



Méthodes non-supervisées pour le traitement d'images hyperspectrales

B. Jaillais¹, K. Meghar¹, L. Nus², S. Miron²,
D. Brie², S.Moussaoui³

¹ StatSC Unité de Statistique, Sensométrie, Chimiométrie, F44300 Nantes.

² Université de Lorraine, CRAN, UMR 7039, F54506 Vandœuvre-lès-Nancy.

³ Ecole Centrale Nantes, LS2N, F-44321 Nantes.

Objectifs

- Description des deux projets :
 - ❖ ALOHA : Analyse en Ligne de données Hyperspectrales pour l'industrie Agroalimentaire
 - ❖ ANR-OPTIFIN : Agence Nationale de la Recherche-OPTimisation des FINitions
- But :
 - ❖ Développer des outils analytiques adaptés à l'analyse en ligne à haut débit d'échantillons pour le traitement d'images hyperspectrales
- Méthodes utilisées :
 - ❖ Analyse en Composantes Principales (ACP), K-means, Factorisation en matrices non-négatives (NMF) en-ligne

Matériels et acquisition des données

- Description des échantillons de bois
- Système d'imagerie hyperspectrale proche infrarouge
- Imageur pushbroom :
 - ❖ Gamme spectrale entre 1000 et 2500 nm
 - ❖ Image de dimensions 288*384*1200 (288 : nombre de longueurs d'onde. 384*1200 : dimensions spatiales)



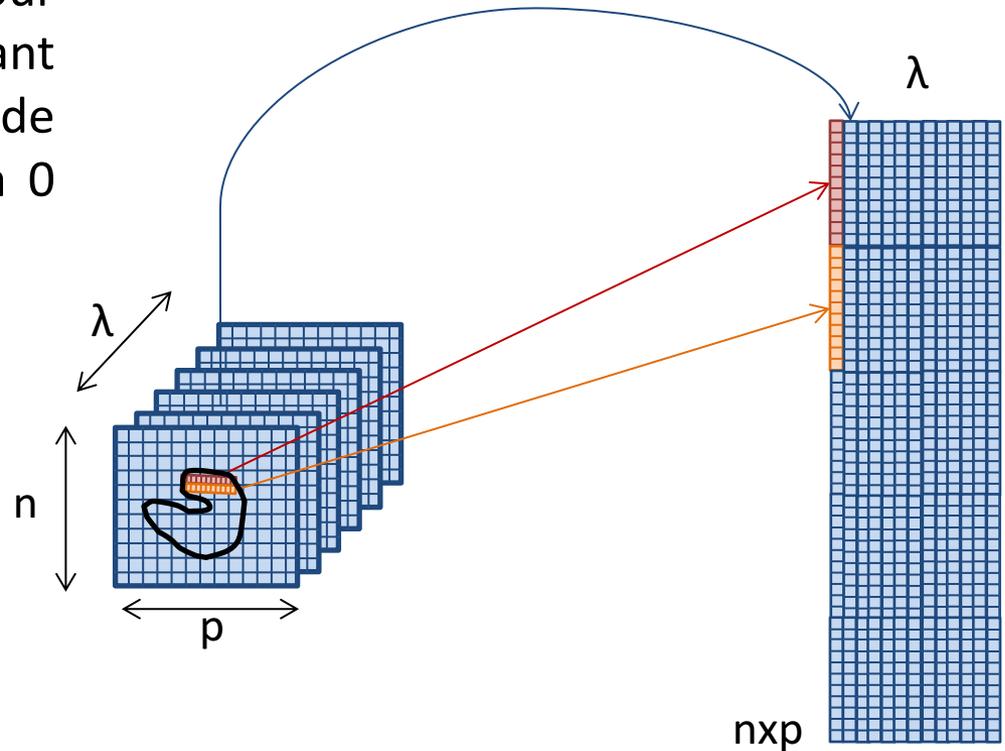
Pré-traitement des données

Sélection ROI

Une segmentation manuelle est effectuée sur l'image moyennée pour chaque cube de données, résultant en une image binaire (valeurs de pixels égales à 1 pour le ROI et à 0 pour le fond).

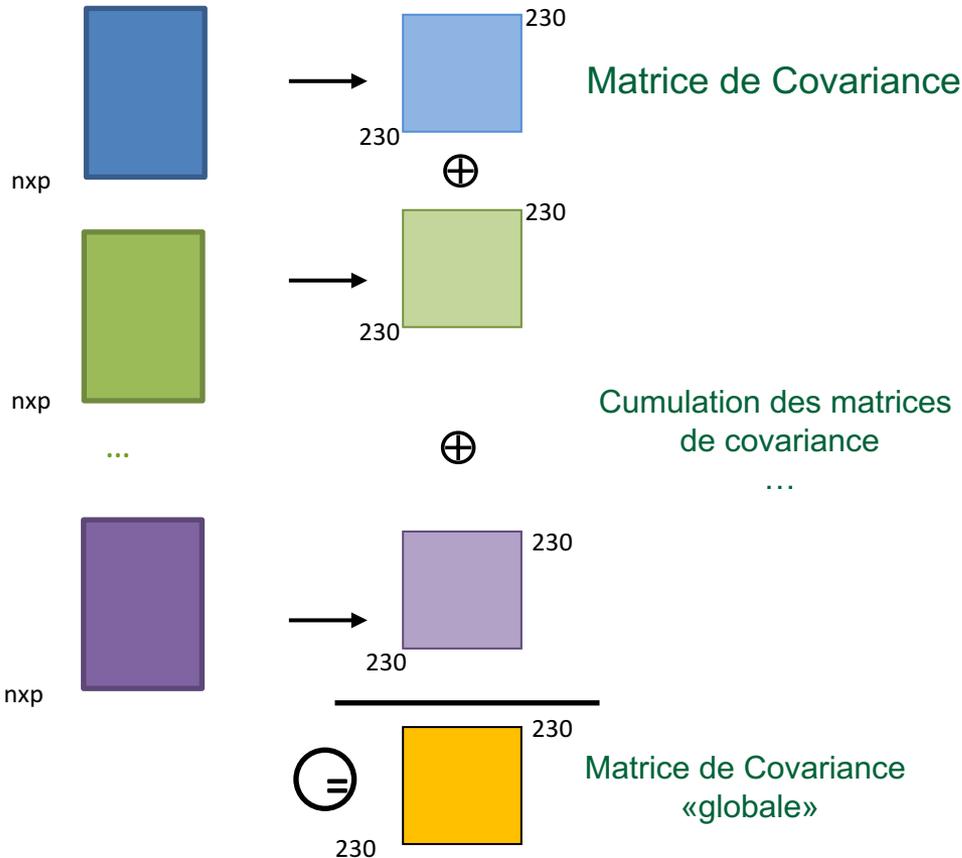


Hypercube déplié

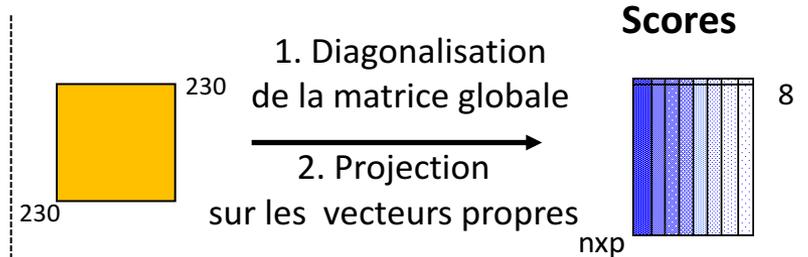


Méthode ACP concaténée

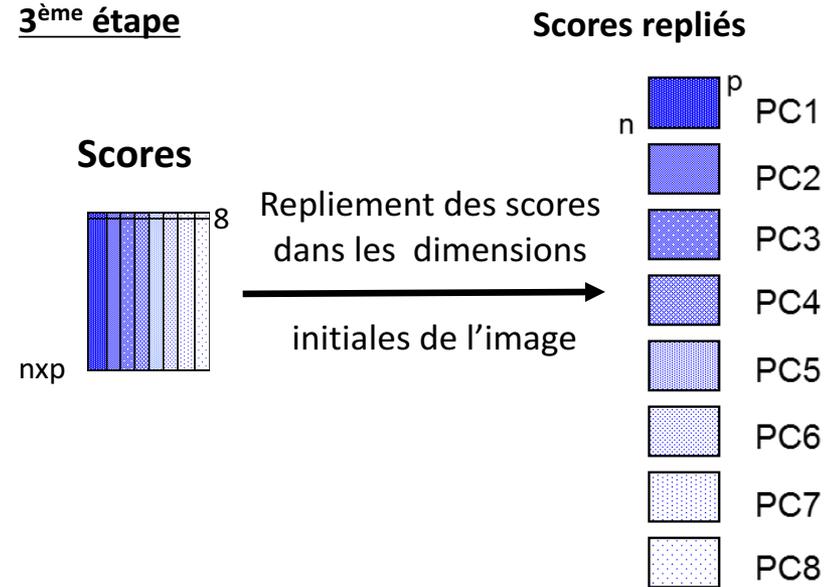
Cubes images dépliés



2^{ème} étape



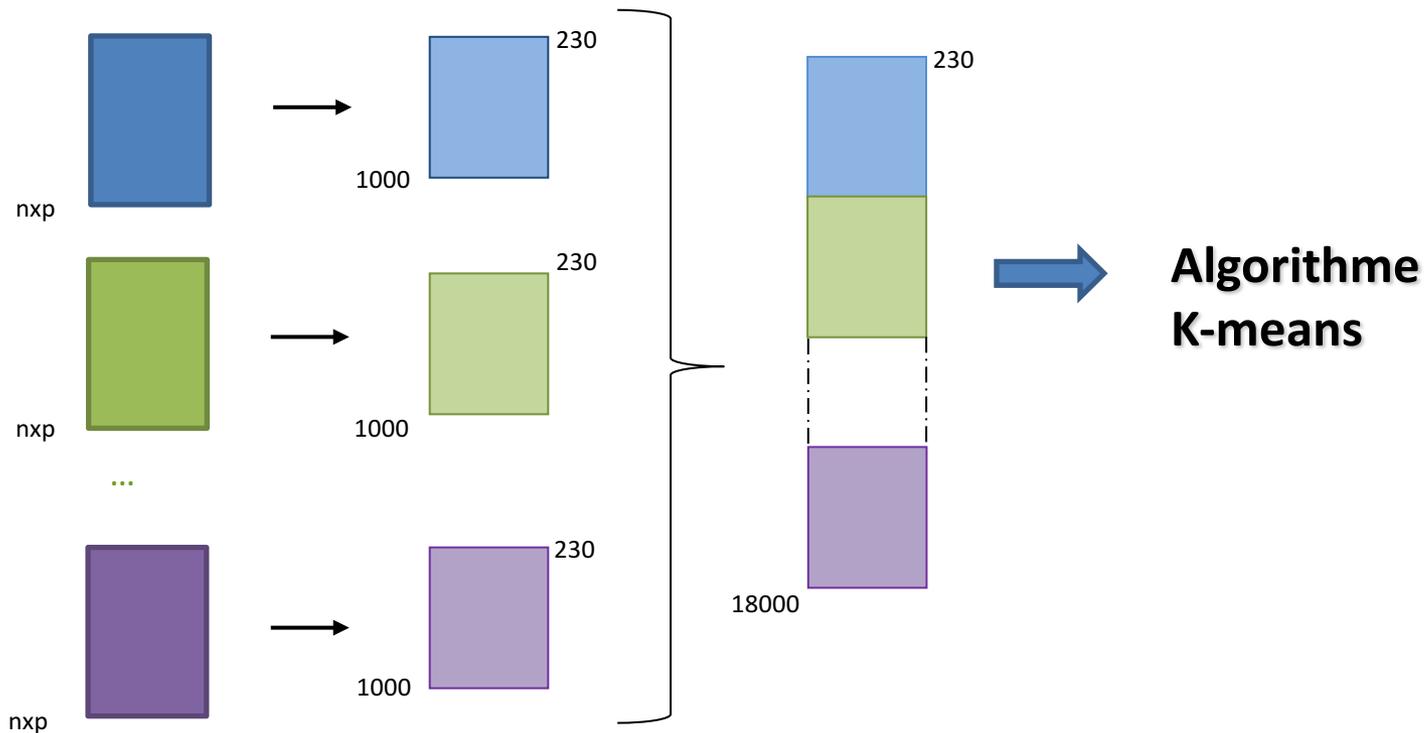
3^{ème} étape



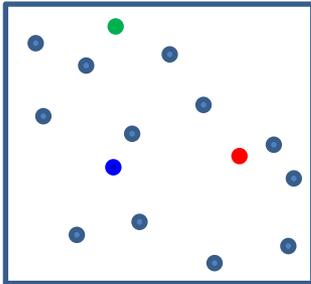
K-means (1)

Échantillonnage aléatoire :
1000 pixels dans ROI/image

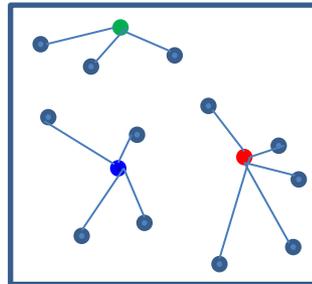
18
cubes
dépliés



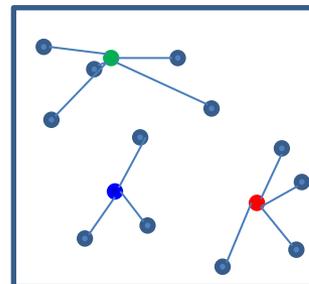
K-means (2)



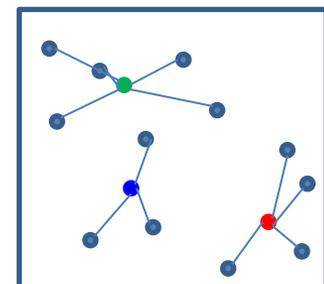
Sélection arbitraire de k centres \Rightarrow k groupes



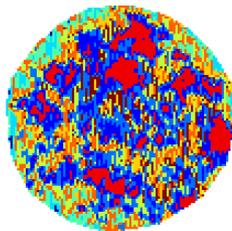
Chaque pixel est assigné à un groupe



Les centres de chaque groupe sont bougés vers la moyenne des points du groupe et les pixels assignés aux nouveaux groupes



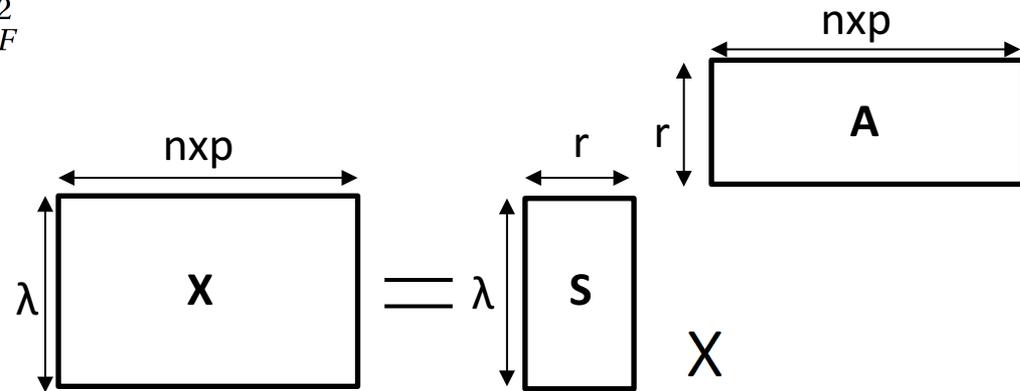
Les centres sont encore bougés à la moyenne des points



Une image en fausses couleurs peut être recréée en assignant une couleur arbitraire à chaque groupe.

Factorisation en matrices non-négatives hors-ligne

- Fonction coût : $\min_{\mathbf{S}, \mathbf{A} \geq 0} \|\mathbf{X} - \mathbf{SA}\|_F^2$
- Image hyperspectrale dépliée :

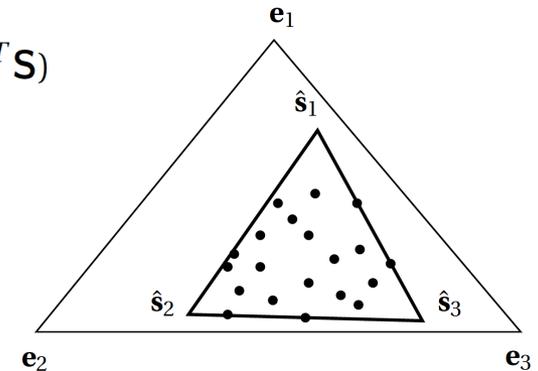


- **Problème 1 : non-unicité de la solution.**

- ❖ Ajout d'une contrainte de volume minimum : $\min_{\mathbf{S}, \mathbf{A} \geq 0} \|\mathbf{X} - \mathbf{SA}\|_F^2 + \mu \ln \det(\mathbf{S}^T \mathbf{S})$

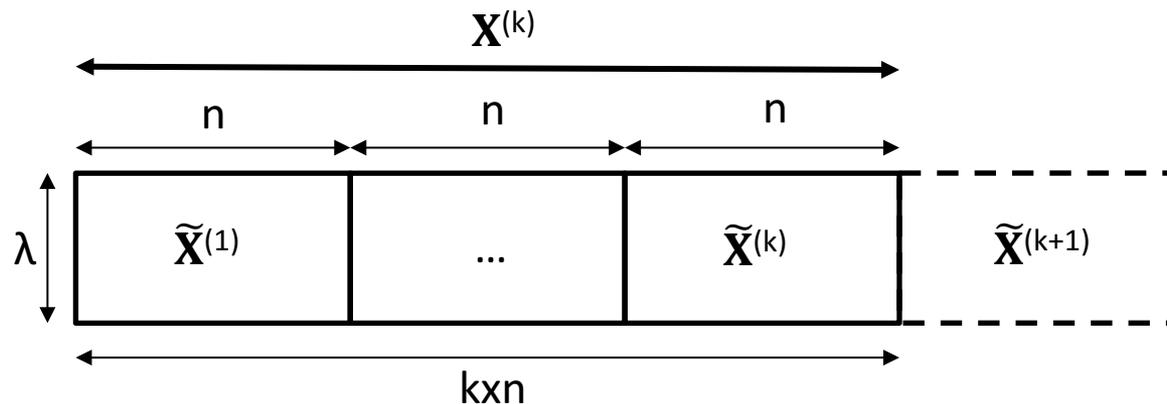
- **Problème 2 : temps de traitement conséquent.**

- ❖ Développement d'un algorithme de factorisation en-ligne de type NMF adapté au rythme de l'acquisition de l'imageur pushbroom. Traite les tranches de l'image successivement.



Factorisation en matrices non-négatives en-ligne (1)

- Estimer et mettre à jour les matrices \mathbf{S} (sources) et \mathbf{A} (abondances) en temps réel (au rythme de l'acquisition).
 - ❖ Prise en charge de grandes séries de données.
 - ❖ Facile à implémenter.
 - ❖ Efficace en terme de temps de traitement.
- Mettre à jour les matrices $\mathbf{S}^{(k)}$ et $\mathbf{A}^{(k)}$ estimées à l'instant k lorsqu'un nouvel échantillon arrive à l'instant $(k + 1)$.



$$\mathbf{X}^{(k+1)} = [\mathbf{X}^{(k)} \tilde{\mathbf{X}}^{(k+1)}]$$

Factorisation en matrices non-négatives en-ligne (2)

- Fonction coût sans contrainte :

$$J^{(k+1)} = \|\mathbf{X}^{(k+1)} - \mathbf{S}^{(k+1)} \mathbf{A}^{(k+1)}\|_F^2$$
$$= \sum_{i=1}^{\lambda} \sum_{j=1}^{k \times n} \left(\mathbf{X}_{ij}^{(k)} - \left(\mathbf{S}^{(k+1)} \mathbf{A}^{(k)} \right)_{ij} \right)^2 + \sum_{i=1}^{\lambda} \sum_{j=1}^n \left(\tilde{\mathbf{X}}_{ij}^{(k+1)} - \left(\mathbf{S}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right)_{ij} \right)^2$$

- Fonction coût avec contrainte :

$$J^{(k+1)} = \alpha J^{(k)} + (1 - \alpha) \tilde{J}^{(k+1)} + \mu \ln \det \left(\mathbf{S}^{(k+1)T} \mathbf{S}^{(k+1)} \right)$$

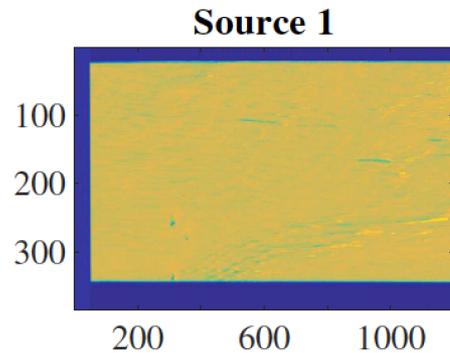
←
Contrainte de
volume minimum

- Algorithme de minimisation par la descente de gradient avec mise à jour multiplicative (Article en cours de soumission pour la conférence Statistical Signal Processing 2018).

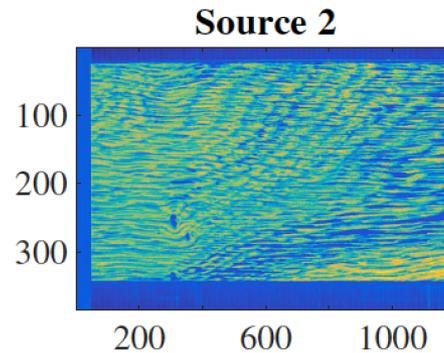
Comparaison des méthodes

	Contraintes	Interprétabilité physique	Variabilité spectrale	Temps de traitement
ACP	Orthogonalité des sources et des abondances	non	non	De l'ordre de 3 min
K-means	Abondances à valeurs binaires et à supports disjoints	non	non	De l'ordre de 3 min
NMF en-ligne	Positivité des sources et abondances	oui	oui	De l'ordre d'1 min

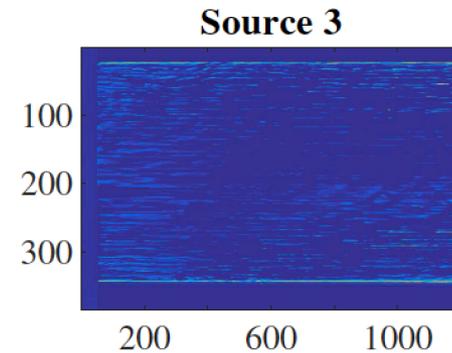
Résultats NMF en-ligne



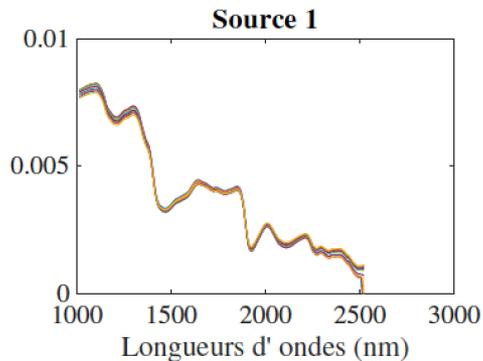
(a) Abondance 1



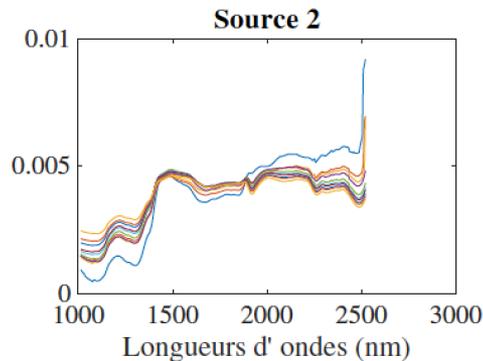
(b) Abondance 2



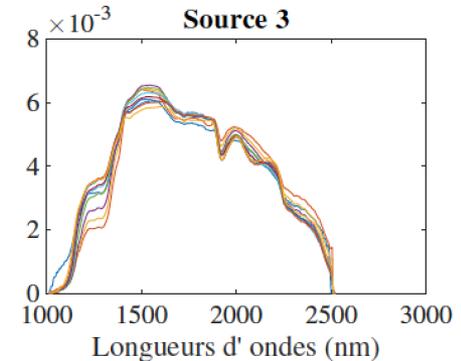
(c) Abondance 3



(d) Source 1

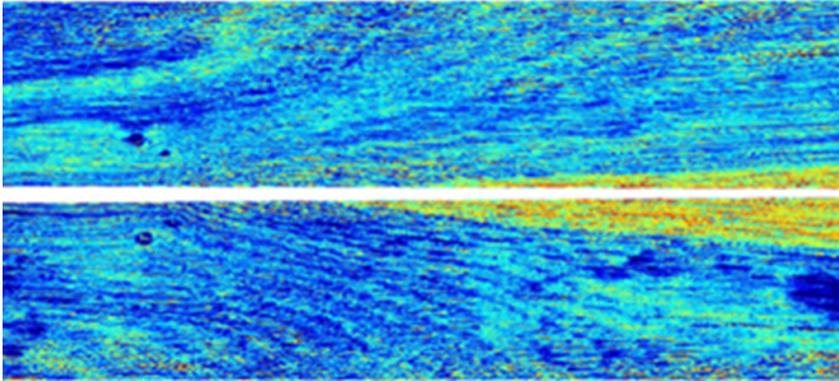


(e) Source 2

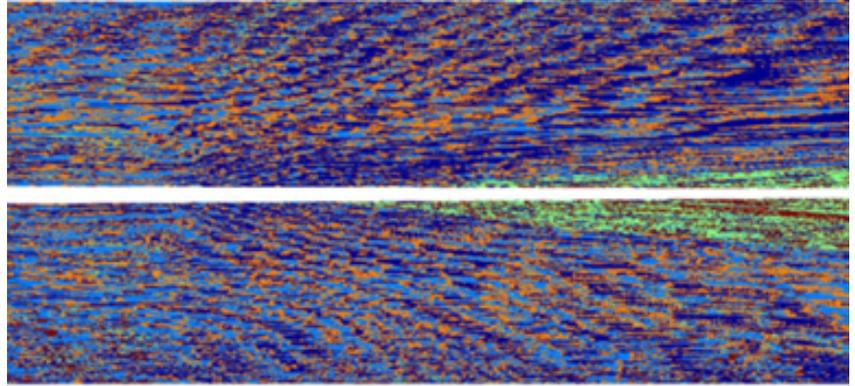


(f) Source 3

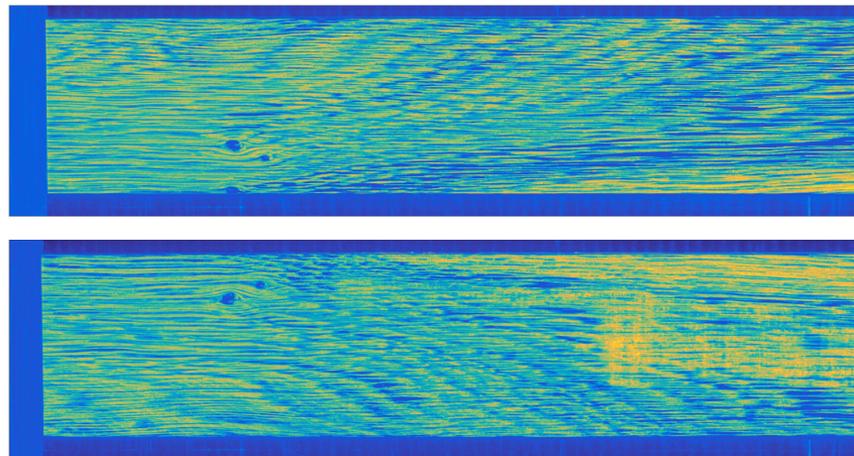
Résultats



ACP



K-means



NMF en-ligne

Conclusions

- Combinaison de l'imagerie hyperspectrale proche infrarouge et de la chimiométrie : visualisation des différents tissus dans le bois indiscernable par une inspection visuelle.
- Résultats similaires pour les trois méthodes : validation de l'approche.
- Les scores PC1, les images k-means et les cartes d'abondance de la NMF en-ligne ont démontré que l'utilisation d'algorithmes non supervisés est pertinente pour le tri des échantillons.
- L'approche NMF en-ligne permet une classification en temps réel directement sur la ligne de production et une interprétation physico-chimique des signatures spectrales des sources.