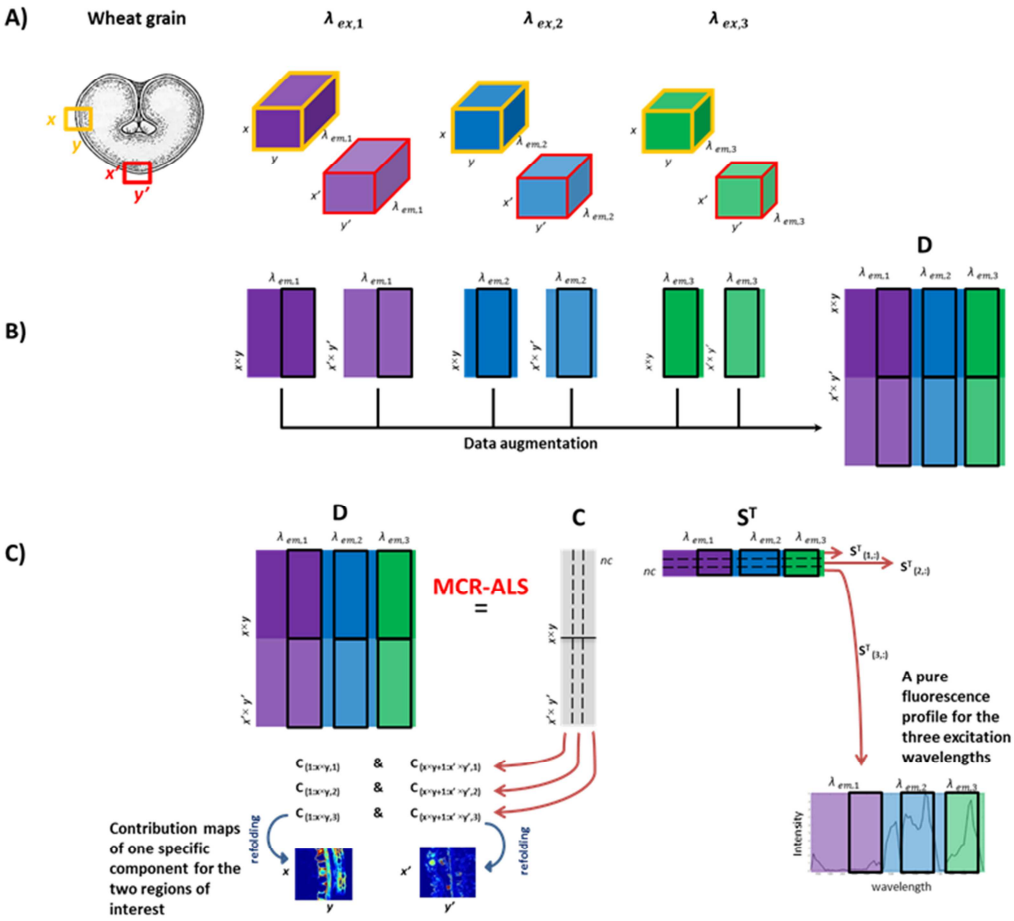


Figure 3. Data arrangements; a) three 3D blocks are three autofluorescence hyperspectral images acquired with three different excitation wavelengths: $\lambda_{ex,1}$, $\lambda_{ex,2}$ and $\lambda_{ex,3}$ which are shown in purple, blue and green, respectively (both dark and light colors are related to orange and red regions of interest of the wheat grain). b) Unfolding of cubes and final data augmentation generating D matrix. c) The augmented data set is used to extract pure concentration and spectral matrices using MCR-ALS. Open black rectangles in the figure represent the common spectral range used when the trilinearity constraint is applied.

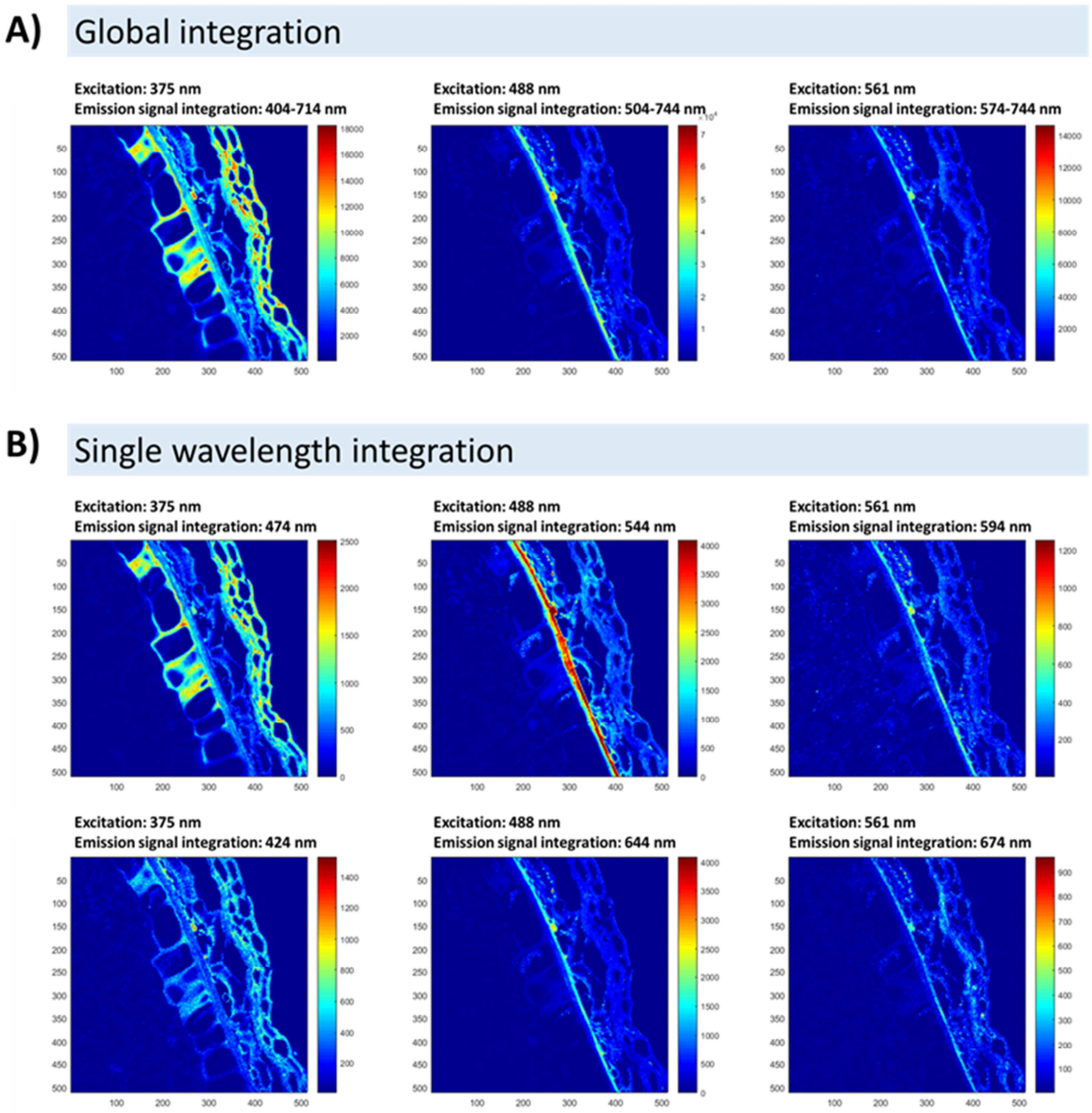
A second scenario has been considered in this work because rotational ambiguity (i.e. uncertainty of MCR-ALS results) cannot be eliminated in the majority of cases only using the non-negativity constraint. Therefore it has been decided to apply additionally the trilinearity constraint in order to reach uniqueness of solutions in MCR-ALS, or at least verify if other

extractions are obtained in these new conditions. However, to analyze the acquired data set under the trilinearity constraint, it must have a trilinear structure at least for one mode. In the case of fluorescence, the shape of the emission spectrum for a given fluorophore is largely independent from excitation wavelength. Therefore the recorded autofluorescence hyperspectral images using different excitation wavelengths are potentially suitable for applying trilinearity constraint in spectral mode. However due to our specific instrumental setup, hyperspectral images have not exactly the same spectral domain, a necessary condition for applying this new constraint in MCR-ALS. As a consequence, a reduced spectral domain (574-714 nm) common to the three unfolded matrices has been considered. This common spectral window is represented by black open rectangles in Figure 3. In this scenario, we can already see that the application of trilinearity constraint has paid a price of losing spectral information below 574 nm.

3. Result and discussions

3.1 Autofluorescence image generation with a conventional approach

As described in section 2.3, the multi-excitation hyperspectral confocal images acquired form a set of three 3D blocks. Their analysis and display are therefore not straightforward. The usual ways to generate a 2D image from a fluorescence imaging data set are the integration of the signal over the whole spectral domain of emission, the integration around the wavelength of a maximum emission or a signal extraction at a particular spectral channel corresponding to a specific emission. Figure 4 shows generated autofluorescence images of the same wheat section using the three excitation wavelengths 375, 488 and 561 nm. More precisely, images in Figure 4a have been obtained considering signal integration on the whole emission domain of each excitation wavelength. Autofluorescent compounds are observed in all tissues corresponding to the wheat grain outer layers and no autofluorescence is observed into the endosperm (see figure 5 for tissue and cell layers annotation if needed). Moreover the use of different wavelengths of excitation seems to be a way to highlight different subparts of the section. Indeed, after excitation at 375 nm, all outer layers fluoresce from the aleurone layer to the epiderm, while the seed coat is highlighted after excitation at 488 nm. Using excitation at 561 nm, some small points are also revealed inside cross cells corresponding to pigments.[44]



318

319 Figure 4. Autofluorescence images of the same wheat section at different excitation wavelength
320 A) when emission is integrated over the whole spectral domain B) for emission at given
321 wavelengths.

322

323 Autofluorescence images in Figure 4b have been generated considering only emission at
324 given wavelengths. These particular wavelengths have been selected because it was possible to
325 generate different emission images highlighting different subparts of the wheat section.

Variations in intensities in the different images suggest that each cell type shows some specific fluorescence spectral response. In autofluorescence imaging, no chemical probes are used and all molecules in the sample that are intrinsically fluorescent give emission in different parts of the spectral domain. Autofluorescence is often considered as a spurious signal but it is not. It is in fact a good opportunity to observe simultaneously different molecular contributions.[15] All images shown in figure 4 obviously reveal the presence of different autofluorescent behaviors of the different tissues. However even if specific sub-parts of the grain section are highlighted at given wavelength of integration, there is no evidence that it corresponds to the contribution of only one molecule, simply because of the huge bandwidth of the considered spectroscopy. The complexity of the data structure is even more striking when looking at spectra extracted from specific areas of the considered wheat section. Figure 5 presents manual selections of pixels in cell walls of different grain tissues such as aleurone cells (red line), seed coat (orange line), cross cells (blue line), and epicarp (green line) . Corresponding mean spectra of selected pixels are also presented in the figure.

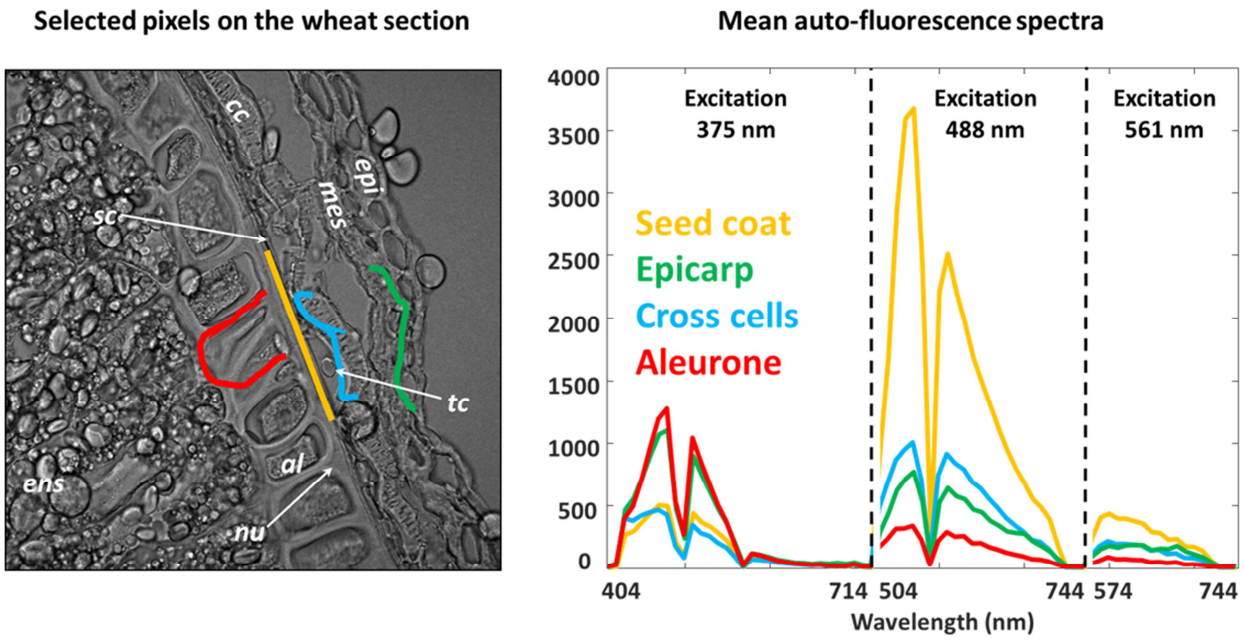


Figure 5. Manual selection of pixels in cell walls of different grain tissues (aleurone = red line, seed coat = orange line, cross cells = blue line and epicarp = green line) and corresponding mean spectra of emission for different excitation wavelengths. Annotations: *epi* = epicarp, *mes* =

mesocarp, *cc* = cross cell, *tc* = tube cell, *sc* = seed coat, *nu* = nucellar epidermis, *al* = aleurone layer and *ens* = endosperm.

Although these four specific tissues are known to contain different fluorescent molecules or different concentration of several ones, the global shape of emission spectra is almost the same, only relative intensity changes being observed for different excitation wavelengths. In conclusion, a simple integration approach does not allow us to obtain unbiased information about the fluorescent molecules present in the grain tissues due to a huge spectral overlap of fluorescence.[3,6,45] As a consequence, it was decided to apply the signal unmixing approach MCR-ALS in order to extract pure fluorescence signatures of molecules in the next section of the paper.

3.1 Signal unmixing using MCR-ALS using the whole spectral domain

All acquired data sets of each specific stage (more than 1.3 million of emission spectra) were analyzed together considering a column-wise data augmentation. In other words, we had eight augmented data sets corresponding to the eight development stages. Each data set has been analyzed separately using MCR-ALS under proper constraint. The chemical rank for each data set has been determined using singular value decomposition (SVD).[22] However it was sometimes necessary to calculate different MCR models with different number of components when SVD information was not so clear. Considering the principle of parsimony, the final MCR model was selected as the one with the smallest chemical rank, a good model fit and chemically meaningful resolved profiles. A mean LOF value of 0.9% and a mean R^2 value of 0.96 were observed for MCR-ALS models which are good figures of merit considering the signal to noise ratio. As summarized in table 1, a total of seven components were extracted with MCR-ALS from the eight development stages when using the whole spectral domain and non-negativity constraint only (i.e. following the first scenario). Some components were present throughout the whole grain development and others only at specific stages. This demonstrates the great potential of such methodology being able to detect various molecular contributions even when the spectral overlap is maximal. Indeed, on the basis of previous results (figure 4 and 5), it was very difficult to envisage observing so many spectral contributions on such complex biological samples.

	150 °DAF	270 °DAF	350 °DAF	470 °DAF	560 °DAF	630 °DAF	780 °DAF	Mature
Component A	✔	✔	✔					
Component B	✔	✔	✔	✔	✔	✔	✔	✔
Component C	✔	✔	✔	✔	✔	✔	✔	✔
Component D					✔	✔	✔	
Component E				✔		✔	✔	
Component F				✔	✔	✔	✔	✔
Component G	✔	✔	✔	✔	✔	✔	✔	✔

Table 1. The seven extracted components, A, B, C, D, E, F and G from the eight development stages. A tick sign indicates the presence of a component at a particular stage.

MCR-ALS extracted spectral profiles of all components A-G at each development stages are shown in figure 6. Correlation coefficients were used in order to set out representations of the same component from the different development stages (i.e. different MCR-ALS models).Each panel of a component is divided in three separated parts corresponding to the three excitation wavelengths 375, 488 and 561 nm respectively highlighted in purple, blue and green. Figure 7 presents corresponding extracted contribution maps of these components. If needed, all these high resolution images can be retrieved in the supplementary material section in Matlab figure format.

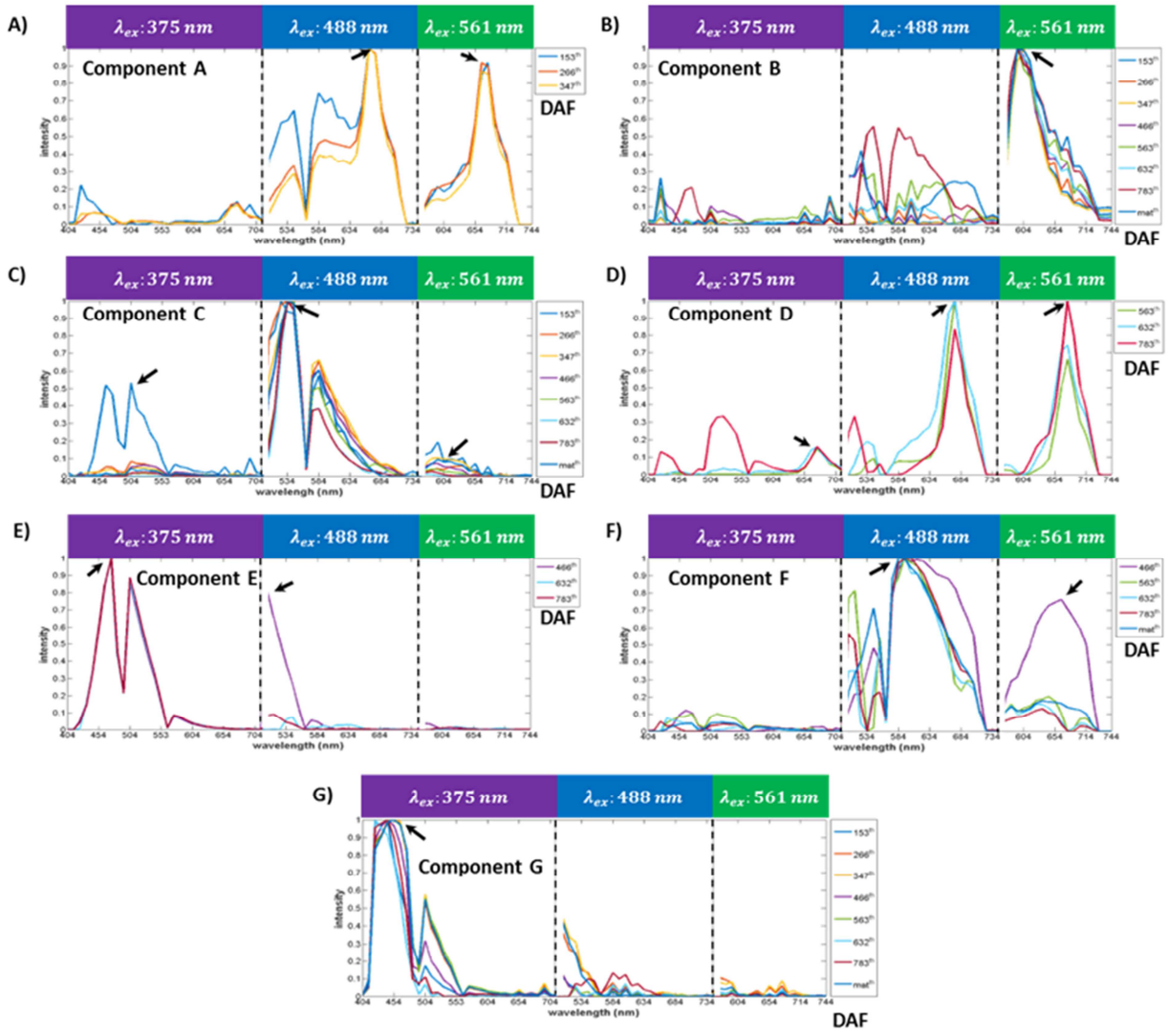


Figure 6. Extracted spectral profiles of components A-G. Subparts of each panel correspond with the three excitation wavelengths. Arrows indicate selected maxima of emission.

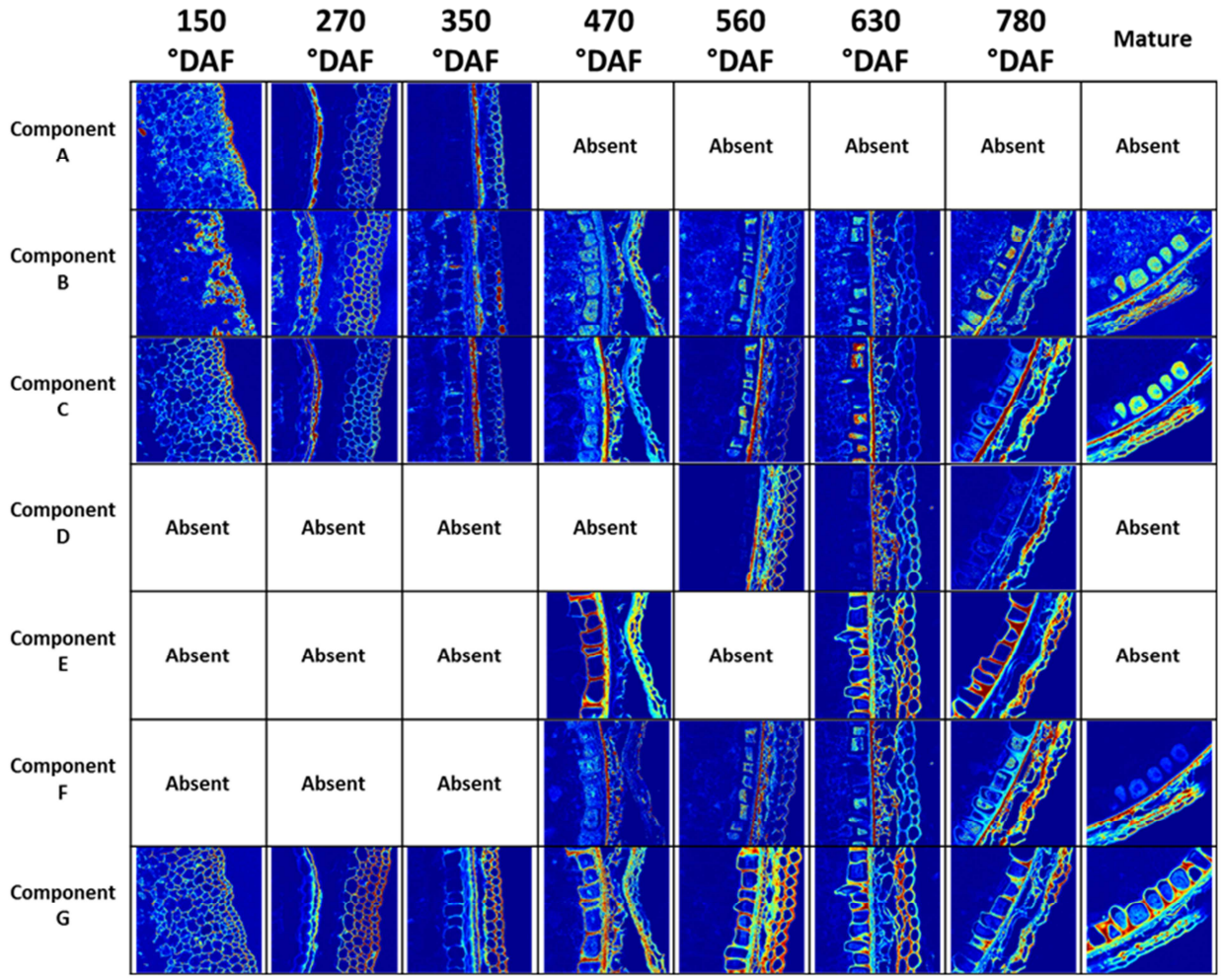


Figure 7. Examples of extracted contribution maps of all components in the wheat grain. Minimum and maximum values in each contribution maps are represented with cold and hot colors respectively.

From a general point of view, the MCR-ALS approach extracted specific spectral profiles for each component. Molecular identification in fluorescence spectroscopy is naturally based on the selection of maximum emission wavelengths on pure extracted profiles but our approach also allow to us to observe hidden components at particular excitation wavelengths. Moreover presence or absence of emission can be used jointly for spectral interpretation. The extracted components are examined in more details here.

Component A (Figure 6A) is present from 150 °DAF to 350 °DAF (as shown in table 1). Its maximum of emission is observed at 664 nm after excitations at 488 and 561 nm (black arrows). Moreover this component exhibits no significant emission when excited at 375 nm. A broad emission is also observed between 500 and 630 nm with an excitation at 488 nm. Note that the emission drop observed at 561 nm corresponds to the wavelength excluded by the dichroic mirror for the three excitation wavelengths. Corresponding contribution maps of component A are observed on the first row of Figure 7. Component A is mainly found in the chloroplasts of cross cells, and in the outer cell wall of the epicarp for the samples at 150° and 270 °DAF. At 350 °DAF this component particularly highlights the cross cells and the seed coat. Considering spectral profile and subcellular locations of this component, it might be assigned to pigments like flavonoids, carotenoids and chlorophyll.[3,7,46–48]

Component B (Figure 6B) is detected from 150 °DAF to mature stage (as shown in table 1). Its maximum of emission is observed at 604 nm for an excitation at 561 nm. Similar to component A, this component have no significant emission for an excitation at 375 nm. Concerning the excitation at 488 nm, a moderate emission is observed in the 500-630 nm spectral region. Corresponding contribution maps are presented in the second row of figure 7 for all stages of development. The location of this component depends on the development stage. At early stages (from 150 to 350 °DAF), it is mainly found inside cells of the epicarp and mesocarp and also inside cross cells. From 560 to 630 °DAF, it is mainly located inside cross cells and in the outer cell wall of the seed coat which is covered by a cuticle. From 780 °DAF to the mature stage, this component is mainly observed in the seed coat, inside aleurone cells and in the cell walls of the epicarp and mesocarp. The multiple location of the component in all stages makes the spectral interpretation of this component difficult. However considering the excitation-emission bands observed and the very similar shape of the component for all stages, this fluorescence might be attributed to flavonoid pigments.

Component C (Figure 6C and 7) is present at all investigated stages. The maximum of emission is located at 544 nm after excitation at 488 nm. Almost no fluorescence is observed for excitation at 561 nm and 375 nm. The only exception is a moderate emission observed at 150 °DAF after an excitation at 375 nm. Component C is mainly found in epicarp cell wall at 150°DAF, in the seed coat, the tube cell walls, and the nucellar epidermis from 350 °DAF to 630 °DAF. In addition, it is detected inside aleurone cells and on all pericarp cell walls at the mature

stage. The assignment of this fluorescence to only one chemical compound is not possible due to the different subcellular locations. Moreover the excitation/emission profile does not correspond to a known autofluorescence compound.[3] It should be stressed that an extracted contribution of the MCR-ALS approach can potentially contain different molecules which are correlated with concentration, which may account for the difficulty of its interpretation.

Component D (Figure 6D and 7) appears from 560 °DAF to the 780 °DAF. The maximum of emission is located at 674 nm for excitations at 561 nm and 488 nm as indicated by black arrows. During these development stages, the compound is present in all pericarp cell walls and in the seed coat with a gradual decrease over time. Unfortunately considering both spectral behavior and localization, no attribution of this compound can be proposed at this time.

Component E (Figure 6E and 7) is present in the cell walls at 470, 630 and 780 °DAF. The maximum of emission is observed at 474 nm after excitation at 375 nm as indicated by black arrows. In this case, no significant fluorescence is observed for excitations at 561 nm and 488 nm except for 460 °DAF where a remaining fluorescence is found in the blue region. For the three stages, the component is mainly found in the cell walls of the aleurone layer, nucellus epidermis and those of the mesocarp and epicarp. Considering the cell wall locations and the spectral signature, this contribution might be attributed to hydroxycinnamic acids like ferulic or para-coumaric ones, even though hydroxycinnamic acids have been detected in cell walls before 460 °DAF.[47] The fluorescence of these compounds is known to vary for instance with pH, types of bonds, and different molecular environments.

Component F (Figure 6F and 7) is present from 470 °DAF to the mature stage. The shape of its emission profile is changing between different stages. Consequently, several wavelengths of maximum emission are observed for this component. In addition there is almost no fluorescence with an excitation at 375 nm. From 470 °DAF to 630 °DAF, the component is located in the seed coats, in the cell walls of cross and tube cells and in the outer cell wall of epicarp. It is noteworthy that both seed coat and epicarp are covered by a cuticle. After 780 °DAF, all pericarp cell walls show an important amount of this component. Considering both spectral and spatial information, this contribution could correspond to some lignin and/or cuticle compounds. Indeed it has already been reported that the autofluorescence of lignins and cuticles was due to multiple fluorophores.[2,47]

Component G (Figure 6G and 7) is present in all investigated stages. The maximum of emission is observed at 450 nm. Fluorescence of this compound is found in the cell walls of all outer tissues of the grain at all stages. Only some variations of intensities are observed. For example, the fluorescence is weaker in the nucellar epidermis compared to the other tissues. It increases in the cell walls in the aleurone layer from 350 °DAF. Considering the spectral pattern of this compound and its localization, it can be assigned to hydroxycinnamic acids like ferulic or para-coumaric ones.[35,49]

Analyzing multi-excitation hyperspectral autofluorescence imaging data sets with the MCR-ALS approach has therefore a good potential to obtain more information about molecules and their localization than with conventional data analysis techniques even if all components are not interpreted at this time. In this first scenario of MCR-ALS optimization, however only non-negativity constraint has been applied which may give rise to some ambiguities on extracted spectral profiles and corresponding contribution maps. It is in this sense that the impact of using the trilinearity constraint in MCR-ALS has to be evaluated in the next section.

3.1 Signal unmixing using MCR-ALS using trilinearity constraint

Trilinearity constraint can potentially guaranty accuracy of unique extractions from MCR-ALS. However, due to the trilinearity constraint, the augmented data set should contain some similar factors in their modes while the factors should be independent from experimental conditions. Thus profiles of a component in different data matrices are forced to have common shape during each ALS optimization but may differ by a scaling factor. During the ALS iterative optimization procedure, the constraint forces the shape of the loading vectors to be the same in all sub-matrices. In order to analyze the above mentioned data sets using MCR-ALS under trilinearity constraint, it was necessary to have the same spectral domain of emission. The spectral region of emission (574-714 nm) common to the three excitations was then considered. The common part of three recorded data sets were augmented row-wise and analyzed using MCR-ALS under non-negativity in both modes and trilinearity in spectral mode. Under these new conditions, only six components have been extracted with MCR-ALS with a mean LOF value of 4% and a mean R^2 value of 0.99. Their presence and absence were exactly similar to the ones given in Table 1 except for component G which was absent. In fact, component G was

effectively present in grain sections but the reduction of the spectral domain prior MCR in this second scenario has led to the loss of specific spectral information. MCR-ALS extracted spectral profiles under non-negativity and trilinearity constraints of the six components A-F are shown in figure 8. Generally speaking, we can observe that for a given component, extracted spectral profiles are more consistent between stages than the ones in Figure 6. Indeed, the trilinearity constraint drastically decreases rotational ambiguity of solutions.

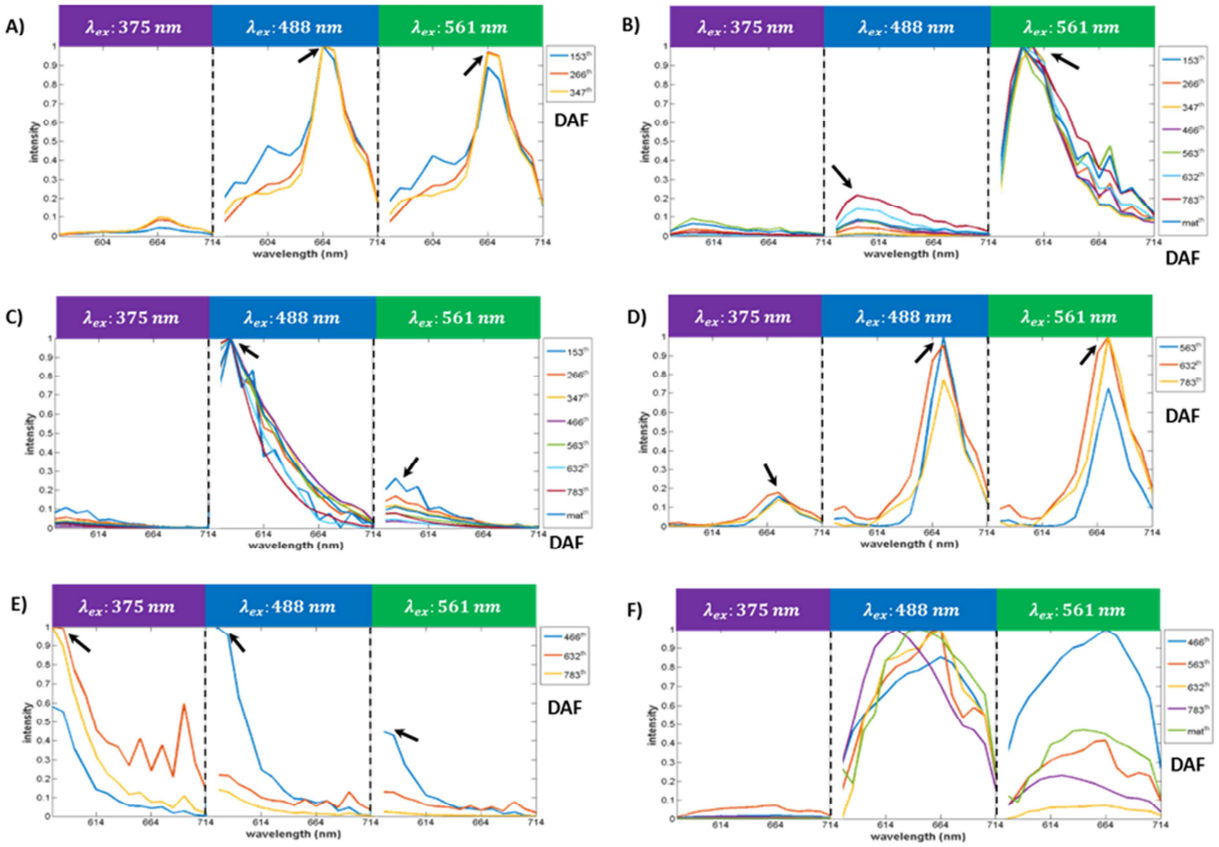


Figure 8. Extracted spectral profiles of component A-F when trilinearity and non-negativity constraints are applied.

Moreover maxima of emission of components A, B and D are exactly the same as the ones observed in Figure 6. It is slightly different for component C with a maximum of emission now observed at 584 nm while it was 544 nm with non-negativity only. However this difference is not due to a real spectral shift of this band. Indeed the maximum emission of this component is

always centered on the same wavelength but the restricted spectral domain imposed by the use of trilinearity constraint only allows us to extract the bottom of this band. We observe the same situation for component E with a maximum of emission at 474 nm with the application of non-negativity constraint becoming 574 nm with trilinearity. As already pointed out, emission profile of component F changes throughout wheat grain development certainly due to different chemical environments. Figure 9 shows extracted contribution maps when trilinearity constraint is applied.

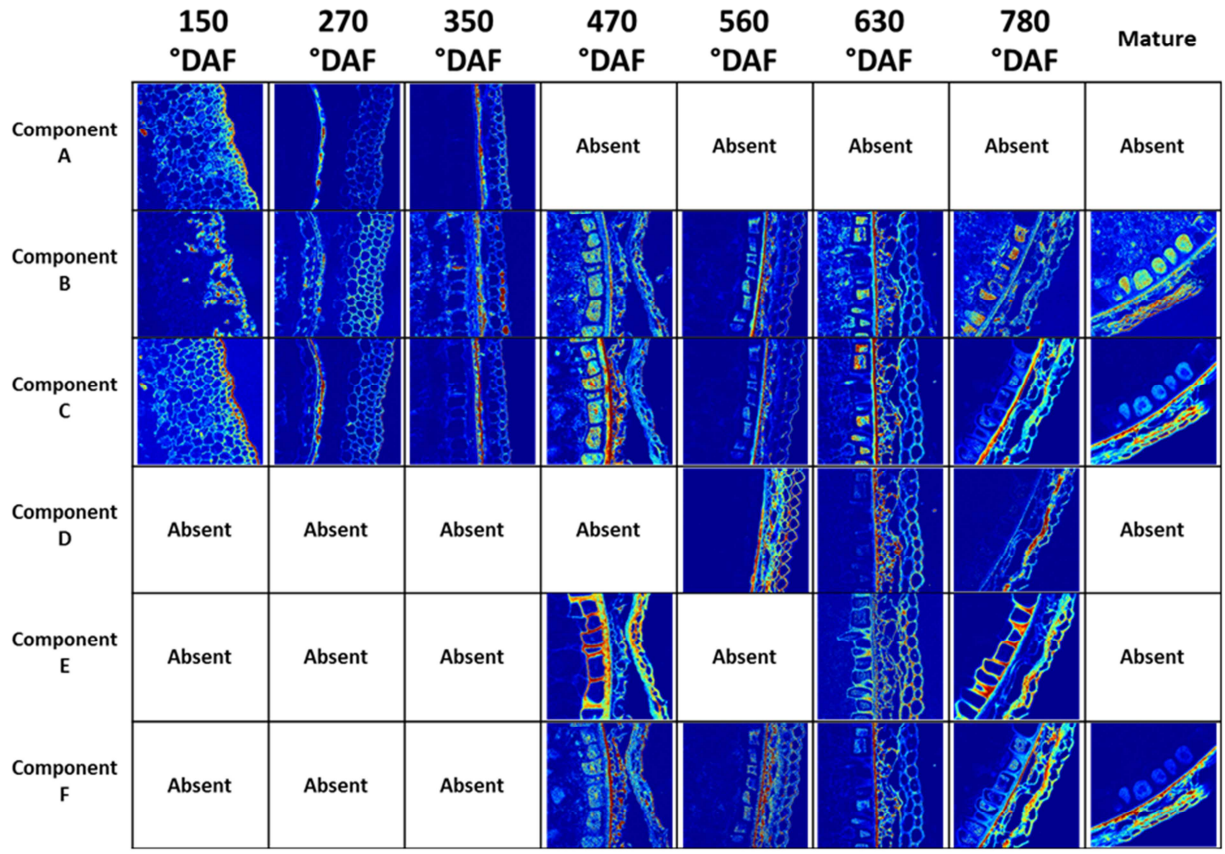


Figure 9. Extracted contribution maps of all components in the wheat grain when trilinearity and non-negativity constraints are applied. Minimum and maximum values in each contribution maps are represented with cold and hot colors respectively.

Compared with Figure 7, we can see that the localization of all components for all stages are almost the same as it was with the first scenario, except for component G which is naturally

absent when trilinearity is considered. However a deeper analysis (particularly by zooming in on concentration maps) highlights better contrasted images when trilinearity constraint is applied. Indeed in these new conditions, rotational ambiguity is decreased and purer concentration profiles can be extracted. In other words, more zones with absence of components are observed. Less ambiguity in spectral profiles allows us to generate less biased images of the biological samples. From a data analysis perspective, it is obvious that applying the trilinearity constraint is a good way to extract better spectral profiles and corresponding contribution maps, component G being absent only because of our specific instrumental setup.

4. Conclusion

This work demonstrates the great potential of multi-excitation hyperspectral autofluorescence imaging for the exploration of complex biological samples. In the present work, fluorescence properties were followed after UV and visible excitations in wheat grain outer layer during development. Our instrumental setup allowed us to obtain fluorescence information from this large spectral domain with a high spatial resolution without any labelling of the samples with additional fluorophores. There is a great potential in such spectral imaging techniques for comparing set of samples by statistical approaches. A set of 40 images containing more than 10 million emission spectra was acquired to analyze wheat grains for eight development stages and one of the challenges of the work was to define a strategy of data analysis for multiset hyperspectral images. Having understood the limits of simple integration approaches on such spectral data sets, a signal unmixing technique has been evaluated. Two implementations of the MCR-ALS approach were proposed to extract pure component spectra and contribution maps according to development stages. In the first case, the full spectral range was taken into account while in the second one, trilinearity constraint was applied on the common spectral emission range of the three excitation wavelengths. In both cases, augmented data sets were obtained by merging the repetition of images of a given development stage. Pure components could be compared for the two MCR-ALS implementations except for the component found in the specific emission ranges of the UV excitation. Four to six pure components were detected depending on the stages. Components were assigned to hydroxycinnamic acids, lignin and cuticle compounds. Pigments compound were also highlighted and found in several tissues of the

wheat grain. Some compounds were difficult to interpret requiring additional investigation of the natural fluorophores encountered in plants. To the best of our knowledge, it is the first time MCR-ALS with trilinearity constraint is used to analyze simultaneously multiple hyperspectral autofluorescence data sets and extract pure spectral profiles and corresponding contributions maps of different molecules present in the sample. Despite the natural spectral overlap of chemical species and the complexity of biological samples, it is possible to unmix signals and extract more information about pure components with this original concept. With these results, we are convinced that the combination of multi-excitation hyperspectral autofluorescence imaging and MCR-ALS approach represent a real breaking point in the analysis of complex biological samples.

Acknowledgments

This work has been supported by Campus France (PHC Gundishapur N°40890RA).

References

(1) Dokken, K. M.; Davis, L. C.; Marinkovic, N. S. Use of Infrared Microspectroscopy in Plant Growth and Development. *Applied Spectroscopy Reviews* **2005**, *40* (4), 301–326. https://doi.org/10.1080/05704920500230898.

(2) Donaldson, L.; Radotić, K.; Kalauzi, A.; Djikanović, D.; Jeremić, M. Quantification of Compression Wood Severity in Tracheids of Pinus Radiata D. Don Using Confocal Fluorescence Imaging and Spectral Deconvolution. *Journal of Structural Biology* **2010**, *169* (1), 106–115. https://doi.org/10.1016/j.jsb.2009.09.006.

(3) García-Plazaola, J. I.; Fernández-Marín, B.; Duke, S. O.; Hernández, A.; López-Arbeloa, F.; Becerril, J. M. Autofluorescence: Biological Functions and Technical Applications. *Plant Science* **2015**, *236*, 136–145. https://doi.org/10.1016/j.plantsci.2015.03.010.

(4) Gierlinger, N.; Schwanninger, M. The Potential of Raman Microscopy and Raman Imaging in Plant Research. *Spectroscopy* **2007**, *21* (2), 69–89. https://doi.org/10.1155/2007/498206.

(5) Gierlinger, N. Revealing Changes in Molecular Composition of Plant Cell Walls on the Micron-Level by Raman Mapping and Vertex Component Analysis (VCA). *Frontiers in Plant Science* **2014**, *5*, 306. https://doi.org/10.3389/fpls.2014.00306.

(6) Talamond, P.; Verdeil, J.-L.; Conéjéro, G. Secondary Metabolite Localization by Autofluorescence in Living Plant Cells. *Molecules* **2015**, *20* (3), 5024–5037. https://doi.org/10.3390/molecules20035024.

(7) Gillbro, T.; Cogdell, R. J. Carotenoid Fluorescence. *Chemical Physics Letters* **1989**, *158* (3–4), 312–316. https://doi.org/10.1016/0009-2614(89)87342-7.

(8) Hutzler, P.; Fischbach, R.; Heller, W.; Jungblut, T. P.; Reuber, S.; Schmitz, R.; Veit, M. Tissue Localization of Phenolic Compounds in Plants by Confocal Laser Scanning Microscopy. *Journal of Experimental Botany* **1998**, *49* (323), 953–965.

(9) Conéjéro, G.; Noirot, M.; Talamond, P.; Verdeil, J.-L. Spectral Analysis Combined with Advanced Linear Unmixing Allows for Histolocalization of Phenolics in Leaves of Coffee Trees. *Frontiers in Plant Science* **2014**, *5*. https://doi.org/10.3389/fpls.2014.00039.

(10) Knievel, D. C.; Abdel-Aal, E.-S. M.; Rabalski, I.; Nakamura, T.; Hucl, P. Grain Color Development and the Inheritance of High Anthocyanin Blue Aleurone and Purple Pericarp in Spring Wheat (Triticum Aestivum L.). *Journal of Cereal Science* **2009**, *50* (1), 113–120. https://doi.org/10.1016/j.jcs.2009.03.007.

(11) Adom, K. K.; Sorrells, M. E.; Liu, R. H. Phytochemicals and Antioxidant Activity of Milled Fractions of Different Wheat Varieties. *Journal of Agricultural and Food Chemistry* **2005**, *53* (6), 2297–2306. https://doi.org/10.1021/jf048456d.

(12) Furukawa, T.; Maekawa, M.; Oki, T.; Suda, I.; Iida, S.; Shimada, H.; Takamure, I.; Kadowaki, K. The Rc and Rd Genes Are Involved in Proanthocyanidin Synthesis in Rice Pericarp: Rc and Rd Genes for Red Rice Seed. *The Plant Journal* **2006**, *49* (1), 91–102. https://doi.org/10.1111/j.1365-313X.2006.02958.x.

(13) Fulcher, R. G.; O'Brien, T. P.; Lee, J. W. Studie on the Aleurone Layer 1.Conventional and Fluorescence Microscopy of the Cell Wall with Emphasis on Phenol-Carbohydrates Complexes in Wheat. *Australian Journal of Biological Sciences* **1972**, *25*, 23–34.

(14) Beaugrand, J.; Crônier, D.; Thiebeau, P.; Schreiber, L.; Debeire, P.; Chabbert, B. Structure, Chemical Composition, and Xylanase Degradation of External Layers Isolated from Developing Wheat Grain. *Journal of Agricultural and Food Chemistry* **2004**, *52*, 7108–7117. https://doi.org/10.1021/jf049529w.

(15) Corcel, M.; Devaux, M.-F.; Guillon, F.; Barron, C. Comparison of UV and Visible Autofluorescence of Wheat Grain Tissues in Macroscopic Images of Cross-Sections and Particles. *Computers and Electronics in Agriculture* **2016**, *127*, 281–288. https://doi.org/10.1016/j.compag.2016.06.016.
(16) Saadi, A.; Lempereur, I.; Sharonov, S.; Autran, J. C.; Manfait, M. Spatial Distribution of Phenolic Materials in Durum Wheat Grain as Probed by Confocal Fluorescence Spectral Imaging. *Journal of Cereal Science* **1998**, *28* (2), 107–114. https://doi.org/10.1006/jcrs.1998.0195.
(17) Harris, P. J.; Trethewey, J. A. K. The Distribution of Ester-Linked Ferulic Acid in the Cell Walls of Angiosperms. *Phytochemistry Reviews* **2010**, *9* (1), 19–33. https://doi.org/10.1007/s11101-009-9146-4.
(18) Jensen, S.; Munck, L.; Martens, H. The Botanical Constituents of Wheat and Wheat Milling Fractions. I. Quantification by Autofluorescence. *Cereal Chemistry* **1982**, *59*, 477–484.
(19) Baldwin, P. M.; Bertrand, D.; Novales, B.; Bouchet, B.; Collobert, G.; Gallant, D. J. Chemometric Labeling of Cereal Tissues in Multichannel Fluorescence Microscopy Images Using Discriminant Analysis. *Analytical Chemistry* **1997**, *69* (21), 4339–4348. https://doi.org/10.1021/ac970145x.
(20) Courcoux, P. Simultaneous Decomposition of Multivariate Images Using Three-Way Data Analysis Application to the Comparison of Cereal Grains by Confocal Laser Scanning Microscopy. *Chemometrics and Intelligent Laboratory Systems* **2002**, *62*, 103–113.
(21) Piqueras, S.; Duponchel, L.; Tauler, R.; de Juan, A. Monitoring Polymorphic Transformations by Using in Situ Raman Hyperspectral Imaging and Image Multiset Analysis. *Analytica Chimica Acta* **2014**, *819*, 15–25. https://doi.org/10.1016/j.aca.2014.02.027.
(22) Zhang, X.; Tauler, R. Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to Remote Sensing Hyperspectral Imaging. *Analytica Chimica Acta* **2013**, *762*, 25–38. https://doi.org/10.1016/j.aca.2012.11.043.
(23) Felten, J.; Hall, H.; Jaumot, J.; Tauler, R.; de Juan, A.; Gorzsás, A. Vibrational Spectroscopic Image Analysis of Biological Material Using Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS). *Nature Protocols* **2015**, *10* (2), 217–240. https://doi.org/10.1038/nprot.2015.008.
(24) Piqueras, S.; Duponchel, L.; Tauler, R.; de Juan, A. Resolution and Segmentation of Hyperspectral Biomedical Images by Multivariate Curve Resolution-Alternating Least Squares. *Analytica Chimica Acta* **2011**, *705* (1–2), 182–192. https://doi.org/10.1016/j.aca.2011.05.020.
(25) Amigo, J. M.; Ravn, C. Direct Quantification and Distribution Assessment of Major and Minor Components in Pharmaceutical Tablets by NIR-Chemical Imaging. *European Journal of Pharmaceutical Sciences* **2009**, *37* (2), 76–82. https://doi.org/10.1016/j.ejps.2009.01.001.
(26) Golshan, A.; Abdollahi, H.; Beyramysoltan, S.; Maeder, M.; Neymeyr, K.; Rajkó, R.; Sawall, M.; Tauler, R. A Review of Recent Methods for the Determination of Ranges of Feasible Solutions Resulting from Soft Modelling Analyses of Multivariate Data. *Analytica Chimica Acta* **2016**, *911*, 1–13. https://doi.org/10.1016/j.aca.2016.01.011.
(27) Rajkó, R.; Abdollahi, H.; Beyramysoltan, S.; Omidikia, N. Definition and Detection of Data-Based Uniqueness in Evaluating Bilinear (Two-Way) Chemical Measurements. *Analytica Chimica Acta* **2015**, *855*, 21–33. https://doi.org/10.1016/j.aca.2014.12.017.

(28) Omidikia, N.; Beyramysoltan, S.; Mohammad Jafari, J.; Tavakkoli, E.; Akbari Lakeh, M.; Alinaghi, M.; Ghaffari, M.; Khodadadi Karimvand, S.; Rajkó, R.; Abdollahi, H. Closure Constraint in Multivariate Curve Resolution: Closure Constraint in MCR-ALS. *Journal of Chemometrics* **2017**, e2975. https://doi.org/10.1002/cem.2975.
(29) Beyramysoltan, S.; Abdollahi, H.; Rajkó, R. Newer Developments on Self-Modeling Curve Resolution Implementing Equality and Unimodality Constraints. *Analytica Chimica Acta* **2014**, *827*, 1–14. https://doi.org/10.1016/j.aca.2014.03.019.
(30) Beyramysoltan, S.; Rajkó, R.; Abdollahi, H. Investigation of the Equality Constraint Effect on the Reduction of the Rotational Ambiguity in Three-Component System Using a Novel Grid Search Method. *Analytica Chimica Acta* **2013**, *791*, 25–35. https://doi.org/10.1016/j.aca.2013.06.043.
(31) Ghaffari, M.; Abdollahi, H. Duality Based Interpretation of Uniqueness in the Trilinear Decompositions. *Chemometrics and Intelligent Laboratory Systems* **2018**, *177*, 17–25. https://doi.org/10.1016/j.chemolab.2018.04.007.
(32) de Juan, A.; Rutan, S. C.; Tauler, R.; Massart, D. L. Comparison between the Direct Trilinear Decomposition and the Multivariate Curve Resolution-Alternating Least Squares Methods for the Resolution of Three-Way Data Sets. *Chemometrics and Intelligent Laboratory Systems* **1998**, *40* (1), 19–32. https://doi.org/10.1016/S0169-7439(98)00003-3.
(33) de Juan, A.; Jaumot, J.; Tauler, R. Multivariate Curve Resolution (MCR). Solving the Mixture Analysis Problem. *Anal. Methods* **2014**, *6* (14), 4964–4976. https://doi.org/10.1039/C4AY00571F.
(34) Saiyed, I. M.; Bullock, P. R.; Sapirstein, H. D.; Finlay, G. J.; Jarvis, C. K. Thermal Time Models for Estimating Wheat Phenological Development and Weather-Based Relationships to Wheat Quality. *Canadian Journal of Plant Science* **2009**, *89* (3), 429–439. https://doi.org/10.4141/CJPS07114.
(35) Chateigner-Boutin, A.-L.; Lapierre, C.; Alvarado, C.; Yoshinaga, A.; Barron, C.; Bouchet, B.; Bakan, B.; Saulnier, L.; Devaux, M.-F.; Girousse, C.; et al. Ferulate and Lignin Cross-Links Increase in Cell Walls of Wheat Grain Outer Layers during Late Development. *Plant Science* **2018**, *276*, 199–207. https://doi.org/10.1016/j.plantsci.2018.08.022.
(36) Tauler, R. Multivariate Curve Resolution Applied to Second Order Data. *Chemometrics and Intelligent Laboratory Systems* **1995**, *30* (1), 133–146. https://doi.org/10.1016/0169-7439(95)00047-X.
(37) Tauler, R.; Smilde, A.; Kowalski, B. Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution. *Journal of Chemometrics* **1995**, *9* (1), 31–58. https://doi.org/10.1002/cem.1180090105.
(38) Cattell, R. B. Parallel Proportional Profiles and Other Principles for Determining the Choice of Factors by Rotation. *Psychometrika* **1944**, *9* (4), 267–283.
(39) Stegeman, A.; Sidiropoulos, N. D. On Kruskal's Uniqueness Condition for the Candecomp/Parafac Decomposition. *Linear Algebra and its Applications* **2007**, *420* (2–3), 540–552. https://doi.org/10.1016/j.laa.2006.08.010.
(40) Culzoni, M. J.; Goicoechea, H. C.; Ibañez, G. A.; Lozano, V. A.; Marsili, N. R.; Olivieri, A. C.; Pagani, A. P. Second-Order Advantage from Kinetic-Spectroscopic Data Matrices in the Presence of Extreme Spectral Overlapping. *Analytica Chimica Acta* **2008**, *614* (1), 46–57. https://doi.org/10.1016/j.aca.2008.03.013.

(41) Windig, W.; Stephenson, D. A. Self-Modeling Mixture Analysis of Second-Derivative near-Infrared Spectral Data Using the SIMPLISMA Approach. *Analytical Chemistry* **1992**, *64* (22), 2735–2742. https://doi.org/10.1021/ac00046a015.
(42) Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11* (1), 137–148. https://doi.org/10.1080/00401706.1969.10490666.
(43) de Juan, A.; Tauler, R. Factor Analysis of Hyphenated Chromatographic Data. *Journal of Chromatography A* **2007**, *1158* (1–2), 184–195. https://doi.org/10.1016/j.chroma.2007.05.045.
(44) Lachman, J.; Martinek, P.; Kotíková, Z.; Orsák, M.; Šulc, M. Genetics and Chemistry of Pigments in Wheat Grain – A Review. *Journal of Cereal Science* **2017**, *74*, 145–154. https://doi.org/10.1016/j.jcs.2017.02.007.
(45) Donaldson, L. Softwood and Hardwood Lignin Fluorescence Spectra of Wood Cell Walls in Different Mounting Media. *IAWA Journal* **2013**, *34* (1), 3–19. https://doi.org/10.1163/22941932-00000002.
(46) Morrison, I. N. Ultrastructure of the Cuticular Membranes of the Developing Wheat Grain. *Revue canadienne de botanique* **1975**, *53* (18), 2077–2087.
(47) Fernández, S.; Osorio, S.; Heredia, A. Monitoring and Visualising Plant Cuticles by Confocal Laser Scanning Microscopy. *Plant Physiology and Biochemistry* **1999**, *37* (10), 789–794. https://doi.org/10.1016/S0981-9428(00)86692-9.
(48) Baker, E. Phenolic Constituents of Tomato Fruit Cuticles. *Phytochemistry* **1980**, *19*, 1415–1419.
(49) Antoine, C.; Peyron, S.; Mabille, F.; Lapierre, C.; Bouchet, B.; Abecassis, J.; Rouau, X. Individual Contribution of Grain Outer Layers and Their Cell Wall Structure to the Mechanical Properties of Wheat Bran. *Journal of Agricultural and Food Chemistry* **2003**, *51* (7), 2026–2033. https://doi.org/10.1021/jf0261598.