

# Extraction des signatures de composés purs constituant la couche externe du grain d'orge à partir d'images de fluorescence

Abdelkamel ELHAFID<sup>1</sup>, Danielle NUZILLARD<sup>1</sup>, Marie-Françoise DEVAUX<sup>2</sup>, Nicolas PETROCHILOS<sup>1</sup>, Fabien BELLOIR<sup>1</sup>

<sup>1</sup>CReSTIC, UFR Sciences Exactes et Naturelles, B.P. 1039, 51687 REIMS Cedex 2

<sup>2</sup>INRA URPOI, Rue de la Géraudière BP 71627 44316 Nantes Cedex 03, France  
abdelkamel.elhafid@univ-reims.fr, danielle.nuzillard@univ-reims.fr,  
nicolas.petrochilos@univ-reims.fr, fabien.belloir@univ-reims.fr, devaux@nantes.inra.fr

**Résumé** – L'identification des différents types de tissus après fragmentation ou transformation des grains de céréale et de leurs dérivés est un enjeu majeur de l'industrie agro-alimentaire. L'objectif de notre travail est l'extraction des signatures spectrales des principaux composés purs des tissus de la couche externe d'un grain d'orge et l'estimation de leur carte de répartition. Les données sont disponibles sous forme d'un cube constitué d'une pile d'images acquises par microscopie confocale en fluorescence. Elles sont traitées soit en considérant le cube dans sa globalité, soit en sélectionnant visuellement des pixels selon des lignes de coupe dans chaque tissu identifiable. Le traitement met en oeuvre des méthodes de séparation aveugle de sources, il tient compte de l'autocorrélation des données, de la positivité des signatures et de leur carte de répartition spatiale. Des signatures de référence permettent de valider les résultats.

**Abstract** – Automatic identification of cereal grain tissues is studied with the objective to follow them after fragmentation or other transformations. In our work, we aim at extracting the spectral signature of the various pure components present in the tissue of the external layer of the barley grain. As well, we desire to estimate their concentration mapping. The data is available in the form of a cube (3-th order tensor), which each layer is an image measured by fluorescence confocal microscopy. We process either the data cube as a whole, or by selecting a slice along the spectral dimension. The processing is done by blind source separation methods, using together the inner correlation of the data, and the positivity of the data model (spectral signature, and concentration mapping). Reference signatures allow an experimental confirmation.

## 1 Description du cube de données et du modèle associé

### 1.1 Les données

Les grains de céréale sont constitués de plusieurs tissus qui se superposent. La partie centrale renferme de l'amidon, les tissus des couches externes servent de protection (figure 1). Ceux-ci contiennent des composés naturellement fluorescents : cutine, acide férulique, lignine, dont les signatures spectrales se recouvrent partiellement.

L'identification de ces différents composés ainsi que leur carte de répartition est très importante puisqu'elle peut permettre de suivre les différents types de tissus après fragmentation ou transformation des grains et de leurs dérivés.

Le cube de données utilisé est constitué de 19 images multispectrales acquises en microscopie confocale de fluorescence (figure 2). Chaque image est le résultat de la mesure de l'autofluorescence des parois végétales d'un grain d'orge pour une longueur d'onde d'excitation et un filtre d'émission donné. Chaque point de mesure est un spectre de mélange de composés purs avec différentes proportions.

A partir de ces images spectrales de mélange l'objec-

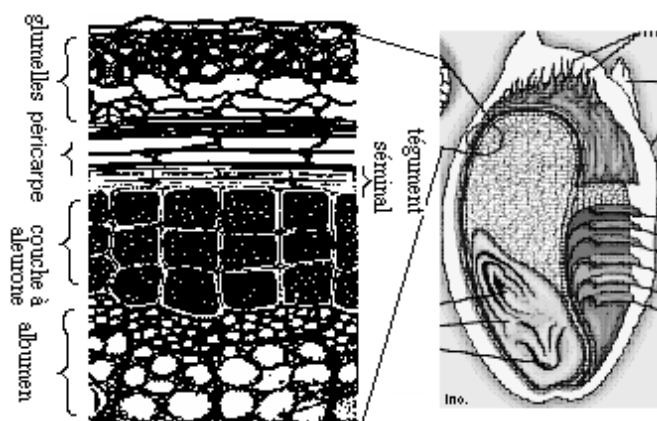


FIG. 1: Coupe transversale d'un grain d'orge.

tif est d'extraire les profils spectraux des composés purs ainsi que leur carte de répartition. On procède de deux manières, soit en séparant les 19 images placées dans une matrice d'observation  $\mathbf{X}$  de taille  $(256 \times 512 \times 19)$  (figure 2), soit en considérant un ensemble de 20 pixels sélectionnés selon des lignes de coupe dans les tissus externes formant une matrice d'observation  $\mathbf{X}$  de taille  $(20 \times 19)$  (figure 3).

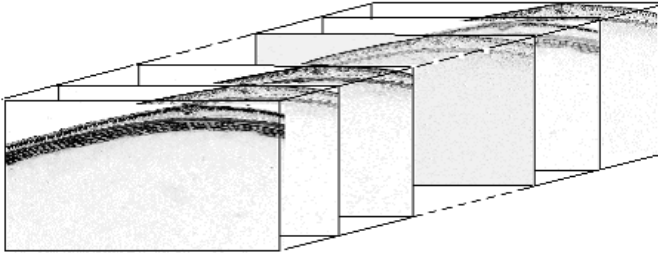


FIG. 2: *Cube d'images multispectrales du grain d'orge.*

## 1.2 Analyse des images

Parmi les techniques utilisées avec ce type de données, la méthode introduite par P. Courcoux *et al.* [1] basée sur l'analyse en composantes principales (ACP) des images de données, a permis de révéler les différents tissus fluorescents à partir de cinq composantes principales. Chaque composante met en évidence de manière contrastée plusieurs tissus simultanément : par exemple la couche à aleurone en positif et le péricarpe en négatif. La méthode ne fournit pas les profils spectraux des composés purs. Notre objectif est de déterminer ces profils spectraux et d'améliorer la reconnaissance de la répartition des tissus. A l'aide d'une méthode de séparation aveugle de sources (SAS), suivie par un post-traitement tenant compte des contraintes physiques des données il est possible d'obtenir des solutions physiquement correctes.

Afin de réduire le bruit de mesure, il est judicieux de ne retenir que des régions d'intérêt ne contenant que les tissus externes.

La qualité du microscope mis en oeuvre est suffisamment bonne pour négliger la déformation introduite par la fonction de transfert instrumentale. Sachant que les lois physiques de la spectroscopie sont linéaires, les images spectrales enregistrées sont une somme pondérée des images des composés purs. Le choix du modèle instantané et linéaire couramment utilisé en SAS, est justifié. On suppose :

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

où  $\mathbf{X}$  est la matrice d'observation construite à partir du cube de données. Par la suite chaque image est reformattée pour placer les éléments de la matrice des données en ligne. On note que tous les éléments de  $\mathbf{X}$ ,  $\mathbf{A}$  et  $\mathbf{S}$  sont à valeurs positives.  $\mathbf{A}$  est la matrice de mélange où chaque colonne représente la signature d'un composé pur et  $\mathbf{S}$  est la matrice source qui représente l'image de leurs concentrations. Vu la structure des couches externes du grain d'orge, les cartes de concentrations des composés purs présentent une corrélation spatiale non nulle.

Puisque la plupart des molécules fluorescentes ont des bandes fluorescentes larges et non structurées et que les valeurs des intensités des spectres sont positives, l'indépendance mathématique des sources ne peut pas être vérifiée. Ainsi les méthodes basées sur les statistiques d'ordre supérieur, type ICA [2, 3], ne sont pas appropriées. Elles fournissent néanmoins des résultats intéressants. Par contre la prise en compte de la corrélation intrinsèque des profils de concentration et leur indépendance extrinsèque est plus réaliste, ce qui justifie l'utilisation de l'algorithme (SOBI)

[4]. Cependant celui-ci ne suffit pas à donner une solution finale. Il est nécessaire de forcer la positivité par une procédure itérative.

## 1.3 Analyse des spectres

Une seconde approche consiste à limiter l'observation à un ensemble de pixels sélectionnés selon des lignes de coupe dans les tissus externes du grain d'orge. A chaque pixel de cette ligne correspond un vecteur spectral exprimé en longueur d'onde, ( $\lambda = c/\nu$ ), qui constitue une ligne de la matrice d'observation  $\mathbf{X}$  (figure 3). Dans ce cas les sources recherchées  $\mathbf{S}$  sont les signatures des composés purs et la matrice  $\mathbf{A}$  est une matrice de proportion de ces composés purs dans les vecteurs de mélanges observés. Les hypothèses de mélange instantané et linéaire restent valables.

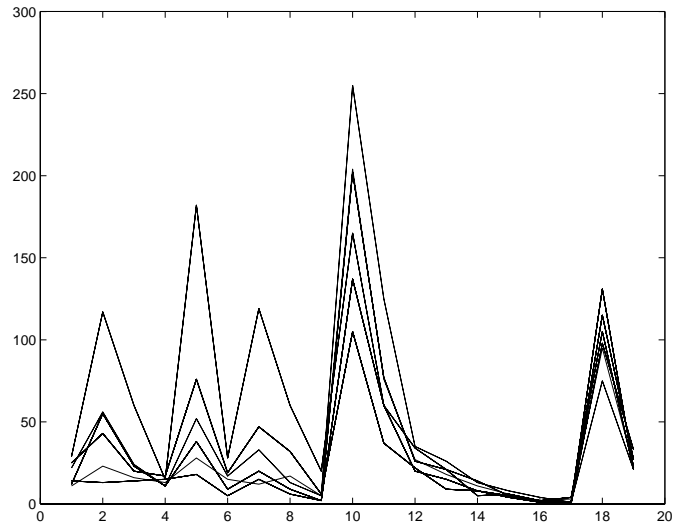


FIG. 3: *Profils de mélange extraits pour des pixels sélectionnés selon une ligne de coupe choisie dans la région du tégument séminal.*

Pour cette approche spectrale, les données étant parcimonieuses (figure 3), on applique l'algorithme "Non-negative matrix factorization" (NMF) [5], qui consiste à factoriser la matrice  $\mathbf{X}$  en deux matrices positives  $\mathbf{A}$  et  $\mathbf{S}$ .

## 2 Les algorithmes utilisés

### 2.1 SOBI positif

Une étude préliminaire a montré que l'utilisation des algorithmes utilisant les statistiques d'ordre supérieur ne fournit pas de "bons résultats spectraux" du fait du faible nombre de points de mesure dont on dispose (19 points de mesure spectrale). Notre choix s'est porté sur l'algorithme SOBI [4] qui a déjà été utilisé avec succès dans le domaine des signaux spectraux [6, 7] même avec un faible nombre de points de mesure. Cet algorithme est ensuite suivi par une procédure itérative dite de positivité pour prendre en compte la réalité physique des données [6].

Cette procédure répète les étapes suivantes :

1. Les valeurs négatives des sources  $\hat{\mathbf{S}}_n$  sont remplacées par des zéros.
2. La matrice de mélange  $\hat{\mathbf{A}}_{n+1}$  est évaluée par pseudo-inversion à droite de  $\hat{\mathbf{S}}_n$  :

$$\hat{\mathbf{A}}_{n+1} = \mathbf{X} \hat{\mathbf{S}}_n^T (\hat{\mathbf{S}}_n \hat{\mathbf{S}}_n^T)^{-1} \quad (2)$$

3. Les valeurs négatives de la matrice  $\hat{\mathbf{A}}_{n+1}$  sont remplacées par des zéros.
4. La matrice des sources  $\hat{\mathbf{S}}_{n+1}$  est évaluée par pseudo-inversion à gauche de  $\hat{\mathbf{A}}_{n+1}$  :

$$\hat{\mathbf{S}}_{n+1} = (\hat{\mathbf{A}}_{n+1}^T \hat{\mathbf{A}}_{n+1})^{-1} \hat{\mathbf{A}}_{n+1}^T \mathbf{X} \quad (3)$$

5. Tant que le couple  $(\hat{\mathbf{S}}_{n+1}, \hat{\mathbf{A}}_{n+1})$  évolue, on itère une nouvelle fois.

## 2.2 Factorisation en matrice non négative (NMF)

L'algorithme de factorisation en matrice non négative [5] est un algorithme itératif, qui permet de factoriser une matrice de donnée positive (tous ses éléments sont positifs)  $\mathbf{X}$  en deux matrices positives  $\mathbf{A}$  et  $\mathbf{S}$  en minimisant une fonction coût par rapport à  $\mathbf{A}$  et  $\mathbf{S}$  sous les contraintes  $\forall (i,j) \in \mathbb{R}^2, \mathbf{A}_{ij} \geq 0$ , et  $\mathbf{S}_{ij} \geq 0$ . Dans notre cas on choisit le carré de la distance Euclidienne entre  $\mathbf{X}$  et le produit  $\mathbf{AS}$ ,  $\|\mathbf{X} - \mathbf{AS}\|_{\text{FRO}}^2$ . De multiples techniques de minimisation sont possibles, mais afin de respecter le critère de positivité, Lee et al proposent d'utiliser les mises à jour suivantes jusqu'à la convergence :

$$\mathbf{S}_{ij} \leftarrow \mathbf{S}_{ij} \frac{(\mathbf{A}^T \mathbf{X})_{ij}}{(\mathbf{A}^T \mathbf{AS})_{ij}} \quad (4)$$

$$\mathbf{A}_{ij} \leftarrow \mathbf{A}_{ij} \frac{(\mathbf{XS}^T)_{ij}}{(\mathbf{ASS}^T)_{ij}} \quad (5)$$

Il a été démontré dans [5] que ces mises à jour font varier la fonction coût de manière non strictement décroissante jusqu'à atteindre un minimum (global ou local).

On note que par construction (4) et (5) ne peuvent donner que des valeurs positives pour  $\mathbf{S}$  et  $\mathbf{A}$  à partir de valeurs positives.

L'application de cet algorithme sur des données de céréales de blé ayant les mêmes caractéristiques que l'orge a fourni des résultats très satisfaisants [8].

## 3 Résultats

### 3.1 SOBI positif

L'algorithme SOBI a été appliqué sur le cube des 19 images, en prenant quatre matrices de corrélation des données. L'expérience montre que le choix optimum du nombre de ces matrices est compris entre 4 et 8. Cela a permis d'extraire quatre sources décorréées qui mettent en évidence des régions opposées et présentent des valeurs négatives. Ensuite la procédure de positivité a été itérée jusqu'à la convergence.

De plus les signatures des composés de référence (figure 4) ont été obtenues expérimentalement à partir d'autres matériaux biologiques caractérisés à l'INRA de Nantes. Ces signatures permettent de comparer l'estimation des signatures à valeurs positives (figure 5) et d'accepter les cartes de répartition correspondantes (figure 6).

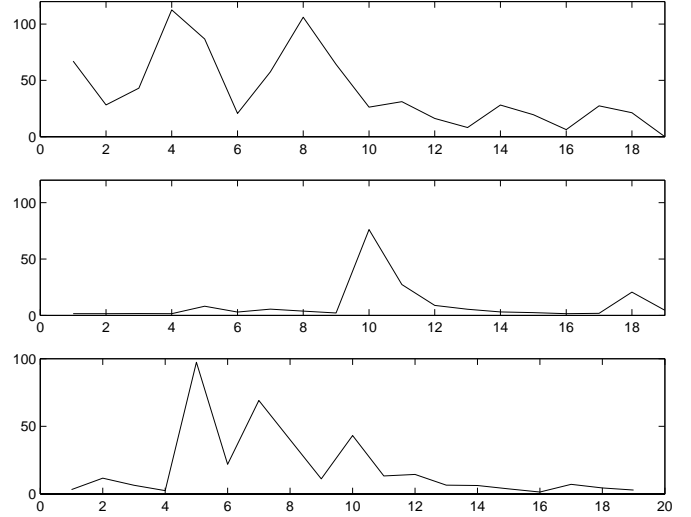


FIG. 4: Signatures de référence des composés purs (cutine, acide férulique et lignine).

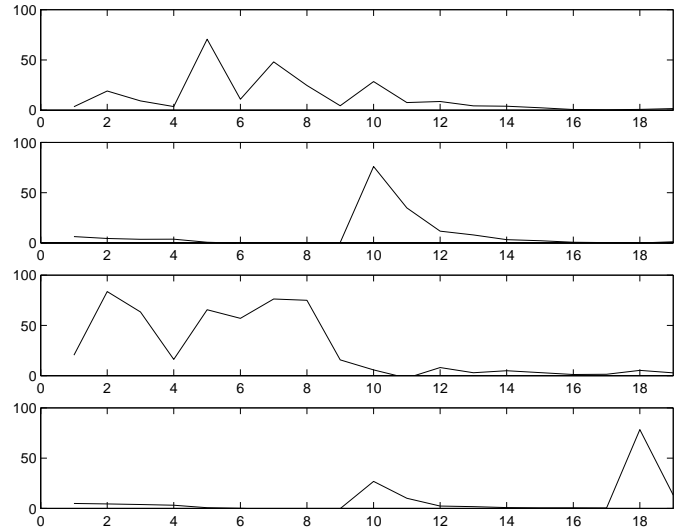


FIG. 5: Estimation des signatures des composés purs par l'algorithme SOBI + positivité: (a) lignine, (b)+(d) acide férulique, (c) cutine.

Le profil (5.a) correspond à un profil de type lignine. Le graphe (5.c) correspond à la signature de la cutine. L'image (figure 6.a) correspond à la répartition de la lignine, celle-ci est obtenue dans l'ensemble des tissus péricarpe+glumelle. L'image (figure 6.c) correspond à la cutine et révèle la région la plus externe du grain et de manière moins intense les glumelles et le péricarpe. Une combinaison linéaire des graphes des figures (5.b) et (5.d) correspond à la signature de l'acide férulique visible sur le graphe (4.b). La région correspond à la couche à aleurone et aux glumelles.

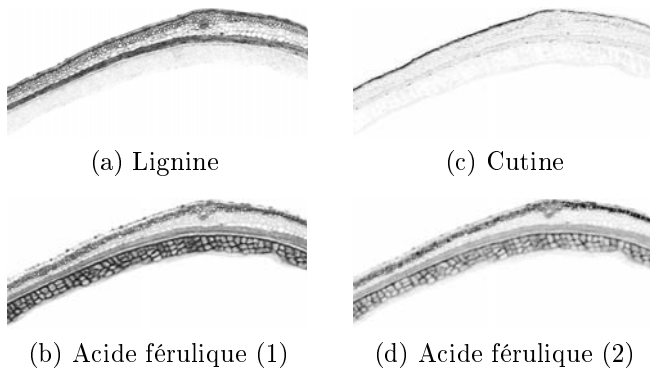


FIG. 6: Estimation des concentrations des composés purs par la méthode SOBI + positivité.

Les cartes de concentration associées aux composés purs (figure 5) sont présentées sur la figure 6 en respectant l'ordre de la figure 5. Les deux images (figure 6.b) et (figure 6.d) sont fortement corrélées puisqu'elles correspondent à la concentration d'un seul composé pur : l'acide férulique.

### 3.2 Factorisation en matrice non négative (NMF)

Pour chaque tissu identifiable (péricarpe, couche à aleurone, cutine, glumelle), des pixels ont été sélectionnés visuellement selon une ligne de coupe. Dans l'exemple présenté ci-dessous, les données à séparer sont des vecteurs appartenant à la région du tégument séminal (figure 3).

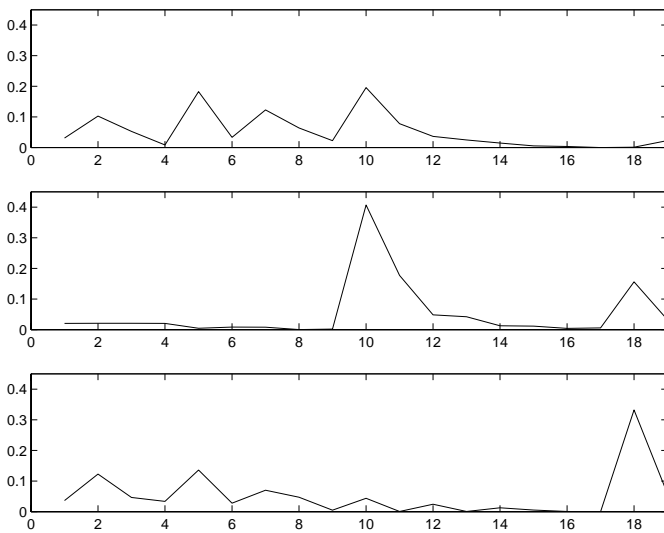


FIG. 7: Estimation des signatures des composés purs du tégument séminal par la méthode NMF : (a) lignine, (b) acide férulique, (c) résidus.

Le traitement du tégument séminal met en évidence des profils de type lignine (figure 7.a), acide férulique (figure 7.b) et un résidu en (figure 7.c). La cutine n'étant pas présente dans le jeu de points choisi, il est normal de ne pas la retrouver après séparation. L'algorithme NMF a pu être appliqué ici car le jeu de données est réduit. Il est difficilement applicable directement sur les images du

cube car il ne prend pas en compte l'autocorrélation des données et il introduit des indéterminations aux endroits où l'image ne contient qu'un faible signal.

## 4 Conclusions et perspectives

Ce travail a porté sur l'analyse d'un cube de données multispectrales d'une section d'un grain d'orge pour en extraire les composés purs ainsi que leur carte de répartition. La combinaison linéaire des composés purs permet de remonter aux tissus.

La prise en compte de la positivité des données aide à fournir des résultats physiquement interprétables. Ainsi le cube a été considéré dans son ensemble et les images ont été séparées grâce à l'algorithme SOBI + positivité de manière à fournir les cartes de répartition des composés purs ainsi que leurs signatures spectrales. Parallèlement, les données sélectionnées suivant une ligne de coupe ont été séparées en vecteurs spectraux dans le même but.

Cette méthodologie sera appliquée à un ensemble de nouvelles observations d'images de grains broyés dans le but de comprendre les mécanismes de fragmentation des tissus.

## Références

- [1] P. Courcoux, M.-F. Devaux, and B. Bouchet, "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*, vol. 62, pp. 103–113, Mai 2002.
- [2] P. Common, "Independent component analysis, a new concept?," *Signal Processing, Special Issue on High-Order Statistics*, vol. 36, pp. 287–314, Avril 1996.
- [3] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, pp. 161–164, Janvier 1996.
- [4] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. on Signal Processing*, vol. 45, pp. 434–444, Février 1997.
- [5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 434–444, Octobre 1999.
- [6] D. Nuzillard and J. Nuzillard, "Blind source separation applied to nonorthogonal signals," in *1st Workshop ICA '99*, (Aussois, (France)), pp. 25–30, 11-15 Janvier 1999.
- [7] D. Nuzillard, S. Bourg, and J. Nuzillard, "Model-free analysis of mixture by NMR," *Journal of Magnetic Resonance*, vol. 133, pp. 358–363, août 1998.
- [8] C. Gobinet, A. Elhafid, V. Vrabie, R. Huez, and D. Nuzillard, "About importance of positivity constraints for source separation in fluorescence spectroscopy," in *Eusipco '05*, (Antalya, (Turkey)), 5 Septembre 2005.