



HAL
open science

Méthodes rapides de traitement d'images hyperspectrales. Application à la caractérisation en temps réel du matériau bois

Ludivine Nus

► **To cite this version:**

Ludivine Nus. Méthodes rapides de traitement d'images hyperspectrales. Application à la caractérisation en temps réel du matériau bois. Traitement du signal et de l'image [eess.SP]. Université de Lorraine, 2019. Français. NNT : 2019LORR0163 . tel-02478101

HAL Id: tel-02478101

<https://hal.univ-lorraine.fr/tel-02478101>

Submitted on 13 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Méthodes rapides de traitement d'images hyperspectrales. Application à la caractérisation en temps réel du matériau bois

THÈSE

Présentée publiquement le 12 Décembre 2019 pour l'obtention du

Doctorat de l'Université de Lorraine
(Mention automatique, traitement du signal et génie informatique)

par

Ludivine NUS

Composition du jury

<i>Rapporteurs :</i>	Jean-Yves Tourneret	Professeur à l'Université de Toulouse
	Jean-Philippe Ovarlez	Directeur de recherche à l'ONERA
<i>Examineurs :</i>	Christine Fernandez-Maloigne	Professeur à l'Université de Poitiers
	Nicolas Gillis	Associate professor à l'Université de Mons
	David Brie	Professeur à l'Université de Lorraine
	Sebastian Miron	Maître de conférences à l'Université de Lorraine
<i>Invité :</i>	Eric Masson	Responsable Pôle R&D CRITT Bois

Centre de Recherche en Automatique de Nancy

CRAN UMR 7039, Université de Lorraine, CNRS
Faculté des Sciences et Technologies, BP 70239, 54506 Vandœuvre Cedex
E-mail : ludivine.nus@univ-lorraine.fr

Remerciements

Je tiens à exprimer mes profonds et sincères remerciements à mes directeurs de thèse, **David Brie**, professeur à l'Université de Lorraine et **Sebastian Miron**, maître de conférences à l'Université de Lorraine, pour tout le soutien apporté lors de mon doctorat. J'ai apprécié leur excellent encadrement et leur immense connaissance, ainsi que leur motivation, leur patience et leur confiance. Je garderai un excellent souvenir de notre collaboration.

Outre mes directeurs, je voudrais remercier le reste de mon comité de thèse d'avoir examiné mon travail et de m'avoir fait part de leurs commentaires. Mes sincères remerciements vont à Mme. **Christine Fernandez-Maloigne**, professeur à l'Université de Poitiers, M. **Jean-Philippe Ovarlez**, directeur de recherche à l'ONERA, M. **Jean-Yves Tourneret**, professeur à l'Université de Toulouse et M. **Nicolas Gillis**, associate professor à l'Université de Mons, pour avoir pris le temps d'évaluer ma thèse.

Je voudrais également exprimer ma reconnaissance à **Saïd Moussaoui**, professeur à l'École centrale de Nantes ainsi qu'à **Benoît Jaillais**, ingénieur de recherche à l'INRA, pour la collaboration fructueuse engagée dans le projet ALOHA. Je leur exprime ma gratitude quant aux discussions et à leurs aides apportées sur les travaux de ma thèse ainsi que pour les invitations à des visites scientifiques dans leurs laboratoires.

Je remercie également nos partenaires du projet ANR-OPTIFIN ainsi que M. **Eric Masson**, responsable du pôle recherche et développement du Crittbois et M. **Cédric Carteret**, professeur au LCPME.

Je voudrais remercier les membres permanents du laboratoire : **El-Hadi Djermoune**, **Konstantin Usevich**, **Julien Flamant**, **Eddy Robin** et **Thomas Aiguier** pour la qualité des discussions scientifiques et la convivialité.

Merci aux anciens doctorants **Yingying Song** et **Mamadou Diop** ainsi que **Clémence Prévost**, doctorante actuelle, pour leur soutien, leur présence et leur aide généreuse, pour tout le plaisir que nous avons partagé au cours de ces dernières années. Vous êtes devenus de vrais amis et grâce à vous, l'ambiance était au rendez-vous !

Je remercie également les équipes Réseaux et Télécommunications et Génie Mécanique et Productique de l'IUT Nancy-Brabois : **Muriel Lombard**, **Magalie Thomassin**, **Nicolas Krommenacker**, **Michaël David** et autres pour le temps qu'ils m'ont consacré et leur confiance.

Un grand merci à tout le personnel des services administratifs : **Sabine Huraux, Monique Billon, Séverine Heissler** et **Christelle Kondratow** pour leur gentillesse, leur dévouement et leur travail.

Enfin, je souhaite remercier ma famille et mes amis pour leur soutien indéfectible au cours de ma thèse.
Un merci tout particulier à ma mère et à mon père pour avoir toujours cru en moi.

Résumé

Cette thèse aborde le démixage en-ligne d'images hyperspectrales acquises par un imageur *pushbroom*, pour la caractérisation en temps réel du matériau bois. La première partie de cette thèse propose un modèle de mélange en-ligne fondé sur la factorisation en matrices non-négatives. À partir de ce modèle, trois algorithmes pour le démixage séquentiel en-ligne, fondés respectivement sur les règles de mise à jour multiplicatives, le gradient optimal de Nesterov et l'optimisation ADMM (*Alternating Direction Method of Multipliers*) sont développés. Ces algorithmes sont spécialement conçus pour réaliser le démixage en temps réel, au rythme d'acquisition de l'imageur *pushbroom*. Afin de régulariser le problème d'estimation (généralement mal posé), deux sortes de contraintes sur les *endmembers* sont utilisées : une contrainte de dispersion minimale ainsi qu'une contrainte de volume minimal. Une méthode pour l'estimation automatique du paramètre de régularisation est également proposée, en reformulant le problème de démixage hyperspectral en-ligne comme un problème d'optimisation bi-objectif. Dans la seconde partie de cette thèse, nous proposons une approche permettant de gérer la variation du nombre de sources, *i.e.* le rang de la décomposition, au cours du traitement. Les algorithmes en-ligne préalablement développés sont ainsi modifiés, en introduisant une étape d'apprentissage d'une bibliothèque hyperspectrale, ainsi que des pénalités de parcimonie permettant de sélectionner uniquement les sources actives. Enfin, la troisième partie de ces travaux consiste en l'application de nos approches à la détection et à la classification des singularités du matériau bois.

Mots-clés : Démixage hyperspectral en-ligne, imagerie hyperspectrale *pushbroom*, factorisation en matrices non-négatives, contrainte de volume minimal, estimation du paramètre de régularisation, bibliothèque hyperspectrale, suivi du rang

Abstract

This PhD dissertation addresses the problem of on-line unmixing of hyperspectral images acquired by a pushbroom imaging system, for real-time characterization of wood. The first part of this work proposes an on-line mixing model based on non-negative matrix factorization. Based on this model, three algorithms for on-line sequential unmixing, using multiplicative update rules, the Nesterov optimal gradient and the ADMM optimization (Alternating Direction Method of Multipliers), respectively, are developed. These algorithms are specially designed to perform the unmixing in real time, at the pushbroom imager acquisition rate. In order to regularize the estimation problem (generally ill-posed), two types of constraints on the endmembers are used: a minimum dispersion constraint and a minimum volume constraint. A method for the unsupervised estimation of the regularization parameter is also proposed, by reformulating the on-line hyperspectral unmixing problem as a bi-objective optimization. In the second part of this manuscript, we propose an approach for handling the variation in the number of sources, *i.e.* the rank of the decomposition, during the processing. Thus, the previously developed on-line algorithms are modified, by introducing a hyperspectral library learning stage as well as sparse constraints allowing to select only the active sources. Finally, the third part of this work consists in the application of these approaches to the detection and the classification of the singularities of wood.

Keywords: On-line hyperspectral unmixing, pushbroom hyperspectral imaging, non-negative matrix factorization, minimal volume constraint, regularization parameter estimation, hyperspectral library, rank tracking

Table des matières

Abréviations	ix
Notations	xi
Liste des figures	xii
Liste des tableaux	xvii
1 Introduction générale	1
1.1 Contexte industriel et objectifs	1
1.2 La spectroscopie proche infrarouge	3
1.2.1 Principe de la spectroscopie proche infrarouge	3
1.2.2 Instrumentation	5
1.2.3 Spectroscopie proche infrarouge pour l'analyse du bois	6
1.3 L'imagerie hyperspectrale	8
1.3.1 Principe de l'imagerie hyperspectrale	8
1.3.2 Techniques d'acquisition en imagerie hyperspectrale	10
1.3.2.1 Balayage spatial	10
1.3.2.2 Balayage spectral	11
1.3.2.3 Acquisition instantanée	12
1.3.3 Domaines d'application de l'imagerie hyperspectrale	12
1.3.4 Présentation du système d'imagerie hyperspectrale <i>pushbroom</i>	13
1.4 Démélange d'image hyperspectrale	14
1.4.1 Modèle de mélange linéaire	15
1.4.1.1 Formulation du problème	15
1.4.1.2 Identifiabilité du modèle	16
1.4.2 Méthodes de démélange hyperspectral	22
1.4.2.1 Méthodes géométriques avec pixel pur	23
1.4.2.2 Méthodes géométriques sans pixel pur	24
1.4.2.3 La factorisation en matrices non-négatives	25
1.5 Choix méthodologiques et limites des approches existantes	30
1.6 Contributions de ce travail de thèse	33
1.7 Publications	35

2	Démélange hyperspectral en-ligne pour les systèmes industriels d'imagerie <i>pushbroom</i>	36
2.1	Introduction	36
2.2	Démélange hyperspectral en-ligne pour le système d'acquisition <i>pushbroom</i>	38
2.2.1	Acquisition <i>pushbroom</i>	38
2.2.2	Modèle de données	39
2.2.3	Fonction objectif	40
2.3	Régularisation de la solution à l'aide d'une contrainte de volume minimal	41
2.3.1	Identifiabilité	41
2.3.2	Contraintes de volume	41
2.4	Approches proposées	45
2.4.1	Minimisation de la fonction objectif à l'aide des mises à jour multiplicatives	45
2.4.2	Minimisation de la fonction objectif à l'aide de l'optimisation ADMM	48
2.4.3	Minimisation de la fonction objectif à l'aide du gradient optimal de Nesterov	49
2.4.4	Implantation des algorithmes et complexité de calcul	51
2.4.5	Analyse de convergence	54
2.5	Résultats expérimentaux pour le démélange hyperspectral en-ligne	56
2.5.1	Préservation du rang lors de l'utilisation de MDC et MVC	56
2.5.2	Vitesse de convergence	58
2.5.3	Influence des paramètres	61
2.5.4	Application sur des données réelles	66
2.6	Estimation du paramètre de régularisation	72
2.6.1	Le démélange hyperspectral comme un problème d'optimisation bi-objectif : le choix de l'hyperparamètre μ	73
2.6.1.1	Critère de courbure maximale	74
2.6.1.2	Critère de distance minimale	75
2.6.2	Évaluation des critères MCC et MDist sur des données réelles et simulées	76
2.6.2.1	Forme d'une courbe de réponse	76
2.6.2.2	Influence des conditions initiales	77
2.6.2.3	Évaluation des critères MCC et MDistC sur des données simulées	77
2.6.2.4	Évaluation des critères MCC et MDistC sur une image hyperspectrale réelle	82
2.7	Conclusions	84
3	Démélange hyperspectral semi-supervisé en-ligne avec suivi du nombre de <i>endmembers</i>	85
3.1	Introduction	85
3.2	Algorithme OSS-ADMM	87
3.2.1	Fonction objectif	87
3.2.2	Mises à jour des paramètres d'intérêt	88
3.2.3	Implantation de l'algorithme OSS-ADMM et complexité de calcul	88
3.3	Résultats expérimentaux	90
3.3.1	Influence des hyperparamètres v , γ et ω	90
3.3.2	OSS-ADMM appliqué aux images hyperspectrales simulées	94
3.3.3	OSS-ADMM pour l'analyse de la diffusion de l'eau au sein d'un produit céréalier	98

3.4	Conclusions	105
4	Analyse de planches de chêne	106
4.1	Introduction	106
4.2	Présentation des pièces de bois	107
4.2.1	Échantillons à traiter	107
4.2.2	Définition des singularités	109
4.3	Démélange hyperspectral en-ligne	111
4.3.1	Traitement des pièces de bois à l'aide de OMDC-ADMM	111
4.3.2	Traitement des pièces de bois à l'aide de OSS-ADMM	117
4.3.3	Segmentation des images hyperspectrales	123
4.4	Conclusions	126
5	Conclusions et perspectives	127
	Bibliographie	143

Abréviations

PIR	Proche InfraRouge
UV	UltraViolet
IR	InfraRouge
CCD	<i>Charge Coupled Device</i>
InGaAs	<i>Indium Gallium Arsenide</i>
FOV	<i>Field Of View</i>
MVT	<i>Minimum Volume Transform</i>
VolMin	Volume Minimal
NMF	<i>Non-negative Matrix Factorization</i>
OSP	<i>Orthogonal Subspace Projection</i>
PPI	<i>Pixel Purity Index</i>
VCA	<i>Vertex Component Analysis</i>
MNF	<i>Maximum Noise Fraction</i>
NCLS	<i>Non-Negativity Constrained Least Squares</i>
FCLS	<i>Fully Constrained Least Squares</i>
SGM	<i>Scaled Gradient Method</i>
CCA	<i>Convex Cone Analysis</i>
ICE	<i>Iterative Constrained Endmembers</i>
MVSA	<i>Minimum Volume Simplex Analysis</i>
MVES	<i>Minimum Volume Enclosing Simplex</i>
SISAL	<i>Simplex Identification via Split Augmented Lagrangian</i>
SPICE	<i>Sparsity-Promoting Iterative Constrained Endmembers</i>
ADMM	<i>Alternating Direction Method of Multipliers</i>

NeNMF	<i>Nesterov Non-negative Matrix Factorization</i>
MU	<i>MUltiplicatif</i>
ALS	<i>Alternating Least Square</i>
HALS	<i>Hierarchical Alternating Least Square</i>
KKT	<i>Karush Kuhn Tucker</i>
INMF	<i>Incremental Non-negative Matrix Factorization</i>
MCC	<i>Maximum Curvature Criterion</i>
MDistC	<i>Minimum Distance Criterion</i>
SGD	<i>Stochastic Gradient Descent</i>
2D	<i>Bidimensionnel</i>
MVC	<i>Minimum Volume Constraint</i>
MDC	<i>Minimum Dispersion Constraint</i>
OMDC-MU	<i>On-line Minimum Dispersion Constraint-MUltiplicative</i>
OMVC-MU	<i>On-line Minimum Volume Constraint-MUltiplicative</i>
OMDC-ADMM	<i>On-line Minimum Dispersion Constraint-Alternating Direction Method of Multipliers</i>
OMDC-Ne	<i>On-line Minimum Dispersion Constraint-Nesterov</i>
OMVC-Ne	<i>On-line Minimum Volume Constraint-Nesterov</i>
BMDC-ADMM	<i>Batch Minimum Dispersion Constraint-Alternating Direction Method of Multipliers</i>
GCV	<i>Generalized Cross-Validation</i>
SAD	<i>Spectral Angle Distance</i>
RMSE	<i>Root Mean Square Error</i>
SNR	<i>Signal to Noise Ratio</i>
RAM	<i>Random Access Memory</i>
OSS-ADMM	<i>On-line Semi-Supervised-Alternating Direction Method of Multipliers</i>
FW	<i>Free Water</i>
BW	<i>Bound Water</i>
CANDECOMP	<i>CANonical DECOMPosition</i>
PARAFAC	<i>PARAllel FACtor analysis</i>

Notations

Notations générales

\mathbb{R}	: l'ensemble des nombres réels
\mathbb{R}_+	: l'ensemble des nombres réels non négatifs
$\mathbb{R}_+^{I \times J}$: l'ensemble des matrices non-négatives de dimension $I \times J$
$\mathbb{I}_{\mathbb{R}_+}$: la fonction indicatrice de \mathbb{R}_+
$[\cdot]_+$: projection sur \mathbb{R}_+
s, S	: scalaire
\mathbf{v}	: vecteur
\mathbf{v}^T	: transposée du vecteur \mathbf{v}
$\text{diag}(\mathbf{v})$: matrice diagonale dont la diagonale est le vecteur \mathbf{v}
$\mathbf{1}$: vecteur colonne/matrice composé que de 1
\mathbf{M}	: matrice
\mathbf{I}	: matrice identité
\mathbf{M}^T	: transposée de la matrice \mathbf{M}
\mathbf{M}^{-1}	: matrice inverse de la matrice \mathbf{M}
$\mathbf{M}(:, j) = \mathbf{m}_j$: $j^{\text{ème}}$ colonne de la matrice \mathbf{M}
$\mathbf{M}(i, :) = \mathbf{m}^i$: $i^{\text{ème}}$ ligne de la matrice \mathbf{M}
$\mathbf{M}(i, j) = m_{ij}$: élément de la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne de la matrice \mathbf{M}
$\mathbf{M} \geq \mathbf{0}$: matrice \mathbf{M} dont les éléments sont non-négatifs
$\text{cône}\{\mathbf{M}\}$: cône convexe généré par les colonnes de \mathbf{M}
$\text{conv}\{\mathbf{M}\}$: enveloppe convexe des colonnes de \mathbf{M} (simplexe)
$\text{Vol}(\mathbf{M})$: mesure liée ou proportionnelle au volume du simplexe $\text{conv}\{\mathbf{M}\}$
$\text{rang}(\mathbf{M})$: rang de la matrice \mathbf{M}
$\text{trace}(\mathbf{M})$: trace de la matrice \mathbf{M}
$\det(\mathbf{M})$: déterminant de la matrice \mathbf{M}
$\nabla_{\mathbf{M}} \mathcal{J} = \frac{\partial \mathcal{J}}{\partial \mathbf{M}}$: gradient du critère \mathcal{J} par rapport à \mathbf{M}
$\log(\cdot)$: logarithme
$\ \cdot\ _2$: norme ℓ_2 d'un vecteur
$\ \cdot\ _{q,p}$: norme $\ell_{q,p}$ d'une matrice
$\ \cdot\ _F$: norme de Frobenius d'une matrice
\odot	: produit matriciel de Hadamard

Liste des figures

1.1	Coupe transversale d'un rondin de bois	2
1.2	Différence aubier/duramen pour le chêne	2
1.3	Effet de l'aubier sur du parquet	2
1.4	Différents effets induits par l'interaction d'un rayonnement sur un échantillon	3
1.5	Spectre électromagnétique	4
1.6	Les quatre modes normaux de vibration du CO_2 ; de gauche à droite : élongation symétrique, élongation anti-symétrique, déformation de l'angle dans le plan, déformation de l'angle perpendiculaire au plan	4
1.7	Liaisons chimiques à différentes bandes d'absorption dans le proche infrarouge	5
1.8	Principaux composants d'un spectrographe	6
1.9	Bandes d'absorption PIR (1100-2500 nm) pour les principaux composants du bois	8
1.10	Spectres proche infrarouge pour différentes essences de bois	8
1.11	Structure d'une image hyperspectrale	9
1.12	De l'imagerie monochrome à l'hyperspectrale	9
1.13	Système d'acquisition <i>whiskbroom</i>	11
1.14	Système d'acquisition <i>pushbroom</i>	11
1.15	Système d'acquisition à balayage spectral	11
1.16	Système d'acquisition instantanée	12
1.17	Composants du système d'imagerie hyperspectrale	13
1.18	Imageur <i>pushbroom</i> FX17 de chez Specim	14
1.19	Pixels purs et pixel mélangé au sein d'une image hyperspectrale	14
1.20	Principe du modèle linéaire	15
1.21	Principe du modèle non-linéaire	15
1.22	Représentation géométrique pour $R = 3$. Version non normalisée (à gauche) : cône convexe. Version normalisée (à droite) : simplexe. Les points représentent les \mathbf{x}_p	17
1.23	Illustration d'un problème non-unique. Les simplexes en pointillé représentent $\text{conv}\{\bar{\mathbf{S}}\}$ et englobent toutes les données \mathbf{x}_p	18
1.24	Illustration de la condition de séparabilité dans le cas où $R = 3$. À gauche, le domaine des \mathbf{A} et à droite le domaine des \mathbf{X} . On suppose que l'on se situe face aux cônes (vers l'origine). Les points \mathbf{x}_p situés au centre du cône font référence aux pixels mélangés	19
1.25	Illustration de la condition de dispersion dans le cas où $R = 3$. À gauche, le domaine des \mathbf{A} et à droite le domaine des \mathbf{X}	19
1.26	Illustration de la condition de régularité non satisfaite. Le cercle extérieur \mathcal{C}^* est le cône dual de \mathcal{C} défini tel que $\mathcal{C}^* = \{\mathbf{x} \in \mathbb{R}^R \mathbf{x}^T \mathbf{1} \geq \ \mathbf{x}\ _2\}$	20

1.27	L'intuition du volume minimal dans le cas où $R = 3$. Les points représentent les \mathbf{x}_p . En pointillé est représenté un simplexe englobant les données et en ligne continue le simplexe ayant le volume minimal permettant ainsi d'identifier \mathbf{S}	21
1.28	Données issues d'une image hyperspectrale de bois (acquise par l'imageur FX17) projetées sur les deux premières composantes principales	30
1.29	Acquisition des données hyperspectrales <i>pushbroom</i>	31
2.1	Acquisition des données par un imageur <i>pushbroom</i>	39
2.2	Structure des données pour l'acquisition <i>pushbroom</i>	39
2.3	Image hyperspectrale <i>pushbroom</i> dépliée	39
2.4	Courbes de niveau	43
2.5	Fonctions de régularisation MVC et MDC en fonction de l'angle β	44
2.6	Cartes d'abondance simulées	56
2.7	Signatures spectrales des trois <i>endmembers</i> (End)	57
2.8	Valeurs de MDC et MVC en fonction de μ , pour le mélange de la Figure 2.6	57
2.9	Simplexes estimés en fonction de μ pour MVC et MDC, pour le mélange de la Figure 2.6	58
2.10	Cartes d'abondance simulées	59
2.11	Vitesses de convergence du SAD et du RMSE en fonction de N_{iter} , pour le mélange de la Figure 2.10	60
2.12	OMDC-ADMM vs. BMDC-ADMM : vitesses de convergence du SAD en fonction de N_{iter} , pour le mélange de la Figure 2.10	61
2.13	Évolution du SAD en fonction de α , pour le mélange de la Figure 2.10	62
2.14	Vitesse de convergence du SAD en fonction de N_{iter} , pour différentes valeurs de α , pour le mélange de la Figure 2.10	62
2.15	Vitesse de convergence du SAD en fonction du nombre d'échantillons, pour différentes valeurs de α , pour le mélange de la Figure 2.10	62
2.16	Évolution des <i>endmembers</i> au cours des tranches de l'image hyperspectrale	63
2.17	Vitesse de convergence du SAD en fonction du nombre d'échantillons, pour différentes valeurs de α , dans le cas de sources non stationnaires	64
2.18	Vitesse de convergence du SAD en fonction du nombre d'échantillons, pour différentes valeurs de ρ , pour le mélange de la Figure 2.10	64
2.19	Évolution du SAD et du RMSE en fonction de ϵ , pour le mélange de la Figure 2.10	65
2.20	Résultat du démixage hyperspectral pour le mélange de la Figure 2.10, pour $\epsilon = 0.0001$	65
2.21	Résultat du démixage hyperspectral pour le mélange de la Figure 2.10, pour $\epsilon = 1$	65
2.22	Vérité terrain : cartes d'abondance et signatures spectrales pour l'image <i>Samson</i> . De gauche à droite : le sol, les arbres et l'eau	66
2.23	<i>Endmembers</i> vrais et estimés par OMDC-ADMM, OMDC-Ne et OMVC-Ne, pour l'image <i>Samson</i>	67
2.24	Cartes d'abondance vraies et estimées par OMDC-ADMM, OMDC-Ne et OMVC-Ne, pour l'image <i>Samson</i>	68
2.25	Vérité terrain : cartes d'abondance et signatures spectrales pour l'image <i>Jasper Ridge</i> . De gauche à droite : les arbres, l'eau, le sol et la route	69
2.26	<i>Endmembers</i> estimés tranche par tranche par OMDC-ADMM, pour l'image <i>Jasper Ridge</i>	70

2.27	Spectres de l'eau estimés tranche par tranche par OMDC-ADMM pour $\alpha = 0.9$, pour l'image <i>Jasper Ridge</i>	71
2.28	<i>Endmembers</i> estimés pour l'image <i>Jasper Ridge</i>	71
2.29	Cartes d'abondance vraies et estimées par OMDC et BMDC-ADMM, pour l'image <i>Jasper Ridge</i>	72
2.30	Exemple d'une courbe de réponse présentant une courbure maximale non-unique	75
2.31	Représentation de la courbe de réponse. Le point idéal est désigné par Υ et le point optimal par Φ	75
2.32	Courbe de réponse estimée pour le mélange de la Figure 2.6	76
2.33	Évolution du simplexe estimé pour différentes valeurs de μ , pour le mélange de la Figure 2.6	76
2.34	<i>Endmembers</i> estimés pour différentes valeurs de μ , pour le mélange de la Figure 2.6	77
2.35	Courbes de réponse estimées pour différentes valeurs de μ et différentes initialisations, pour le mélange de la Figure 2.6	77
2.36	Courbe de réponse estimée pour le mélange de la Figure 2.6 avec le point de courbure maximale (croix rouge)	78
2.37	Résultat du démélange hyperspectral pour le mélange de la Figure 2.6, pour $\mu = 0.0006$ (estimé par MCC)	78
2.38	Courbe de réponse estimée pour le mélange de la Figure 2.6 avec le point Φ (croix rouge) ayant la distance minimale du point idéal Υ	79
2.39	Résultat du démélange hyperspectral pour le mélange de la Figure 2.6, pour $\mu = 0.1$ (estimé par MDistC)	79
2.40	SAD et RMSE en fonction de μ , pour le mélange de la Figure 2.6	80
2.41	Courbe de réponse estimée pour le mélange de la Figure 2.10	80
2.42	Cartes d'abondance estimées pour le mélange de la Figure 2.10, pour $\mu = 0.001$ (estimé par MCC)	81
2.43	Cartes d'abondance estimées pour le mélange de la Figure 2.10, pour $\mu = 0.2$ (estimé par MDistC)	81
2.44	Cartes d'abondance estimées pour le mélange de la Figure 2.10, pour $\mu = 0$	81
2.45	SAD et RMSE en fonction de μ , pour le mélange de la Figure 2.10	82
2.46	Cartes d'abondance vraies (expertisées) pour l'image <i>Jasper Ridge</i>	82
2.47	Cartes d'abondance estimées pour l'image <i>Jasper Ridge</i> , pour $\mu = 0.018$ (estimé par MCC)	82
2.48	Cartes d'abondance estimées pour l'image <i>Jasper Ridge</i> , pour $\mu = 0.2$ (estimé par MDistC)	83
2.49	Spectres des <i>endmembers</i> expertisés et estimés pour l'image <i>Jasper Ridge</i>	83
2.50	SAD et RMSE en fonction de μ , pour l'image <i>Jasper Ridge</i>	83
3.1	Illustration des matrices d'abondance. Deux scénarios sont considérés : à gauche, rang constant au cours du traitement <i>i.e.</i> $\text{rang}(\tilde{\mathbf{S}}^{(k)}) = R$. À droite : rang variable au cours du traitement <i>i.e.</i> $\text{rang}(\tilde{\mathbf{S}}^{(k)}) \leq R$	85
3.2	Vérité terrain : spectres des trois <i>endmembers</i> (End)	90
3.3	Bibliothèque hyperspectrale \mathbf{B} composée de quatre <i>endmembers</i>	90

3.4	Évolution de la désactivation de End4 en fonction du nombre d'échantillons, pour différents couples (ν, γ)	91
3.5	Vitesse de convergence du critère $\mathcal{J}_{sparse2}$ en fonction du nombre d'échantillons, pour différents couples (ν, γ)	92
3.6	Cartes d'abondance simulées ayant des supports constant par morceaux	92
3.7	Rang de la décomposition relatif au mélange de la Figure 3.6	93
3.8	Cartes d'abondance générées par OSS-ADMM pour le mélange de la Figure 3.6, avec $\nu = 0.002$ et $\gamma = 0$	93
3.9	Cartes d'abondance générées par OSS-ADMM pour le mélange de la Figure 3.6, avec $\nu = 0.002$ et $\gamma = 0.009$	93
3.10	Évolution du RMSE en fonction de ω , pour le mélange de la Figure 3.6	94
3.11	Cartes d'abondance générées par OSS-ADMM pour le mélange de la Figure 3.6, avec $\omega = 0$	94
3.12	Cartes d'abondance et signatures spectrales associées pour les images simulées 1, 2 et 3 .	95
3.13	Cartes d'abondance et signatures spectrales associées pour l'image simulée 4	95
3.14	Cartes d'abondance générées par OMDC-ADMM et rang de la décomposition, pour le mélange de l'image 1	96
3.15	Composition de la bibliothèque hyperspectrale B pour le traitement des images 1 à 4 . .	97
3.16	Cartes d'abondance et rang de la décomposition estimés par OSS-ADMM , pour les mélanges des images 1 à 4	98
3.17	Ajout de deux gouttes d'eau au centre du biscuit	99
3.18	Bibliothèque hyperspectrale estimée par BMDC-ADMM pour le traitement des images hyperspectrales de biscuit	99
3.19	Cartes d'abondance et spectres de l'amidon et de fibres cellulosiques estimés par OSS-ADMM pour les images hyperspectrales de biscuit	100
3.20	Résultats du démixage hyperspectral par OSS-ADMM pour les images hyperspectrales de biscuit acquises aux temps $t = 1, 15, 25, 35, 45,$ et 55 min	101
3.21	Zoom sur un des pics caractéristiques de l'eau liée et de l'eau libre, à $t = 1$ min	102
3.22	Pénétration de l'eau au sein du biscuit. À gauche : $t = 1$ s, à droite $t = 1+$ quelques secondes	103
3.23	Tissus hydratés non homogènes du biscuit	103
3.24	Cartes d'abondance estimées par OSS-ADMM pour l'image hyperspectrale de biscuit à $t = 1$ min, pour $\nu = 10^{-5}$ et $\gamma = 0.045$	104
3.25	Rang de la décomposition estimé par OSS-ADMM en fonction du nombre d'échantillons pour l'image hyperspectrale de biscuit à $t = 1$ min	105
4.1	Étapes de traitement des pièces de bois	106
4.2	Échantillon 1	107
4.3	Échantillon 2	107
4.4	Échantillon 3	108
4.5	Échantillon 4	108
4.6	Échantillon 5	108
4.7	Échantillon 6	108

4.8	Échantillon 7	108
4.9	Échantillon 8	109
4.10	Échantillon 9	109
4.11	Différents types de nœuds	110
4.12	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 1	111
4.13	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 2	112
4.14	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 3	112
4.15	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 4	113
4.16	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 5	113
4.17	RCartes d'abondance générées par OMDC-ADMM pour l'échantillon 6	114
4.18	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 7	114
4.19	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 8	115
4.20	Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 9	115
4.21	Composition de la bibliothèque hyperspectrale B pour le traitement des images de pièces de bois	117
4.22	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 1	118
4.23	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 2	118
4.24	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 3	119
4.25	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 4	119
4.26	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 5	120
4.27	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 6	120
4.28	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 7	121
4.29	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 8	121
4.30	Cartes d'abondance générées par OSS-ADMM pour l'échantillon 9	122
4.31	Segmentation des cartes d'abondance de l'échantillon 1 : 5 classes	123
4.32	Segmentation des cartes d'abondance de l'échantillon 2 : 5 classes	123
4.33	Segmentation des cartes d'abondance de l'échantillon 3 : 10 classes	123
4.34	Segmentation des cartes d'abondance de l'échantillon 4 : 4 classes	124
4.35	Segmentation des cartes d'abondance de l'échantillon 5 : 5 classes	124
4.36	Segmentation des cartes d'abondance de l'échantillon 6 : 5 classes	124
4.37	Segmentation des cartes d'abondance de l'échantillon 7 : 5 classes	124
4.38	Segmentation des cartes d'abondance de l'échantillon 8 : 10 classes	125
4.39	Segmentation des cartes d'abondance de l'échantillon 9 : 5 classes	125

Liste des tableaux

1.1	Caractéristiques de l'imageur FX17	14
2.1	Complexité de calcul des algorithmes en-ligne par échantillon k	54
2.2	Valeurs asymptotiques du SAD et du RMSE liées au temps et aux nombres d'itérations, pour le mélange de la Figure 2.10	60
2.3	SAD et RMSE pour OMDC-ADMM, OMDC-Ne et OMVC-Ne, pour l'image <i>Samson</i>	67
2.4	SAD et RMSE pour OMDC-ADMM, BMDC-ADMM et VCA, pour l'image <i>Jasper Ridge</i>	70
2.5	SAD et RMSE pour NMF, ℓ_1 -NMF et $\ell_{1/2}$ -NMF, pour l'image <i>Jasper Ridge</i>	70
4.1	Récapitulatif des singularités détectées par l'algorithme OMDC-ADMM	116
4.2	Dimension de chaque image hyperspectrale de bois et temps de traitement par l'algo- rithme OMDC-ADMM	116
4.3	Temps de traitement de chaque échantillon par OSS-ADMM	122

Chapitre 1

Introduction générale

1.1 Contexte industriel et objectifs

Cette thèse a été menée au Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039), Université de Lorraine, CNRS. Elle a été financée par le projet ANR-OPTIFIN (Agence Nationale de la Recherche, OPTImisation des FINitions) qui regroupe 2 laboratoires : le CRAN et le LCPME (Laboratoire de Chimie Physique et Microbiologie pour les Matériaux et l'Environnement), un centre technique (Critt Bois) et 2 partenaires industriels (BERRY Wood et OGF Industrie).

Le projet ANR-OPTIFIN a pour objectif le développement de méthodes d'analyse permettant de prédire et de classifier la qualité des rendus de pièces de bois en fonction d'une analyse de surface en temps réel. En effet, le bois est un matériau dont le rendu naturel est apprécié dans de nombreuses applications (parquet, ameublement, agencement intérieur ou extérieur). Cependant son hétérogénéité et sa variabilité de surface rendent souvent problématique la maîtrise et l'harmonie des rendus, générant parfois des non-qualités. Ce rendu et cette harmonie sont majoritairement regardés au travers d'une finition appliquée sur le bois (vernis ou peinture). La qualité de la finition et le rendu final sont impactés par le couple bois-finition. La texture physique, l'acidité, la présence de molécules pouvant migrer ou réagir avec la finition sont des facteurs potentiellement impactants. Ces difficultés techniques liées à l'hétérogénéité et à la variabilité des surfaces du bois génèrent donc aujourd'hui d'importants coûts de tri manuel et/ou de la non-qualité. Les solutions techniques industrialisées à ce jour se cantonnent à détecter des singularités très marquées, telles que les nœuds. En revanche, certaines singularités du bois comme l'aubier présente une grande variabilité colorimétrique alors que leur impact est important sur la qualité et le rendu de la surface finie. Cette variabilité explique les faibles performances des techniques de vision couleur sur ce type de singularité.

L'aubier caractérise la partie vivante du bois et se situe entre le duramen (appelé également bois dur ou bois parfait) et l'écorce comme représenté sur la Figure 1.1.

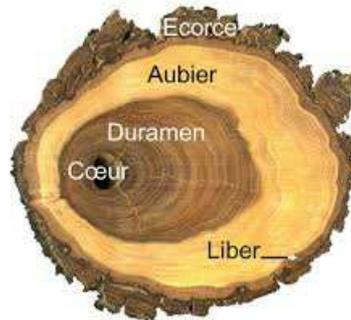


FIGURE 1.1 : Coupe transversale d'un rondin de bois

Image extraite de : [Rakotovao 12]

Pour certaines essences de bois, telles que le chêne, l'aubier est parfois difficilement discernable du duramen comme nous pouvons le constater sur la Figure 1.2. Cela représente un réel problème notamment dans le domaine du parquet, car l'aubier peut former, après finition, des lignes blanches plus ou moins importantes, dégradant ainsi l'esthétique du revêtement (Figure 1.3). Ainsi, la détection fiable de l'aubier présente un enjeu économique important.

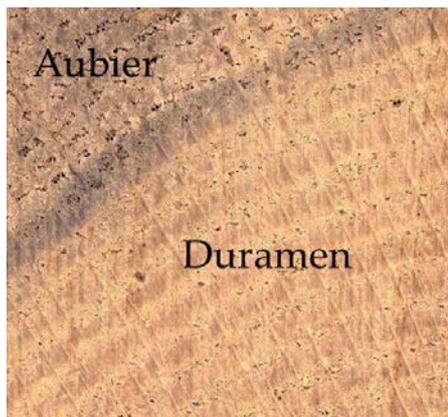


FIGURE 1.2 : Différence aubier/duramen pour le chêne

Image extraite de :

<http://bernard.langellier.pagesperso-orange.fr/arb/chene-ct.htm>

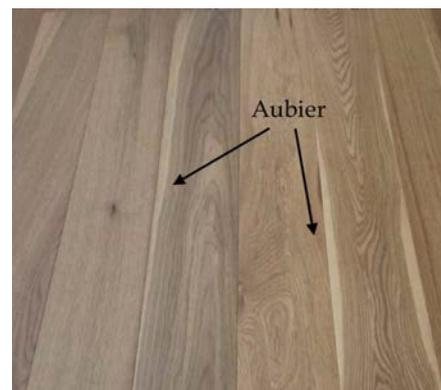


FIGURE 1.3 : Effet de l'aubier sur du parquet

Image extraite de : <https://www.tropical-woods.fr>

D'après [Schwanninger 11], la spectroscopie dans le proche infrarouge (PIR) est une méthode de détection pertinente pour analyser le matériau bois en général. Ainsi, l'une des solutions proposées dans le cadre de ce projet est d'installer directement sur la ligne de production des caméras hyperspectrales PIR permettant d'acquérir en temps réel des images hyperspectrales, *i.e.* au rythme où les pièces de bois défilent sur un convoyeur.

D'un point de vue traitement de données, le principal objectif du projet OPTIFIN est de développer des algorithmes rapides pour le traitement des images ainsi acquises, permettant l'extraction des paramètres d'intérêt, nécessaires à la caractérisation des différents échantillons de bois. Les méthodes de démixage hyperspectral semblent être une solution adéquate pour ce type de problème. Néanmoins, la plupart

d'entre elles se limitent à un traitement « hors-ligne » (ou *batch*), ce qui ne permet pas une analyse en temps réel des pièces de bois, indispensable au tri industriel. Ainsi, nous proposons le développement de méthodes séquentielles de démélange hyperspectral, permettant d'estimer et de mettre à jour au rythme de l'acquisition les paramètres d'intérêt.

1.2 La spectroscopie proche infrarouge

1.2.1 Principe de la spectroscopie proche infrarouge

La spectroscopie consiste à établir une relation entre la réponse de l'échantillon à un rayonnement électromagnétique et les caractéristiques de l'échantillon que l'on souhaite étudier [Genot 14]. L'interaction entre le rayonnement et l'échantillon conduit à différents effets comme représentés sur la Figure 1.4. Le rayonnement peut être absorbé, dispersé, réfléchi ou transmis à travers la matière.

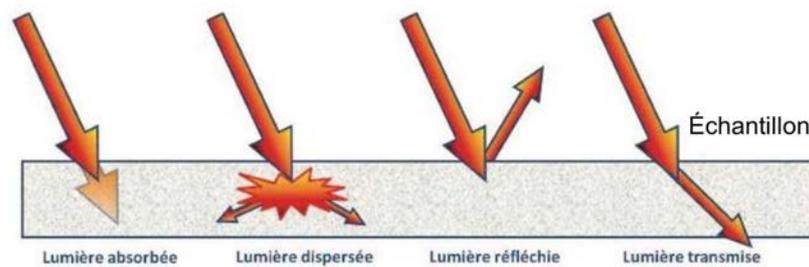


FIGURE 1.4 : Différents effets induits par l'interaction d'un rayonnement sur un échantillon

Image extraite de : <http://www.cinq.ulaval.ca/imagerie>

Selon la gamme spectrale utilisée (rayons X, ultraviolets (UV), visible, infrarouge (IR), micro-ondes) et les modes d'interaction (absorption ou émission, par exemple) considérés, différentes informations peuvent être récupérées. La gamme spectrale choisie permet de caractériser le matériau à différentes échelles. Par exemple, les spectroscopies à rayons X, UV et visibles reflètent principalement les transitions électroniques, tandis que les spectroscopies IR et Raman sont utilisées pour caractériser les modes de vibration des liaisons chimiques au sein d'une molécule. Si l'on souhaite obtenir des informations sur les modes de rotation d'une molécule, il est nécessaire de choisir les micro-ondes comme gamme spectrale. Pour l'analyse de pièces de bois, c'est la spectroscopie PIR qui est la plus largement utilisée.

Le rayonnement IR est subdivisé en trois régions : les domaines du proche, du moyen et du lointain IR. En PIR, la gamme de longueurs d'onde du spectre électromagnétique retenue varie entre 800 et 2 500 nm (Figure 1.5).

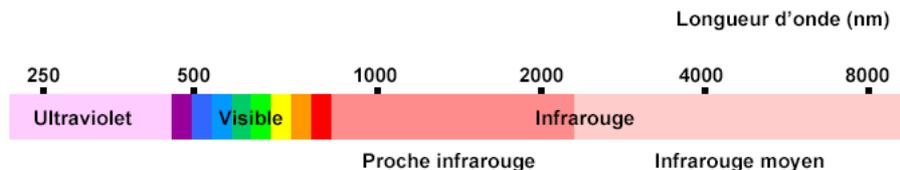


FIGURE 1.5 : Spectre électromagnétique

Image extraite de : www.geoportail.gouv.fr

La spectroscopie IR mesure la transmittance ou la réflectance des rayons IR par un échantillon. Pour la plupart des applications sur le bois, les spectres IR sont collectés selon le mode de réflectance. En effet, le bois est un matériau épais et opaque qui empêche le rayonnement de le traverser [Leblon 13]. En réflectance ou en transmittance, la spectroscopie IR est une spectroscopie vibrationnelle d'absorption (contrairement à la spectroscopie Raman, par exemple, qui repose sur l'émission). Chaque photon du rayonnement électromagnétique incident, est, en fonction de sa longueur d'onde, à l'origine des vibrations des molécules. Si la fréquence de l'onde incidente coïncide avec la fréquence de vibration d'un des modes normaux de la molécule, alors une partie de l'énergie de l'onde électromagnétique est absorbée par la matière et seule une petite partie est réfléchi.

Une molécule composée de N atomes possède $3N$ degrés de liberté ; en effet, la molécule a trois mouvements de translation, selon les trois dimensions et trois mouvements de rotation autour de son centre de gravité. Lorsque cette dernière reçoit un rayonnement, des mouvements complexes entre les atomes vont se produire : ces mouvements sont la combinaison de $3N - 6$ degrés de liberté pour lesquels les atomes vibrent selon des mouvements harmoniques simples nommés *modes normaux de vibration* [Colthup 12]. Dans le cas d'une molécule linéaire, comme le dioxyde de carbone (CO_2), $3N - 5$ modes normaux de vibration sont observés comme le schématise la Figure 1.6.

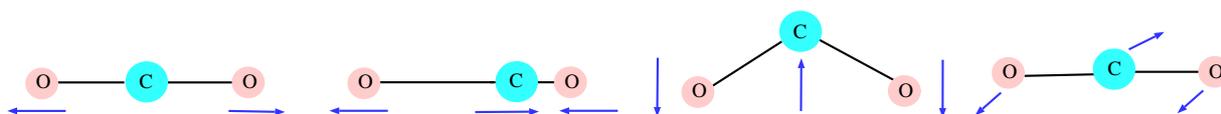


FIGURE 1.6 : Les quatre modes normaux de vibration du CO_2 ; de gauche à droite : élongation symétrique, élongation anti-symétrique, déformation de l'angle dans le plan, déformation de l'angle perpendiculaire au plan

Image extraite de : [Mazet 05]

En d'autres termes, un photon dont l'énergie est égale à l'énergie nécessaire à la molécule pour passer d'un état de basse énergie à un état excité sera absorbé [Osborne 06]. Une bande d'absorption sera donc observée sur le spectre à une énergie caractéristique de la liaison et du mouvement de vibration. Ainsi, un spectre IR peut être interprété comme une « image » des molécules présentes dans l'échantillon.

Suivant les différentes régions du domaine IR utilisées, des phénomènes différents sont observés. Dans l'IR lointain, ce sont les rotations moléculaires qui sont étudiées tandis que dans l'IR moyen ce sont les vibrations des molécules composées d'atomes lourds (et les rotations-vibrations). Les bandes d'ab-

sorption observées dans le PIR sont des harmoniques ou des combinaisons de bandes fondamentales observées dans l'infrarouge moyen. Elles sont plus larges que dans le moyen infrarouge et peuvent donc être plus difficiles à interpréter. Les bandes d'absorption les plus importantes présentes dans la région PIR sont liées aux groupes fonctionnels -CH, -NH et -OH (Figure 1.7).

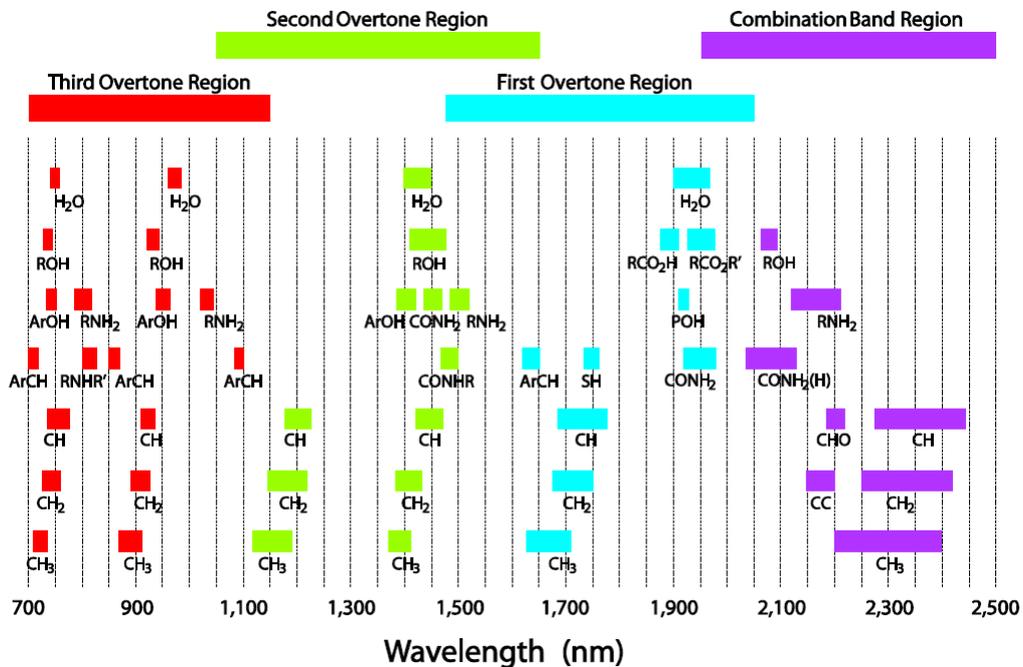


FIGURE 1.7 : Liaisons chimiques à différentes bandes d'absorption dans le proche infrarouge

Image extraite de : www.ipgphotonics.com

1.2.2 Instrumentation

Le spectrographe est le matériel de mesure utilisé en spectroscopie. Le spectrographe est généralement composé d'un dispositif de dispersion (un prisme ou un réseau) permettant de décomposer la lumière et d'un capteur linéaire (Figure 1.8). Ce dernier, composé de cellules photosensibles, permet de convertir un rayonnement électromagnétique en un signal électrique analogique.

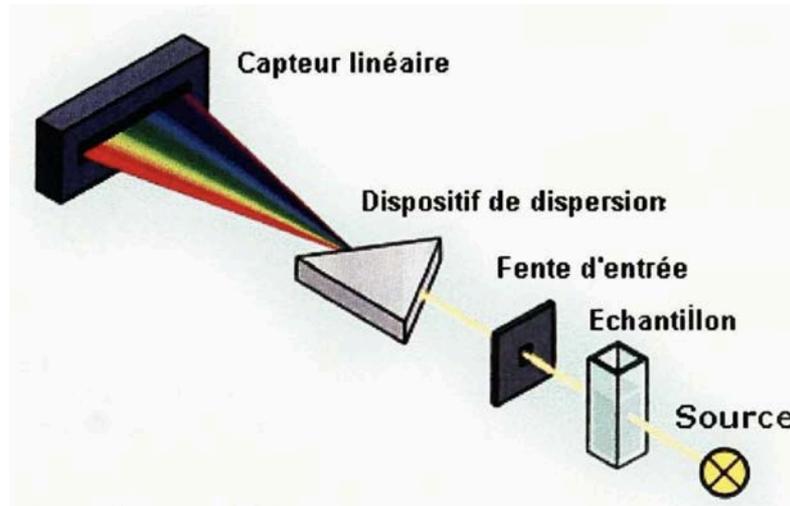


FIGURE 1.8 : Principaux composants d'un spectrographe

Image extraite de : [Girod 08]

Suivant la gamme spectrale désirée, les cellules photosensibles utilisées ne seront pas les mêmes ; si les cellules à Dispositif à Transfert de Charge (CCD pour *Charged Coupled Device*) sont efficaces par exemple dans le domaine du visible, elles sont peu sensibles dans le proche infrarouge. On leur préfère alors des cellules dites InGaAs (*Indium Gallium Arsenide*) beaucoup plus adaptées, mais aussi plus onéreuses [Qin 17]. Concernant la source lumineuse, ce sont les lampes halogènes qui sont le plus largement utilisées dans le domaine PIR, grâce à leur spectre plat dans cette gamme de longueurs d'onde.

1.2.3 Spectroscopie proche infrarouge pour l'analyse du bois

La spectroscopie PIR est un outil puissant pour l'analyse qualitative et quantitative de molécules présentes dans un échantillon. Mais qu'en est-il de son utilisation pour l'analyse du matériau bois ? Tout d'abord, nous allons nous intéresser à la composition chimique de ce dernier. La composition chimique élémentaire du bois varie suivant les espèces, mais globalement, le bois est constitué d'environ 50% de carbone (C), 43% d'oxygène (O), 6% d'hydrogène (H), 1% d'azote (N) et 1% de minéraux (principalement Calcium, Potassium, Sodium, Magnésium, Fer, Manganèse) [Pettersen 84]. Parmi ses principaux constituants organiques, nous pouvons citer :

- La lignine (de 20 à 30%) : molécules qui apportent une rigidité, une imperméabilité à l'eau et une grande résistance à la décomposition.
- La cellulose (environ 50%) : molécules qui assurent à la fois la solidité (face à la traction ou à la compression) et l'élasticité.
- L'hémicellulose (15 à 25 %) : molécules proches de la cellulose qui contribuent au renforcement du bois.
- Les extractibles : molécules de faible masse moléculaire dont la nature et la teneur donnent les propriétés spécifiques du bois comme la couleur, l'odeur et la durabilité biologique.

En outre, le matériau bois contient de l'eau. On distingue trois types d'eau dans le bois :

- L'eau libre : elle est contenue dans les cellules (la cellule s'apparente à un tube, son espace intérieur est appelé le vide cellulaire). Elle est facile à extraire du bois ; on parle dans ce cas de ressuyage.
- L'eau liée : elle est à l'intérieur des parois des cellules et ne peut s'extraire que sous forme de vapeur ; il s'agit alors d'un séchage.
- L'eau de constitution : elle entre dans la composition chimique de la matière, elle disparaît lors de la destruction du bois.

La spectroscopie PIR devient de plus en plus importante pour la caractérisation chimique du bois. En effet, elle est rapide, non-destructive et ne requiert aucune transformation du bois avant analyse (comme le broyage, par exemple). Elle a été utilisée pour déterminer les quantités de lignine [Ono 03], de cellulose [McLellan 91], d'hémicellulose [Yao 10] et des extractibles [Gierlinger 02]. La spectroscopie PIR permet également de déterminer les caractéristiques secondaires du bois, qui dépendent directement de sa composition chimique, comme l'angle et la morphologie des fibres qui le composent [Schimleck 02, Jones 05], ses propriétés mécaniques (traction, flexion) [Kelley 04, Kohan 12, Via 13], le bois de compression [Chen 07] et la différenciation aubier/duramen [Stirling 13]. D'autres singularités, plutôt en surface du bois et visibles à l'œil nu peuvent être mises en évidence telles que les nœuds, les poches de résine [Mora 11], les trous d'insectes, la pourriture *etc.* [Sandak 16]. De plus, de nombreuses recherches ont été menées concernant l'influence du taux d'humidité du bois sur sa densité [Hoffmeyer 95, Leblon 13] et ses variations dimensionnelles [Baillères 02] ; plus précisément, [Berthold 98, Svedas 00, Yang 14] ont étudié l'influence de l'eau libre et de l'eau liée sur les propriétés physico-chimiques du bois. Par exemple, l'évacuation de l'eau libre ne modifie pas les dimensions et les propriétés du bois, exceptée la densité (elle diminue lorsque le taux d'humidité diminue). En revanche, l'évacuation de l'eau liée est bien plus délicate et peut mener à des déformations (que l'on appelle retrait) et même des ruptures du bois, en particulier si le processus de séchage est mal conduit. Ainsi, l'étude du taux d'humidité est un réel enjeu. La Figure 1.9 représente l'attribution des bandes d'absorption dans le proche infrarouge pour les principaux constituants organiques du bois. Pour davantage de détails, le lecteur peut se référer à [Schwanninger 11].

La Figure 1.10 représente des spectres de réflectance en proche infrarouge de différentes essences de bois ; ces derniers apportent, entre autres, des informations quant au positionnement des groupements relatifs à la cellulose, à la lignine et à l'eau. Bien que la plupart des spectres soient corrélés, des différences significatives sont constatées permettant ainsi de différencier les différentes essences. De plus, les spectres de réflectance en proche infrarouge fournissent de riches informations pour la discrimination de l'aubier, par exemple. Selon diverses études [Bertaud 04, Sandberg 09], l'aubier contient moins de lignine, plus de cellulose et présente un taux d'humidité plus élevé que le duramen. En effet, le duramen est souvent significativement plus sec que l'aubier et cette différence de teneur en eau peut être observée sur leur spectre respectif.

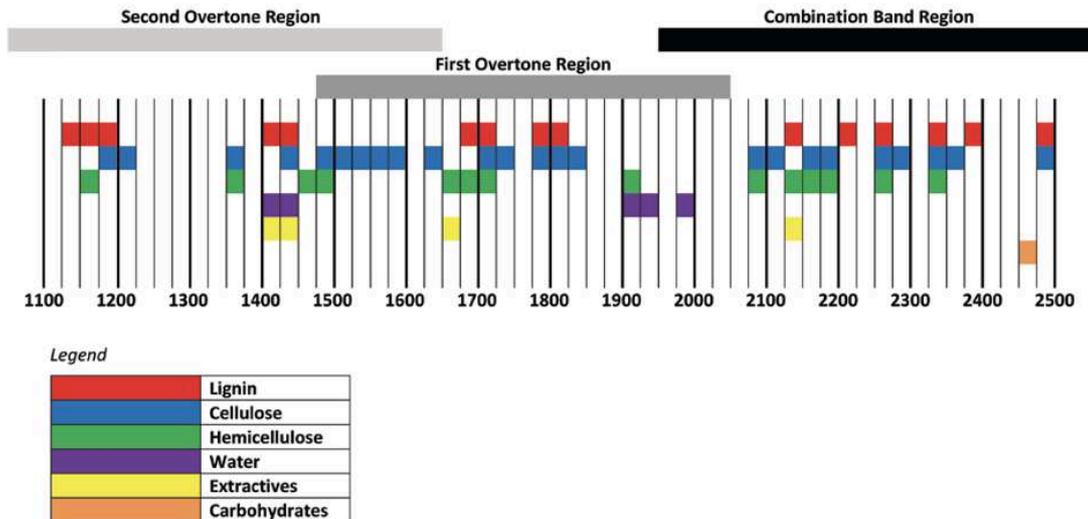


FIGURE 1.9 : Bandes d'absorption PIR (1100-2500 nm) pour les principaux composants du bois

Image extraite de : [Leblon 13]

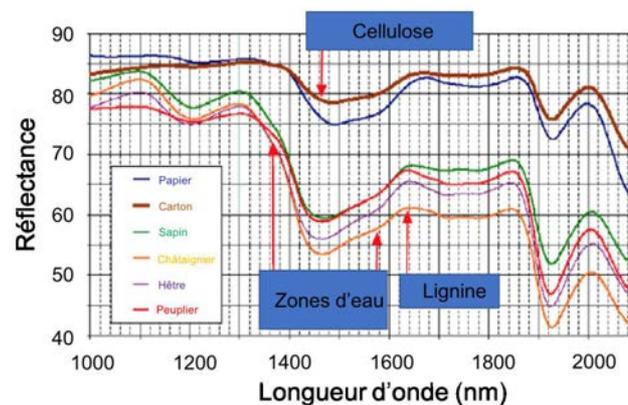


FIGURE 1.10 : Spectres proche infrarouge pour différentes essences de bois

Image extraite de : [Belmerhnia 17]

1.3 L'imagerie hyperspectrale

1.3.1 Principe de l'imagerie hyperspectrale

À la différence de la spectroscopie, qui fournit uniquement une mesure spectrale, l'imagerie hyperspectrale intègre une dimension spatiale à la mesure. C'est en quelque sorte l'union des techniques de spectroscopie et d'imagerie. Les données hyperspectrales, illustrées sur la Figure 1.11, peuvent être vues comme une fonction de trois variables : deux variables spatiales x et y donnant la position du pixel dans l'image et une variable spectrale L correspondant à la longueur d'onde de la lumière analysée. À chaque pixel de l'image est associé une signature spectrale, représentée généralement par la réflectance (ou encore l'absorbance) en fonction des longueurs d'onde. Ainsi, il est possible d'obtenir la répartition

spatiale d'une molécule, d'un matériau ou d'un tissu grâce à sa signature spectrale unique. Les images hyperspectrales sont donc des données agencées en cube.

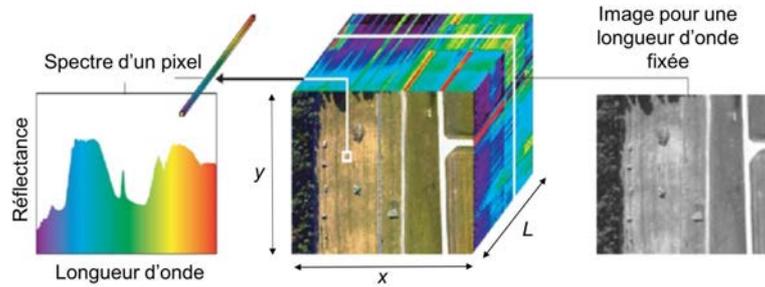


FIGURE 1.11 : Structure d'une image hyperspectrale

Image extraite de : [Ravel 17]

Le terme hyperspectral fait référence au nombre de bandes spectrales qu'il est possible d'acquérir sur le spectre électromagnétique. On différencie généralement les images hyperspectrales des images monochromes et multispectrales. Comme nous pouvons le voir sur la Figure 1.12, les images monochromes sont acquises suivant une unique bande spectrale (couvrant une large gamme de longueurs d'onde) et fournissent l'intensité lumineuse mesurée pour chaque pixel de l'image. Les images multispectrales sont acquises sur un faible nombre de bandes spectrales, généralement irrégulièrement espacées (3 en imagerie couleur et ≈ 10 en imagerie multispectrale). Les images hyperspectrales, quant à elles, sont acquises suivant plusieurs centaines de longueurs d'onde, de même largeur et régulièrement espacées (on parle communément de bandes spectrales étroites et contiguës), permettant d'obtenir un spectre quasi continu pour chaque pixel.

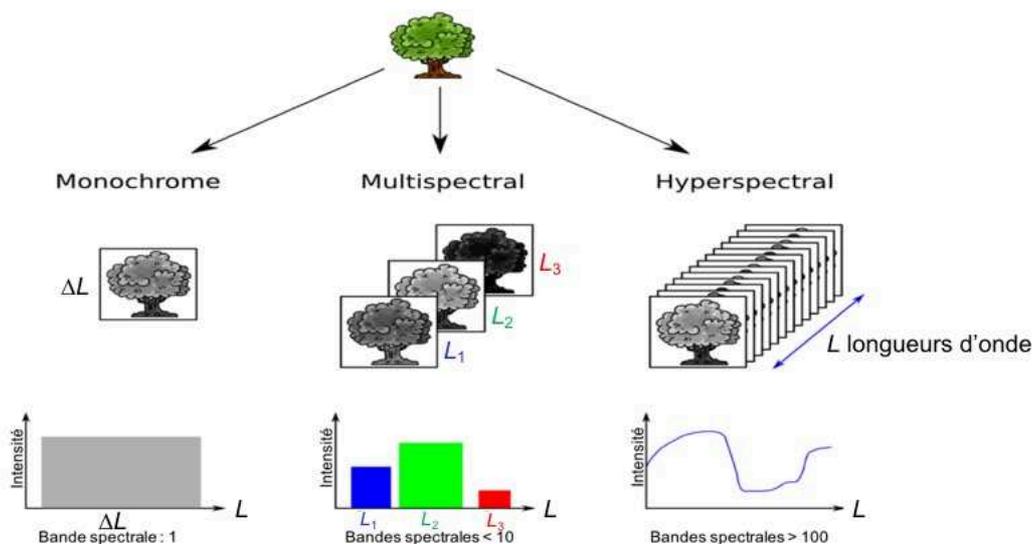


FIGURE 1.12 : De l'imagerie monochrome à l'hyperspectrale

Image extraite de : [Boulais 17]

Les images hyperspectrales sont acquises par des capteurs de plus en plus performants. Trois caractéristiques principales permettent de quantifier les performances de ces capteurs, et donc, d'évaluer la capacité des données obtenues à restituer l'information du milieu observé :

- La résolution spatiale : cette caractéristique permet, comme pour les images conventionnelles, d'évaluer la finesse des détails d'une image. Elle désigne l'aptitude du capteur à distinguer de petits objets adjacents.
- La résolution spectrale : cette caractéristique permet d'évaluer la finesse du spectre obtenu pour chaque pixel. Elle désigne l'aptitude du capteur à distinguer deux longueurs d'onde proches de la lumière qui lui parvient. Elle quantifie la distance entre deux bandes spectrales adjacentes.
- La résolution radiométrique : cette caractéristique permet d'évaluer la finesse de l'intensité mesurée en chaque point du spectre. Elle désigne l'aptitude du capteur à discriminer des petites variations d'énergie électromagnétique. C'est le nombre de bits utilisés pour quantifier l'image qui régit le nombre maximum de niveaux d'intensité disponibles.

1.3.2 Techniques d'acquisition en imagerie hyperspectrale

Les techniques d'acquisition en imagerie hyperspectrale se regroupent en trois catégories [Lu 14] :

1. Les méthodes à balayage spatial.
2. Les méthodes à balayage spectral.
3. Les méthodes d'acquisition sans balayage dites instantanées.

1.3.2.1 Balayage spatial

Les méthodes à balayage spatial consistent à acquérir en une seule fois toute l'information spectrale d'un ou plusieurs pixels. L'acquisition de la scène à imager se fait soit ligne par ligne (méthode *pushbroom*) ou pixel par pixel (méthode *whiskbroom*) en déplaçant soit l'imageur soit son champ de vue (FOV pour *Field Of View*). Le système d'imagerie *whiskbroom* utilise un capteur linéaire pour former le cube de données. Comme le montre la Figure 1.13, la scène est observée pixel par pixel le long des deux dimensions spatiales et s'effectue par le biais d'un miroir rotatif. La lumière réfléchiée est ensuite projetée à travers un élément dispersif et enregistrée sur une barrette CCD, dont chaque cellule correspond à une longueur d'onde. La scène est donc observée point par point selon les deux dimensions spatiales (x et y), ce qui engendre un ensemble de spectres à empiler pour former le cube hyperspectral. La durée d'acquisition d'une scène est généralement longue car le balayage se fait selon chaque dimension spatiale x et y .

La technique d'acquisition *pushbroom*, à la différence de la méthode *whiskbroom* est une technique de balayage linéaire qui permet d'acquérir simultanément la totalité d'une ligne de la scène, ce qui génère une tranche bidimensionnelle de l'image hyperspectrale qui a pour dimension spatiale x et pour dimension spectrale L (Figure 1.14). La lumière réfléchiée est dispersée sur une matrice CCD et l'ensemble du cube hyperspectral est obtenu en déplaçant l'imageur le long de l'autre dimension spatiale y ou alors en déplaçant directement les échantillons à imager sur un convoyeur. Par rapport au système d'imagerie *whiskbroom*, le temps d'acquisition de la scène est généralement plus court.

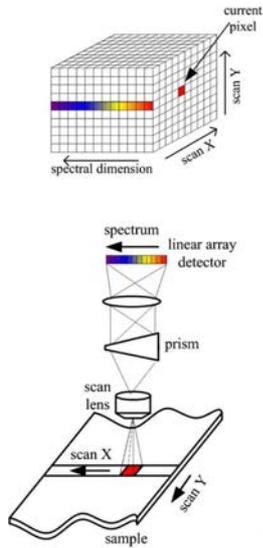


FIGURE 1.13 : Système d'acquisition *whiskbroom*

Image extraite de : [Li 13]

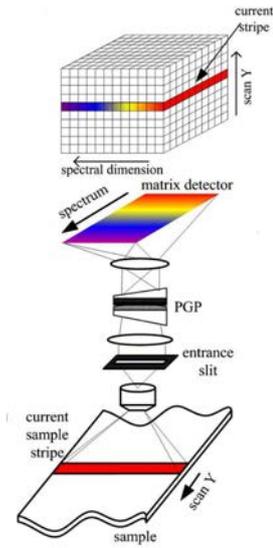


FIGURE 1.14 : Système d'acquisition *pushbroom*

Image extraite de : [Li 13]

1.3.2.2 Balayage spectral

Les méthodes à balayage spectral (ou *staring*) consistent à acquérir successivement une zone à imager suivant une longueur d'onde précise comme le montre la Figure 1.15. Chaque longueur d'onde est sélectionnée par un filtre optique passe-bande, le filtre optique est alors changé pour capturer l'image à la longueur d'onde suivante et ainsi de suite jusqu'à l'acquisition complète du cube hyperspectral. Cette méthode nécessite d'avoir des filtres optiques très sélectifs.

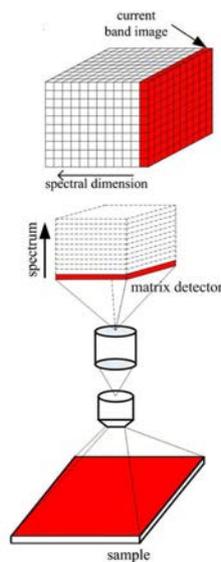


FIGURE 1.15 : Système d'acquisition à balayage spectral

Image extraite de : [Li 13]

1.3.2.3 Acquisition instantanée

L'imagerie hyperspectrale instantanée (ou *snapshot*) permet l'acquisition simultanée de l'information spatiale et spectrale d'une zone, en une seule exposition et sans balayage. Comme le montre la Figure 1.16, la technique d'imagerie *snapshot* capture la totalité du cube hyperspectral en procédant au « découpage » de l'image en bandes et en réorganisant ces bandes en une seule ligne à l'aide d'un système optique complexe [Hagen 12]. Cette technique est beaucoup plus complexe à utiliser que les méthodes à balayage pour des résolutions moindres ce qui restreint son utilisation.

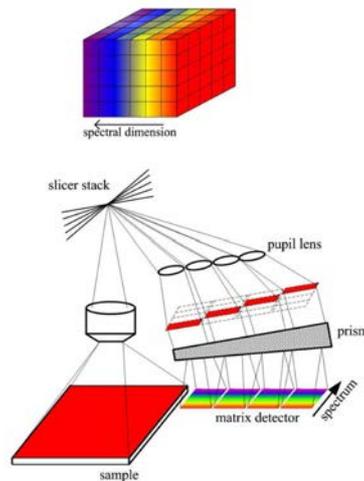


FIGURE 1.16 : Système d'acquisition instantané

Image extraite de : [Li 13]

1.3.3 Domaines d'application de l'imagerie hyperspectrale

L'imagerie hyperspectrale est issue de la télédétection [Wolfe 97] et a été explorée pour diverses applications allant de l'observation de la Terre, à l'espace et à l'astronomie [Hege 04, Goetz 09, Berné 10]. Elle est également utilisée pour des applications de « proximité » ; nous entendons par « proximité » toute exploitation ne nécessitant pas un capteur aéroporté ou satellisé. Parmi ces applications, nous pouvons citer, sans être exhaustifs :

- L'agriculture [Lleó 11] : surveillance des cultures, évaluation de la maturation des fruits.
- L'écologie [Abd-Elrahman 11] : qualité des eaux, détection de polluants.
- L'étude d'oeuvres d'art [De Viguerie 16] : identification de pigments, datation de tableaux.
- La médecine [Lu 14] : détection de tumeurs, vision améliorée du champ opératoire.
- La biologie [Zimmermann 03] : détection de tissus sains ou pathologiques.

Un autre domaine d'application très prometteur est l'analyse de matériaux, en particulier le contrôle, la qualité et la sécurité [Gowen 07, Feng 12, Qin 17]. Ces types d'applications sont qualifiés d'imagerie hyperspectrale industrielle. D'un point de vue technologique, l'imagerie hyperspectrale industrielle soulève de nouveaux défis. La question qui se pose le plus souvent est la suivante : est-il possible de développer des méthodes de traitement de données rapides pouvant être utilisées dans des systèmes industriels d'imagerie hyperspectrale en temps réel ? Dans le cadre du projet OPTIFIN, les diverses pièces de bois à analyser sont acheminées sur un convoyeur. Ainsi, la technique d'acquisition par balayage spatial *pushbroom* convient naturellement pour inspecter les matériaux en mouvement. Nous proposons donc de traiter le problème de démélange hyperspectral ligne par ligne (donc tranche par tranche), de manière **réursive** et en **temps réel**. Dans ce contexte, la récursivité doit agir le long d'une des dimensions spatiales de l'image, que l'on assimilera au temps.

1.3.4 Présentation du système d'imagerie hyperspectrale *pushbroom*

Le système d'imagerie hyperspectrale mis en place dans le cadre de ces travaux est composé d'un système d'éclairage (l'échantillon est éclairé de manière diffuse par des lampes halogènes), d'une caméra hyperspectrale comprenant un spectrographe et d'un ordinateur doté des logiciels adéquats pour l'acquisition et le traitement des données. L'ensemble spectrographe/caméra reste fixe ; l'échantillon se déplace sous l'imager à l'aide d'un rail linéaire, comme le montre la Figure 1.17.

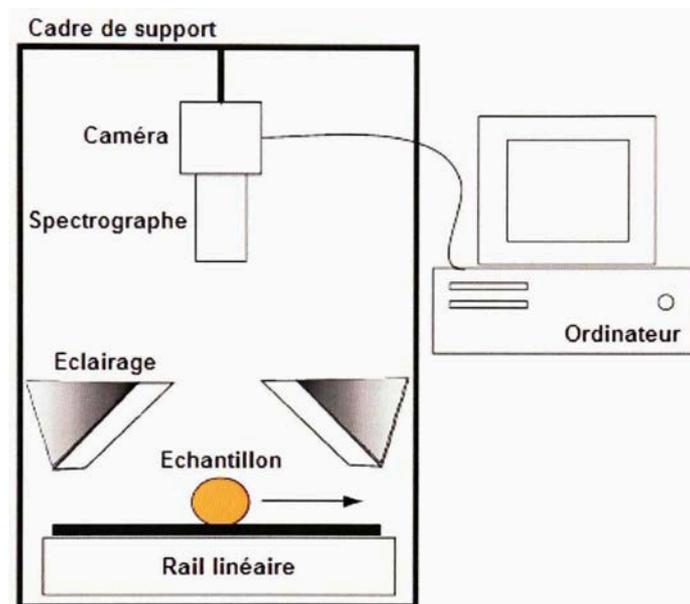


FIGURE 1.17 : Composants du système d'imagerie hyperspectrale

Image extraite de : [Girod 08]

Dans le cadre de notre étude, nous avons choisi la caméra hyperspectrale PIR *pushbroom* FX17 de chez Specim, représentée sur la Figure 1.18. Ses principales caractéristiques techniques sont regroupées dans le Tableau 1.1.



FIGURE 1.18 : Imageur *pushbroom* FX17 de chez Specim

Image extraite de : <http://www.specim.fi>

Gamme spectrale	900 à 1700 nm
Nombre de bandes spectrales	224
Champs de vue	38°
Type de capteur	InGaAs
Taille d'un pixel	15 × 15 μm
Échantillonnage spatial	640 pixels
Résolution spectrale	8 nm
Fréquence d'images	670 images/seconde

TABLE 1.1 : Caractéristiques de l'imageur FX17

1.4 Démélange d'image hyperspectrale

Dans l'idéal, chaque pixel d'une image hyperspectrale est caractérisé par une signature spectrale relative à un matériau unique. Dans la pratique, en raison de la faible résolution spatiale des capteurs utilisés ou en raison des propriétés physiques de la surface étudiée, il arrive souvent qu'un seul pixel contienne plusieurs matériaux ; on le caractérise alors comme étant un « pixel mélangé », ou encore « mixel », par opposition aux pixels purs qui ne contiennent qu'un seul composant. Les « mixels » présentent une signature spectrale ne correspondant à aucun matériau naturel ou artificiel présent dans l'image, mais plutôt à un mélange des spectres des différents matériaux purs présents dans le pixel (Figure 1.19).

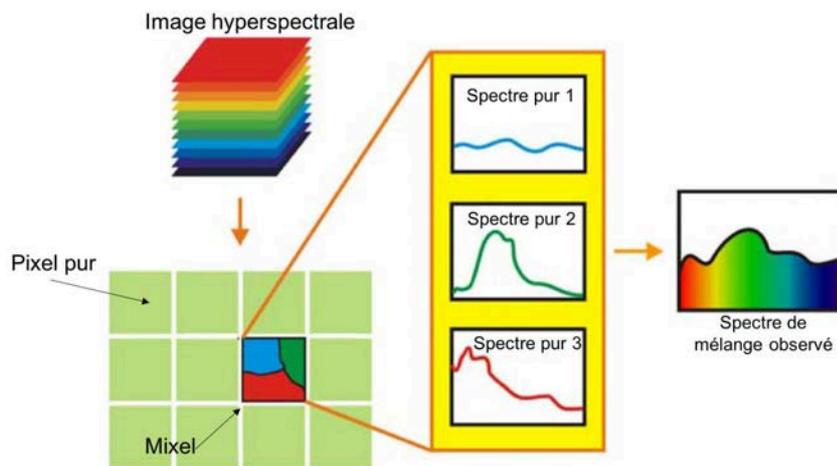


FIGURE 1.19 : Pixels purs et pixel mélangé au sein d'une image hyperspectrale

Image extraite de : [Zhang 14]

Afin d'analyser ce type d'images, il est nécessaire d'effectuer un « démélange hyperspectral ». Cette procédure, sur laquelle est fondé notre travail, permet la décomposition d'un spectre de pixel mélangé en un ensemble de spectres de matériaux purs, appelés *endmembers*, et un ensemble d'abondances indiquant la quantité de chaque composant spectral dans le pixel considéré. Pour résoudre ce problème, on utilise un modèle de mélange qui décrit la manière dont sont combinés les *endmembers* au sein de l'image.

1.4.1 Modèle de mélange linéaire

1.4.1.1 Formulation du problème

Toutes les méthodes de démélange reposent sur un modèle de mélange prédéfini. Selon l'application et/ou les données utilisées, celui-ci peut être linéaire ou non-linéaire. Le modèle linéaire est le plus couramment utilisé car il s'agit d'un modèle relativement simple, dont les applications sont nombreuses. Ce modèle convient lorsque les composants purs sont disposés côte à côte dans le pixel ; il est réaliste lorsque la surface imagée est relativement plane et uniformément éclairée (Figure 1.20). Dans le cas contraire, c'est à dire lorsque la scène observée est un mélange intime de matériaux purs, engendrant des effets de dispersion, le modèle non-linéaire est le plus adéquat pour représenter le mélange spectral (Figure 1.21).

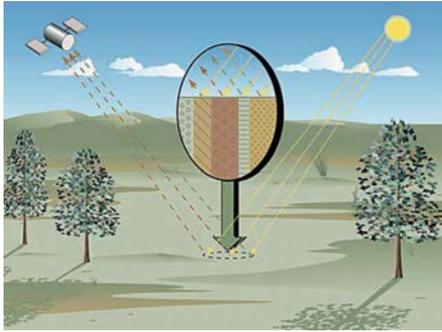


FIGURE 1.20 : Principe du modèle linéaire

Image extraite de : [Keshava 02]

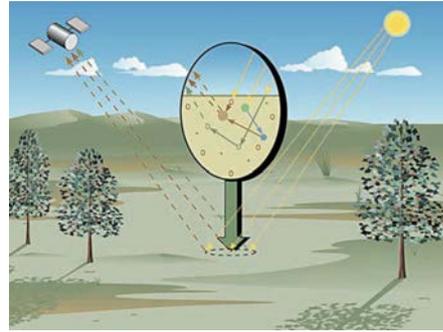


FIGURE 1.21 : Principe du modèle non-linéaire

Image extraite de : [Keshava 02]

Dans la suite de cette thèse, nous considérons uniquement le modèle de mélange linéaire. Ce modèle, qui décrit le cas de matériaux spatialement séparés, peut être vu comme une approximation au premier ordre d'une transformation non-linéaire à l'horizon d'observation suffisamment petit. Dans le cas du modèle linéaire, le spectre observé au niveau d'un mixel est un spectre linéairement mélangé et il représente une moyenne pondérée (combinaison linéaire) des différents spectres des matériaux purs présents dans ce mixel. La pondération affectée à chaque spectre d'un matériau pur représente l'abondance de ce dernier dans le mixel.

Selon le modèle de mélange linéaire, sans bruit additif, le spectre $\mathbf{x}_p = [x_1, \dots, x_L]^T$ d'un pixel mélangé p , où L est le nombre de longueurs d'onde, s'exprime comme la combinaison linéaire de R signatures spectrales pures \mathbf{s}_r , $r = 1, \dots, R$ tel que :

$$\mathbf{x}_p = \sum_{r=1}^R \mathbf{s}_r a_{rp}, \quad (1.1)$$

où \mathbf{s}_r est le spectre du $r^{\text{ème}}$ *endmember* et le scalaire a_{rp} est l'abondance du $r^{\text{ème}}$ *endmember* dans le pixel p . Si l'on considère tous les pixels P de l'image, l'équation (1.1) peut s'écrire sous la forme matricielle suivante :

$$\mathbf{X} = \mathbf{S}\mathbf{A}, \quad (1.2)$$

où $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_P] \in \mathbb{R}_+^{L \times P}$ est la matrice des observations, $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_R] \in \mathbb{R}_+^{L \times R}$ est la matrice

contenant dans ses colonnes les spectres des *endmembers* et $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_P] \in \mathbb{R}_+^{R \times P}$ est la matrice contenant dans ses colonnes les coefficients d'abondance de chaque *endmember* pour un pixel donné. Pour des raisons d'interprétation physique, les matrices \mathbf{S} et \mathbf{A} sont contraintes à être non-négatives, *i.e.* $\mathbf{S} \geq \mathbf{0}$ et $\mathbf{A} \geq \mathbf{0}$. En effet, la matrice \mathbf{S} contient les spectres de réflectance spécifiques des matériaux de la scène imagée, ses coefficients ne peuvent donc pas être négatifs et la matrice des abondances \mathbf{A} contient les proportions des *endmembers* dans chaque pixel, ses coefficients sont donc eux aussi non-négatifs. De plus, une contrainte peut être imposée sur les colonnes de la matrice \mathbf{A} qui consiste à ce que la somme des abondances, pour chaque pixel de l'image, soit égale à 1.

1.4.1.2 Identifiabilité du modèle

Reprenons le modèle de mélange linéaire de l'équation (1.2). Le problème de démixage hyperspectral devient un problème de factorisation de la matrice \mathbf{X} en deux matrices \mathbf{S} et \mathbf{A} . Cependant, si elle existe, cette factorisation n'est généralement pas unique. En effet, considérons une matrice carrée inversible \mathbf{Q} de dimension $R \times R$. Pour chaque couple (\mathbf{S}, \mathbf{A}) , un nouveau couple $(\bar{\mathbf{S}}, \bar{\mathbf{A}})$ peut être défini par :

$$\begin{aligned}\bar{\mathbf{S}} &= \mathbf{S}\mathbf{Q}, \\ \bar{\mathbf{A}} &= \mathbf{Q}^{-1}\mathbf{A},\end{aligned}$$

vérifiant lui aussi $\mathbf{X} = \bar{\mathbf{S}}\bar{\mathbf{A}}$. Ainsi, si une factorisation existe, il existe en général une infinité, et donc plusieurs mélanges différents peuvent conduire aux mêmes observations. Le modèle de mélange spectral est donc *non-identifiable*. Trouver une solution au problème, *i.e.* trouver une factorisation de \mathbf{X} en un produit de matrices non-négatives, ne garantit aucunement que cette factorisation soit la solution du problème de démixage hyperspectral. D'un point de vue applicatif, cette non-unicité est problématique en termes d'interprétabilité. La contrainte de non-négativité sur les matrices $\bar{\mathbf{S}}$ et $\bar{\mathbf{A}}$ limite le nombre de possibilités pour la matrice de transformation \mathbf{Q} et dans certains cas, cette contrainte rend la décomposition unique ; on dit alors que le modèle est identifiable. Formellement, on définit l'identifiabilité du modèle linéaire de la façon suivante :

Définition 1 [Identifiabilité du modèle de factorisation]

La factorisation $\mathbf{X} = \mathbf{S}\mathbf{A}$ est identifiable si $\forall \bar{\mathbf{S}}, \bar{\mathbf{A}}$ tel que $\mathbf{X} = \bar{\mathbf{S}}\bar{\mathbf{A}}$, nous avons :

$$\begin{aligned}\bar{\mathbf{S}} &= \mathbf{S}\mathbf{\Omega}\mathbf{E}, \\ \bar{\mathbf{A}} &= \mathbf{E}^{-T}\mathbf{\Omega}^T\mathbf{A},\end{aligned}$$

où $\mathbf{\Omega}$ est une matrice de permutation et \mathbf{E} est une matrice diagonale inversible de rang plein avec les éléments diagonaux positifs (matrice de mise à l'échelle). En d'autres termes, la factorisation est considérée unique si elle l'est à une matrice de permutation et/ou d'échelle près.

Les contraintes de non-négativité sur les matrices \mathbf{S} et \mathbf{A} permettent d'appréhender l'identifiabilité du problème de démixage hyperspectral sous un nouvel angle. En effet, les conditions d'unicité relatives au modèle $\mathbf{X} = \mathbf{S}\mathbf{A}$ peuvent s'interpréter de manière géométrique. Tous les spectres \mathbf{x}_p avec $p = 1, \dots, P$ peuvent être représentés dans un cône convexe :

$$\mathbf{x}_p \in \text{cône} \{\mathbf{S}\}, p = 1, \dots, P, \quad (1.3)$$

où $\text{cône} \{\mathbf{S}\} = \{\mathbf{x} \in \mathbb{R}^L | \mathbf{x} = \mathbf{S}\boldsymbol{\theta}, \boldsymbol{\theta} \geq \mathbf{0}\}$ représente le cône convexe dont les arrêtes sont les colonnes de \mathbf{S} . Il existe une autre interprétation géométrique très utile du modèle mais qui nécessite, en plus de la non-négativité, la contrainte de somme à 1 des colonnes de la matrice \mathbf{A} i.e. $\mathbf{A}^T \mathbf{1} = \mathbf{1}$, où $\mathbf{1}$ est un vecteur composé uniquement de 1. Si cette hypothèse est valide, (1.3) peut s'écrire à nouveau comme :

$$\mathbf{x}_p \in \text{conv} \{\mathbf{S}\}, p = 1, \dots, P,$$

où $\text{conv} \{\mathbf{S}\} = \{\mathbf{x} \in \mathbb{R}^L | \mathbf{x} = \mathbf{S}\boldsymbol{\theta}, \boldsymbol{\theta} \geq \mathbf{0}, \boldsymbol{\theta}^T \mathbf{1} = 1\}$ représente l'enveloppe convexe des colonnes de \mathbf{S} . Si $\text{rang}(\mathbf{S}) = R$, alors les colonnes de \mathbf{S} sont les sommets de cette enveloppe convexe. Cette enveloppe est également appelée simplexe à condition que les colonnes de \mathbf{S} soient affinement indépendantes, i.e., que les vecteurs $\{\mathbf{s}_r - \mathbf{s}_R\}_{r=1, \dots, R-1}$ soient linéairement indépendants. La Figure 1.22 montrent les deux interprétations géométriques possibles.

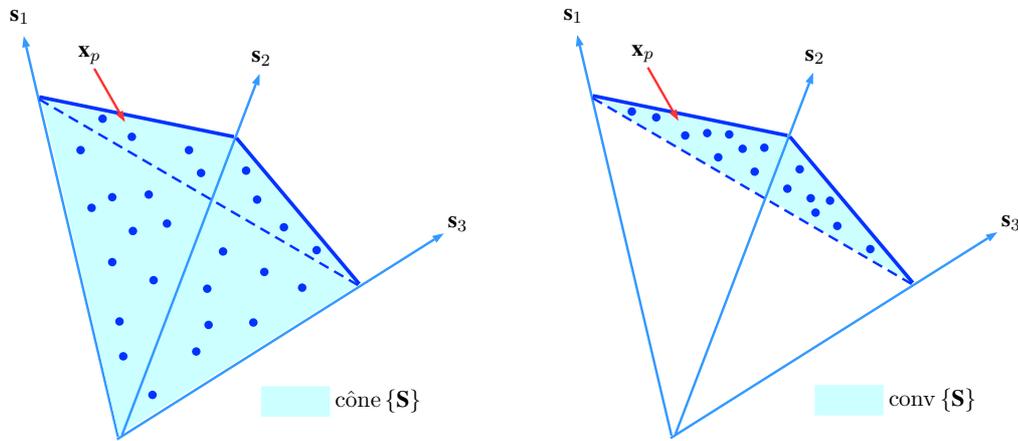


FIGURE 1.22 : Représentation géométrique pour $R = 3$. Version non normalisée (à gauche) : cône convexe. Version normalisée (à droite) : simplexe. Les points représentent les \mathbf{x}_p

Ainsi, il est évident que le problème de démélange hyperspectral est analogue à la recherche des sommets d'un simplexe. Néanmoins, malgré les contraintes de non-négativité qui restreignent le champ des solutions possibles, les autres formes d'indétermination, que l'on peut assimiler à des indéterminations de rotation ne sont pas sans conséquences puisqu'elles peuvent mettre en échec le démélange hyperspectral. Considérons $\mathbf{S} \geq \mathbf{0}$ et $\mathbf{A} \geq \mathbf{0}$; alors $\mathbf{x}_p \in \mathbb{R}_+^L \forall p$. Ainsi, toute matrice $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1, \dots, \bar{\mathbf{s}}_R]$ qui satisfait $\text{conv} \{\mathbf{X}\} \subseteq \text{conv} \{\bar{\mathbf{S}}\} \subseteq \mathbb{R}_+^L$ est une solution du problème de factorisation de \mathbf{X} tel que $\mathbf{X} = \bar{\mathbf{S}}\bar{\mathbf{A}}$, avec $\bar{\mathbf{A}}^T \mathbf{1} = \mathbf{1}$. La Figure 1.23 illustre l'existence de plusieurs solutions d'un point de vue géométrique.

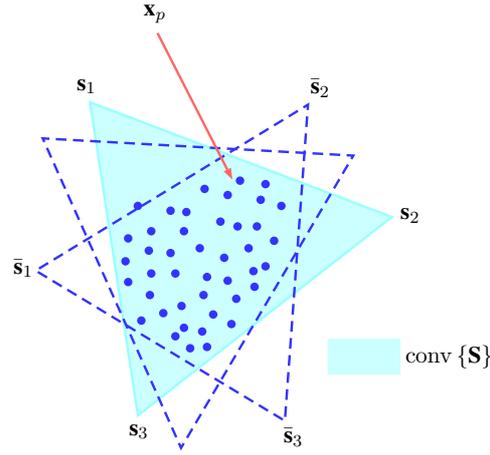


FIGURE 1.23 : Illustration d'un problème non-unique. Les simplexes en pointillé représentent $\text{conv}\{\bar{\mathbf{S}}\}$ et englobent toutes les données \mathbf{x}_p

D'après la Figure 1.23, on remarque que si les données \mathbf{x}_p sont « suffisamment dispersées » dans le simplexe $\text{conv}\{\mathbf{S}\}$ et qu'il n'y a pas de spectres purs, alors le nombre de simplexes $\text{conv}\{\bar{\mathbf{S}}\}$ susceptible de contenir l'ensemble des données diminue. Le cas extrême est quand il n'existe qu'un seul simplexe $\text{conv}\{\mathbf{S}\}$ qui peut englober toutes les données. Dans ce cas, on dit que le modèle est identifiable ou que la factorisation est unique. À partir de ces observations, des conditions d'identifiabilité ont été proposées et démontrées dans la littérature. Elles sont fondées sur les notions de *séparabilité* et de *dispersion suffisante*.

Définition 2 [Donoho 04, Laurberg 08] [*Séparabilité - pixels purs*]

Une matrice non-négative $\mathbf{A} \in \mathbb{R}_+^{R \times P}$ satisfait la condition de séparabilité si :

$$\text{cône}\{\mathbf{A}\} = \text{cône}\{\mathbf{e}_1, \dots, \mathbf{e}_R\} = \mathbb{R}_+^R$$

où \mathbf{e}_r est le $r^{\text{ème}}$ vecteur canonique de \mathbb{R}_+^R . De façon équivalente, \mathbf{A} satisfait la condition de séparabilité si pour tout $r = 1, \dots, R$, il existe un indice p_r tel que :

$$\mathbf{A}(:, p_r) = \alpha_r \mathbf{e}_r,$$

où $\alpha_r > 0$ est un scalaire. Lorsque la contrainte de somme à 1 est imposée sur la matrice \mathbf{A} , alors $\alpha_r = 1$.

C'est une condition relativement forte car cela signifie que certaines colonnes de la matrice \mathbf{A} touchent les vecteurs canoniques de l'orthant positif, comme on peut le voir sur la Figure 1.24. De manière correspondante, dans le domaine des données, on constate que $\mathbf{x}_{p_r} = \alpha_r \mathbf{s}_r$ existe pour tout $r = 1, \dots, R$, i.e. certains points \mathbf{x}_p touchent les angles extrêmes de $\text{cône}\{\mathbf{S}\}$. Cela revient à supposer que chaque *endmember* au sein de l'image hyperspectrale est représenté par au moins un pixel pur.

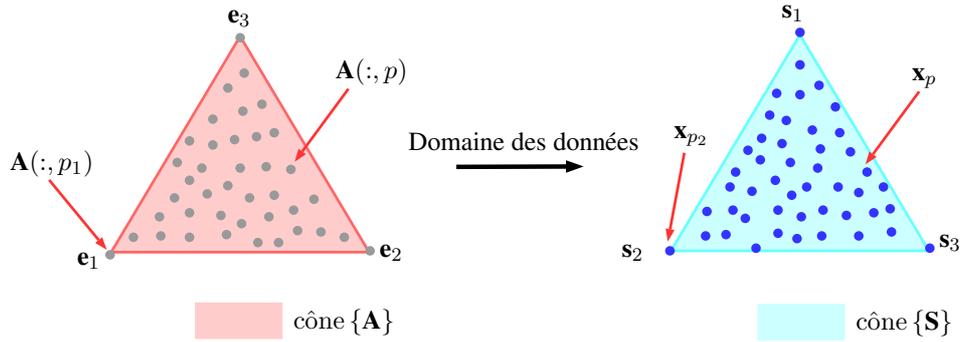


FIGURE 1.24 : Illustration de la condition de séparabilité dans le cas où $R = 3$. À gauche, le domaine des \mathbf{A} et à droite le domaine des \mathbf{X} . On suppose que l'on se situe face aux cônes (vers l'origine). Les points \mathbf{x}_p situés au centre du cône font référence aux pixels mélangés

Définition 3 [Laurberg 08, Huang 14, Fu 15, Fu 18a] [*Dispersion suffisante*]

Une matrice non-négative $\mathbf{A} \in \mathbb{R}_+^{R \times P}$ est considérée comme suffisamment dispersée si les deux conditions suivantes sont satisfaites :

1. Les colonnes de \mathbf{A} sont assez dispersées pour que : $\mathcal{C} \subseteq \text{cône}\{\mathbf{A}\}$, où \mathcal{C} est un cône de second ordre défini par $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^R \mid \mathbf{x}^T \mathbf{1} \geq \sqrt{R-1} \|\mathbf{x}\|_2\}$.
2. Pour toute matrice \mathbf{Q} unitaire à l'exception des matrices de permutation, on a $\text{cône}\{\mathbf{A}\} \not\subseteq \text{cône}\{\mathbf{Q}\}$.

Le cône de second ordre \mathcal{C} est un cône qui est tangent à toutes les facettes de l'orthant non-négatif. La première condition signifie que si la matrice \mathbf{A} est suffisamment dispersée, alors toutes les facettes de l'orthant non-négatif sont touchées par au moins l'une de ses colonnes. De manière correspondante, dans le domaine des données, il n'est pas nécessaire que les points \mathbf{x}_p touchent les angles extrêmes définis par \mathbf{s}_r (Figure 1.25).

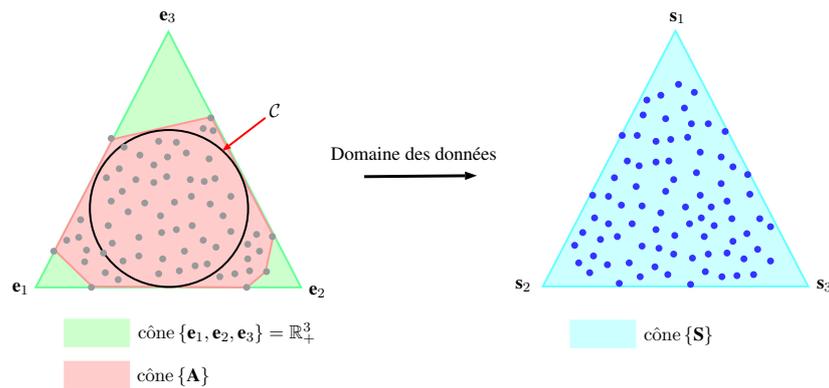


FIGURE 1.25 : Illustration de la condition de dispersion dans le cas où $R = 3$. À gauche, le domaine des \mathbf{A} et à droite le domaine des \mathbf{X}

La deuxième condition est une condition de régularité [Huang 14, Fu 15, Fu 18a] qui stipule que cône $\{\mathbf{A}\}$ doit être légèrement plus grand que \mathcal{C} , pas seulement tangent à \mathcal{C} . La Figure 1.26 montre un cas où la condition de régularité n'est pas satisfaite. La condition de dispersion suffisante est bien moins restrictive que la condition de séparabilité.

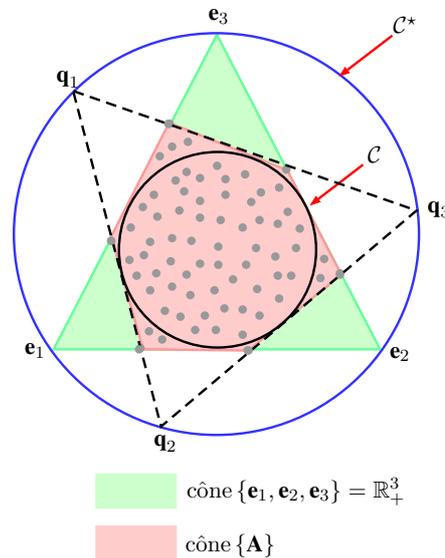


FIGURE 1.26 : Illustration de la condition de régularité non satisfaite. Le cercle extérieur \mathcal{C}^* est le cône dual de \mathcal{C} défini tel que $\mathcal{C}^* = \{\mathbf{x} \in \mathbb{R}^R \mid \mathbf{x}^T \mathbf{1} \geq \|\mathbf{x}\|_2\}$

Une première condition suffisante d'identifiabilité, fondée sur la notion de séparabilité a été proposée par Donoho et Stodden en 2004 [Donoho 04] et sous une forme légèrement différente, par Laurberg *et al.* en 2008 [Laurberg 08]. Le modèle (1.2) est identifiable si :

- La matrice des abondances \mathbf{A} satisfait à la condition de séparabilité - pixels purs (selon la **Définition 2**).
- La matrice des *endmembers* \mathbf{S} contient au moins un élément nul sur chacune de ses colonnes et à des lignes différentes. Géométriquement, cela suppose qu'il y a des lignes de la matrice \mathbf{S} qui touchent les facettes de l'orthant non-négatif.

Finalement, Huang *et al.* [Huang 14] ont démontré une condition suffisante d'identifiabilité moins restrictive, fondée sur la notion de dispersion suffisante. Ils ont montré que la factorisation (1.2) est identifiable si :

- Les matrices \mathbf{S} et \mathbf{A} sont toutes les deux suffisamment dispersées (selon la **Définition 3**).

Toutefois, ces conditions suffisantes ne sont pas toujours vérifiées dans la pratique. En effet, elles nécessitent que les matrices \mathbf{S} et \mathbf{A} contiennent toutes les deux des éléments nuls. Dans certaines applications, telles que l'imagerie hyperspectrale, qui fait l'objet de ce travail de thèse, la matrice \mathbf{S} contenant les signatures spectrales peut ne contenir aucun élément nul. Ainsi, la question que

l'on peut se poser est : est-il possible de garantir l'identifiabilité de \mathbf{S} et \mathbf{A} dans une telle situation ? La réponse est affirmative, néanmoins, cela demande l'ajout de contraintes supplémentaires au problème.

En 1994, Craig a proposé de minimiser le volume du simplexe contenant les données (MVT pour *Minimum Volume Transform*) [Craig 94]. Afin de garantir que le simplexe de volume minimal (que l'on appellera par la suite *VolMin*) permet d'identifier les « bons » *endmembers*, Craig a proposé la condition suivante :

- Si $\mathbf{A} \geq \mathbf{0}$, $\mathbf{A}^T \mathbf{1} = \mathbf{1}$ et les points de données \mathbf{x}_p sont suffisamment dispersés dans le simplexe $\text{conv}\{\mathbf{S}\}$, alors le simplexe ayant le volume minimal permet d'identifier la matrice \mathbf{S} .

Cette intuition est illustrée sur la Figure 1.27.

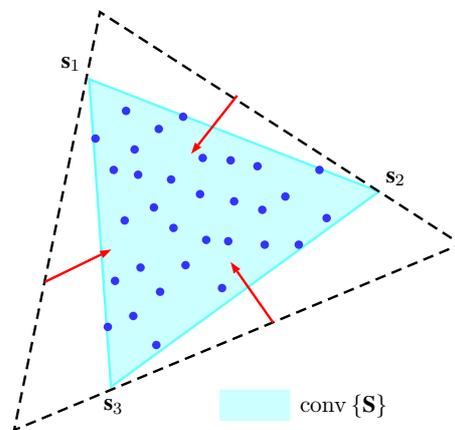


FIGURE 1.27 : L'intuition du volume minimal dans le cas où $R = 3$. Les points représentent les \mathbf{x}_p . En pointillé est représenté un simplexe englobant les données et en ligne continue le simplexe ayant le volume minimal permettant ainsi d'identifier \mathbf{S}

Néanmoins, la condition proposée par [Craig 94] était seulement une conjecture ; c'est pourquoi, de nombreuses recherches ont été menées au fil des années pour prouver l'identifiabilité du critère VolMin [Lin 15, Fu 15]. Formellement, le critère VolMin peut s'écrire de la manière suivante :

$$\begin{aligned} & \underset{\mathbf{S}, \mathbf{A}}{\text{minimiser}} \text{Vol}(\mathbf{S}), \\ & \text{sous contraintes } \mathbf{X} = \mathbf{S}\mathbf{A}, \mathbf{S} \geq \mathbf{0}, \mathbf{A} \geq \mathbf{0}. \end{aligned}$$

où le terme $\text{Vol}(\mathbf{S})$ désigne une mesure liée ou proportionnelle au volume du simplexe $\text{conv}\{\mathbf{S}\}$. Dans la littérature, de nombreuses expressions pour $\text{Vol}(\mathbf{S})$ ont été utilisées [Berman 04, Li 08, Zhou 11a, Li 12, Fu 15, Li 16] et nous en discuterons en détail dans le chapitre suivant. À titre d'exemple, une mesure couramment utilisée est $\text{Vol}(\mathbf{S}) = \det(\mathbf{S}^T \mathbf{S})$.

Récemment, des travaux [Fu 18a] ont considérablement approfondi la compréhension de l'identifiabilité de VolMin. Ils ont montré que :

- Si $\text{Vol}(\mathbf{S}) = \det(\mathbf{S}^T \mathbf{S})$, $\text{rang}(\mathbf{S}) = \text{rang}(\mathbf{A}) = R$ et si la matrice \mathbf{A} respecte la condition de dispersion suffisante (selon la **Définition 3**) alors VolMin garantit l'identifiabilité du modèle de factorisation (1.2).

Contrairement aux **Définitions 2** et **3** établies précédemment, où \mathbf{S} doit contenir des éléments nuls pour garantir l'identifiabilité, aucune restriction spécifique sur \mathbf{S} n'est nécessaire lors de l'utilisation du critère VolMin. Cette condition d'identifiabilité est beaucoup plus générale et plus relaxée que celles établies précédemment.

1.4.2 Méthodes de démixage hyperspectral

Les méthodes de démixage hyperspectral peuvent être partagées en quatre principales classes [Parente 10, Bioucas-Dias 11] et décrites succinctement de la manière suivante :

- Les méthodes basées sur des concepts géométriques : elles sont les plus populaires dans le domaine du démixage hyperspectral. En effet, à partir des contraintes naturelles du modèle, nous avons montré précédemment que le nuage de points formant une image hyperspectrale est contenu dans un simplexe. Les sommets du simplexe représentent les spectres (*endmembers*) recherchés. Les méthodes géométriques fondées sur l'hypothèse de pixels purs ont pour objectif la recherche des sommets du simplexe. Lorsque la condition de pixels purs n'est pas satisfaite, d'autres méthodes géométriques interviennent [Plaza 12] : elles consistent à étendre l'espace de recherche des *endmembers* au delà du nuage de points, tout en gardant un simplexe qui circonscrit les données aussi compacte que possible (simplexe ayant un volume minimal). L'ensemble de ces méthodes sont considérées comme non-supervisées, elles ne requièrent aucune connaissance *a priori* ; notons, néanmoins, que le nombre de *endmembers* R à extraire doit être connu.
- Les méthodes statistiques : elles sont utilisées lorsque les données sont fortement mélangées. Ces dernières sont puissantes mais de complexité de calcul supérieure à celle des méthodes géométriques. Parmi elles, on retrouve les méthodes basées sur l'analyse en composantes indépendantes (*e.g.* [Hyvärinen 00]) : ces méthodes supposent l'indépendance statistique des sources. En imagerie hyperspectrale, cette hypothèse n'est pas systématiquement respectée du fait de la nature des spectres dans certains cas très corrélés (notamment lorsqu'ils appartiennent à une même grande classe par exemple végétation, bois *etc.*). D'autres méthodes se basent sur des approches bayésiennes (*e.g.* [Eches 10]) : elles exploitent les connaissances *a priori* sur les inconnues du problème. Pour ce faire, des distributions de probabilité *a priori* sont utilisées pour modéliser les informations disponibles sur les abondances et les *endmembers*. Cependant, un inconvénient majeur des méthodes bayésiennes est la non connaissance exacte des distributions de probabilité *a priori* en pratique.
- Les méthodes basées sur la régression parcimonieuse : lorsqu'une bibliothèque spectrale est disponible, le démixage peut être réalisé en identifiant les *endmembers* à partir de cette bibliothèque (*e.g.* [Bioucas-Dias 10]). Le démixage est possible en intégrant des termes de régularisation pour imposer la parcimonie. L'ensemble des *endmembers* est identifié en réduisant, de manière récursive, la bibliothèque spectrale. Ces méthodes sont appelées semi-supervisées car,

généralement, uniquement les abondances ont besoin d'être calculées puisque les signatures spectrales sont connues.

- Les méthodes basées sur la Factorisation en Matrices Non-négatives (NMF pour *Non-negative Matrix Factorization*) qui exploitent la non-négativité des données. Elles consistent, de manière générale, à factoriser la matrice des données en un produit de deux matrices non-négatives de rang R , caractérisant les abondances et les *endmembers*. L'estimation des paramètres se fait conjointement. L'inconvénient majeur de ces méthodes est la non-unicité de la solution. Pour y remédier, de nombreuses méthodes fondées sur la NMF ajoutent des contraintes supplémentaires au modèle, telles que des contraintes géométriques de type simplexe de volume minimal ou encore la parcimonie ce qui font d'elles des méthodes « hybrides ».

Dans la suite, nous détaillons essentiellement les méthodes géométriques et les méthodes fondées sur la NMF. Ce choix nous paraît judicieux car les méthodes géométriques permettent d'appréhender au mieux les données hyperspectrales et la NMF nous garantit la non-négativité des paramètres d'intérêt, indispensable pour l'interprétation physique des résultats. Par ailleurs, la NMF s'adapte bien au problème de démélange hyperspectral en-ligne. Les méthodes parcimonieuses sont écartées car, dans le cadre de cette thèse, le démélange hyperspectral aveugle est concerné *i.e.* aucune bibliothèque spectrale n'est disponible et aucune connaissance *a priori* des données n'est nécessaire. Les méthodes statistiques ne sont également pas abordées de par leur complexité de calcul élevée ainsi que de leur faible performance particulièrement en imagerie hyperspectrale due à la forte corrélation des *endmembers*.

1.4.2.1 Méthodes géométriques avec pixel pur

Les méthodes géométriques supposant l'existence de pixels purs pour chaque composant, ont principalement pour but de retrouver les sommets du simplexe englobant le nuage de points d'une image. Ces sommets, avec l'hypothèse en question, appartiennent au nuage de points. Parmi ces méthodes, les plus populaires sont : la méthode OSP (*Orthogonal Subspace Projection*) [Harsanyi 94], la méthode PPI (*Pixel Purity Index*) [Boardman 95], la méthode baptisée N-FINDR [Winter 99] et la méthode VCA (*Vertex Component Analysis*) [Nascimento 05].

La méthode OSP est basée sur une succession de projections orthogonales des données pour identifier itérativement les sommets du simplexe. Le premier sommet retenu est le pixel observé ayant la plus grande distance avec l'ensemble des pixels restant. Les sommets suivants sont ensuite obtenus en projetant les données sur un vecteur orthogonal au sous-espace engendré par les colonnes de la matrice des *endmembers* déjà identifiée. Le nouveau sommet retenu correspond à la projection extrême. Cette étape de projection/identification est ensuite répétée jusqu'à ce que tous les sommets du simplexe soient identifiés.

La méthode PPI effectue d'abord une réduction de la dimension des données en utilisant la méthode MNF (*Maximum Noise Fraction*) [Green 88]. Ensuite, elle effectue plusieurs projections de l'ensemble du nuage de points (qui correspond aux données dans l'espace réduit) sur des vecteurs de norme unité générés aléatoirement (appelés *skewers*), et retient, pour chaque projection réalisée, les pixels qui ont des valeurs extrêmes. Les pixels les plus purs recherchés correspondent aux points qui ont le plus grand score

(défini par le nombre de fois où un pixel a été retenu comme point extrême). Cette méthode nécessite un nombre élevé d'itérations.

La méthode N-FINDR cherche à déterminer les pixels qui maximisent le volume du simplexe contenant les données. Après une étape de réduction de la dimension des données, cette méthode sélectionne aléatoirement dans le nuage de points, des pixels considérés comme les *endmembers* (comme étant des sommets potentiels), puis le volume formé par ces pixels est calculé. Ce volume est calculé à nouveau en remplaçant l'un des *endmembers* par un autre pixel ; si ce remplacement engendre une augmentation du volume, alors le pixel est considéré comme un *endmember* et ainsi de suite. Cette procédure s'achève lorsqu'il n'y a plus aucune possibilité de remplacement des *endmembers*. Cette méthode dépend fortement des pixels initiaux sélectionnés pour le calcul du volume du premier simplexe.

La méthode VCA est considérée parmi les plus efficaces de cette catégorie et est très similaire à la méthode OSP. Elle est basée sur le même principe de projection/identification. Cependant elle inclut un processus de caractérisation du bruit pour réduire la sensibilité au bruit de la projection.

L'ensemble de ces méthodes n'estiment pas conjointement les *endmembers* et les abondances. En effet, cela représente deux étapes indépendantes. Une fois les spectres des *endmembers* connus, des méthodes de régression sont utilisées pour estimer les coefficients d'abondance telles que la méthode NCLS (*Non-Negativity Constrained Least Squares*) [Chang 00], la méthode FCLS (*Fully Constrained Least Squares*) [Heinz 01], la méthode SGM (*Scaled Gradient Method*) [Theys 09] ou encore une approche bayésienne proposée par [Dobigeon 08].

Pour davantage de détails sur l'ensemble de ces méthodes de démixage ainsi que d'autres méthodes appartenant à cette catégorie, le lecteur peut se référer à [Bioucas-Dias 12]. Ces méthodes sont très efficaces dans la situation où l'on dispose de pixels purs ; néanmoins, lorsque ce n'est pas le cas, leurs performances se dégradent considérablement. En effet, l'hypothèse sur laquelle reposent ces méthodes est une exigence forte qui n'est malheureusement pas applicable sur de nombreuses images hyperspectrales dont la résolution spatiale est trop faible par exemple. Dans ce cas de figure, d'autres approches géométriques ont été développées et sont présentées ci-après.

1.4.2.2 Méthodes géométriques sans pixel pur

En s'inspirant de l'approche de minimisation du volume proposé par Craig en 1994 [Craig 94], la majorité des méthodes développées dans cette catégorie recherchent à générer des *endmembers* virtuels (qui ne sont pas disponibles dans l'image considérée). En d'autres termes, cette approche consiste à minimiser le volume du simplexe contenant les données.

Parmi les méthodes de démixage hyperspectral les plus connues, qui minimisent le volume du simplexe, nous pouvons citer : la méthode CCA (*Convex Cone Analysis*) [Ifarraguerri 99], la méthode ICE (*Iterative Constrained Endmembers*) [Berman 04], la méthode MVSA (*Minimum Volume Simplex Analysis*) [Li 08], la méthode MVES (*Minimum Volume Enclosing Simplex*) [Chan 09] ainsi que la méthode SISAL (*Simplex Identification via Split Augmented Lagrangian*) [Bioucas-Dias 09]. L'approche adoptée par ces méthodes est une optimisation non-convexe et est ainsi plus difficile à résoudre par rapport aux

approches présentées dans la catégorie précédente. Pour ce faire, chacune de ces méthodes a sa propre stratégie.

La méthode CCA par exemple, cherche à déterminer les points limitrophes du cône ou de la région convexe contenant l'ensemble des points de l'image. En pratique, cette méthode estime les spectres des *endmembers* représentés par les sommets de la région convexe en se basant sur les vecteurs propres de la matrice de corrélation de l'image. Il est à préciser que cette méthode adopte une approche similaire à celle proposée par [Craig 94].

La méthode ICE, après une étape de réduction de dimension, minimise une fonction représentée par la distance euclidienne entre les données considérées \mathbf{X} et le produit recherché entre les matrices \mathbf{S} et \mathbf{A} (conformément au modèle linéaire présenté en (1.2)), à laquelle elle ajoute un terme de régularisation lié à la distance entre tous les sommets du simplexe. En pratique, au sein de chaque itération, pour chaque pixel, l'algorithme ICE utilise des routines de programmation quadratiques pour estimer les abondances, ce qui rend cette méthode inadéquate pour des grands ensembles de données. Une autre version de la méthode ICE, qui s'intitule SPICE (*Sparsity-Promoting ICE*) ajoute un terme favorisant la parcimonie et permet ainsi de sélectionner le nombre approprié de *endmembers*.

La méthode MVSA est basée sur des méthodes d'optimisation dites *Sequential Quadratic Programming*. La méthode SISAL propose une approche basée sur la résolution séquentielle de sous-problèmes d'optimisation convexe à l'aide d'un algorithme du Lagrangien augmenté. Ces méthodes ont la particularité de permettre une violation de la contrainte de non-négativité. Cette caractéristique prend en compte le fait que des vecteurs observés peuvent être en dehors du simplexe à cause du bruit. L'introduction d'une version plus souple de la contrainte de non-négativité rend les méthodes SISAL et MVSA plus robustes et plus efficaces pour résoudre le problème d'optimisation. De plus, pour réduire le risque que ces méthodes convergent dans des minima locaux, les auteurs proposent dans les deux cas d'initialiser leurs algorithmes par la matrice des *endmembers* estimée grâce à la méthode VCA.

La méthode MVES propose une reformulation du problème d'optimisation non-convexe permettant ainsi une approche de minimisation cyclique utilisant une méthode de programmation linéaire. Cependant, les performances de ces méthodes et leur efficacité de calcul sont souvent limitées en raison des calculs complexes nécessaires pour définir le simplexe de volume minimal, leur sensibilité à l'initialisation ainsi que l'absence d'analyse de performance rigoureuse.

1.4.2.3 La factorisation en matrices non-négatives

Initialement proposée par Paatero et Tapper avec des travaux fondés sur la factorisation en matrices positives [Paatero 94], la NMF a connu un véritable succès depuis les années 90, suite aux travaux de Lee et Seung [Lee 99] comme une nouvelle technique de démélange hyperspectral aveugle. Elle ne requiert pas la condition de pixels purs et détermine conjointement les *endmembers* et les abondances. Le principe de la NMF consiste à factoriser une matrice d'observation non-négative $\mathbf{X} \in \mathbb{R}_+^{L \times P}$ selon :

$$\mathbf{X} \approx \mathbf{S}\mathbf{A},$$

sous contraintes de non-négativité des matrices $\mathbf{S} \in \mathbb{R}_+^{L \times R}$ et $\mathbf{A} \in \mathbb{R}_+^{R \times P}$, respectivement la matrice des *endmembers* et des abondances avec R , le nombre de sources que nous cherchons à estimer. Afin de réaliser cette factorisation, on minimise/maximise une fonction objectif. Cette dernière est en général un critère (distance, similarité, divergence, *etc.*) mesurant la distance entre la matrice des observations \mathbf{X} et le produit entre les deux matrices \mathbf{S} et \mathbf{A} . La mesure la plus couramment utilisée est la norme de Frobenius, telle que :

$$\mathcal{J}(\mathbf{S}, \mathbf{A}) = \frac{1}{2} \|\mathbf{X} - \mathbf{S}\mathbf{A}\|_F^2, \quad (1.4)$$

où \mathcal{J} représente le critère ou la fonction objectif à optimiser. Plusieurs autres mesures de distance peuvent être utilisées telles que la divergence de Kullback Leibler ou encore la divergence de Itakura Saito [Févotte 11].

La méthode NMF peut être formulée en tant que problème d'optimisation qui minimise une fonction objectif sous les contraintes de non-négativité, comme suit :

$$\begin{aligned} & \underset{\mathbf{S}, \mathbf{A}}{\text{minimiser}} \mathcal{J}(\mathbf{S}, \mathbf{A}), \\ & \text{sous contraintes } \mathbf{S} \geq \mathbf{0}, \mathbf{A} \geq \mathbf{0}. \end{aligned} \quad (1.5)$$

Par la suite, afin de déterminer les deux matrices \mathbf{S} et \mathbf{A} qui minimisent la fonction objectif (1.4), on utilise un algorithme de minimisation alternée qui consiste à estimer alternativement chacune des matrices \mathbf{S} et \mathbf{A} , jusqu'à la convergence, tout en respectant les contraintes de non-négativité. Les estimations s'écrivent de la manière générale suivante :

$$\begin{aligned} \mathbf{A}_{t+1} &= \arg \min_{\mathbf{A} \geq \mathbf{0}} \mathcal{J}(\mathbf{S}_t, \mathbf{A}), \\ \mathbf{S}_{t+1} &= \arg \min_{\mathbf{S} \geq \mathbf{0}} \mathcal{J}(\mathbf{S}, \mathbf{A}_{t+1}), \end{aligned}$$

où t est l'indice de l'itération. Dans la littérature, il existe de nombreux algorithmes itératifs permettant de parvenir à cette tâche et ils se classent de la manière suivante :

- Les méthodes fondées sur le gradient projeté.
- Les méthodes fondées sur les règles de mise à jour multiplicatives.
- Les méthodes fondées sur les moindres carrés alternés avec contrainte de non-négativité.
- Les méthodes fondées sur les directions alternées (ADMM pour *Alternating Direction Method of Multipliers*)

Les méthodes de gradient constituent la base de résolution des problèmes en traitement du signal et des images dès lors que la fonction à optimiser est différentiable. Elles consistent, à partir d'un point courant, à migrer dans la direction de la plus grande pente du critère. À chaque itération de cet algorithme, les règles de mise à jour suivantes sont utilisées :

$$\begin{aligned}\mathbf{A} &\leftarrow [\mathbf{A} - \eta_{\mathbf{A}} \nabla_{\mathbf{A}} \mathcal{J}]_+, \\ \mathbf{S} &\leftarrow [\mathbf{S} - \eta_{\mathbf{S}} \nabla_{\mathbf{S}} \mathcal{J}]_+, \end{aligned}$$

où $[\cdot]_+$ correspond à la projection des mises à jour sur l'orthant positif. Les termes $\eta_{\mathbf{A}}$ et $\eta_{\mathbf{S}}$ sont des pas positifs dont l'objectif est de décroître la fonction \mathcal{J} . Les nombreuses méthodes basées sur le gradient projeté se différencient entre elles quant à la manière de choisir les pas $\eta_{\mathbf{A}}$ et $\eta_{\mathbf{S}}$. Parmi les plus connues, nous pouvons citer l'algorithme de gradient projeté de Lin [Lin 07b], qui utilise un pas commun pour \mathbf{S} et \mathbf{A} et ce pas est mis à jour en suivant la règle modifiée de *Armijo* [Bertsekas 97]. En revanche, la recherche du pas optimal est une procédure qui demande beaucoup de temps. Ainsi, d'autres méthodes ont été développées pour tenter de minimiser le coût de calcul généré par cette procédure ; les lecteurs intéressés peuvent se référer aux articles suivants : [Zdunek 06, Kim 07, Han 09, Guan 12a]. Par exemple, [Guan 12a] propose la méthode NeNMF (*Nesterov-NMF*), qui utilise la méthode du gradient optimal de Nesterov pour optimiser alternativement un facteur avec le second fixe. En particulier, à chaque itération, les facteurs sont mis à jour alternativement à l'aide de la méthode du gradient projeté et où la taille du pas est déterminée par la constante de *Lipschitz*. Des résultats ont montré que l'algorithme NeNMF surpasse de nombreuses méthodes de gradient projeté en terme d'efficacité et de vitesse de convergence. En effet, il a été montré que NeNMF permet d'atteindre une vitesse de convergence optimale de l'ordre de $\mathcal{O}(\frac{1}{i^2})$ où i est l'indice de l'itération.

L'algorithme multiplicatif (MU), initialement proposé par Lee et Seung [Lee 99, Lee 01], est une extension de l'algorithme du gradient, où le pas est exprimé en fonction des matrices \mathbf{S} et \mathbf{A} de façon à obtenir une forme multiplicative. Dans le cas où la fonction objectif est basée sur la norme de Frobenius (1.4), les règles de mise à jour des deux matrices s'expriment sous la forme suivante :

$$\begin{aligned}\mathbf{A} &\leftarrow \mathbf{A} \odot \frac{\mathbf{S}^T \mathbf{X}}{\mathbf{S}^T \mathbf{S} \mathbf{A}}, \\ \mathbf{S} &\leftarrow \mathbf{S} \odot \frac{\mathbf{X} \mathbf{A}^T}{\mathbf{S} \mathbf{A} \mathbf{A}^T}, \end{aligned} \quad (1.6)$$

où \odot est le produit matriciel de Hadamard. L'avantage de l'algorithme multiplicatif, en le comparant avec l'algorithme du gradient projeté, est qu'il n'y a pas de pas à fixer ou à calculer à chaque itération. En outre, la non-négativité est assurée par les règles de mise à jour (à condition que les valeurs initiales pour \mathbf{S} et \mathbf{A} soient non-négatives) sans avoir recours à la projection à chaque itération. En revanche, cette méthode d'optimisation souffre parfois d'une convergence lente.

L'algorithme des moindres carrés alternés (ALS pour *Alternating Least Square*) a été proposé pour la première fois par [Paatero 94]. Le principe de cet algorithme consiste à estimer alternativement une des deux matrices \mathbf{A} et \mathbf{S} de la manière suivante :

$$\begin{aligned}\mathbf{A} &\leftarrow [(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{X}]_+, \\ \mathbf{S} &\leftarrow [\mathbf{X} \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1}]_+ . \end{aligned}$$

Le principal inconvénient de l'algorithme ALS est qu'il ne peut souvent pas échapper aux minima locaux. Afin d'atténuer ce problème, des conditions de régularisation et/ou des pénalité supplémentaires ont été introduites *e.g.* dans [Cichocki 09], ainsi que de nouvelles fonctions objectif permettant de dériver des algorithmes hiérarchiques locaux ALS (HALS pour *Hierarchical ALS*).

Les approches MU et ALS sont sans aucun doute les plus populaires en majeure partie due à leur remarquable simplicité. Néanmoins, certaines propriétés théoriques ne sont toujours pas bien comprises, notamment la convergence ou non vers un point stationnaire. Le problème (1.5) est bilinéaire ; il est soit convexe en \mathbf{S} (si \mathbf{A} fixé) ou en \mathbf{A} (si \mathbf{S} fixé) mais pas en les deux simultanément. Ainsi, il est impossible de garantir l'optimalité globale. Les méthodes NMF existantes considèrent un point stationnaire comme une solution optimale locale ; un point est dit stationnaire s'il respecte les conditions de Karush Kuhn Tucker (KKT) [Boyd 04]. La stationnarité est importante car c'est une condition nécessaire d'un minimum local. Par exemple, les travaux menés par [Lee 01] ont montré que l'algorithme MU converge vers un point stationnaire mais des exemples numériques présentés par [Gonzalez 05] ont montré le contraire. Pour pallier ce problème, [Lin 07a] a proposé de légères modifications des mises à jour MU existantes et a prouvé cette fois-ci leur convergence vers un point stationnaire. La stratégie utilisée consiste en l'addition de petites valeurs au dénominateur de chacune des mises à jour MU, évitant ainsi les divisions par zéro. Néanmoins, cette nouvelle version ne permet pas d'accélérer la vitesse de convergence. Précisons que des versions accélérées des méthodes MU et ALS ont été proposées dans [Gillis 12]. Quant aux méthodes basées sur le gradient projeté, par exemple l'algorithme de [Lin 07b] et [Guan 12a], plusieurs résultats sur la convergence vers un point stationnaire ont été démontrés. Ces résultats se basent principalement sur les travaux de [Grippo 00] qui affirment que si uniquement deux blocs de variables sont mis à jour (à savoir dans notre cas \mathbf{S} et \mathbf{A}) alors tout point limite généré par une séquence d'itérations est un point stationnaire de (1.5). Bien que ces algorithmes soient plus difficiles à mettre en oeuvre, ils offrent des garanties théoriques plus intéressantes que les algorithmes multiplicatifs. Le lecteur intéressé peut se référer à [Cichocki 09].

Pour finir, le quatrième groupe concerne les méthodes fondées sur l'algorithme ADMM [Boyd 11]. ADMM est basée sur le Lagrangien augmenté et repose sur le concept de partitionnement des variables qui permet de séparer la fonction objectif à minimiser en plusieurs termes indépendants. ADMM a été récemment appliquée aux problèmes NMF [Xu 12, Sun 14, Hajinezhad 16, Huang 16]. Elle consiste en l'introduction de deux variables auxiliaires que l'on appelle \mathbf{U} et \mathbf{V} et considère le modèle équivalent suivant :

$$\begin{aligned} & \underset{\mathbf{S}, \mathbf{A}}{\text{minimiser}} \quad \frac{1}{2} \|\mathbf{X} - \mathbf{SA}\|_F^2, \\ & \text{sous contraintes } \mathbf{S} - \mathbf{U} = \mathbf{0}, \mathbf{A} - \mathbf{V} = \mathbf{0}, \mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}. \end{aligned} \quad (1.7)$$

Le Lagrangien augmenté relatif à (1.7) s'écrit sous la forme suivante :

$$\begin{aligned} \mathcal{L}(\mathbf{S}, \mathbf{A}, \mathbf{U}, \mathbf{V}, \mathbf{\Pi}, \mathbf{\Lambda}) &= \frac{1}{2} \|\mathbf{X} - \mathbf{SA}\|_F^2 + \frac{\rho}{2} \|\mathbf{A} - \mathbf{V} + \mathbf{\Pi}\|_F^2 - \frac{\rho}{2} \|\mathbf{\Pi}\|_F^2 \\ &+ \frac{\rho}{2} \|\mathbf{S} - \mathbf{U} + \mathbf{\Lambda}\|_F^2 - \frac{\rho}{2} \|\mathbf{\Lambda}\|_F^2 + \mathbb{I}_{\mathbb{R}_+}(\mathbf{V}) + \mathbb{I}_{\mathbb{R}_+}(\mathbf{U}), \end{aligned}$$

où $\mathbb{I}_{\mathbb{R}_+}$ représente la fonction indicatrice de \mathbb{R}_+ imposant la positivité des abondances et des *endmembers*. Notons que cette formulation du Lagrangien augmenté est issue de la version réduite de ADMM qui permet d'écrire le problème plus simplement [Boyd 04]. $\mathbf{\Pi}$ et $\mathbf{\Lambda}$ sont les multiplicateurs de Lagrange et ρ est un terme de pénalité relatif aux contraintes $\mathbf{S} - \mathbf{U} = \mathbf{0}$ et $\mathbf{A} - \mathbf{V} = \mathbf{0}$. Les mises à jour des différents paramètres s'obtiennent en minimisant successivement le Lagrangien augmenté \mathcal{L} par rapport à \mathbf{S} , \mathbf{A} , \mathbf{U} et \mathbf{V} , puis en mettant à jour les multiplicateurs $\mathbf{\Pi}$ et $\mathbf{\Lambda}$. Le nombre d'algorithmes NMF fondés sur ADMM a augmenté considérablement ces dernières années en raison de ses remarquables propriétés. En effet, elle surpasse les algorithmes MU, ALS et même NeNMF en ce qui concerne l'efficacité et la précision de la reconstruction ainsi que la vitesse de convergence [Xu 12, Sun 14, Goldstein 14].

Cependant, la NMF est non-unique (voir section 1.4.1.2). Ainsi, des contraintes supplémentaires au modèle doivent être introduites. De nombreuses méthodes NMF contraintes ont été proposées, en imposant des contraintes sur la matrice des *endmembers* et/ou sur la matrice des abondances. Le problème NMF régularisé peut s'exprimer sous la forme de la fonction objectif suivante :

$$\mathcal{J}(\mathbf{S}, \mathbf{A}) = \frac{1}{2} \|\mathbf{X} - \mathbf{SA}\|_F^2 + \mu f_{\mathbf{S}}(\mathbf{S}) + \tau f_{\mathbf{A}}(\mathbf{A}),$$

où $f_{\mathbf{S}}(\mathbf{S})$ et $f_{\mathbf{A}}(\mathbf{A})$ sont des contraintes sur \mathbf{S} et \mathbf{A} respectivement et μ et τ sont les paramètres de régularisation. Généralement, ces contraintes sont de type simplexe de volume minimal [Miao 07] ou encore de parcimonie [Yang 11] et de lissage [Pauca 06]. Sans être exhaustifs, nous pouvons citer la méthode NMF avec contrainte de variance minimale [Huck 10], qui encourage les spectres des *endmembers* à être plats, en préservant les éventuelles singularités spectrales (pics et variations nettes). Dans [Wang 13c], une méthode NMF avec une contrainte de dissimilarité des *endmembers* est proposée, supposant que les *endmembers* doivent être lisses et différents les uns des autres. Une autre approche nommée *Piecewise Smoothness NMF with Sparseness Constraint* proposée par [Jia 09] impose une contrainte de lissage par morceaux sur les spectres et une contrainte de parcimonie sur les abondances forçant ainsi certaines valeurs à zéro. Parmi les différentes méthodes soumises à des contraintes de parcimonie, la méthode NMF avec contrainte de type norme $\ell_{1/2}$ est une approche très populaire [Qian 11]. Géométriquement, la parcimonie sur les abondances revient à satisfaire la condition de dispersion suffisante (voir **Définition 3**). Les méthodes proposées dans [Miao 07, Schachtner 09, Zhou 11a, Fu 16, Ang 18, Fu 18b, Leplat 19] incorporent une contrainte de type volume minimal. Une autre méthode NMF incorpore une contrainte de dispersion minimale entre les *endmembers* [Yu 07]. L'idée est de minimiser la distance entre les *endmembers* (donc chaque sommet du simplexe) et leur barycentre. Récemment, [Li 16] a développé un algorithme combinant la NMF avec une méthode reposant sur l'hypothèse de pixels purs (e.g. VCA). Il propose un nouveau terme de volume minimal consistant à « pousser » le simplexe estimé par la NMF vers le simplexe estimé en amont par la méthode de pixel pur, qui, étant donné qu'il est défini par des vecteurs observés, se trouvera à l'intérieur du vrai simplexe. De plus, il impose une contrainte de parcimonie sur la matrice des abondances de type norme $\ell_{2,1}$, qui permet, en cas de surestimation du nombre de sources, de sélectionner uniquement celles qui sont actives. Cette méthode s'apparente à une approche de type régression parcimonieuse où la bibliothèque spectrale est établie grâce à une méthode géométrique avec pixels purs.

1.5 Choix méthodologiques et limites des approches existantes

Le choix concernant l'approche la plus adaptée pour notre problème de démixage hyperspectral est conditionné par la nature des données à traiter. Comme nous pouvons l'observer sur la Figure 1.28, les données hyperspectrales de bois avec lesquelles nous travaillons sont très mélangées et ainsi nous éliminons naturellement les méthodes géométriques avec pixels purs comme choix possible pour un démixage hyperspectral en-ligne.

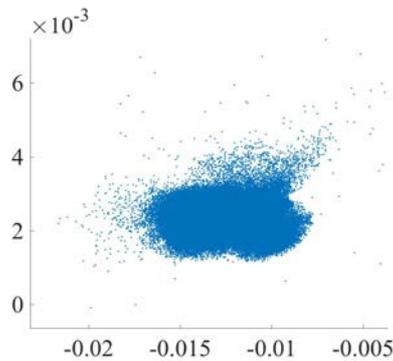


FIGURE 1.28 : Données issues d'une image hyperspectrale de bois (acquise par l'imageur FX17) projetées sur les deux premières composantes principales

Il a été décidé, dans le cadre de cette thèse, de s'orienter vers des méthodes qui exploitent la non-négativité des données ainsi que des variables latentes (*endmembers* et abondances). Les méthodes privilégiées ont donc été des méthodes basées sur la NMF, qui, par ailleurs, ont déjà été utilisées dans le cas de traitement de données en-ligne. Cette piste n'est bien sûr pas la seule possible, les méthodes telles que MVSA ou encore MVSE auraient pu être envisagées ; néanmoins, de par leur implémentation algorithmique, leur extension au modèle en-ligne que nous considérons et que nous présentons dans la suite n'est pas une tâche facile. De plus, notre choix s'est orienté vers la NMF pour les raisons suivantes :

1. Simplicité d'implémentation : les mises à jour des différents paramètres d'intérêt ont des expressions simples.
2. Simplicité d'adaptation au problème de type temps réel : de par son critère (1.4), la NMF peut être étendue aisément au traitement en-ligne/séquentiel des échantillons.
3. Temps de traitement en accord avec les contraintes industrielles temps-réel grâce à une complexité de calcul faible et à une vitesse de convergence rapide (à condition de choisir une méthode d'optimisation adéquate).
4. Possibilité d'introduire facilement un large choix de contraintes en fonction des données à traiter.
5. Méthode de démixage hyperspectral non-supervisée.

Bien que la NMF « classique » possède de nombreux avantages, son fonctionnement n'est pas directement transposable à un traitement en-ligne, qui est le problème abordé dans cette thèse. La Figure 1.29 représente schématiquement l'acquisition d'une image hyperspectrale *pushbroom*. Chaque

tranche/échantillon, représenté en pointillé, est une image de dimension $L \times P$ où L fait référence au nombre de longueurs d'onde et P à la dimension spatiale (une ligne de la scène). Les échantillons sont acquis le long d'une des dimensions spatiales de l'image, que l'on assimilera au temps K .

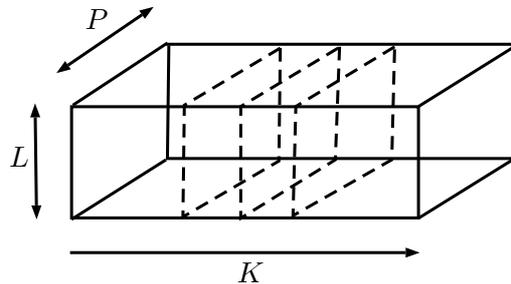


FIGURE 1.29 : Acquisition des données hyperspectrales *pushbroom*

Le processus en-ligne que nous considérons doit répondre simultanément aux trois objectifs suivants :

1. S'adapter au rythme de l'acquisition de l'imageur *pushbroom*, *i.e.* estimer les matrices des *end-members* \mathbf{S} et des abondances \mathbf{A} à chaque nouvelle tranche, en temps réel, et sans accroître la complexité de calcul.
2. Prendre en compte la variabilité spectrale (adaptabilité) ; ceci est important pour suivre les changements de contenu dynamiques et peut être réalisé en imposant un facteur d'oubli sur les échantillons passés.
3. Assurer la cohérence/le suivi des signatures spectrales estimées d'un échantillon à l'autre ; cela permet de garantir une certaine stabilité de la solution, tout en préservant un degré de variabilité.

Différentes stratégies peuvent être envisagées pour répondre à l'objectif 1. Il est possible d'appliquer le processus NMF classique tranche par tranche ; cela revient simplement à considérer un ensemble de problèmes indépendants des moindres carrés et, dans ce cas, la configuration en-ligne n'a aucun intérêt particulier puisque l'on perd la cohérence entre les différentes estimations de \mathbf{S} . Une autre solution possible est d'exécuter, de manière répétée, le processus NMF classique en entier à chaque fois qu'un nouvel échantillon est reçu. L'arrivée de chaque nouvel échantillon entraîne une augmentation de la dimension de la matrice des données \mathbf{X} . La complexité de calcul de la NMF est de $\mathcal{O}(LPR)$ par itération, ce qui implique que le coût de calcul augmente linéairement avec la dimension des données. De plus, l'utilisation de cette stratégie nécessite une demande importante en stockage. Prenons exemple sur l'algorithme MU ; comme l'indiquent les équations en (1.6), les formules de mise à jour multiplicatives utilisent les matrices \mathbf{S} , \mathbf{A} et \mathbf{X} à chaque itération, ce qui nécessite de stocker tous les échantillons précédents. Par conséquent, une quantité énorme de mémoire doit être affectée au stockage des matrices \mathbf{A} et \mathbf{X} . Il est donc inconcevable d'employer cette solution dans le cas d'un traitement en temps réel, compte tenu des temps de lecture et écriture. Une solution intermédiaire serait d'appliquer la NMF par fenêtre glissante, ce qui permet à la fois de prendre en compte la variabilité spectrale ainsi que la cohérence entre les échantillons. Néanmoins, l'horizon de mémoire est limité par la taille de la fenêtre glissante. Suite à cette discussion, il semble que la NMF standard ne permet pas de répondre à l'ensemble de nos objectifs. Il est donc essentiel d'étendre le modèle NMF classique à un modèle plus complexe adapté au traitement

temps réel.

Apportons maintenant davantage de détails sur l'objectif 2. En imagerie hyperspectrale, étudier la variabilité spectrale est un point essentiel. La variabilité spectrale peut être induite par différents phénomènes physiques tels que des changements d'illumination au cours de l'acquisition de l'image, des effets atmosphériques, ou encore la variabilité intrinsèque des matériaux. Cette dernière est probablement la plus importante en terme d'impact et est la cause qui nous concerne le plus fortement. Elle est purement dépendante du matériau et correspond généralement à la variation d'un paramètre caché, ce qui n'est pas pris en compte dans le modèle de mélange linéaire classique. En effet, dans la plupart des modèles et algorithmes de démixage hyperspectral, un matériau est supposé être parfaitement représenté par un spectre unique. C'est en fait une hypothèse très forte puisque les différents matériaux présentent toujours une certaine variabilité intra-classe, ce qui peut modifier localement les spectres des matériaux purs, quel que soit le processus de mélange [Zare 13, Halimi 15]. Dans ce cas de figure, le modèle de mélange est toujours linéaire pour un échantillon donné, mais nous considérons que les *endmembers* peuvent varier légèrement d'un échantillon à l'autre.

Plus généralement, la variabilité intrinsèque est causée par les variations physico-chimiques au sein des matériaux. Les études menées sur l'analyse des spectres proche infrarouge du bois ont montré que les signatures spectrales extraites ont tendance à être corrélées car la composition chimique et la structure des composants sont très similaires. Par conséquent, les principales différences entre les spectres se limitent généralement à des variations d'amplitude ou à des légers décalages en longueurs d'onde. La moindre variation doit être prise en considération car elle peut être le signe d'un changement de densité ou d'une liaison chimique différente et ainsi révéler une singularité. De ce fait, l'algorithme en-ligne que nous considérons doit permettre de suivre les changements dynamiques des sources au cours du traitement et cela est un véritable point clé pour l'analyse de matériaux en temps réel.

La problématique considérée est un sujet d'actualité qui suscite un vif intérêt dans la communauté du traitement du signal. Ces dernières années, divers algorithmes en-ligne fondés sur la NMF ont été proposés pour pallier les limitations de cette dernière ; ils consistent à mettre à jour séquentiellement les *endmembers* et les abondances à mesure que la taille des données augmente, tout en maintenant une complexité de calcul faible et contrôlée. Ainsi, ils sont parfaitement adaptés au traitement d'images hyperspectrales en temps réel. Ils peuvent être classés en deux catégories : la première catégorie d'algorithmes suppose que les *endmembers* sont stationnaires. Les applications sont nombreuses telles que la reconnaissance faciale, l'annotation d'images, ou encore la classification de documents [Mairal 10, Wang 11, Guan 12b, Zhao 16]. La deuxième catégorie d'algorithmes [Cao 07, Bucak 09, Zhou 11b, Lefevre 11, Wang 13b, Wang 13a, Wu 14], plus adaptée au suivi dynamique, suppose que les *endmembers* peuvent varier au cours du temps. C'est l'intérêt de cette seconde catégorie que nous allons exploiter dans cette thèse pour le démixage hyperspectral en-ligne. L'algorithme qui a retenu particulièrement notre attention est celui proposé par [Bucak 09] et nommé *Incremental Non-negative Matrix Factorization* (INMF). Les mises à jour s'effectuent de manière récursive, en considérant les estimations passées, ce qui permet d'intégrer la notion de suivi à l'algorithme. De plus, elle est capable de contrôler de manière adaptative la contribution de chaque échantillon en intégrant dans la fonction objectif des coefficients de pondération. Ces derniers permettent d'établir un compromis entre la cohérence des sources et la variabilité spectrale.

Bien que cet algorithme semble approprié à notre application, il repose sur des règles de mise à jour multiplicatives et souffre ainsi d'une convergence lente et donc d'un temps de traitement susceptible de devenir conséquent en fonction de la taille des données à traiter. De plus, la méthode INMF n'intègre aucune contrainte supplémentaire au modèle ; la solution est généralement non-unique et cela représente un véritable problème en termes d'interprétabilité des résultats. Ainsi, notre objectif est de tirer profit de l'idée sous-jacente de l'algorithme INMF tout en palliant ses divers inconvénients. Nos contributions sont détaillées dans la section suivante.

1.6 Contributions de ce travail de thèse

Dans cette thèse, nous abordons le problème de démélange séquentiel de sources positives dans des mélanges additifs. Cette problématique est rencontrée dans les systèmes d'imagerie hyperspectrale temps-réel (ou en-ligne) mais également dans le traitement de données de grande dimension. Pour ce type de problème, les approches classiques de type NMF présentent un certain nombre de limitations. L'une de ces limitations est la complexité de calcul qui est bien trop élevée et rend la NMF inadaptée pour un traitement industriel en temps réel. Une seconde limitation est la non prise en compte du suivi et de la variabilité spectrale des sources dans le modèle.

Le premier objectif de cette thèse est de développer des algorithmes NMF en-ligne palliant les limitations de la méthode NMF. En s'inspirant de l'algorithme INMF, nous proposons deux principales méthodes de démélange hyperspectral en-ligne fondées sur le gradient optimal de Nesterov [Guan 12a] ou sur ADMM [Boyd 11], adaptées au rythme d'acquisition de l'imageur *pushbroom*. À notre connaissance, de telles méthodes utilisées pour le démélange hyperspectral en-ligne (ainsi que le suivi dynamique des sources) n'ont jamais été rapportées dans la littérature. Nous montrerons qu'elles surpassent la méthode INMF aussi bien en ce qui concerne la vitesse de convergence que de la performance en démélange hyperspectral. Afin de rendre la factorisation unique et d'identifier sans ambiguïté les paramètres d'intérêt, nous proposons d'utiliser une contrainte de volume minimal sur les *endmembers*, permettant de régulariser l'espace des solutions. À ce stade de l'étude, nous n'avons aucune connaissance *a priori* sur les signatures spectrales à extraire et exploitables dans un algorithme ; nous n'avons donc aucune raison de penser que les spectres sont lisses, parcimonieux ou encore significativement différents les uns des autres.

Comme nous l'avons vu précédemment, toute contrainte supplémentaire au modèle est liée à un paramètre de régularisation (hyperparamètre) ; la valeur choisie pour ce dernier doit être réglée de manière adéquate. Par exemple, une contrainte trop forte sur le volume décrit par les *endmembers* force le simplexe estimé à devenir très petit (voire nul) et cela se répercute sur les signatures spectrales qui tendent vers la colinéarité. *A contrario*, une contrainte trop faible laisse place à une multitude de solutions. Une des méthodes les plus connues pour l'estimation de la valeur de l'hyperparamètre est la *courbe en L* [Hansen 92]. Cette dernière a été principalement appliquée à des cas de déconvolution avec une régularisation de type Tikhonov. C'est une courbe de réponse sur une échelle log-log du coût de régularisation par rapport à l'adéquation aux données ; cela permet de caractériser l'ensemble des solutions obtenues en faisant varier les valeurs de l'hyperparamètre. La valeur optimale de ce dernier correspond à la courbure maximale de la courbe en L. Cependant, cette approche présente certains inconvénients : elle est non-convexe et la courbure maximale n'est pas garantie d'être unique. En se basant

sur les travaux menés par [Song 16], nous proposons de reformuler notre problème NMF en-ligne avec contrainte de volume minimal comme un problème de minimisation bi-objectif. Afin d'estimer la valeur optimale du paramètre de régularisation, nous appliquons le critère de courbure maximale (MCC pour *Maximum Curvature Criterion*) ainsi que le critère de distance minimale (MDistC pour *Minimum Distance Criterion*). Contrairement aux approches de type courbe en L qui utilisent une échelle log-log pour tracer la courbe de réponse, MCC et MDistC sont directement appliqués sur une échelle linéaire.

Le second objectif de cette thèse est de résoudre le problème de la variation du rang (nombre de sources) inter-échantillons, lors du traitement séquentiel en-ligne. En effet, nous supposons que le nombre de *end-members* est le même pour chaque échantillon. Cette hypothèse est tout à fait concevable dans le cas du traitement des pièces de bois. Néanmoins, il existe des applications où cette hypothèse n'est pas valable, où une ou plusieurs sources sont susceptibles d'apparaître ou de disparaître au cours du traitement. Ainsi, afin d'adapter nos algorithmes de démixage hyperspectral en-ligne à des applications plus complexes, nous proposons de remplacer la contrainte de volume minimal par un nouveau terme de régularisation consistant à « pousser » les signatures spectrales estimées à chaque tranche, à être proches d'une bibliothèque hyperspectrale générée au préalable par apprentissage ou par connaissance *a priori*. Notons que l'on autorise toujours la variabilité spectrale au cours du traitement, on contraint essentiellement les sources à ne pas s'éloigner d'une solution que l'on considère admissible. Pour gérer les situations où une ou plusieurs sources disparaissent ou apparaissent, une contrainte de parcimonie est imposée sur les abondances, permettant de sélectionner uniquement les sources actives. Cette nouvelle version de nos algorithmes est caractérisée comme semi-supervisée puisque des connaissances *a priori* sont intégrées dans le modèle. Afin de valider les performances des méthodes proposées, elles sont testées sur une application industrielle originale consistant en le suivi de la diffusion de l'eau au sein de produits céréaliers.

Le troisième objectif de cette thèse est l'utilisation des méthodes développées pour la caractérisation du matériau bois. Nous montrons que l'imagerie hyperspectrale proche infrarouge combinée à nos algorithmes de démixage hyperspectral en-ligne est un duo puissant pour la détection robuste de nombreuses singularités qui, pour certaines, ne sont pas détectables par les systèmes de vision industriels existant.

La suite du manuscrit est organisée comme suit :

- Le Chapitre 2 présente et analyse cinq algorithmes en-ligne fondés sur les règles de mise à jour multiplicatives (adaptation de INMF), le gradient optimal de Nesterov et l'ADMM pour le démixage en-ligne des images hyperspectrales acquises par l'imageur industriel *pushbroom*. Nous proposons de combiner à ces méthodes deux termes de régularisation : une contrainte de dispersion minimale ainsi qu'une contrainte de volume minimal. Une étude comparative entre ces deux contraintes est menée. Par ailleurs, une méthode pour l'estimation du paramètre de régularisation est également présentée.
- Le Chapitre 3 traite du problème de l'évolution du nombre de sources lors d'un processus séquentiel, en-ligne. Nous présentons les modifications apportées aux algorithmes en-ligne préalablement développés, permettant de prendre en compte la variation du nombre de sources entre les échantillons. Nous montrons leur efficacité sur un ensemble d'images hyperspectrales réelles.

- Le Chapitre 4 présente les résultats obtenus quant à la détection des singularités du matériau bois. Divers échantillons annotés par des experts sont présentés et traités par nos méthodes.
- Le Chapitre 5 est réservé à la conclusion de nos travaux ainsi qu'aux perspectives.

1.7 Publications

Article de conférence internationale en cours de révision

[C1] Ludivine NUS, Sebastian Miron & David Brie. *A semi-supervised rank tracking algorithm for on-line unmixing of hyperspectral images*. In 18th IEEE International Conference on Acoustics, Speech, and Signal Processing, Barcelona, Spain, 2020.

Articles de conférences internationales

[C2] Ludivine NUS, Sebastian Miron & David Brie. *On-line blind unmixing for hyperspectral pushbroom imaging systems*. In 20th IEEE Workshop on Statistical Signal Processing, Freiburg, Germany, 2018.

[C3] Ludivine NUS, Sebastian Miron & David Brie. *Estimation of the regularization parameter of an on-line NMF with minimum volume constraint*. In 10th IEEE Workshop on Sensor Array and Multichannel Signal Processing, Sheffield, UK, 2018.

[C4] Ludivine NUS, Sebastian Miron & David Brie. *On-line NMF with minimum volume constraint for hyperspectral pushbroom imaging systems and the estimation of the regularization parameter*. In 18th Conference European Network for Business and Industrial Statistics, Nancy, France, 2018.

Articles de conférences nationales

[C5] Benoît Jaillais, Karima Meghar, Ludivine Nus, Sebastian Miron, David Brie & Said Moussaoui. *Unsupervised processing of hyperspectral images*, In CHIMIOMETRIE XIX, Paris, France, 2018.

[C6] Ludivine NUS, Sebastian Miron & David Brie. *Démélange d'images hyperspectrales à l'aide de la NMF en-ligne avec contrainte de dispersion minimale*. In 27^{ème} Colloque GRETSI Traitement du Signal et des Images, Lille, France, 2019.

Chapitre 2

Démélange hyperspectral en-ligne pour les systèmes industriels d'imagerie *pushbroom*

2.1 Introduction

L'imagerie hyperspectrale a fait l'objet d'une attention considérable au cours de la dernière décennie car elle combine la puissance de l'imagerie numérique et de la spectroscopie. Chaque pixel d'une image hyperspectrale fournit des informations spectrales locales sur une scène d'intérêt, suivant un grand nombre de bandes contiguës. Ces informations peuvent être utilisées pour caractériser des objets avec une grande précision dans un certain nombre de domaines comme l'agriculture, la médecine, la défense ou encore l'inspection industrielle. Plusieurs techniques ont été mises au point pour l'acquisition d'images hyperspectrales. Elles peuvent être classées en trois principales catégories [Li 13, Willett 14] : les méthodes d'acquisition à balayage spatial (*whiskbroom* et *pushbroom*), à balayage spectral (*staring*) et instantanée (*snapshot*). Les systèmes *pushbroom* sont utilisés dans de nombreux domaines tels que la sécurité alimentaire [Chen 96, Gowen 07, Qin 17] ou encore le tri de matériaux [Tatzer 05] car ils permettent l'acquisition d'images au rythme que les produits défilent sur un convoyeur. La principale caractéristique des systèmes d'imagerie *pushbroom* est que les images hyperspectrales sont acquises tranche par tranche, de manière séquentielle dans le temps. L'objectif de ce chapitre est d'aborder le démélange hyperspectral en temps réel et de manière récursive d'images hyperspectrales capturées par les systèmes d'imagerie *pushbroom*.

De par les propriétés de non-négativité des images hyperspectrales (observations, sources et abondances), la NMF a été largement utilisée pour résoudre le problème de démélange hyperspectral. Cependant, dans le cas d'un processus industriel en-ligne, où l'acquisition des données s'effectue séquentiellement, la NMF n'est pas l'outil adapté pour le démélange, comme expliqué dans le chapitre précédent. Pour pallier les limitations de la NMF standard, plusieurs algorithmes en-ligne fondés sur la NMF ont été proposés ; ils permettent de mettre à jour séquentiellement les *endmembers* et les abondances à mesure que la taille des données augmente. À cet égard, ils sont adaptés au traitement de flux de données hyperspectrales en continu, car ils permettent de maintenir une complexité de calcul faible et contrôlée. Ces algorithmes peuvent être regroupés en deux principales catégories, en fonction des hypothèses retenues sur les *endmembers*. Dans [Mairal 10, Wang 11, Guan 12b, Zhao 16], il est supposé que les *endmembers* ne varient pas d'un échantillon à l'autre. Les méthodes proposées dans les articles mentionnés ci-dessus

utilisent la descente de gradient stochastique (SGD pour *Stochastic Gradient Descent*) comme stratégie pour réduire la complexité de calcul. Au lieu de considérer tous les échantillons disponibles jusqu'à l'instant considéré, ces méthodes se contentent, à chaque itération, d'un petit lot d'échantillons choisis aléatoirement parmi les données disponibles, ce qui réduit considérablement la charge de calcul. Elles ont été appliquées principalement pour la reconnaissance faciale et l'annotation d'images [Guan 12b], la classification de documents [Wang 11] ainsi que la retouche d'images (par exemple, la suppression de texte) [Mairal 10]. L'algorithme développé par [Zhao 16] utilise l'optimisation *Alternating Direction Method of Multipliers*. Cet algorithme est efficace pour la séparation de l'arrière et du premier plan de séquences d'images issues de vidéos.

Cependant, dans le cadre de cette thèse, un des points importants est le suivi des changements dynamiques des *endmembers* au cours du temps. C'est pourquoi, la seconde catégorie d'algorithmes NMF en-ligne [Cao 07, Bucak 09, Lefevre 11, Zhou 11b, Wang 13b, Wang 13a, Wu 14], qui suppose que les *endmembers* peuvent évoluer entre deux échantillons successifs, est davantage appropriée à notre contexte. Les travaux présentés dans [Cao 07] supposent que les matrices des *endmembers* entre deux échantillons successifs sont reliées par une transformation linéaire, ce qui facilite la mise à jour des paramètres. L'algorithme *Incremental Non-negative Matrix Factorization* (INMF) [Bucak 09] considère que la matrice des *endmembers* évolue lentement entre deux acquisitions consécutives, ce qui permet de réduire les coûts de calculs. De plus, cette méthode est capable de contrôler de manière adaptative la contribution de chaque échantillon en intégrant dans la fonction objectif des coefficients de pondération. Dans le contexte du démixage hyperspectral, cela permet de prendre en compte la variabilité spectrale. Ce schéma est maintenant largement utilisé dans les algorithmes NMF en-ligne. Cependant, comme dans le cas hors-ligne, l'unicité de la solution n'est généralement pas garantie. Afin de surmonter ce problème, divers algorithmes NMF en-ligne avec contraintes ont été développés tels que [Zhou 11b] qui intègre une contrainte de simplexe de volume minimal sur la matrice des abondances, ou encore les méthodes [Wang 13b, Wang 13a, Wu 14] qui utilisent des contraintes de parcimonie et/ou de lissage sur la matrice des *endmembers* et/ou la matrice des abondances. Dans [Lefevre 11], un algorithme NMF en-ligne est proposé, basé sur la divergence d'Itakura Saito, utilisé principalement pour la séparation de sources audios.

Les approches développées dans ce chapitre sont liées aux méthodes issues de la seconde catégorie, puisque qu'elles abordent le problème de traitement de données en-ligne tout en considérant l'évolution possible des *endmembers* au cours du temps. Cependant, la plupart des méthodes mentionnées utilisent des règles de mise à jour multiplicatives, connues pour être très sensibles à l'initialisation et souffrent souvent d'une vitesse de convergence lente [Lin 07b]. Ainsi, nous proposons deux algorithmes NMF en-ligne plus rapides et plus efficaces que ceux fondés sur les mises à jour multiplicatives ; l'un fondé sur l'algorithme ADMM et le second sur le gradient optimal de Nesterov. Compte tenu du caractère non-unique de la décomposition NMF, une contrainte de volume minimal est intégrée aux deux méthodes. Dans la littérature, de nombreuses formulations existent pour cette contrainte ; une étude comparative sera menée permettant de choisir la ou les fonctions de régularisation les plus adaptées à notre modèle en-ligne. De plus, nous proposons un algorithme en-ligne fondé sur des règles de mise à jour multiplicatives avec une contrainte de volume minimal ; il s'agit d'une simple adaptation de l'algorithme de [Zhou 11b] au système d'imagerie hyperspectrale *pushbroom*. Étant donné que nous n'avons pas trouvé

dans la littérature d'autres algorithmes adaptés au traitement en-ligne d'images hyperspectrales, il servira de référence dans ce document. Les originalités de ces algorithmes comparées à l'état de l'art et en particulier aux méthodes proposées dans [Bucak 09, Zhou 11b, Zhao 16] sont les suivantes :

1. L'utilisation de ces algorithmes dans le contexte du démélange en temps réel de données hyperspectrales *pushbroom*.
2. L'intégration d'une capacité de suivi qui permet de détecter les changements de contenu dynamiques.
3. L'ajout d'une contrainte de dispersion ou de volume minimal pour régulariser le problème.

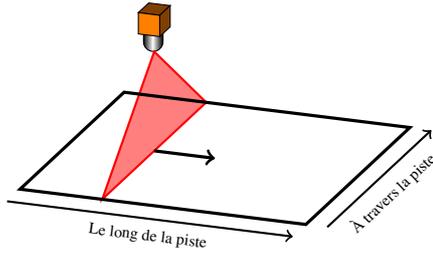
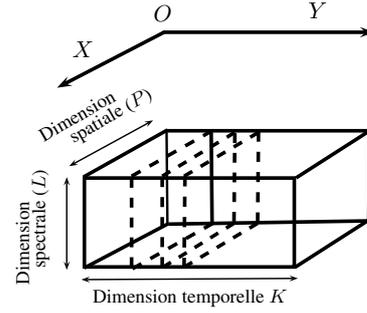
Ce chapitre est organisé de la manière suivante : la Section 2 est consacrée à la formulation du problème NMF en-ligne pour le schéma d'acquisition de l'imageur *pushbroom*. La Section 3 présente une étude comparative des différentes contraintes de volume minimal. Nous soulignons en particulier les différences entre la dispersion minimale et les contraintes basées sur le déterminant. La Section 4 présente la dérivation des différentes méthodes en-ligne proposées ; la convergence et la complexité de calcul sont également discutées. De nombreux résultats sur des données simulées sont donnés en Section 5 afin d'évaluer la vitesse de convergence des algorithmes ainsi que l'influence des divers paramètres. Les méthodes proposées sont également appliquées sur des images hyperspectrales réelles. Dans la Section 6, nous proposons une méthode pour estimer le paramètre de régularisation relatif à la contrainte de dispersion minimale ; pour ce faire, nous formulons notre modèle de démélange hyperspectral en-ligne comme un problème d'optimisation bi-objectif. Des expériences sur des données simulées et réelles permettent d'évaluer les performances de la méthode proposée. Nous concluons le chapitre dans la Section 7.

2.2 Démélange hyperspectral en-ligne pour le système d'acquisition *pushbroom*

2.2.1 Acquisition *pushbroom*

La principale caractéristique des systèmes d'imagerie *pushbroom* est que les images hyperspectrales sont acquises tranches par tranches par des capteurs 2D, séquentiellement dans le temps (ce qui représente la dimension spatiale dite le long de la piste). La Figure 2.1 illustre l'acquisition d'une image hyperspectrale par un imageur *pushbroom*. Chaque tranche (représentée en pointillée sur la Figure 2.2) est une image de dimensions $L \times P$, où L fait référence à la dimension spectrale (nombre de longueurs d'onde) et P représente la dimension spatiale (une ligne de la scène). L'image hyperspectrale est créée en déplaçant le capteur dans la direction spatiale OY ou en déplaçant les objets directement sur un convoyeur (par conséquent, l'imageur est fixe).

L'objectif du démélange hyperspectral aveugle, en-ligne, est de produire des estimations en temps réel des *endmembers* et des abondances, pour chaque nouvelle tranche qui arrive, à la vitesse d'acquisition du système *pushbroom*.

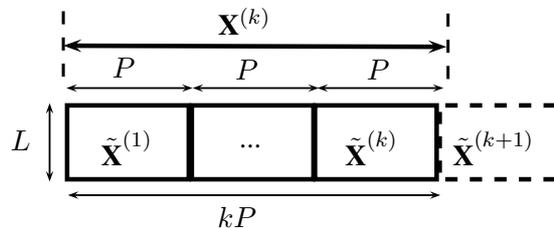
FIGURE 2.1 : Acquisition des données par un imageur *pushbroom*FIGURE 2.2 : Structure des données pour l'acquisition *pushbroom*

2.2.2 Modèle de données

Les méthodes de démélange hyperspectral en-ligne que nous avons développé sont basées sur le modèle NMF. Pour une matrice non-négative $\mathbf{X} \in \mathbb{R}_+^{L \times P}$, la NMF consiste à estimer deux matrices non-négatives $\mathbf{S} \in \mathbb{R}_+^{L \times R}$ et $\mathbf{A} \in \mathbb{R}_+^{R \times P}$ où R est le nombre de *endmembers*, tel que [Lee 99]:

$$\mathbf{X} \approx \mathbf{S}\mathbf{A}.$$

En imagerie hyperspectrale, les P colonnes de \mathbf{X} représentent les échantillons enregistrés suivant L longueurs d'onde. \mathbf{S} est une matrice contenant les R signatures spectrales normalisées (les *endmembers*) et \mathbf{A} est une matrice contenant sur ses colonnes les coefficients de mélange (les abondances) de tous les échantillons enregistrés. Le principe des méthodes en-ligne proposées est de mettre à jour alternativement la matrice des *endmembers* et des abondances estimées à l'instant k lorsqu'un nouvel échantillon (tranche) arrive à l'instant $k + 1$. Une façon de représenter le problème consiste à déplier l'image hyperspectrale, comme illustrée sur la Figure 2.3, où $\tilde{\mathbf{X}}^{(1)}$ caractérise la première tranche de l'image hyperspectrale et $\tilde{\mathbf{X}}^{(k)}$, la $k^{\text{ème}}$ tranche. L'ensemble des données à l'instant $k + 1$, *i.e.*, $\mathbf{X}^{(k+1)}$, peut être représenté comme la concaténation des k premiers échantillons avec le nouvel échantillon entrant *i.e.*, $\mathbf{X}^{(k+1)} = [\mathbf{X}^{(k)} \tilde{\mathbf{X}}^{(k+1)}]$. De la même manière, nous définissons $\mathbf{S}^{(k+1)} = [\mathbf{S}^{(k)} \tilde{\mathbf{S}}^{(k+1)}]$ et $\mathbf{A}^{(k+1)} = [\mathbf{A}^{(k)} \tilde{\mathbf{A}}^{(k+1)}]$.

FIGURE 2.3 : Image hyperspectrale *pushbroom* dépliée

Ainsi, le modèle de mélange en-ligne, pour le $k + 1^{\text{ème}}$ échantillon, est donné par :

$$\tilde{\mathbf{X}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)}.$$

2.2.3 Fonction objectif

La manière la plus simple d'estimer les paramètres du modèle NMF consiste à minimiser le critère des moindres carrés entre les données et le modèle. Soit $\mathcal{J}^{(k)}$ la fonction objectif qui correspond aux k premiers échantillons :

$$\mathcal{J}^{(k)}(\mathbf{S}^{(k)}, \mathbf{A}^{(k)}) = \frac{1}{2} \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(\ell)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2.$$

Lorsque le $k + 1^{\text{ème}}$ échantillon, $\tilde{\mathbf{X}}^{(k+1)}$ arrive, la fonction objectif peut être décomposée de la manière suivante :

$$\mathcal{J}^{(k+1)}(\mathbf{S}^{(k+1)}, \mathbf{A}^{(k+1)}) = \frac{1}{2} \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(\ell)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2 + \frac{1}{2} \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2. \quad (2.1)$$

Sans hypothèses supplémentaires, la fonction objectif (2.1) est simplement un ensemble de k problèmes indépendants des moindres carrés et dans ce cas, la configuration en-ligne n'a aucun intérêt particulier. Cependant, une hypothèse naturelle est que les *endmembers* ne varient que légèrement entre chaque échantillon, *i.e.* $\tilde{\mathbf{S}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(k)} \forall k$. Ainsi, la fonction objectif (2.1) peut être exprimée comme suit :

$$\mathcal{J}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} | \mathbf{A}^{(k)}) = \mathcal{J}^{(k)}(\tilde{\mathbf{S}}^{(k+1)} | \mathbf{A}^{(k)}) + \tilde{\mathcal{J}}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}),$$

où,

$$\begin{aligned} \mathcal{J}^{(k)}(\tilde{\mathbf{S}}^{(k+1)} | \mathbf{A}^{(k)}) &= \frac{1}{2} \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2, \\ \tilde{\mathcal{J}}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}) &= \frac{1}{2} \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2. \end{aligned}$$

Afin d'ajouter une capacité de suivi à l'algorithme (un facteur d'oubli), un coefficient de pondération α ($0 \leq \alpha \leq 1$) est inclus dans la fonction objectif tel que formulé par l'expression (2.2) :

$$\mathcal{J}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}) = \alpha \mathcal{J}^{(k)} + (1 - \alpha) \tilde{\mathcal{J}}^{(k+1)}. \quad (2.2)$$

Une version de cette fonction objectif est utilisée par l'algorithme INMF proposé par [Bucak 09]. Néanmoins, sans contrainte supplémentaire, la solution du problème de minimisation du critère (2.2) n'est pas unique. Le caractère unique de la NMF repose sur la parcimonie des matrices $\tilde{\mathbf{S}}^{(k+1)}$ et $\tilde{\mathbf{A}}^{(k+1)}$. En particulier, si $\tilde{\mathbf{S}}^{(k+1)}$ et/ou $\tilde{\mathbf{A}}^{(k+1)}$ ont uniquement des entrées non nulles, la factorisation NMF n'est pas unique (voir la section 1.4.1.2). Pour réduire l'ensemble des solutions admissibles, nous proposons d'imposer sur la matrice des *endmembers* $\tilde{\mathbf{S}}^{(k+1)}$ une contrainte géométrique forçant le simplexe défini par les *endmembers* à circonscrire les données aussi étroitement que possible. Pour le moment, nous désignons cette contrainte comme étant $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)})$ et nous l'intégrons dans la fonction objectif de la manière suivante :

$$\mathcal{J}_{\text{vol}}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}) = \alpha \mathcal{J}^{(k)} + (1 - \alpha) \tilde{\mathcal{J}}^{(k+1)} + \mu \text{Vol}(\tilde{\mathbf{S}}^{(k+1)}),$$

où $\mu \geq 0$ contrôle le compromis entre l'adéquation aux données et la pénalité. Le choix de la contrainte pour $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)})$ est discuté dans la section suivante. Ainsi, les approches proposées ont pour objectif de résoudre le problème d'optimisation

$$\underset{\tilde{\mathbf{S}}^{(k+1)} \geq \mathbf{0}, \tilde{\mathbf{A}}^{(k+1)} \geq \mathbf{0}}{\text{minimiser}} \mathcal{J}_{\text{vol}}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}) \quad (2.3)$$

pour un choix particulier de la contrainte de volume minimal.

2.3 Régularisation de la solution à l'aide d'une contrainte de volume minimal

La mesure classique pour la contrainte de volume minimal est le déterminant, *i.e.* $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)}) = \det(\tilde{\mathbf{S}}^{(k+1)})$ [Strang 06]. Cependant, le déterminant est défini seulement si la matrice $\tilde{\mathbf{S}}^{(k+1)}$ est une matrice carrée. Dans de nombreux cas, notamment le nôtre, cette exigence n'est pas satisfaite. Ainsi, un autre choix représentatif de $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)})$ est [Schachtner 09, Zhou 11a] :

$$\text{Vol}(\tilde{\mathbf{S}}^{(k+1)}) = \det(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)}), \quad (2.4)$$

ou ses variantes présentées dans [Miao 07, Chan 09]. L'expression (2.4) est mathématiquement justifiable car elle peut s'interpréter comme le volume au carré du simplexe défini par les *endmembers* [Peng 07, Theorem 7].

2.3.1 Identifiabilité

De nombreux travaux ont été menés au cours de la dernière décennie pour comprendre l'identifiabilité du modèle NMF. Les résultats dans [Donoho 04, Laurberg 08, Huang 14] ont montré que si les matrices $\tilde{\mathbf{S}}^{(k+1)}$ et $\tilde{\mathbf{A}}^{(k+1)}$ contiennent toutes les deux des éléments nuls, selon un certain motif, alors le modèle est unique. Cependant, en imagerie hyperspectrale, les *endmembers* $\tilde{\mathbf{S}}^{(k+1)}$ sont susceptibles d'être strictement positifs et, par conséquent, le modèle n'est pas identifiable. C'est dans ce cas que des approches incluant des contraintes de volume minimal peuvent être utilisées avec succès. Sur la base des travaux récents de [Fu 16], nous donnons une condition suffisante d'identifiabilité pour notre modèle en-ligne :

Définition 4 [Condition suffisante d'identifiabilité]

Si $\forall k$, $\text{rang}(\tilde{\mathbf{S}}^{(k)}) = \text{rang}(\tilde{\mathbf{A}}^{(k)}) = R$ et $\tilde{\mathbf{A}}^{(k)}$ est suffisamment dispersée (voir section 1.4.1.2), alors le problème d'optimisation (2.3) a une solution unique.

Cette condition stipule qu'il n'y a pas de perte de rang entre les tranches de l'image hyperspectrale.

2.3.2 Contraintes de volume

Afin de simplifier les notations, les indices $k + 1$ seront omis dans cette section. L'utilisation de la contrainte de volume minimal (2.4) rend le sous-problème en $\tilde{\mathbf{S}}$ non-convexe et peut être plus difficile à résoudre. Pour pallier les problèmes de minimisation du déterminant, d'autres substituts pour la

régularisation du volume peuvent être envisagés. Une formulation alternative pour $\text{Vol}(\tilde{\mathbf{S}})$ proposée par [Bioucas-Dias 09] est :

$$\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right) \right). \quad (2.5)$$

Le choix du logarithme du déterminant plutôt que du déterminant lui-même est principalement motivé par des raisons algorithmiques, car il simplifie considérablement les règles de mise à jour de $\tilde{\mathbf{S}}$. En effet, chacune des expressions (2.4) et (2.5) est dérivable par rapport à $\tilde{\mathbf{S}}$ et peut s'écrire comme suit :

$$\begin{aligned} \frac{\partial \det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right)}{\partial \tilde{\mathbf{S}}} &= 2 \det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right) \tilde{\mathbf{S}} \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right)^{-1}, \\ \frac{\partial \log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right) \right)}{\partial \tilde{\mathbf{S}}} &= 2 \tilde{\mathbf{S}} \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right)^{-1}. \end{aligned}$$

On constate que l'utilisation du logarithme permet de simplifier significativement l'expression de la dérivée. Afin d'éviter de fortes valeurs négatives de (2.5) lorsque $\tilde{\mathbf{S}}$ devient proche de la singularité *i.e.* $\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right) \rightarrow 0$, une version modifiée a été proposée par [Fu 16, Leplat 19] :

$$\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right), \quad (2.6)$$

où \mathbf{I} est la matrice identité et $\epsilon > 0$ est une faible valeur. Alors que (2.5) est une fonction concave, il apparaît que (2.6) a un point stationnaire en $\mathbf{0}$ et est convexe au voisinage de $\mathbf{0}$. En effet, considérons $\{\lambda_i^2, i = 1, \dots, R\}$ les valeurs propres ordonnées de $\tilde{\mathbf{S}}^T \tilde{\mathbf{S}}$, les λ_i sont les valeurs singulières de $\tilde{\mathbf{S}}$. Alors on a :

$$\begin{aligned} \log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right) &= \sum_{i=1}^R \log \left(\lambda_i^2 + \epsilon \right) = f(\lambda_i), \\ \frac{\partial f(\lambda_i)}{\partial \lambda_i} &= \frac{2\lambda_i}{\lambda_i^2 + \epsilon}. \end{aligned}$$

Clairement, on constate que $\lim_{\lambda_i \rightarrow 0} \frac{\partial f(\lambda_i)}{\partial \lambda_i} = 0$. De plus, $\frac{\partial^2 f(\lambda_i)}{\partial \lambda_i^2} = \frac{2\epsilon - 2\lambda_i^2}{(\lambda_i^2 + \epsilon)^2} \geq 0$ si $\lambda_i^2 \leq \epsilon$. Ainsi, au voisinage de 0 nous pouvons écrire :

$$\begin{aligned} \log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right) &= R \log(\epsilon) + \sum_{i=1}^R \log \left(\frac{\lambda_i^2}{\epsilon} + 1 \right) \\ &\approx R \log(\epsilon) + \frac{1}{\epsilon} \sum_{i=1}^R \lambda_i^2 \\ &= R \log(\epsilon) + \frac{1}{\epsilon} \text{trace} \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right). \end{aligned}$$

En d'autres termes, au voisinage des $\lambda_i = 0$, $\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right)$ se comporte essentiellement comme $\text{trace} \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right)$, l'étendue du voisinage étant contrôlée par la valeur de ϵ . Cela a d'importantes conséquences : tandis que $\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right) \right)$ favorise la perte de rang à mesure que le paramètre de régularisation μ augmente [Fazel 03], l'utilisation de $\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right)$ préserve une solution de rang plein sur les

colonnes de $\tilde{\mathbf{S}}$ même pour des grandes valeurs de μ .

La Figure 2.4 représente les courbes de niveau obtenues pour $\sum_{i=1}^2 \lambda_i^2$, $\sum_{i=1}^2 \log(\lambda_i^2 + \epsilon)$ et $\sum_{i=1}^2 \log(\lambda_i^2)$ avec λ_i variant de 0.1 à 1 et pour $R = 2$. Pour des grandes valeurs de ϵ (e.g. 0.1) et pour λ_i proche de 0, nous constatons que $\sum_{i=1}^2 \log(\lambda_i^2 + \epsilon)$ se conduit de manière similaire à $\sum_{i=1}^2 \lambda_i^2$ tandis qu'il se comporte comme $\sum_{i=1}^2 \log(\lambda_i^2)$ pour des valeurs de ϵ plus faibles (e.g. 0.05). Les flèches sur les Figures 2.4 (b) et 2.4 (d) indiquent la limite pour laquelle $\log(\det(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I}))$ est convexe.

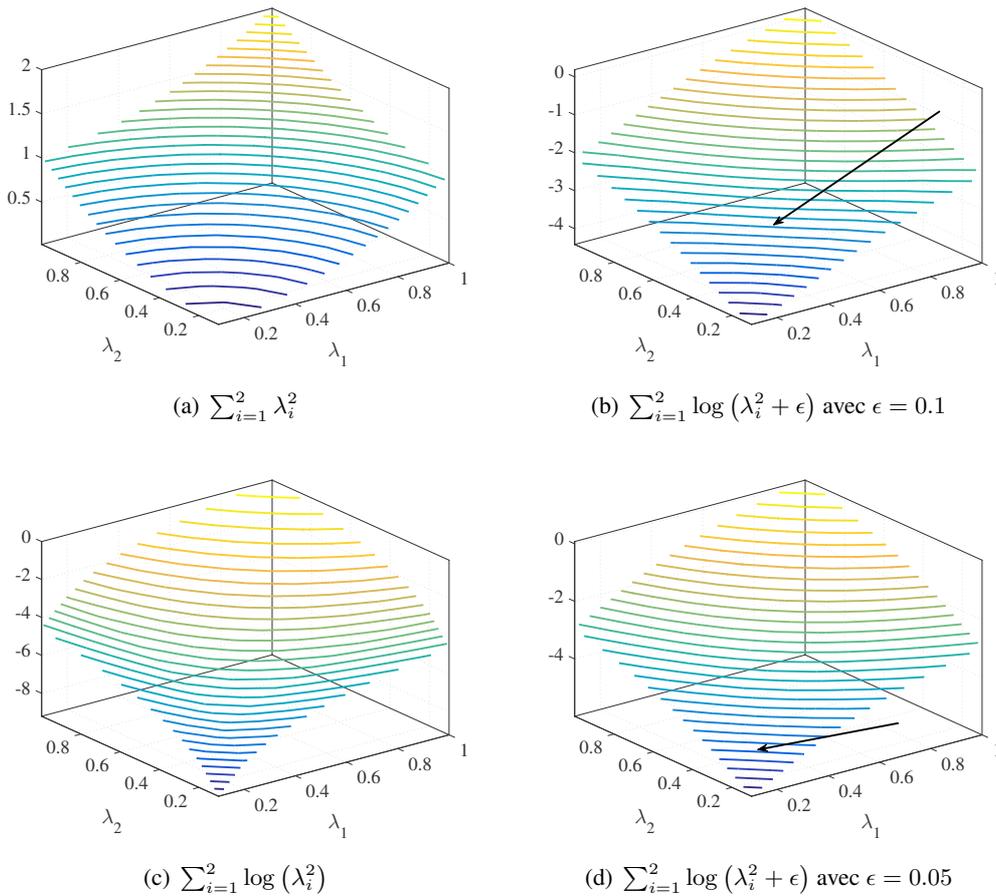


FIGURE 2.4 : Courbes de niveau

Finalement, un substitut strictement convexe pour $\text{Vol}(\tilde{\mathbf{S}})$ a été proposé par [Berman 04] comme :

$$\text{trace}(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \mathbf{D}) = \text{trace}(\tilde{\mathbf{S}} \mathbf{D} \tilde{\mathbf{S}}^T), \quad (2.7)$$

où $\mathbf{D} = \mathbf{I} - \frac{1}{R} \mathbf{1} \mathbf{1}^T$ ($\mathbf{1}$ est un vecteur de dimension $R \times 1$ contenant uniquement des 1). Cette contrainte peut s'interpréter comme une mesure de la dispersion des *endmembers* par rapport à leur centre de gravité. L'expression (2.7) est facilement dérivable par rapport à $\tilde{\mathbf{S}}$ et s'exprime selon :

$$\frac{\partial \text{trace}(\tilde{\mathbf{S}} \mathbf{D} \tilde{\mathbf{S}}^T)}{\partial \tilde{\mathbf{S}}} = 2 \tilde{\mathbf{S}} \mathbf{D},$$

où \mathbf{D} est fixe à chaque itération et pour chaque nouvel échantillon.

Dans la suite, nous nommons la contrainte $\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right)$ par MVC (pour *Minimum Volume Constraint*) et $\text{trace} \left(\tilde{\mathbf{S}} \mathbf{D} \tilde{\mathbf{S}}^T \right)$, par MDC (pour *Minimum Dispersion Constraint*). Afin d'illustrer la différence entre ces deux fonctions de régularisation, considérons une matrice $\tilde{\mathbf{S}}$ dont les colonnes sont deux vecteurs unitaires :

$$\tilde{\mathbf{S}} = \begin{bmatrix} 1 & \cos(\beta) \\ 0 & \sin(\beta) \end{bmatrix},$$

où β est un angle variant entre 0 et $\frac{\pi}{2}$. Si $\beta = 0$, les deux vecteurs sont colinéaires (le volume est nul); *a contrario*, si $\beta = \frac{\pi}{2}$, les deux vecteurs deviennent orthogonaux (le volume est maximal et égal à 1). Nous avons tracé sur la Figure 2.5, MVC pour différentes valeurs de ϵ et MDC, en fonction de β variant de 0 à $\frac{\pi}{2}$. Les courbes ont été normalisées pour définir leur valeur minimale à 0 et leur valeur maximale à 1. Pour des faibles valeurs de ϵ , MVC favorise des solutions de rang déficient. Pour des grandes valeurs de ϵ , la perte de rang n'est plus promue et MDC et MVC conservent une solution de rang plein sur les colonnes de $\tilde{\mathbf{S}}$. Cette propriété de préservation du rang est essentielle pour garantir que la condition suffisante d'identifiabilité établie en section 2.3.1 soit satisfaite.

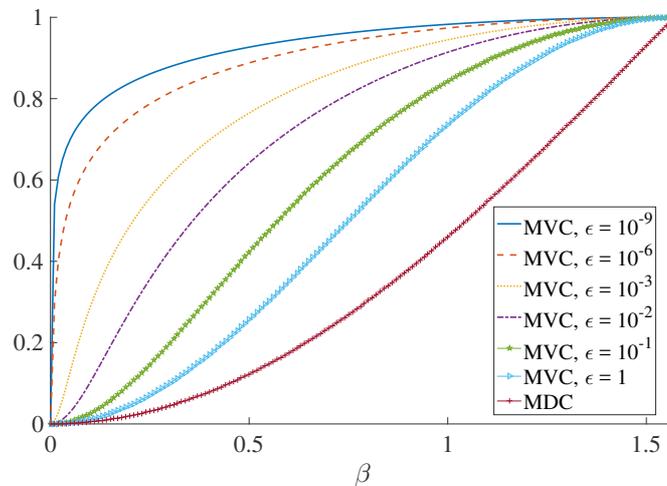


FIGURE 2.5 : Fonctions de régularisation MVC et MDC en fonction de l'angle β

Il n'existe pas de « meilleur choix » concernant la contrainte de volume minimal. La seule exigence, dans notre cas, est que le rang de la décomposition soit conservé d'une tranche de l'image hyperspectrale à l'autre. Ainsi, les contraintes $\log \left(\det \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right) \right)$ (à condition que ϵ soit compris dans un certain intervalle) et $\text{trace} \left(\tilde{\mathbf{S}} \mathbf{D} \tilde{\mathbf{S}}^T \right)$ semblent parfaitement adaptées à cette tâche d'estimation. Bien entendu, elles possèdent chacune des propriétés différentes. Par exemple, bien que MVC ne soit pas strictement convexe, elle peut être directement interprétée comme le volume du simplexe. *A contrario*, MDC est strictement convexe et offre des propriétés d'optimisation intéressantes, mais ne représente qu'une approximation du volume du simplexe. Ainsi, dans la suite, nous proposons des variantes de nos algorithmes utilisant les deux types de pénalités ; diverses simulations nous guideront quant aux choix de celle la plus adaptée.

2.4 Approches proposées

Dans cette section, nous présentons les algorithmes en-ligne que nous avons développé, fondés respectivement sur les règles de mise à jour multiplicatives, sur l'ADMM ainsi que le gradient optimal de Nesterov, tous dérivés à partir de la fonction objectif contrainte ci après :

$$\begin{aligned} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= \alpha \frac{1}{2} \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2 + (1 - \alpha) \frac{1}{2} \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2 \\ &+ \mu \text{Vol} \left(\tilde{\mathbf{S}}^{(k+1)} \right), \end{aligned} \quad (2.8)$$

où

$$\text{Vol} \left(\tilde{\mathbf{S}}^{(k+1)} \right) = \text{trace} \left(\tilde{\mathbf{S}}^{(k+1)} \mathbf{D} \tilde{\mathbf{S}}^{(k+1)T} \right), \quad (2.9)$$

ou

$$\text{Vol} \left(\tilde{\mathbf{S}}^{(k+1)} \right) = \log \left(\det \left(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} + \epsilon \mathbf{I} \right) \right). \quad (2.10)$$

2.4.1 Minimisation de la fonction objectif à l'aide des mises à jour multiplicatives

Les règles de mise à jour multiplicatives (MU) sont une extension de l'algorithme du gradient projeté; en considérant la fonction objectif (2.8), la forme générale des mises à jour s'écrit de la manière suivante :

$$\begin{aligned} \tilde{\mathbf{A}}^{(k+1)} &\leftarrow \left[\tilde{\mathbf{A}}^{(k+1)} - \eta_{\tilde{\mathbf{A}}^{(k+1)}} \nabla_{\tilde{\mathbf{A}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \right]_+, \\ \tilde{\mathbf{S}}^{(k+1)} &\leftarrow \left[\tilde{\mathbf{S}}^{(k+1)} - \eta_{\tilde{\mathbf{S}}^{(k+1)}} \nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \right]_+, \end{aligned}$$

où η est le pas de la descente de gradient et les termes $\nabla_{\tilde{\mathbf{A}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)}$ et $\nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)}$ représentent respectivement le gradient de $\mathcal{J}_{vol}^{(k+1)}$ par rapport à $\tilde{\mathbf{A}}^{(k+1)}$ et $\tilde{\mathbf{S}}^{(k+1)}$. Dans le cas multiplicatif, le pas est exprimé en fonction des matrices $\tilde{\mathbf{A}}^{(k+1)}$ et $\tilde{\mathbf{S}}^{(k+1)}$ [Lee 01], ce qui garantit la non-négativité des paramètres tout en évitant une projection à chaque itération. Dans ce cas, les règles de mise à jour MU peuvent s'écrire sous la forme générale suivante :

$$\tilde{\mathbf{A}}^{(k+1)} \leftarrow \tilde{\mathbf{A}}^{(k+1)} \odot \frac{\nabla_{\tilde{\mathbf{A}}^{(k+1)}}^- \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)}{\nabla_{\tilde{\mathbf{A}}^{(k+1)}}^+ \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)}, \quad (2.11)$$

$$\tilde{\mathbf{S}}^{(k+1)} \leftarrow \tilde{\mathbf{S}}^{(k+1)} \odot \frac{\nabla_{\tilde{\mathbf{S}}^{(k+1)}}^- \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)}{\nabla_{\tilde{\mathbf{S}}^{(k+1)}}^+ \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)}. \quad (2.12)$$

Le terme $\nabla^- \mathcal{J}_{vol}^{(k+1)}$ regroupe les quantités du gradient précédées par un signe négatif tandis que $\nabla^+ \mathcal{J}_{vol}^{(k+1)}$ regroupe les quantités du gradient précédées par un signe positif. Comme $\|\mathbf{X}\|_F^2 = \text{trace}(\mathbf{X}^T \mathbf{X})$, la fonction objectif $\mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)$ peut s'écrire sous la forme équivalente suivante :

$$\begin{aligned} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= \frac{1}{2} \alpha \sum_{l=1}^k \text{trace} \left(\left(\tilde{\mathbf{X}}^{(l)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(l)} \right)^T \left(\tilde{\mathbf{X}}^{(l)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(l)} \right) \right) \\ &+ \frac{1}{2} (1 - \alpha) \text{trace} \left(\left(\tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right)^T \left(\tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right) \right) \\ &+ \mu \text{Vol} \left(\tilde{\mathbf{S}}^{(k+1)} \right). \end{aligned}$$

Étant donné qu'aucune contrainte n'est appliquée sur la matrice $\tilde{\mathbf{A}}^{(k+1)}$, $\nabla_{\tilde{\mathbf{A}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)}$ s'écrit de la même manière pour l'utilisation des contraintes (2.9) et (2.10) :

$$\nabla_{\tilde{\mathbf{A}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) = \frac{1}{2} (1 - \alpha) \left(-2\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{X}}^{(k+1)} + 2\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right),$$

où

$$\begin{aligned} \nabla_{\tilde{\mathbf{A}}^{(k+1)}}^- \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= (1 - \alpha) \tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{X}}^{(k+1)}, \\ \nabla_{\tilde{\mathbf{A}}^{(k+1)}}^+ \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= (1 - \alpha) \tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)}. \end{aligned}$$

En revanche, $\nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)}$ s'exprime sous deux formes distinctes :

- Si $\text{Vol} \left(\tilde{\mathbf{S}}^{(k+1)} \right) = \text{trace} \left(\tilde{\mathbf{S}}^{(k+1)} \mathbf{D} \tilde{\mathbf{S}}^{(k+1)T} \right)$, alors :

$$\begin{aligned} \nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= \frac{1}{2} \alpha \sum_{l=1}^k \left(-2\tilde{\mathbf{X}}^{(l)} \tilde{\mathbf{A}}^{(l)T} + 2\tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(l)} \tilde{\mathbf{A}}^{(l)T} \right) \\ &+ \frac{1}{2} (1 - \alpha) \left(-2\tilde{\mathbf{X}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T} + 2\tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T} \right) \\ &+ 2 \mu \tilde{\mathbf{S}}^{(k+1)} \mathbf{D}, \end{aligned}$$

où $\mathbf{D} = \mathbf{I} - \frac{1}{R} \mathbf{1}\mathbf{1}^T$.

- Si $\text{Vol} \left(\tilde{\mathbf{S}}^{(k+1)} \right) = \log \left(\det \left(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} + \epsilon \mathbf{I} \right) \right)$, alors :

$$\begin{aligned} \nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= \frac{1}{2} \alpha \sum_{l=1}^k \left(-2\tilde{\mathbf{X}}^{(l)} \tilde{\mathbf{A}}^{(l)T} + 2\tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(l)} \tilde{\mathbf{A}}^{(l)T} \right) \\ &+ \frac{1}{2} (1 - \alpha) \left(-2\tilde{\mathbf{X}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T} + 2\tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T} \right) \\ &+ 2 \mu \tilde{\mathbf{S}}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} + \epsilon \mathbf{I} \right)^{-1}. \end{aligned}$$

En se basant sur la forme générale (2.11), la mise à jour de $\tilde{\mathbf{A}}^{(k+1)}$ peut s'exprimer comme suit :

$$\tilde{\mathbf{A}}^{(k+1)} = \tilde{\mathbf{A}}^{(k)} \odot \frac{\tilde{\mathbf{S}}^{(k)T} \tilde{\mathbf{X}}^{(k+1)}}{\tilde{\mathbf{S}}^{(k)T} \tilde{\mathbf{S}}^{(k)} \tilde{\mathbf{A}}^{(k)} + \delta \mathbf{I}},$$

où δ est une faible valeur évitant les divisions par zéro. $\tilde{\mathbf{A}}^{(k)}$ et $\tilde{\mathbf{S}}^{(k)}$ sont les abondances et les *end-members* estimés à l'échantillon k . En effet, bien que l'on suppose en théorie que $\tilde{\mathbf{S}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(k)}$, les paramètres à l'instant $k+1$ sont mis à jour grâce aux résultats précédemment obtenus à l'instant k , ce qui permet un suivi de la variabilité des sources hyperspectrales d'un échantillon à l'autre.

La mise à jour de $\tilde{\mathbf{S}}^{(k+1)}$ peut s'exprimer sous deux formes différentes, en fonction de la contrainte choisie. En effet, en se basant sur la forme générale (2.12) :

- Si $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)}) = \text{trace}(\tilde{\mathbf{S}}^{(k+1)} \mathbf{D} \tilde{\mathbf{S}}^{(k+1)T})$, alors :

$$\tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{S}}^{(k)} \odot \frac{\mathbf{N}^{(k+1)} + 2\mu \tilde{\mathbf{S}}^{(k)} \frac{1}{R} \mathbf{1} \mathbf{1}^T}{\tilde{\mathbf{S}}^{(k)} \mathbf{M}^{(k+1)} + 2\mu \tilde{\mathbf{S}}^{(k)} \mathbf{I} + \delta \mathbf{I}},$$

où $\mathbf{N}^{(k+1)} = \alpha \sum_{l=1}^k \tilde{\mathbf{X}}^{(l)} \tilde{\mathbf{A}}^{(l)T} + (1 - \alpha) \tilde{\mathbf{X}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T}$ et $\mathbf{M}^{(k+1)} = \alpha \sum_{l=1}^k \tilde{\mathbf{A}}^{(l)} \tilde{\mathbf{A}}^{(l)T} + (1 - \alpha) \tilde{\mathbf{A}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T}$. Selon [Bucak 09], sous l'hypothèse que $\tilde{\mathbf{S}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(k)}$, nous pouvons écrire $\mathbf{N}^{(k)} \approx \sum_{l=1}^k \tilde{\mathbf{X}}^{(l)} \tilde{\mathbf{A}}^{(l)T}$ et $\mathbf{M}^{(k)} \approx \sum_{l=1}^k \tilde{\mathbf{A}}^{(l)} \tilde{\mathbf{A}}^{(l)T}$. Ainsi, $\mathbf{N}^{(k+1)}$ et $\mathbf{M}^{(k+1)}$ sont définis comme suit :

$$\begin{aligned} \mathbf{N}^{(k+1)} &= \alpha \mathbf{N}^{(k)} + (1 - \alpha) \tilde{\mathbf{X}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T}, \\ \mathbf{M}^{(k+1)} &= \alpha \mathbf{M}^{(k)} + (1 - \alpha) \tilde{\mathbf{A}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)T}. \end{aligned}$$

- Si $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)}) = \log(\det(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} + \epsilon \mathbf{I}))$, alors :

$$\tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{S}}^{(k)} \odot \frac{\mathbf{N}^{(k+1)} (\tilde{\mathbf{S}}^{(k)T} \tilde{\mathbf{S}}^{(k)} + \epsilon \mathbf{I})}{\tilde{\mathbf{S}}^{(k)} \mathbf{M}^{(k+1)} (\tilde{\mathbf{S}}^{(k)T} \tilde{\mathbf{S}}^{(k)} + \epsilon \mathbf{I}) + 2\mu \tilde{\mathbf{S}}^{(k)} \mathbf{I} + \delta \mathbf{I}}. \quad (2.13)$$

Pour établir la relation (2.13), il faut calculer la dérivée de $\log(\det(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} + \epsilon \mathbf{I}))$. Cette dérivée implique une inversion qui ne garantit plus la non-négativité de $\tilde{\mathbf{S}}^{(k+1)}$ et exclut toute dérivation d'une règle de mise à jour multiplicative. Afin de supprimer cette inversion, nous proposons de nous inspirer de [Zhou 11b] qui utilise le gradient naturel au lieu du gradient ordinaire pour mettre à jour $\tilde{\mathbf{S}}^{(k+1)}$. Le gradient naturel se formule comme suit [Amari 98, Theorem 6] :

$$\tilde{\nabla}_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) = \nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) \tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)}. \quad (2.14)$$

Suite à quelques manipulations algébriques et en adaptant (2.14) à notre problème, l'expression (2.13) provient de la mise à jour générale suivante :

$$\tilde{\mathbf{S}}^{(k+1)} \leftarrow \tilde{\mathbf{S}}^{(k+1)} \odot \frac{\tilde{\nabla}_{\tilde{\mathbf{S}}^{(k+1)}}^- \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)}{\tilde{\nabla}_{\tilde{\mathbf{S}}^{(k+1)}}^+ \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right)},$$

où $\tilde{\nabla}_{\tilde{\mathbf{S}}^{(k+1)}}^{(k+1)}$ désigne le gradient naturel.

Nous désignons ces nouvelles méthodes par OMDC-MU (*On-line Minimum Dispersion Constraint-Multiplicative*) et OMVC-MU (*On-line Minimum Volume Constraint-Multiplicative*). Les implantations algorithmiques sont présentées dans la section 2.4.4.

2.4.2 Minimisation de la fonction objectif à l'aide de l'optimisation ADMM

Un des points clés pour accélérer les algorithmes en-ligne est l'utilisation des mises à jour explicites des paramètres d'intérêt. Lorsque l'on utilise l'algorithme ADMM pour minimiser la fonction objectif (2.8), MVC ne permet pas d'obtenir des mises à jour explicites. Ainsi, dans le cas de l'approche ADMM, MDC est essentiellement utilisée. En définissant,

$$\begin{aligned} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) &= \alpha \frac{1}{2} \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2 + (1 - \alpha) \frac{1}{2} \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2 \\ &\quad + \mu \text{trace} \left(\tilde{\mathbf{S}}^{(k+1)} \mathbf{D} \tilde{\mathbf{S}}^{(k+1)T} \right), \end{aligned}$$

et en introduisant deux variables auxiliaires $\tilde{\mathbf{U}}$ et $\tilde{\mathbf{V}}$, le problème (2.3) est équivalent à :

$$\begin{aligned} &\underset{\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}, \tilde{\mathbf{V}}^{(k+1)}, \tilde{\mathbf{U}}^{(k+1)}}{\text{minimiser}} \quad \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) + \mathbb{I}_{\mathbb{R}_+} \left(\tilde{\mathbf{V}}^{(k+1)} \right) + \mathbb{I}_{\mathbb{R}_+} \left(\tilde{\mathbf{U}}^{(k+1)} \right), \\ &\text{sous contraintes } \tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{U}}^{(k+1)} \text{ et } \tilde{\mathbf{A}}^{(k+1)} = \tilde{\mathbf{V}}^{(k+1)}. \end{aligned} \quad (2.15)$$

Par convenance algorithmique, nous utilisons la version réduite du Lagrangien augmenté [Boyd 11] dans laquelle les termes linéaires et quadratiques sont combinés et les variables duales sont mises à l'échelle. Ainsi, le Lagrangien augmenté \mathcal{L} correspondant au problème (2.15) est donné par :

$$\begin{aligned} &\mathcal{L} \left(\tilde{\mathbf{A}}^{(k+1)}, \tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{V}}^{(k+1)}, \tilde{\mathbf{U}}^{(k+1)}, \tilde{\mathbf{\Pi}}^{(k+1)}, \tilde{\mathbf{\Lambda}}^{(k+1)} \right) \\ &= \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) + \frac{\rho}{2} \left\| \tilde{\mathbf{A}}^{(k+1)} - \tilde{\mathbf{V}}^{(k+1)} + \tilde{\mathbf{\Pi}}^{(k+1)} \right\|_F^2 - \frac{\rho}{2} \left\| \tilde{\mathbf{\Pi}}^{(k+1)} \right\|_F^2 \\ &\quad + \frac{\rho}{2} \left\| \tilde{\mathbf{S}}^{(k+1)} - \tilde{\mathbf{U}}^{(k+1)} + \tilde{\mathbf{\Lambda}}^{(k+1)} \right\|_F^2 - \frac{\rho}{2} \left\| \tilde{\mathbf{\Lambda}}^{(k+1)} \right\|_F^2 + \mathbb{I}_{\mathbb{R}_+} \left(\tilde{\mathbf{V}}^{(k+1)} \right) + \mathbb{I}_{\mathbb{R}_+} \left(\tilde{\mathbf{U}}^{(k+1)} \right). \end{aligned} \quad (2.16)$$

Le paramètre ρ contrôle la convergence de la méthode. $\tilde{\mathbf{\Pi}}^{(k+1)}$ et $\tilde{\mathbf{\Lambda}}^{(k+1)}$ sont les versions mises à l'échelle des variables duales correspondant aux contraintes d'égalité $\tilde{\mathbf{A}}^{(k+1)} = \tilde{\mathbf{V}}^{(k+1)}$ et $\tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{U}}^{(k+1)}$, respectivement. L'optimisation ADMM minimise le Lagrangien augmenté (2.16) par rapport aux paramètres $(\tilde{\mathbf{A}}^{(k+1)}, \tilde{\mathbf{V}}^{(k+1)})$ et $(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{U}}^{(k+1)})$, puis met à jour les variables duales $\tilde{\mathbf{\Pi}}^{(k+1)}$ et $\tilde{\mathbf{\Lambda}}^{(k+1)}$. Les mises à jour des paramètres à la tranche $k + 1$ s'expriment comme suit :

$$\tilde{\mathbf{A}}^{(k+1)} = \left((1 - \alpha)\tilde{\mathbf{S}}^{(k)T}\tilde{\mathbf{S}}^{(k)} + \rho\mathbf{I} \right)^{-1} \left((1 - \alpha)\tilde{\mathbf{S}}^{(k)T}\tilde{\mathbf{X}}^{(k+1)} + \rho(\tilde{\mathbf{V}}^{(k)} - \tilde{\mathbf{\Pi}}^{(k)}) \right), \quad (2.17a)$$

$$\tilde{\mathbf{V}}^{(k+1)} = \max(\mathbf{0}, \tilde{\mathbf{A}}^{(k+1)} + \tilde{\mathbf{\Pi}}^{(k)}), \quad (2.17b)$$

$$\tilde{\mathbf{\Pi}}^{(k+1)} = \tilde{\mathbf{\Pi}}^{(k)} + \tilde{\mathbf{A}}^{(k+1)} - \tilde{\mathbf{V}}^{(k+1)}, \quad (2.17c)$$

$$\tilde{\mathbf{S}}^{(k+1)} = \left(\mathbf{N}^{(k+1)} + \rho(\tilde{\mathbf{U}}^{(k)} - \tilde{\mathbf{\Lambda}}^{(k)}) \right) \left(\mathbf{M}^{(k+1)} + 2\mu\mathbf{D} + \rho\mathbf{I} \right)^{-1}, \quad (2.17d)$$

$$\tilde{\mathbf{U}}^{(k+1)} = \max(\mathbf{0}, \tilde{\mathbf{S}}^{(k+1)} + \tilde{\mathbf{\Lambda}}^{(k)}), \quad (2.17e)$$

$$\tilde{\mathbf{\Lambda}}^{(k+1)} = \tilde{\mathbf{\Lambda}}^{(k)} + \tilde{\mathbf{S}}^{(k+1)} - \tilde{\mathbf{U}}^{(k+1)}. \quad (2.17f)$$

Les mises à jour de $\tilde{\mathbf{V}}^{(k+1)}$ et $\tilde{\mathbf{U}}^{(k+1)}$ consistent simplement à mettre à zéro les valeurs négatives relatives respectivement aux quantités $\tilde{\mathbf{A}}^{(k+1)} + \tilde{\mathbf{\Pi}}^{(k)}$ et $\tilde{\mathbf{S}}^{(k+1)} + \tilde{\mathbf{\Lambda}}^{(k)}$. Nous désignons cette nouvelle méthode par OMDC-ADMM (*On-line Minimum Dispersion Constraint-Alternating Direction Method of Multipliers*).

2.4.3 Minimisation de la fonction objectif à l'aide du gradient optimal de Nesterov

La méthode du gradient optimal de Nesterov consiste à mettre à jour les paramètres $\tilde{\mathbf{S}}^{(k+1)}$ et $\tilde{\mathbf{A}}^{(k+1)}$ par une succession de descentes de gradient avec un pas adaptatif dépendant de la constante de *Lipschitz* [Guan 12a]. Cette méthode garantit la convergence vers un minimum local à condition que la fonction objectif à minimiser soit convexe et que son gradient soit Lipschitz-continu. Dans le cas où $\text{Vol}(\tilde{\mathbf{S}}^{(k+1)}) = \text{trace}(\tilde{\mathbf{S}}^{(k+1)}\mathbf{D}\tilde{\mathbf{S}}^{(k+1)T})$, la fonction objectif (2.8) est convexe en $\tilde{\mathbf{S}}^{(k+1)}$ pour la matrice $\tilde{\mathbf{A}}^{(k+1)}$ fixée (puisque MDC est convexe) et inversement. De plus pour toutes matrices $^{(1)}\tilde{\mathbf{A}}, ^{(2)}\tilde{\mathbf{A}}, ^{(1)}\tilde{\mathbf{S}}$ et $^{(2)}\tilde{\mathbf{S}}$ nous avons :

$$\left\| \nabla_{^{(1)}\tilde{\mathbf{A}}}\mathcal{J}_{vol} \left(^{(1)}\tilde{\mathbf{A}} \right) - \nabla_{^{(2)}\tilde{\mathbf{A}}}\mathcal{J}_{vol} \left(^{(2)}\tilde{\mathbf{A}} \right) \right\|_F \leq \Psi_{\tilde{\mathbf{A}}} \left\| ^{(1)}\tilde{\mathbf{A}} - ^{(2)}\tilde{\mathbf{A}} \right\|_F, \quad (2.18)$$

$$\left\| \nabla_{^{(1)}\tilde{\mathbf{S}}}\mathcal{J}_{vol} \left(^{(1)}\tilde{\mathbf{S}} \right) - \nabla_{^{(2)}\tilde{\mathbf{S}}}\mathcal{J}_{vol} \left(^{(2)}\tilde{\mathbf{S}} \right) \right\|_F \leq \Psi_{\tilde{\mathbf{S}}} \left\| ^{(1)}\tilde{\mathbf{S}} - ^{(2)}\tilde{\mathbf{S}} \right\|_F, \quad (2.19)$$

où $\Psi_{\tilde{\mathbf{A}}}$ et $\Psi_{\tilde{\mathbf{S}}}$ sont les constantes de Lipschitz relatives à chaque sous-problème. Les expressions (2.18) et (2.19) peuvent se réécrire sous la forme équivalente suivante :

$$\left\| \nabla_{^{(1)}\tilde{\mathbf{A}}}\mathcal{J}_{vol} \left(^{(1)}\tilde{\mathbf{A}} \right) - \nabla_{^{(2)}\tilde{\mathbf{A}}}\mathcal{J}_{vol} \left(^{(2)}\tilde{\mathbf{A}} \right) \right\|_F \leq (1 - \alpha) \left\| \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right\|_2 \times \left\| ^{(1)}\tilde{\mathbf{A}} - ^{(2)}\tilde{\mathbf{A}} \right\|_F,$$

$$\left\| \nabla_{(1)\tilde{\mathbf{S}}} \mathcal{J}_{vol} \left({}^{(1)}\tilde{\mathbf{S}} \right) - \nabla_{(2)\tilde{\mathbf{S}}} \mathcal{J}_{vol} \left({}^{(2)}\tilde{\mathbf{S}} \right) \right\|_F \leq (\|\mathbf{M}\|_2 + 2\mu\|\mathbf{D}\|_2) \times \left\| {}^{(1)}\tilde{\mathbf{S}} - {}^{(2)}\tilde{\mathbf{S}} \right\|_F,$$

où $\nabla_{\tilde{\mathbf{A}}} \mathcal{J}_{vol} \left(\tilde{\mathbf{A}} \right) = (1 - \alpha) \left(-\tilde{\mathbf{S}}^T \tilde{\mathbf{X}} + \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \tilde{\mathbf{A}} \right)$ et $\nabla_{\tilde{\mathbf{S}}} \mathcal{J}_{vol} \left(\tilde{\mathbf{S}} \right) = \tilde{\mathbf{S}} \mathbf{M} - \mathbf{N} + 2\mu \tilde{\mathbf{S}} \mathbf{D}$. Ainsi, $\nabla_{\tilde{\mathbf{A}}} \mathcal{J}_{vol} \left(\tilde{\mathbf{A}} \right)$ est Lipschitz-continu et la constante de Lipschitz est $\Psi_{\tilde{\mathbf{A}}} = (1 - \alpha) \left\| \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right\|_2$. De manière équivalente, $\nabla_{\tilde{\mathbf{S}}} \mathcal{J}_{vol} \left(\tilde{\mathbf{S}} \right)$ est Lipschitz-continu et la constante de Lipschitz est $\Psi_{\tilde{\mathbf{S}}} = \|\mathbf{M}\|_2 + 2\mu\|\mathbf{D}\|_2$.

Grâce à sa convexité, MDC offre des propriétés d'optimisation intéressantes. En revanche, MVC rend le sous-problème $\tilde{\mathbf{S}}$ non-convexe ; dans ce cas, l'utilisation du gradient optimal de Nesterov ne garantit pas la convergence vers un minimum local.

Le gradient optimal de Nesterov consiste à mettre à jour alternativement les paramètres $\tilde{\mathbf{S}}^{(k+1)}$ et $\tilde{\mathbf{A}}^{(k+1)}$ dans deux boucles internes distinctes, puis les estimations générées suite à cette séquence d'itérations sont de nouveau itérées dans une boucle externe commune. Prenons exemple sur $\tilde{\mathbf{A}}^{(k+1)}$; la méthode construit deux séquences *i.e.* $\tilde{\mathbf{A}}_i^{(k+1)}$ et $\tilde{\mathbf{Y}}_i^{(k+1)}$ et les met à jour alternativement à chaque itération i de la manière suivante :

$$\begin{cases} \tilde{\mathbf{A}}_{i+1}^{(k+1)} = \max \left(\mathbf{0}, \tilde{\mathbf{Y}}_i^{(k+1)} - \frac{1}{\Psi_{\tilde{\mathbf{A}}}} \nabla_{\tilde{\mathbf{A}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{Y}}_i^{(k+1)} \right) \right), \\ \tau_{i+1} = \frac{1 + \sqrt{4\tau_i^2 + 1}}{2}, \\ \tilde{\mathbf{Y}}_{i+1}^{(k+1)} = \tilde{\mathbf{A}}_{i+1}^{(k+1)} + \frac{\tau_i - 1}{\tau_{i+1}} \left(\tilde{\mathbf{A}}_{i+1}^{(k+1)} - \tilde{\mathbf{A}}_i^{(k+1)} \right). \end{cases}$$

La première étape consiste à descendre classiquement le long d'une direction de descente en adoptant un pas de $\frac{1}{\Psi_{\tilde{\mathbf{A}}}}$ et la deuxième étape a pour but d'accélérer l'algorithme par recherche linéaire avec un pas spécifique [Nesterov 83]. La mise à jour de $\tilde{\mathbf{S}}^{(k+1)}$ s'obtient de la même façon :

$$\begin{cases} \tilde{\mathbf{S}}_{i+1}^{(k+1)} = \max \left(\mathbf{0}, \tilde{\mathbf{Z}}_i^{(k+1)} - \frac{1}{\Psi_{\tilde{\mathbf{S}}}} \nabla_{\tilde{\mathbf{S}}^{(k+1)}} \mathcal{J}_{vol}^{(k+1)} \left(\tilde{\mathbf{Z}}_i^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) \right), \\ \tau_{i+1} = \frac{1 + \sqrt{4\tau_i^2 + 1}}{2}, \\ \tilde{\mathbf{Z}}_{i+1}^{(k+1)} = \tilde{\mathbf{S}}_{i+1}^{(k+1)} + \frac{\tau_i - 1}{\tau_{i+1}} \left(\tilde{\mathbf{S}}_{i+1}^{(k+1)} - \tilde{\mathbf{S}}_i^{(k+1)} \right). \end{cases}$$

Nous désignons ces nouvelles méthodes par OMDC-Ne (*On-line Minimum Dispersion Constraint-Nesterov*) et OMVC-Ne (*On-line Minimum Volume Constraint-Nesterov*).

2.4.4 Implantation des algorithmes et complexité de calcul

L'**Algorithme 1** et l'**Algorithme 2** qui font référence aux méthodes OMDC-MU/OMVC-MU et OMDC-ADMM respectivement, comprennent deux boucles principales : une boucle externe qui produit des estimations de l'ensemble des paramètres pour chaque nouvel échantillon et une boucle interne qui permet d'affiner itérativement les estimations en utilisant un nombre d'itérations fixe (N_{iter}). En ce qui concerne les méthodes OMDC-Ne et OMVC-Ne résumées en l'**Algorithme 3**, trois boucles principales sont utilisées : deux boucles internes qui produisent une estimation de $\tilde{\mathbf{A}}$ et $\tilde{\mathbf{S}}$ respectivement en utilisant un nombre d'itérations fixe $Iter_{in}$. Ces estimations sont affinées itérativement dans une autre boucle externe d'itérations $Iter_{ex}$. Cette procédure est répétée pour chaque nouvel échantillon $\tilde{\mathbf{X}}^{(k+1)}$.

Algorithme 1 OMDC-MU / OMVC-MU

Entrées : \mathbf{X} , R , α , μ , δ , ϵ , N_{iter} ;

Initialisation : $k = 0$, $\mathbf{N} = \text{zeros}(L, R)$, $\mathbf{M} = \text{zeros}(R, R)$, $\tilde{\mathbf{S}} = \text{rand}(L, R)$, $\tilde{\mathbf{A}} = \text{rand}(R, P)$, $\mathbf{A} = []$, $\mathbf{S} = []$;

Sorties : \mathbf{A} , \mathbf{S} ;

Tant que Nouvel échantillon $k + 1$ disponible **faire**

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

$$t = 1;$$

Tant que $t < N_{iter}$ **faire**

$$\tilde{\mathbf{A}} \leftarrow \tilde{\mathbf{A}} \odot \frac{\tilde{\mathbf{S}}^T \tilde{\mathbf{X}}}{\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \tilde{\mathbf{A}} + \delta \mathbf{1}};$$

$$\tilde{\mathbf{N}} = \alpha \mathbf{N} + (1 - \alpha) (\tilde{\mathbf{X}} \tilde{\mathbf{A}}^T);$$

$$\tilde{\mathbf{M}} = \alpha \mathbf{M} + (1 - \alpha) (\tilde{\mathbf{A}} \tilde{\mathbf{A}}^T);$$

$$\text{Si MDC : } \tilde{\mathbf{S}} \leftarrow \tilde{\mathbf{S}} \odot \frac{\tilde{\mathbf{N}} + 2\mu \tilde{\mathbf{S}} \frac{1}{R} \mathbf{1} \mathbf{1}^T}{\tilde{\mathbf{M}} + 2\mu \tilde{\mathbf{S}} \mathbf{1} + \delta \mathbf{1}};$$

$$\text{Si MVC : } \tilde{\mathbf{S}} \leftarrow \tilde{\mathbf{S}} \odot \frac{\tilde{\mathbf{N}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I})}{\tilde{\mathbf{M}} (\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I}) + 2\mu \tilde{\mathbf{S}} \mathbf{1} + \delta \mathbf{1}};$$

$$t \leftarrow t + 1;$$

Fin tant que

$$\mathbf{N} = \tilde{\mathbf{N}}; \mathbf{M} = \tilde{\mathbf{M}};$$

$$\mathbf{A} \leftarrow [\mathbf{A}, \tilde{\mathbf{A}}]; \mathbf{S} \leftarrow [\mathbf{S}, \tilde{\mathbf{S}}];$$

Fin tant que

Algorithme 2 OMDC-ADMM

Entrées : $\mathbf{X}, \mathbf{D}, R, \alpha, \mu, \rho, N_{iter}$;**Initialisation :** $k = 0, \mathbf{N} = \text{zeros}(L, R), \mathbf{M} = \text{zeros}(R, R), \tilde{\mathbf{S}} = \text{rand}(L, R), \tilde{\mathbf{V}} = \text{zeros}(R, P), \tilde{\mathbf{U}} = \text{zeros}(L, R), \tilde{\mathbf{\Pi}} = \text{zeros}(R, P), \tilde{\mathbf{\Lambda}} = \text{zeros}(L, R), \mathbf{A} = [], \mathbf{S} = []$;**Sorties :** \mathbf{A}, \mathbf{S} ;**Tant que** Nouvel échantillon $k + 1$ disponible **faire**

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

$$t = 1;$$

Tant que $t < N_{iter}$ **faire**

$$\tilde{\mathbf{A}} = \left((1 - \alpha) \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \rho \mathbf{I} \right)^{-1} \left((1 - \alpha) \tilde{\mathbf{S}}^T \tilde{\mathbf{X}} + \rho (\tilde{\mathbf{V}} - \tilde{\mathbf{\Pi}}) \right);$$

$$\tilde{\mathbf{V}} = \max(\mathbf{0}, \tilde{\mathbf{A}} + \tilde{\mathbf{\Pi}});$$

$$\tilde{\mathbf{\Pi}} \leftarrow \tilde{\mathbf{\Pi}} + \tilde{\mathbf{A}} - \tilde{\mathbf{V}};$$

$$\tilde{\mathbf{N}} = \alpha \mathbf{N} + (1 - \alpha) (\tilde{\mathbf{X}} \tilde{\mathbf{A}}^T);$$

$$\tilde{\mathbf{M}} = \alpha \mathbf{M} + (1 - \alpha) (\tilde{\mathbf{A}} \tilde{\mathbf{A}}^T);$$

$$\tilde{\mathbf{S}} = \left(\tilde{\mathbf{N}} + \rho (\tilde{\mathbf{U}} - \tilde{\mathbf{\Lambda}}) \right) (\tilde{\mathbf{M}} + \rho \mathbf{I} + 2\mu \mathbf{D})^{-1};$$

$$\tilde{\mathbf{U}} = \max(\mathbf{0}, \tilde{\mathbf{S}} + \tilde{\mathbf{\Lambda}});$$

$$\tilde{\mathbf{\Lambda}} \leftarrow \tilde{\mathbf{\Lambda}} + \tilde{\mathbf{S}} - \tilde{\mathbf{U}};$$

$$t \leftarrow t + 1;$$

Fin tant que

$$\mathbf{N} = \tilde{\mathbf{N}}; \mathbf{M} = \tilde{\mathbf{M}};$$

$$\mathbf{A} \leftarrow [\mathbf{A}, \tilde{\mathbf{A}}]; \mathbf{S} \leftarrow [\mathbf{S}, \tilde{\mathbf{S}}];$$

Fin tant que

Algorithme 3 OMDC-Ne / OMVC-Ne

Entrées : $\mathbf{X}, \mathbf{D}, R, \alpha, \mu, \epsilon, Iter_{ex}, Iter_{in}$;

Initialisation : $k = 0, \mathbf{N} = \text{zeros}(L, R), \mathbf{M} = \text{zeros}(R, R), \tilde{\mathbf{S}} = \text{rand}(L, R), \tilde{\mathbf{A}} = \text{rand}(R, P),$
 $\mathbf{A} = [], \mathbf{S} = []$;

Sorties : \mathbf{A}, \mathbf{S} ;

Tant que Nouvel échantillon $k + 1$ disponible **faire**

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

$$t = 1;$$

Tant que $t < Iter_{ex}$ **faire**

$$\tau_0 = 1;$$

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{A}};$$

$$\Psi = (1 - \alpha) \left\| \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \right\|_2;$$

$$i = 1;$$

Tant que $i < Iter_{in}$ **faire**

$$\tilde{\mathbf{A}}_0 = \tilde{\mathbf{A}};$$

$$\tilde{\mathbf{A}} = \max \left(\mathbf{0}, \tilde{\mathbf{Y}} - \frac{1}{\Psi} (1 - \alpha) \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \tilde{\mathbf{Y}} - \tilde{\mathbf{S}}^T \tilde{\mathbf{X}} \right) \right);$$

$$\tau_1 = \frac{1 + \sqrt{4\tau_0^2 + 1}}{2};$$

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{A}} + \frac{\tau_0 - 1}{\tau_1} \left(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0 \right);$$

$$\tau_0 = \tau_1;$$

$$i \leftarrow i + 1;$$

Fin tant que

$$\tau_0 = 1;$$

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{S}};$$

$$\tilde{\mathbf{N}} = \alpha \mathbf{N} + (1 - \alpha) \left(\tilde{\mathbf{X}} \tilde{\mathbf{A}}^T \right);$$

$$\tilde{\mathbf{M}} = \alpha \mathbf{M} + (1 - \alpha) \left(\tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \right);$$

$$\text{Si MDC : } \Psi = \left\| \tilde{\mathbf{M}} \right\|_2 + 2\mu \left\| \mathbf{D} \right\|_2;$$

$$\text{Si MVC : } \Psi = \left\| \tilde{\mathbf{M}} \right\|_2 + 2\mu \left\| \left(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \epsilon \mathbf{I} \right)^{-1} \right\|_2;$$

$$i = 1;$$

Tant que $i < Iter_{in}$ **faire**

$$\tilde{\mathbf{S}}_0 = \tilde{\mathbf{S}};$$

$$\text{Si MDC : } \tilde{\mathbf{S}} = \max \left(\mathbf{0}, \tilde{\mathbf{Z}} - \frac{1}{\Psi} \left(\tilde{\mathbf{Z}} \tilde{\mathbf{M}} - \tilde{\mathbf{N}} + 2\mu \tilde{\mathbf{Z}} \mathbf{D} \right) \right);$$

$$\text{Si MVC : } \tilde{\mathbf{S}} = \max \left(\mathbf{0}, \tilde{\mathbf{Z}} - \frac{1}{\Psi} \left(\tilde{\mathbf{Z}} \tilde{\mathbf{M}} - \tilde{\mathbf{N}} + 2\mu \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}} + \epsilon \mathbf{I} \right)^{-1} \right) \right);$$

$$\tau_1 = \frac{1 + \sqrt{4\tau_0^2 + 1}}{2};$$

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{S}} + \frac{\tau_0 - 1}{\tau_1} \left(\tilde{\mathbf{S}} - \tilde{\mathbf{S}}_0 \right);$$

$$\tau_0 = \tau_1$$

$$i \leftarrow i + 1;$$

Fin tant que

$$t \leftarrow t + 1;$$

Fin tant que

$$\mathbf{N} = \tilde{\mathbf{N}}; \mathbf{M} = \tilde{\mathbf{M}};$$

$$\mathbf{A} \leftarrow \left[\mathbf{A}, \tilde{\mathbf{A}} \right]; \mathbf{S} \leftarrow \left[\mathbf{S}, \tilde{\mathbf{S}} \right];$$

Fin tant que

Nous évaluons la complexité de calcul des algorithmes proposés ; cela est réalisé en considérant uniquement les opérations matricielles de multiplication, qui dominent la complexité des algorithmes. En considérant une tranche de k l'image hyperspectrale de dimension $L \times P$, N_{iter} itérations (et $Iter_{ex}$, $Iter_{in}$ pour OMDC-Ne/OMVC-Ne) et un rang de décomposition R , les complexités de calcul approximatives pour les mises à jour de $\tilde{\mathbf{A}}$ et $\tilde{\mathbf{S}}$ sont regroupées dans le Tableau 2.1. Pour un nombre d'itérations équivalent, nous constatons que les complexités sont comparables ; nous montrerons dans la suite, à travers des simulations, que OMDC-ADMM requière moins d'itérations pour converger par rapport aux autres méthodes, et réduit ainsi significativement le coût de calcul.

	$\tilde{\mathbf{A}}$	$\tilde{\mathbf{S}}$
OMDC-MU	$(RPL + RP + (L + P)R^2)N_{iter}$	$(RPL + RL + (2L + P)R^2)N_{iter}$
OMVC-MU	$(RPL + RP + (L + P)R^2)N_{iter}$	$(RPL + RL + (5L + P)R^2)N_{iter}$
OMDC-ADMM	$(RPL + (L + P)R^2 + R^3)N_{iter}$	$(RPL + (L + P)R^2 + R^3)N_{iter}$
OMDC-Ne	$(RPL + LR^2)Iter_{ex} + (PR^2)Iter_{ex}Iter_{in}$	$(RPL + PR^2)Iter_{ex} + (2LR^2)Iter_{ex}Iter_{in}$
OMVC-Ne	$(RPL + LR^2)Iter_{ex} + (PR^2)Iter_{ex}Iter_{in}$	$(RPL + PR^2)Iter_{ex} + (3LR^2 + R^3)Iter_{ex}Iter_{in}$

TABLE 2.1 : Complexité de calcul des algorithmes en-ligne par échantillon k

2.4.5 Analyse de convergence

En raison de la non-convexité du problème NMF classique, il est généralement hors de portée d'atteindre l'optimum global du problème d'optimisation

$$\underset{\mathbf{S} \geq 0, \mathbf{A} \geq 0}{\text{minimiser}} \|\mathbf{X} - \mathbf{SA}\|_F^2. \quad (2.20)$$

Par conséquent, dans la littérature [Grippo 00, Lin 07a, Xu 12, Guan 12a, Huang 16], la convergence vers un point stationnaire de (2.20) a été étudiée ; la stationnarité est importante car c'est une condition nécessaire d'optimalité. La NMF est considérée comme un cas simple de la méthode de *minimisation par blocs* où seulement deux blocs de variables \mathbf{S} et \mathbf{A} sont mis à jour [Bertsekas 97]. L'analyse de convergence pour les méthodes de minimisation par blocs exige que les sous-problèmes aient des solutions uniques [Powell 73, Bertsekas 97], mais cette propriété n'est pas valable dans le cas de la NMF classique : les sous-problèmes en \mathbf{S} et \mathbf{A} sont convexes, mais ils ne le sont pas strictement, puisque c'est un problème bilinéaire. Par conséquent, ces sous-problèmes peuvent avoir une multitude de solutions optimales. Fort heureusement, dans le cas de deux blocs, les travaux de [Grippo 00] ont montré que cette condition d'unicité n'était pas nécessaire et que tout point limite généré par une séquence d'itérations est un point stationnaire du problème (2.20). Basé sur ces résultats, de nombreux travaux ont été menés pour prouver la convergence vers un point stationnaire des algorithmes hors-ligne fondés sur les règles de mises à jour multiplicative, Nesterov et ADMM [Lin 07a, Guan 12a, Huang 16].

Dans le cas du modèle NMF en-ligne, les principaux résultats sur la convergence vers un point stationnaire concernent les algorithmes qui supposent que les *endmembers* sont stationnaires. Cette hypothèse, bien que forte, permet de fournir des garanties théoriques sur la convergence en utilisant la théorie des processus empiriques, comme cela a été fait dans [Mairal 10, Guan 12b, Zhao 16], ce qui constitue l'état de l'art dans la littérature. Cependant, dans notre cas, nous supposons que les *endmembers*

peuvent légèrement varier entre deux échantillons successifs et dans cette situation, fournir des garanties théoriques sur la convergence est difficile. En revanche, on peut démontrer que nos algorithmes en-ligne convergent à chaque tranche vers un point stationnaire selon [Grippio 00]. Notons que pour OMVC-MU et OMVC-Ne, aucune garantie de convergence ne peut être établie car l'utilisation de la contrainte de volume minimal rend le sous-problème d'estimation des *endmembers* non-convexe. Cependant, nous montrerons à travers des résultats expérimentaux que OMVC-MU et OMVC-Ne sont efficaces et performants pour résoudre le problème (2.3).

Nous détaillons ci-après un résultat partiel sur la convergence de l'algorithme OMDC-ADMM essentiellement, en montrant que tout point généré par une séquence d'itérations est un point stationnaire qui satisfait les conditions de KKT [Boyd 04]. Pour cette analyse, nous considérons le comportement de l'algorithme pour l'échantillon $k + 1$ de l'image. Afin de simplifier les notations, nous regroupons toutes les variables à estimer dans une matrice $\mathbf{W} := (\tilde{\mathbf{A}}, \tilde{\mathbf{S}}, \tilde{\mathbf{V}}, \tilde{\mathbf{U}}, \tilde{\mathbf{\Pi}}, \tilde{\mathbf{\Lambda}})$. En se référant à [Boyd 04, Xu 12] et après quelques manipulations algébriques, nous pouvons facilement montrer que le point \mathbf{W} est un point KKT du problème (2.15) si :

$$(1 - \alpha)\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \tilde{\mathbf{A}} - (1 - \alpha)\tilde{\mathbf{S}}^T \tilde{\mathbf{X}} + \tilde{\mathbf{\Pi}} = \mathbf{0}, \quad (2.21a)$$

$$\tilde{\mathbf{S}} \tilde{\mathbf{M}} - \tilde{\mathbf{N}} + 2\mu \tilde{\mathbf{S}} \tilde{\mathbf{D}} + \tilde{\mathbf{\Lambda}} = \mathbf{0}, \quad (2.21b)$$

$$\tilde{\mathbf{A}} - \tilde{\mathbf{V}} = \mathbf{0}, \quad (2.21c)$$

$$\tilde{\mathbf{S}} - \tilde{\mathbf{U}} = \mathbf{0}, \quad (2.21d)$$

$$\tilde{\mathbf{\Pi}} \leq \mathbf{0} \leq \tilde{\mathbf{V}}, \tilde{\mathbf{\Pi}} \odot \tilde{\mathbf{V}} = \mathbf{0}, \quad (2.21e)$$

$$\tilde{\mathbf{\Lambda}} \leq \mathbf{0} \leq \tilde{\mathbf{U}}, \tilde{\mathbf{\Lambda}} \odot \tilde{\mathbf{U}} = \mathbf{0}. \quad (2.21f)$$

Soit $\mathbf{W}_t := (\tilde{\mathbf{A}}_t, \tilde{\mathbf{S}}_t, \tilde{\mathbf{V}}_t, \tilde{\mathbf{U}}_t, \tilde{\mathbf{\Pi}}_t, \tilde{\mathbf{\Lambda}}_t)$, l'estimation de \mathbf{W} à l'itération t dans l'**Algorithme 2**. D'après la mise à jour (2.17a), la relation suivante peut s'écrire entre $\tilde{\mathbf{A}}_t$ et $\tilde{\mathbf{A}}_{t+1}$:

$$\begin{aligned} & \left((1 - \alpha)\tilde{\mathbf{S}}_t^T \tilde{\mathbf{S}}_t + \rho \mathbf{I} \right) \left(\tilde{\mathbf{A}}_{t+1} - \tilde{\mathbf{A}}_t \right) = \\ & - \left((1 - \alpha)\tilde{\mathbf{S}}_t^T \tilde{\mathbf{S}}_t \tilde{\mathbf{A}}_t - (1 - \alpha)\tilde{\mathbf{S}}_t^T \tilde{\mathbf{X}} + \rho \left(\tilde{\mathbf{A}}_t - \tilde{\mathbf{V}}_t \right) + \rho \tilde{\mathbf{\Pi}}_t \right). \end{aligned} \quad (2.22)$$

Supposons que l'algorithme atteint un point stationnaire *i.e.* $\mathbf{W}_{t+1} = \mathbf{W}_t = \mathbf{W}^* \Rightarrow \tilde{\mathbf{A}}_{t+1} = \tilde{\mathbf{A}}_t = \tilde{\mathbf{V}}_t = \tilde{\mathbf{A}}^*$. En remplaçant dans l'expression (2.22), nous obtenons :

$$(1 - \alpha)\tilde{\mathbf{S}}^{*T} \tilde{\mathbf{S}}^* \tilde{\mathbf{A}}^* - (1 - \alpha)\tilde{\mathbf{S}}^{*T} \tilde{\mathbf{X}} + \rho \tilde{\mathbf{\Pi}}^* = \mathbf{0}.$$

En procédant de la même manière pour les autres paramètres de \mathbf{W} , nous montrons que les quatre premières équations relatives aux conditions KKT (2.21) sont satisfaites pour chaque point limite : $\mathbf{W}^* := (\tilde{\mathbf{A}}^*, \tilde{\mathbf{S}}^*, \tilde{\mathbf{V}}^*, \tilde{\mathbf{U}}^*, \tilde{\mathbf{\Pi}}^*, \tilde{\mathbf{\Lambda}}^*)$. Afin de prouver la condition (2.21e), nous pouvons écrire : $\max(\mathbf{0}, \tilde{\mathbf{A}}^* + \tilde{\mathbf{\Pi}}^*) = \tilde{\mathbf{V}}^*$. Si $\tilde{\mathbf{A}}^* = \tilde{\mathbf{V}}^* = \mathbf{0}$, alors $\max(\mathbf{0}, \tilde{\mathbf{\Pi}}^*) = \mathbf{0}$, ce qui mène à $\tilde{\mathbf{\Pi}}^* < \mathbf{0}$. Si $\tilde{\mathbf{A}}^* = \tilde{\mathbf{V}}^* > \mathbf{0}$, alors $\tilde{\mathbf{\Pi}}^* = \mathbf{0}$. Le même raisonnement s'applique à la condition (2.21f).

L'hypothèse de travail pour établir ce résultat est $\tilde{\mathbf{S}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(k)}$, ce qui signifie que le résultat de la convergence proposé n'est valide que lorsque le régime permanent est atteint. Il ne fournit donc aucune indication sur le comportement transitoire. Ce point sera abordé par des simulations numériques dans les sections 2.5.2 et 2.5.3.

2.5 Résultats expérimentaux pour le démélange hyperspectral en-ligne

Dans cette section, nous avons mené plusieurs expériences sur des images hyperspectrales simulées et réelles. Toutes ces expériences ont été réalisées à l'aide du logiciel Matlab (R2016a) sur un Macbook Pro avec processeur quadricœur à 2.7 GHz et 16 Go de RAM. Ces expériences ont pour objectifs les points suivants :

1. Illustrer la préservation du rang lors de l'utilisation des contraintes de dispersion et de volume minimal.
2. Comparer l'efficacité des différents algorithmes en-ligne que nous proposons en ce qui concerne la vitesse de convergence, le temps de traitement et la qualité d'estimation.
3. Comparer le démélange hyperspectral en-ligne vs. hors-ligne, pour étudier les avantages/inconvénients.
4. Évaluer l'influence des paramètres sur la vitesse de convergence et les performances de démélange. La capacité de suivi des algorithmes en-ligne sera également analysée.
5. Valider les performances du démélange hyperspectral en-ligne sur des données réelles, avec vérité terrain.

2.5.1 Préservation du rang lors de l'utilisation de MDC et MVC

Nous avons réalisé une expérience pour illustrer les propriétés de préservation du rang des deux contraintes de volume, MVC (2.6) et MDC (2.7). Ce point est abordé numériquement à l'aide d'une image hyperspectrale simulée, où les cartes d'abondance sont présentées en Figure 2.6. Cette dernière a pour dimension $119 \times 36 \times 36$, consistant en un mélange non-négatif de trois *endmembers*, ne variant pas dans le temps. 119 correspond au nombre de longueurs d'onde et 36×36 aux dimensions spatiales. Les trois cartes d'abondance sont des matrices de taille 36×36 , générées aléatoirement à partir d'une distribution uniforme et continue sur l'intervalle $[0, 1]$.

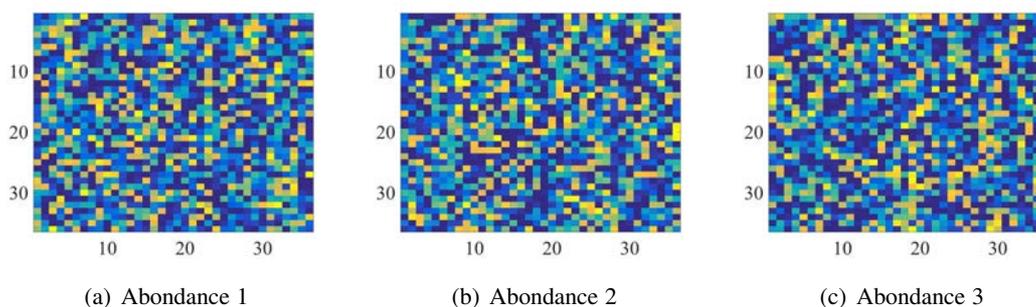
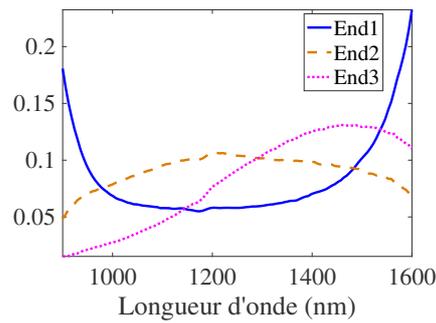
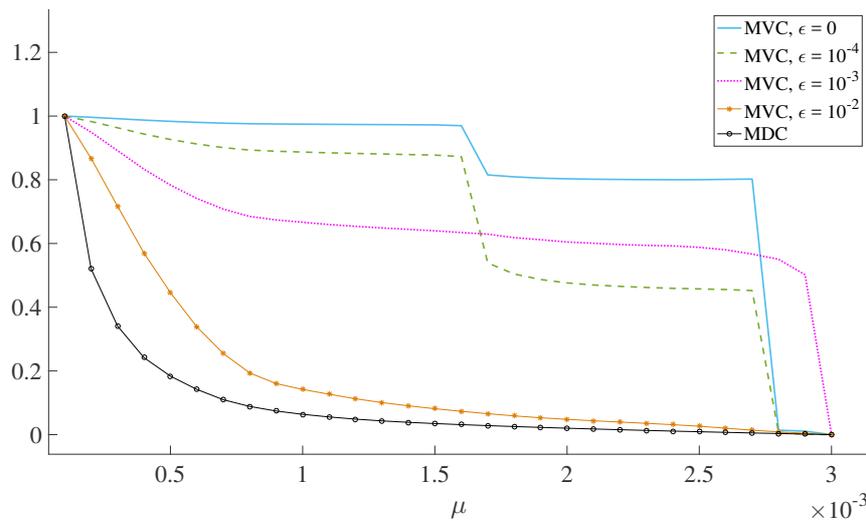


FIGURE 2.6 : Cartes d'abondance simulées

FIGURE 2.7 : Signatures spectrales des trois *endmembers* (End)

Pour un intervalle de μ compris entre 0.0001 et 0.003, les quantités $\text{trace}(\tilde{\mathbf{S}}\mathbf{D}\tilde{\mathbf{S}}^T)$ et $\log(\det(\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} + \epsilon\mathbf{I}))$ ont été calculées, où $\tilde{\mathbf{S}}$ est la matrice des *endmembers* estimée par l'un des algorithmes proposés, à la dernière tranche de l'image hyperspectrale. Pour MVC, 4 différentes valeurs de ϵ ont été choisies : $\epsilon = 0, 10^{-4}, 10^{-3}$ et 10^{-2} . Les courbes normalisées (valeur maximale à 1 et valeur minimale à 0) sont représentées sur la Figure 2.8.

FIGURE 2.8 : Valeurs de MDC et MVC en fonction de μ , pour le mélange de la Figure 2.6

Sur la Figure 2.8, plusieurs comportements peuvent être observés. Pour MVC, avec $\epsilon = 0$, la courbe présente une forme d'escalier et, à mesure que la valeur de ϵ augmente, les paliers tendent à disparaître. En réalité, MVC avec des valeurs faibles de ϵ favorise la perte de rang à mesure que μ augmente ; cela se traduit au niveau des courbes par des variations abruptes, en « escalier ». Lorsque ϵ augmente, la perte de rang n'est plus promue. Sur la Figure 2.9, nous pouvons comparer les simplexes réels \mathcal{D} et estimés \mathcal{S} pour différentes valeurs de μ , pour MVC ($\epsilon = 0, 10^{-3}$ et 10^{-2}) et MDC. Pour $\epsilon = 0$, lorsque μ augmente, le volume du simplexe estimé tend vers zéro, ce qui signifie que les *endmembers* deviennent colinéaires. En revanche, lorsque ϵ augmente, le rang de la décomposition est conservé sur un intervalle de μ plus grand. Par exemple, pour $\epsilon = 10^{-2}$ et $\mu = 0.03$, bien que le simplexe estimé soit inclus dans

le vrai simplexe, le rang est préservé, ce qui évite les instabilités numériques ; cela montre l'intérêt du paramètre ϵ dans le contexte du démélange hyperspectral en-ligne. De même, pour MDC, le simplexe diminue progressivement, tout en préservant le rang de la décomposition.

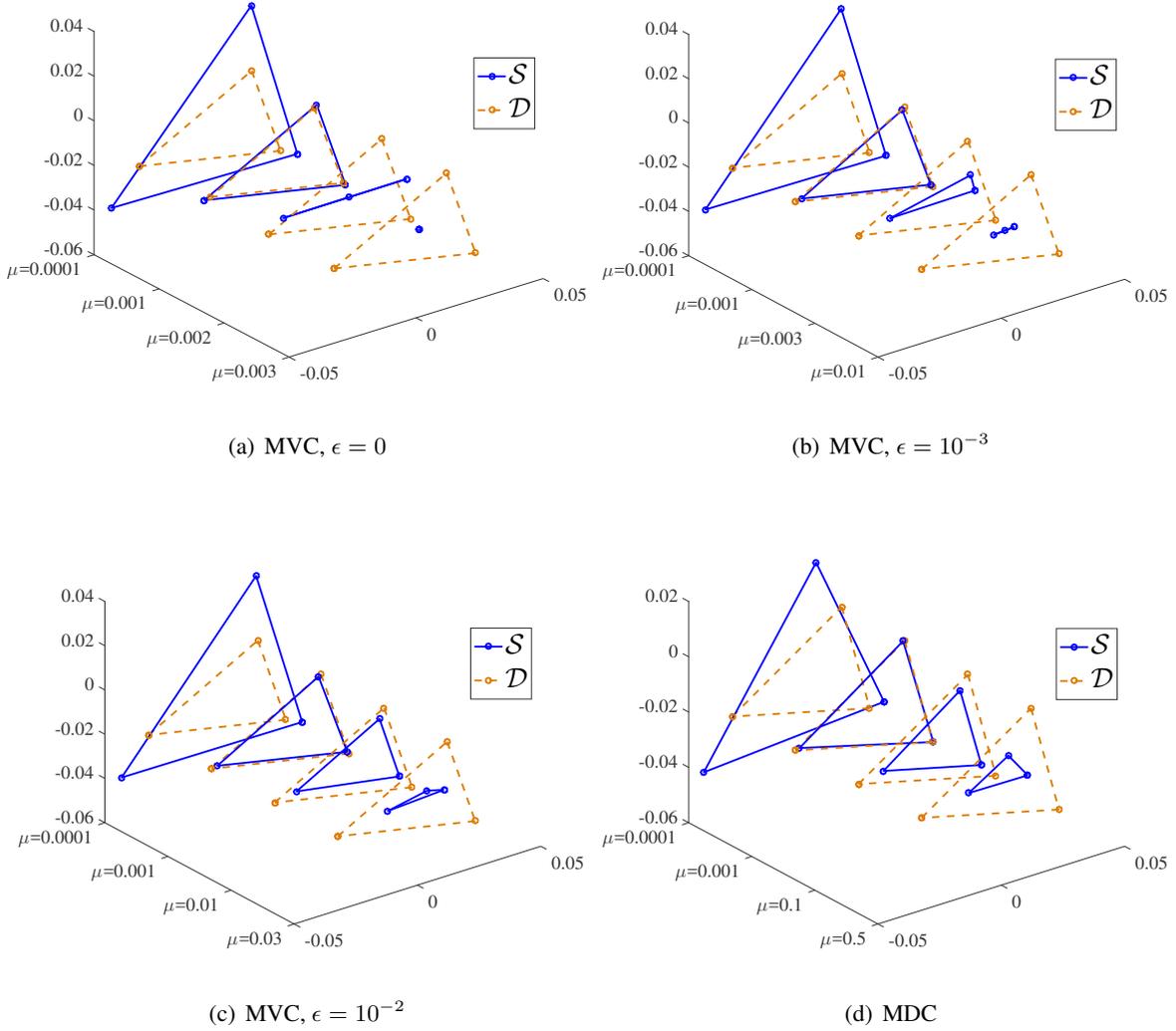


FIGURE 2.9 : Simplexes estimés en fonction de μ pour MVC et MDC, pour le mélange de la Figure 2.6

2.5.2 Vitesse de convergence

L'expérience décrite ci-dessous vise à étudier la vitesse de convergence des différents algorithmes en-ligne au sens de deux critères de performance : *Spectral Angle Distance* (SAD) et *Root Mean Square Error* (RMSE).

SAD mesure la similarité entre les vrais *endmembers* $\tilde{\mathbf{S}}$ et les estimés $\hat{\mathbf{S}}$. Pour une tranche donnée k de l'image, SAD s'exprime de la manière suivante :

$$\frac{1}{R} \sum_{r=1}^R \cos^{-1} \left(\frac{\tilde{\mathbf{s}}_r^T \hat{\mathbf{s}}_r}{\|\tilde{\mathbf{s}}_r\|_2 \|\hat{\mathbf{s}}_r\|_2} \right), \quad (2.23)$$

où R est le nombre de *endmembers*, $\tilde{\mathbf{S}} = [\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_R]$ et $\hat{\tilde{\mathbf{S}}} = [\hat{\tilde{\mathbf{s}}}_1, \dots, \hat{\tilde{\mathbf{s}}}_R]$. Plus le SAD est faible, meilleure est l'estimation.

RMSE mesure l'erreur entre les abondances originales \mathbf{A} et les estimées $\hat{\mathbf{A}}$ et s'écrit :

$$\frac{1}{R} \sum_{r=1}^R \sqrt{\frac{1}{KP} \sum_{d=1}^{KP} (a_{rd} - \hat{a}_{rd})^2}, \quad (2.24)$$

où a_{rd} et \hat{a}_{rd} sont la vérité terrain et les abondances estimées respectivement pour le $r^{\text{ème}}$ *endmember* du pixel d ($d = 1, \dots, KP$).

Précisons que les expressions (2.23) et (2.24) sont les moyennes du SAD et du RMSE obtenus pour l'ensemble des *endmembers*. Pour certaines expériences, les valeurs du SAD et du RMSE sont calculées indépendamment pour chaque *endmember* $r = 1, \dots, R$.

Nous avons simulé une image hyperspectrale de dimension $119 \times 40 \times 40$ dont les cartes d'abondance sont représentées sur la Figure 2.10 ; ces dernières sont binaires. Cette image est composée de trois *endmembers* qui ne varient pas dans le temps, et identiques à ceux illustrés sur la Figure 2.7.

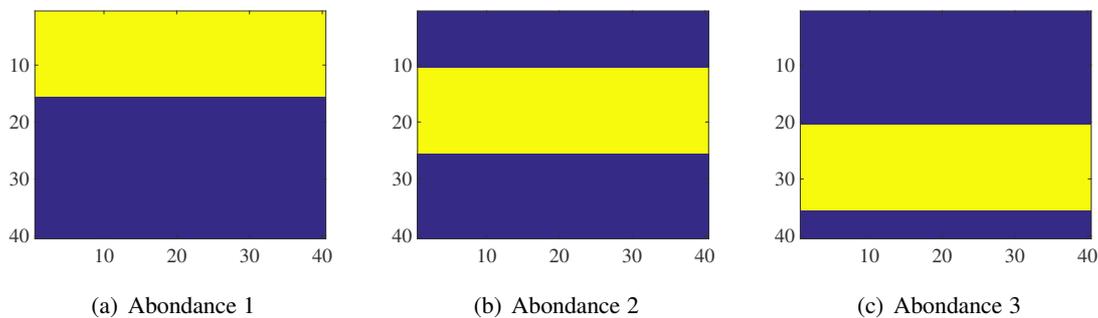


FIGURE 2.10 : Cartes d'abondance simulées

Un faible bruit a été ajouté aux données (SNR = 45 dB). Chaque méthode a traité l'image séquentiellement par tranche de dimension 119×40 . Pour l'ensemble des algorithmes, les paramètres ont été fixés de la manière suivante : $R = 3$, $\alpha = 0.99$ et $\rho = 0.001$ pour OMDC-ADMM. En ce qui concerne l'hyperparamètre μ relatif à la contrainte de dispersion minimale, il a été fixé à 0.001 selon l'approche proposée dans la section 2.6. Afin d'effectuer des comparaisons justes entre tous les algorithmes, μ a également été fixé à 0.001 pour la contrainte de volume minimal (2.6). Le paramètre ϵ a été choisi de manière empirique à 0.4, pour à la fois préserver le rang de la décomposition mais également pour converger vers les mêmes solutions que les algorithmes utilisant MDC. Pour comparer les vitesses de convergence des algorithmes, les valeurs du SAD (estimées à la dernière tranche ; cela est cohérent dans le cas de sources stationnaires) et du RMSE pour différentes valeurs de N_{iter} comprises entre 10 et 5000 ont été calculées. Notons que, pour OMDC-Ne et OMVC-Ne, $N_{iter} = I_{iter_{in}} I_{iter_{ex}}$. Les résultats ont été obtenus en effectuant une moyenne sur 20 essais ; ils sont présentés sur la Figure 2.11.

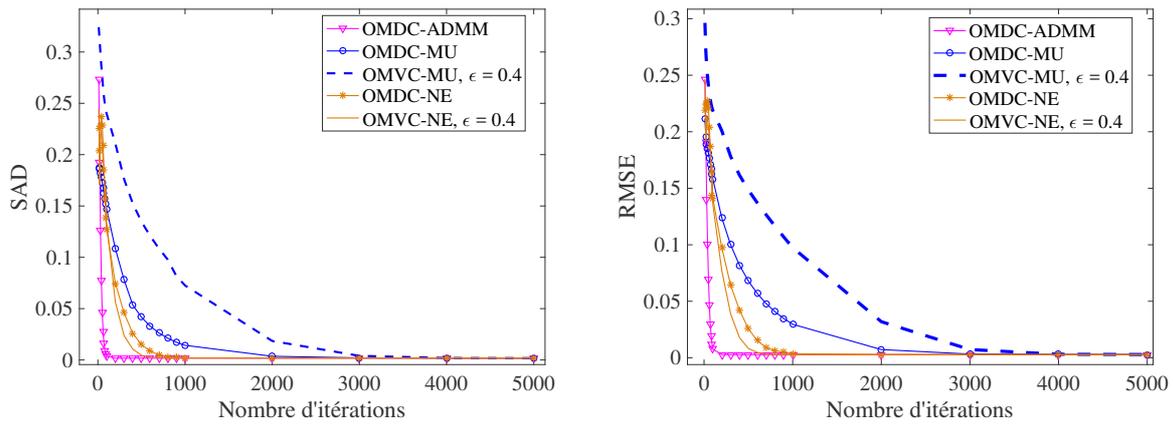


FIGURE 2.11 : Vitesses de convergence du SAD et du RMSE en fonction de N_{iter} , pour le mélange de la Figure 2.10

D'après la Figure 2.11, nous observons que pour toutes les valeurs considérées de N_{iter} , OMDC-ADMM présente la vitesse de convergence la plus rapide. De plus, nous constatons que OMVC-Ne converge plus rapidement que OMDC-Ne. En revanche, OMDC-MU et OMVC-MU présentent les vitesses de convergence les plus faibles ; cela est principalement dû à la sensibilité des règles de mise à jour multiplicatives face aux initialisations aléatoires. Le Tableau 2.2 regroupe les valeurs de N_{iter} pour lesquelles les algorithmes atteignent approximativement le même SAD et RMSE asymptotique, ainsi que le temps de traitement associé.

	SAD	RMSE	N_{iter}	Temps (s)
OMDC-ADMM	0.0018	0.0028	200	0.67
OMDC-Ne	0.0018	0.0028	2000	0.94
OMDC-MU	0.0018	0.0029	4000	6.24
OMVC-MU	0.0018	0.0029	5000	8.06
OMVC-Ne	0.0018	0.0028	800	0.88

TABLE 2.2 : Valeurs asymptotiques du SAD et du RMSE liées au temps et aux nombres d'itérations, pour le mélange de la Figure 2.10

Bien que OMDC-MU et OMVC-MU fournissent de bons résultats, leur temps de traitement est relativement grand (notons que l'image hyperspectrale utilisée est de petite dimension), ce qui les rend inadaptés pour un système de traitement en-ligne. OMDC-ADMM présente le temps de calcul le plus faible (≈ 0.67 s) pour 200 itérations, tandis que OMDC-Ne, par exemple, nécessite 0.94 s (pour $N_{iter} = 2000$) pour traiter l'image hyperspectrale dans sa totalité. Pour 800 itérations, OMVC-Ne a un temps de traitement proche de celui de OMDC-Ne alors que le nombre d'itérations est bien plus faible ; en effet, l'utilisation de la contrainte de volume minimal augmente la complexité de calcul (se référer au Tableau 2.1).

Un des points clés des algorithmes en-ligne pour le démélange hyperspectral est la capacité de traiter des images hyperspectrales de grande dimension, tout en conservant une complexité de calcul constante par échantillon. À cet égard, nous avons comparé les performances de l'algorithme OMDC-ADMM, qui s'est révélé le plus rapide de tous les algorithmes proposés, avec son homologue *batch* que l'on

nomme BMDC-ADMM pour *Batch MDC-ADMM*. Nous avons utilisé la même image hyperspectrale simulée que précédemment (corrompue par un faible niveau de bruit). Les paramètres relatifs à BMDC-ADMM ont été fixés de la manière suivante : $\rho = 0.001$ et $\mu = 0.2$. BMDC-ADMM a été appliqué à la version dépliée de l'image hyperspectrale de dimension 119×1600 , tandis que OMDC-ADMM a traité l'image tranche par tranche. Pour étudier les vitesses de convergence, le SAD a été évalué pour différentes valeurs de N_{iter} allant de 10 à 5000. Précisons que pour une valeur de N_{iter} donnée, le SAD pour OMDC-ADMM a été calculé pour chaque tranche de l'image puis moyenné, afin d'effectuer une comparaison juste entre les deux algorithmes. Les résultats obtenus en effectuant une moyenne sur 20 essais sont illustrés sur la Figure 2.12. Nous pouvons constater que l'algorithme en-ligne converge beaucoup plus rapidement que son équivalent hors-ligne (notons que l'on observe un comportement similaire pour les autres algorithmes, non présentés ici). Par exemple, 200 itérations suffisent à OMDC-ADMM pour produire des estimations précises, alors que BMDC-ADMM nécessite au moins 2000 itérations. Cela a de lourdes conséquences sur la charge de calcul ; en effet, si l'on se réfère au Tableau 2.1 et que l'on considère $R = 3$, $P = 40$, $L = 119$, $K = 40$, $N_{iter} = 200$, alors la complexité de calcul pour OMDC-ADMM est d'approximativement 2.5180×10^8 opérations. Pour BMDC-ADMM, la complexité de calcul pour toute l'image est de l'ordre de $2N_{iter}(RKPL + (L + KP)R^2 + R^3)$; ainsi, pour $N_{iter} = 2000$, elle est d'approximativement 2.3468×10^9 opérations.

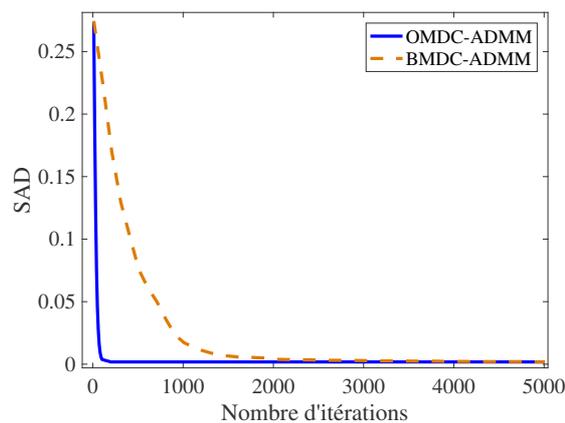


FIGURE 2.12 : OMDC-ADMM vs. BMDC-ADMM : vitesses de convergence du SAD en fonction de N_{iter} , pour le mélange de la Figure 2.10

2.5.3 Influence des paramètres

Pour étudier l'influence des paramètres α , ρ et ϵ , des simulations numériques ont été menées sur l'image hyperspectrale dont les cartes d'abondance sont présentées sur la Figure 2.10. Le paramétrage, de la section précédente, a été effectué de manière à obtenir le meilleur compromis entre le temps de traitement et la qualité des estimations générés par les algorithmes. Étant donné que l'ensemble des paramètres sont couplés, la modification de l'un d'entre eux entraîne alors la modification de tous les autres. C'est pourquoi, nous avons étudié l'influence des paramètres un à la fois, tout en gardant les autres fixes. La Figure 2.13 montre l'évolution du SAD pour différentes valeurs de $\alpha \in [0.1 \ 0.99]$ lorsque $\mu = 0.001$. Nous constatons que la valeur optimale de μ (estimée par la méthode de la section 2.6) est adaptée pour des valeurs de α de 0.8 à 0.99. En revanche, à mesure que α diminue, les valeurs du SAD se dégradent

fortement, indiquant que le problème n'est plus suffisamment régularisé. Les paramètres α et μ sont donc fortement liés. Les vitesses de convergence pour le SAD et en fonction de différentes valeurs de N_{iter} sont représentées sur la Figure 2.14 pour $\alpha = 0.8, 0.9, 0.95$ et 0.99 . À mesure que α augmente, l'algorithme requiert moins d'itérations pour converger vers une solution pertinente. La Figure 2.15, quant à elle, correspond aux valeurs du SAD calculées à chaque tranche de l'image pour $\alpha = 0.8, 0.9, 0.95, 0.97, 0.99$, $N_{iter} = 1000$ (pour que le régime permanent soit atteint dans tous les cas) et $\mu = 0.001$. Au fur et à mesure que α diminue, moins d'échantillons sont nécessaires pour converger vers la bonne solution. Conjointement, le SAD asymptotique diminue ; cela peut s'expliquer par le surapprentissage de chaque tranche introduit par des faibles valeurs de α . Ainsi, α doit être déterminé comme étant le meilleur compromis entre le nombre d'itérations (qui est un point essentiel dans le démélange en-ligne puisqu'il influe le temps de traitement) et la qualité des estimations.

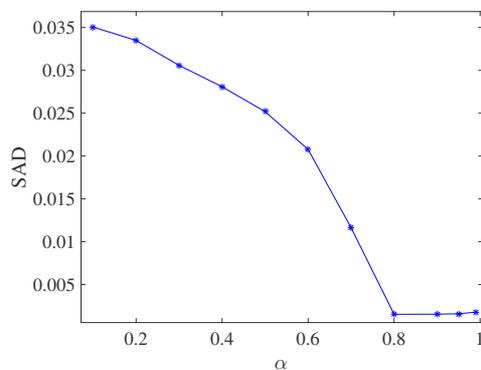


FIGURE 2.13 : Évolution du SAD en fonction de α , pour le mélange de la Figure 2.10

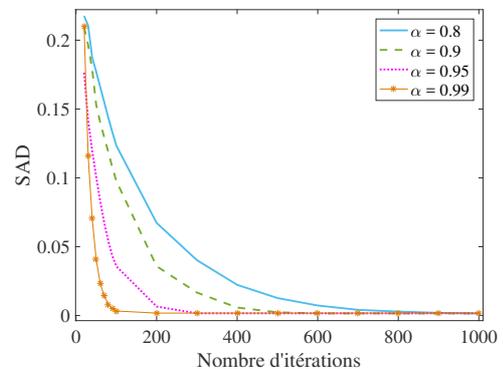


FIGURE 2.14 : Vitesse de convergence du SAD en fonction de N_{iter} , pour différentes valeurs de α , pour le mélange de la Figure 2.10

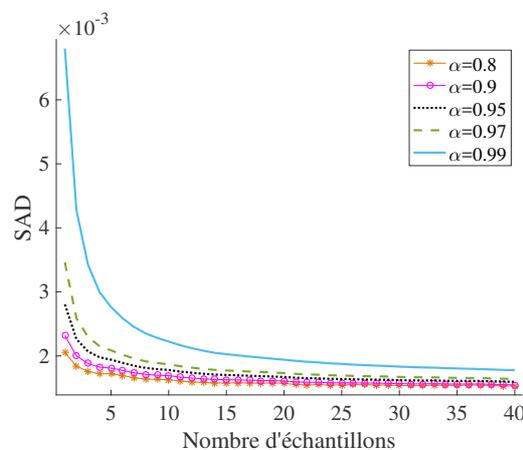


FIGURE 2.15 : Vitesse de convergence du SAD en fonction du nombre d'échantillons, pour différentes valeurs de α , pour le mélange de la Figure 2.10

L'étude de l'influence de α a été réalisée dans le cas de *endmembers* stationnaires. Prenons à présent le

cas de *endmembers* non stationnaires. En effet, en imagerie hyperspectrale, la signature spectrale d'un composant peut présenter une variabilité intra-classe. Ainsi, étudier l'évolution des *endmembers* entre chaque tranche de l'image est parfois cruciale pour comprendre la physique sous-jacente du phénomène. Intégrer une capacité de suivi (via le paramètre α) permet de suivre les changements spectraux au cours du traitement de l'image. Pour résoudre ce problème, une image hyperspectrale de taille $119 \times 40 \times 80$ composée de trois *endmembers* non stationnaires a été simulée. Les cartes d'abondance sont identiques à celles de la Figure 2.10, mais de dimension 40×80 . Les *endmembers* évoluent à chaque tranche selon le modèle suivant : $\tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{S}}^{(k)} + \mathbf{T}^{(k)}$, où $\mathbf{T}^{(k)}$ est un faible bruit Gaussien. La Figure 2.16 représente les vrais *endmembers* aux tranches 10, 50 et 80 de l'image hyperspectrale.

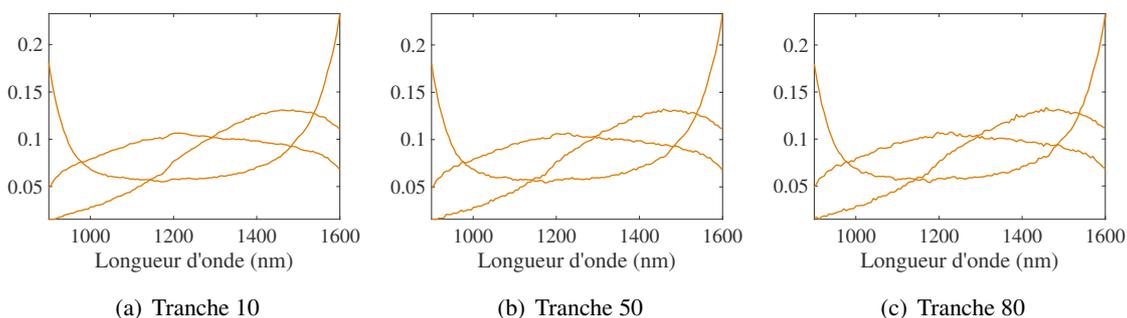


FIGURE 2.16 : Évolution des *endmembers* au cours des tranches de l'image hyperspectrale

Les données simulées ont été corrompues par un bruit avec un SNR = 11 dB. Les paramètres μ et N_{iter} ont été fixés respectivement à 0.01 et 1000. Afin d'évaluer la capacité de nos algorithmes à suivre l'évolution des *endmembers*, nous avons fait varier α de 0.99 à 0.7 et, pour chaque tranche de l'image, nous avons calculé le SAD. Les résultats sont illustrés sur la Figure 2.17. Nous pouvons observer que, de $\alpha = 0.99$ à 0.9, l'algorithme se comporte de la même manière que dans le cas de sources stationnaires, *i.e.* une convergence plus rapide et un SAD asymptotique qui diminue. En effet, pour $\alpha = 0.99$, la contribution des anciens échantillons est bien trop forte pour suivre l'évolution spectrale. En revanche, lorsque $\alpha < 0.9$, bien que la convergence reste plus rapide, nous remarquons une augmentation progressive du SAD ; en effet, plus α est faible, plus la contribution du nouvel échantillon est élevée, ce qui rend l'algorithme beaucoup plus sensible aux variations des spectres. Cela indique qu'il existe une valeur optimale de α atténuant au mieux les erreurs transitoires et les erreurs de suivi.

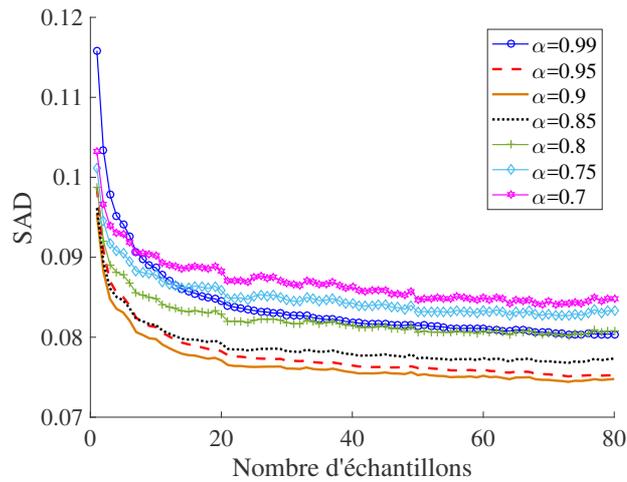


FIGURE 2.17 : Vitesse de convergence du SAD en fonction du nombre d'échantillons, pour différentes valeurs de α , dans le cas de sources non stationnaires

La prochaine expérience vise à montrer l'effet du paramètre ρ sur la convergence de l'algorithme OMDC-ADMM, en examinant l'évolution du SAD. Pour ce faire nous avons fait varier le paramètre ρ (dans le cas de sources stationnaires) de 0.01 à 0.00003 tout en maintenant les valeurs suivantes : $\alpha = 0.99$, $\mu = 0.001$ et $N_{iter} = 200$. La Figure 2.18 montre que, lorsque ρ décroît, le SAD asymptotique diminue, tandis que la vitesse de convergence ne semble pas être affectée. Dès lors que la valeur de ρ devient trop faible, *e.g.* 0.00003, nous constatons une augmentation du SAD, signe que l'algorithme devient instable. Pour le bon fonctionnement de l'algorithme, il existe un intervalle de valeurs adéquates pour le paramètre ρ ; dans notre cas, une valeur de ρ comprise entre 0.001 et 0.00005 permet la stabilité de l'algorithme.

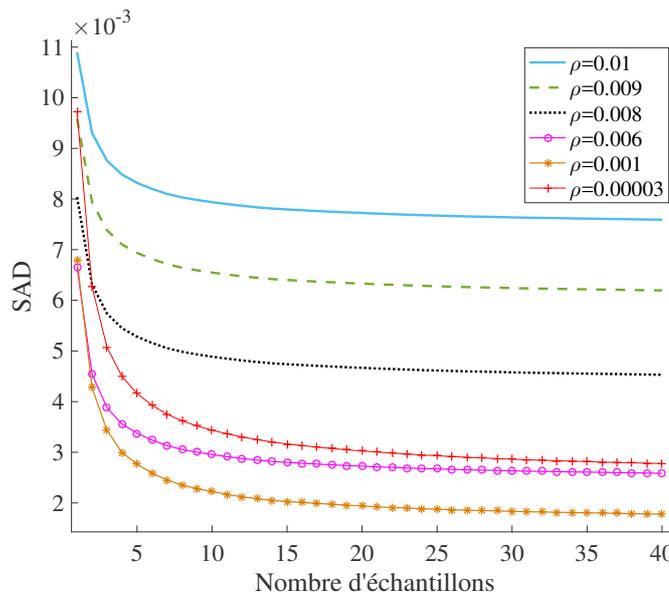


FIGURE 2.18 : Vitesse de convergence du SAD en fonction du nombre d'échantillons, pour différentes valeurs de ρ , pour le mélange de la Figure 2.10

Pour finir, nous avons étudié l'influence de ϵ sur les performances du démélange en utilisant l'algorithme

OMVC-Ne. Les paramètres ont été fixés à $\alpha = 0.99$, $\mu = 0.001$, $Iter_{ex} = 40$ et $Iter_{in} = 20$. Pour différentes valeurs de ϵ comprises entre 0.0001 et 1, le SAD (à partir des *endmembers* estimés à la dernière tranche) et le RMSE ont été calculés pour 20 différentes initialisations aléatoires et les résultats obtenus en effectuant une moyenne sont exposés en Figure 2.19. Nous constatons que pour des faibles et grandes valeurs de ϵ (e.g. 0.0001 et 1), le problème n'est pas correctement régularisé et, dans ce cas, les résultats du démélange hyperspectral ne sont pas pertinents, comme nous pouvons le voir sur les Figures 2.20 et 2.21. La valeur optimale de ϵ (pour la valeur de μ considérée) qui minimise le SAD et le RMSE se situe dans la région « plate » de la courbe. Pour cette expérience, la valeur optimale de ϵ est de 0.4.

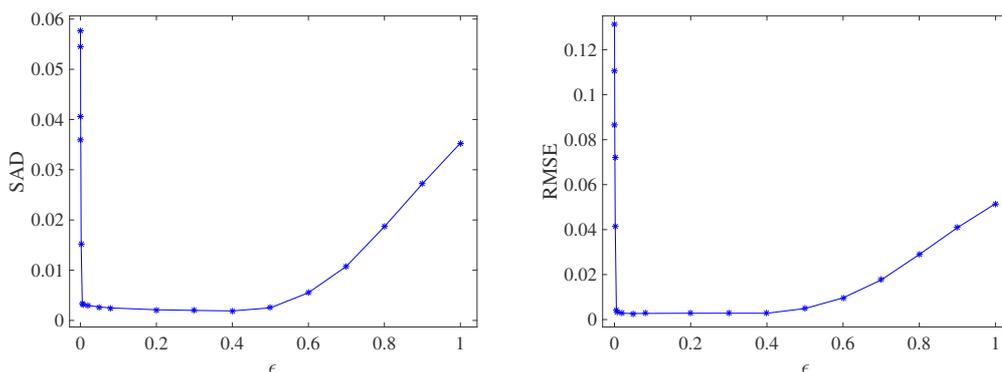


FIGURE 2.19 : Évolution du SAD et du RMSE en fonction de ϵ , pour le mélange de la Figure 2.10

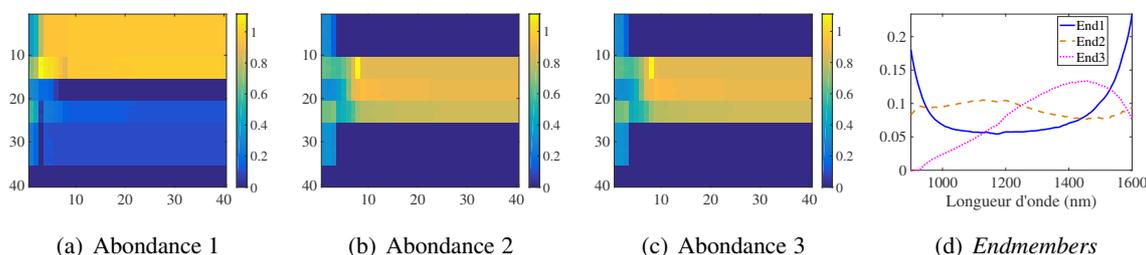


FIGURE 2.20 : Résultat du démélange hyperspectral pour le mélange de la Figure 2.10, pour $\epsilon = 0.0001$

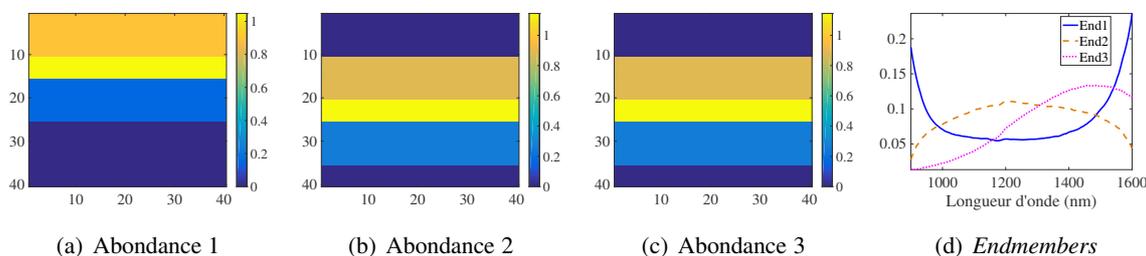


FIGURE 2.21 : Résultat du démélange hyperspectral pour le mélange de la Figure 2.10, pour $\epsilon = 1$

2.5.4 Application sur des données réelles

La dernière expérience a pour objectif de valider les performances du démélange hyperspectral en-ligne sur deux images hyperspectrales réelles avec vérité terrain *Jasper Ridge* et *Samson* disponibles à l'adresse suivante : http://www.escience.cn/people/feiyunZHU/Dataset_GT.html. Pour commencer, nous avons comparé les performances de OMDC-ADMM, OMDC-Ne et OMVC-Ne (nous avons exclu les algorithmes fondés sur les règles de mise à jour multiplicatives, étant donné qu'ils présentent un temps de traitement conséquent) sur l'image hyperspectrale *Samson*. Cette dernière a pour dimension $156 \times 95 \times 95$ où 156 caractérise le nombre de longueurs d'onde (de 401 nm à 889 nm) et 95×95 les dimensions spatiales. Cette image est composée de trois *endmembers* correspondant au sol, aux arbres et à l'eau. La vérité terrain est illustrée sur la Figure 2.22.

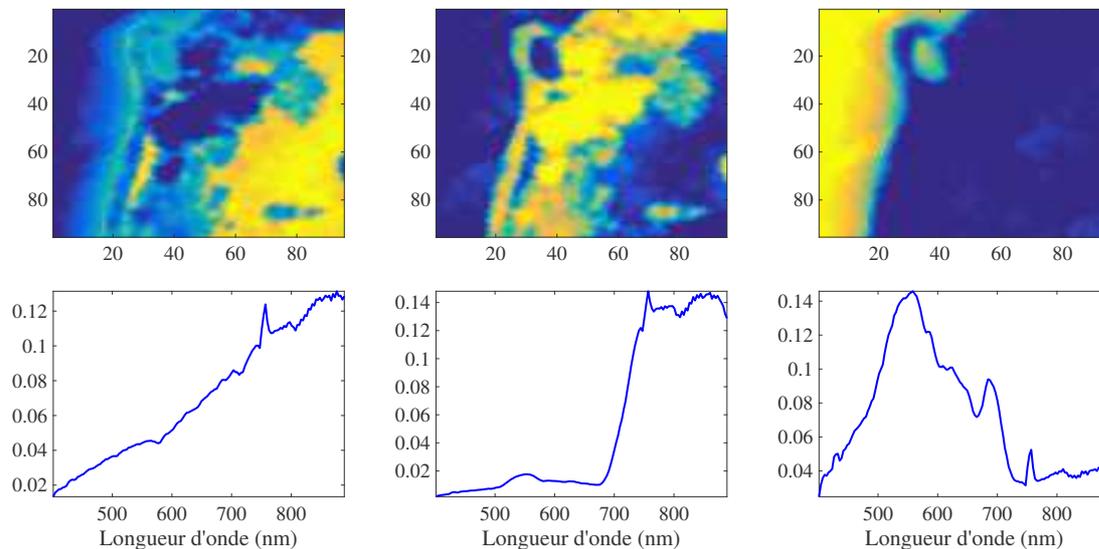


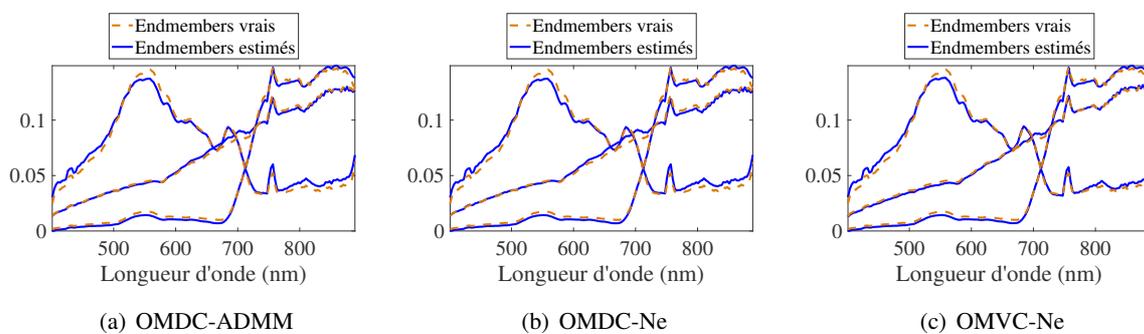
FIGURE 2.22 : Vérité terrain : cartes d'abondance et signatures spectrales pour l'image *Samson*. De gauche à droite : le sol, les arbres et l'eau

L'image hyperspectrale a été traitée séquentiellement le long de l'axe vertical. Notons que le traitement séquentiel le long de l'axe horizontal n'est pas possible car la préservation du rang n'est pas garantie d'une tranche à l'autre. Les divers paramètres de OMDC-ADMM, OMDC-Ne et OMVC-Ne ont été fixés comme suit : $R = 3$, $\alpha = 0.99$ (pour assurer une cohérence forte des sources entre les échantillons), $\mu = 0.05$, $\epsilon = 0.4$ et $\rho = 0.001$. De manière à ce que chaque algorithme atteigne sa meilleure performance, le plus rapidement possible, nous avons fixé le nombre d'itérations de la manière suivante : $N_{iter} = 200$, pour OMDC-ADMM et $Iter_{ex} \times Iter_{in} = 2000$ pour OMDC-Ne et OMVC-Ne. Afin d'évaluer la qualité des *endmembers* et des abondances estimés, le SAD et le RMSE pour toutes les méthodes ont été calculés. Chaque expérience a été répétée 50 fois pour différentes initialisations aléatoires et les résultats moyens pour le SAD et le RMSE sont fournis dans le Tableau 2.3.

<i>Endmembers</i>	OMDC-ADMM		OMDC-Ne		OMVC-Ne	
	SAD	RMSE	SAD	RMSE	SAD	RMSE
Sol	0.0247	0.0645	0.0314	0.0699	0.0177	0.0540
Arbre	0.0321	0.0391	0.0322	0.0447	0.0319	0.0355
Eau	0.0586	0.0383	0.0590	0.0383	0.0503	0.0345
Moyenne	0.0384	0.0473	0.0408	0.0510	0.0333	0.0414
Temps (s)	2.23		2.58		4.94	

TABLE 2.3 : SAD et RMSE pour OMDC-ADMM, OMDC-Ne et OMVC-Ne, pour l'image *Samson*

D'après le Tableau 2.3, nous constatons que l'utilisation de OMDC-ADMM et OMDC-Ne permet de réduire significativement le temps de traitement comparé à OMVC-Ne. En effet, l'utilisation de la contrainte de volume minimal augmente la complexité de calcul, ce qui a pour conséquence un temps de traitement plus important (4.94 s vs. 2.58 s). Il est intéressant de noter que OMVC-Ne fournit des estimations de meilleure qualité ; cela peut s'expliquer par le fait que MVC peut être considéré comme une vraie mesure du volume du simplexe, contrairement à la contrainte de dispersion minimale qui n'est qu'une approximation. Les Figures 2.23 et 2.24 présentent les résultats du démélange des trois algorithmes, comparés à la vérité terrain. Finalement, nous constatons que les trois méthodes proposées ont des performances très similaires en ce qui concerne les estimations. Néanmoins, ce qui les différencie majoritairement est le temps de traitement de l'image. Ainsi, OMDC-ADMM et OMDC-Ne semblent être plus adaptés pour un traitement en-ligne rapide.

FIGURE 2.23 : *Endmembers* vrais et estimés par OMDC-ADMM, OMDC-Ne et OMVC-Ne, pour l'image *Samson*

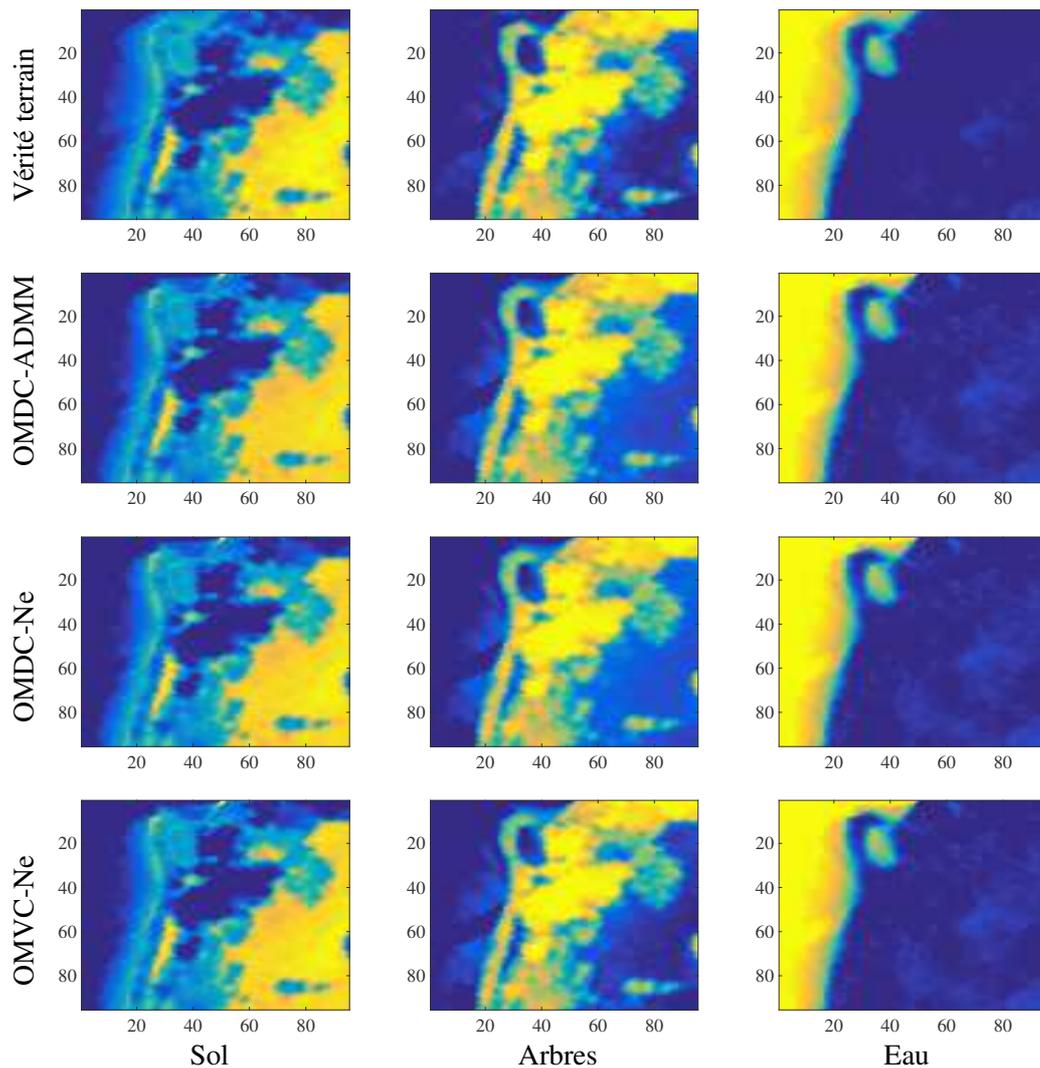


FIGURE 2.24 : Cartes d'abondance vraies et estimées par OMDC-ADMM, OMDC-Ne et OMVC-Ne, pour l'image *Samson*

Pour finir, nous proposons de nous focaliser sur l'algorithme OMDC-ADMM ; les simulations précédentes ont montré que OMDC-ADMM présente le temps de traitement le plus rapide. Pour cette expérience, nous utilisons l'image hyperspectrale *Jasper Ridge*. Cette image a pour dimension $198 \times 100 \times 100$ où 198 caractérise le nombre de longueurs d'onde (de 380 nm à 2500 nm) et 100×100 les dimensions spatiales. Cette image est composée de quatre *endmembers* correspondant aux arbres, à l'eau, au sol et à la route. La vérité terrain est illustrée sur la Figure 2.25.

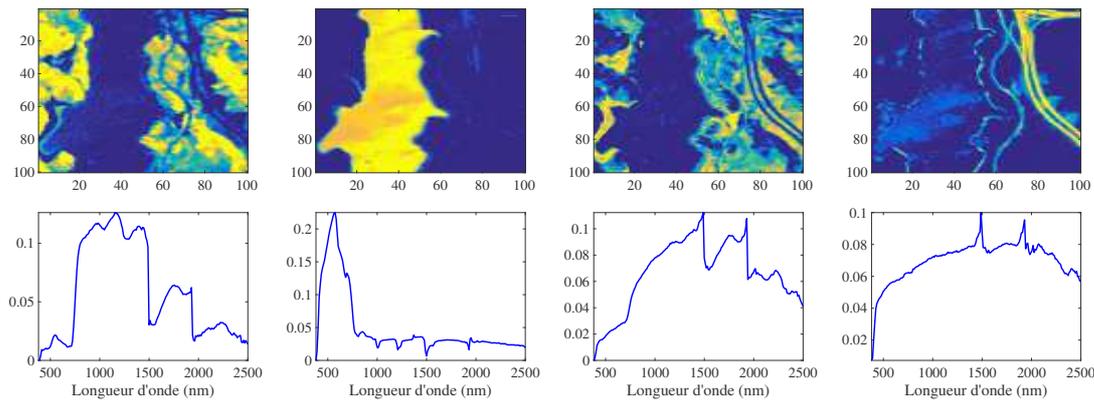


FIGURE 2.25 : Vérité terrain : cartes d'abondance et signatures spectrales pour l'image *Jasper Ridge*.

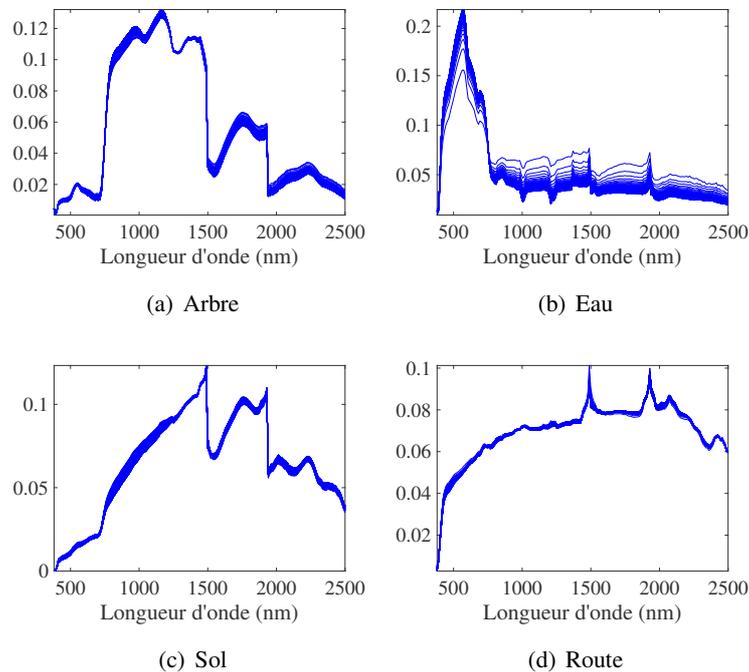
De gauche à droite : les arbres, l'eau, le sol et la route

Nous avons comparé les performances de OMDC-ADMM à sa version hors-ligne BMDC-ADMM mais également à d'autres méthodes de démixage hors-ligne de l'état de l'art : VCA [Nascimento 05], NMF [Lee 99], ℓ_1 -NMF [Hoyer 02] et $\ell_{1/2}$ -NMF [Qian 11]. Les résultats des quatre dernières méthodes proviennent des simulations réalisées dans [Zhu 17]. L'image *Jasper Ridge* a été traitée par OMDC-ADMM séquentiellement le long de l'axe vertical. Les paramètres ont été fixés de la façon suivante : $R = 4$, $\alpha = 0.99$, $\mu = 0.05$, $\rho = 0.001$ et $N_{iter} = 200$. BMDC-ADMM a été appliqué à une version dépliée de l'image hyperspectrale, de dimension 198×10000 avec comme paramètre $\rho = 0.001$, $\mu = 300$ et $N_{iter} = 2000$. Concernant le détail du paramétrage des autres méthodes, le lecteur peut se référer à [Zhu 17]. Le SAD et le RMSE ont été calculés pour toutes les méthodes et chaque expérience a été répétée 50 fois pour obtenir des résultats valides. Notons que pour OMDC-ADMM et BMDC-ADMM, les *endmembers* ont été initialisés aléatoirement tandis que pour les quatre autres méthodes, [Zhu 17] propose de les initialiser avec l'algorithme VCA, afin de générer des résultats reproductibles. Les résultats moyens pour le SAD et le RMSE sont fournis dans les Tableaux 2.4 et 2.5. Ces derniers montrent que BMDC-ADMM surpasse les autres méthodes hors-ligne de l'état de l'art. De plus, OMDC-ADMM et BMDC-ADMM fournissent des estimations proches pour le sol, les arbres et la route. Il est intéressant de noter que nous observons une différence significative entre les deux méthodes en ce qui concerne le SAD de l'eau (0.1113 pour OMDC-ADMM vs. 0.3141 pour BMDC-ADMM). Cette différence peut être expliquée par le phénomène de la variabilité spectrale, qui peut modifier localement le spectre des *endmembers* [Drumetz 16]. Les causes de cette variabilité peuvent être diverses, par exemple, des changements d'illumination au cours de l'acquisition de l'image, la variabilité intrinsèque des composants ou encore les effets atmosphériques. Cet aspect est illustré sur la Figure 2.26, où nous pouvons observer les spectres des *endmembers* estimés par OMDC-ADMM à chaque tranche de l'image. Grâce à sa capacité de suivi, OMDC-ADMM offre la possibilité d'étudier l'évolution dynamique des *endmembers* au cours de l'image. Notons que la signature spectrale de l'eau évolue fortement entre les tranches, contrairement aux trois autres composants. Précisons que dans le Tableau 2.4, les valeurs du SAD présentées pour OMDC-ADMM ont été calculées à partir des spectres moyens obtenus sur l'ensemble des K échantillons (voir Figure 2.28 (a)).

<i>Endmembers</i>	OMDC-ADMM		BMDC-ADMM		VCA	
	SAD	RMSE	SAD	RMSE	SAD	RMSE
Arbre	0.0447	0.0398	0.0905	0.0667	0.2565	0.3268
Eau	0.1113	0.0343	0.3141	0.0521	0.2474	0.3151
Sol	0.0839	0.0897	0.0886	0.0893	0.3584	0.2936
Route	0.0498	0.0787	0.0473	0.0693	0.5489	0.2829
Moyenne	0.0724	0.0606	0.1351	0.0694	0.3528	0.3046

TABLE 2.4 : SAD et RMSE pour OMDC-ADMM, BMDC-ADMM et VCA, pour l'image *Jasper Ridge*

<i>Endmembers</i>	NMF		ℓ_1 -NMF		$\ell_{1/2}$ -NMF	
	SAD	RMSE	SAD	RMSE	SAD	RMSE
Arbre	0.2130	0.1402	0.0680	0.0636	0.0409	0.0707
Eau	0.2001	0.1106	0.3815	0.0660	0.1682	0.1031
Sol	0.1569	0.2557	0.0898	0.2463	0.0506	0.2679
Route	0.3522	0.2450	0.4118	0.2344	0.3670	0.2737
Moyenne	0.2305	0.1879	0.2378	0.1526	0.1567	0.1789

TABLE 2.5 : SAD et RMSE pour NMF, ℓ_1 -NMF et $\ell_{1/2}$ -NMF, pour l'image *Jasper Ridge*FIGURE 2.26 : *Endmembers* estimés tranche par tranche par OMDC-ADMM, pour l'image *Jasper Ridge*

Il est important de signaler que les pics situés à 1500 et 1900 nm du spectre de l'eau estimé par OMDC-ADMM ont une amplitude inversée par rapport à la vérité terrain. Cela se justifie par la valeur de α ; en effet, la capacité de suivi de l'algorithme peut être altérée lorsque la valeur de α est choisie trop grande notamment 0.99. En revanche, nous assurons la cohérence des *endmembers* entre chaque échantillon.

Pour davantage prendre en compte la variabilité spectrale, tout en garantissant des estimations de bonne qualité, nous avons réduit la valeur de α à 0.9, tout en adaptant μ à 0.2. La Figure 2.27 montre la variabilité du spectre de l'eau estimé par OMDC-ADMM pour chaque tranche, où nous pouvons observer cette fois-ci l'apparition des deux pics associés aux longueurs d'onde 1500 et 1900 nm. Rappelons que dans le cas où les *endmembers* varient fortement entre les échantillons, α doit être choisi comme étant le meilleur compromis entre l'erreur transitoire et l'erreur de suivi.

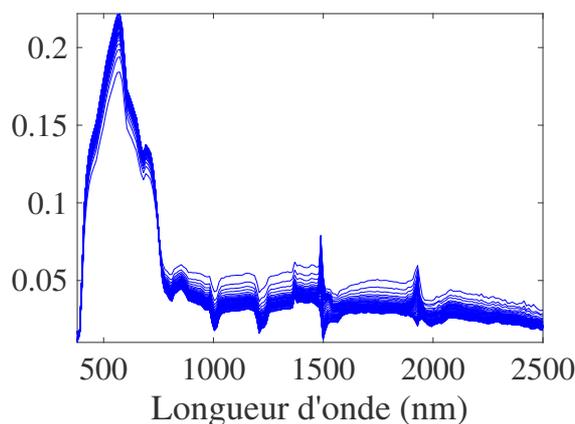


FIGURE 2.27 : Spectres de l'eau estimés tranche par tranche par OMDC-ADMM pour $\alpha = 0.9$, pour l'image *Jasper Ridge*

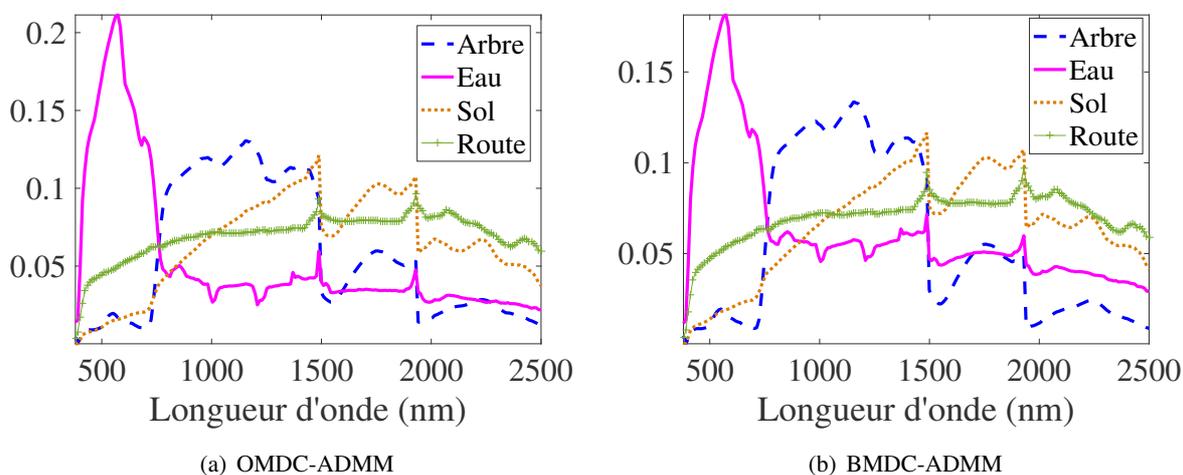


FIGURE 2.28 : *Endmembers* estimés pour l'image *Jasper Ridge*

Pour BMDC-ADMM, cette variabilité spectrale n'est pas explicitement prise en compte, et les *endmembers* estimés peuvent être interprétés comme des spectres moyens pour l'ensemble de l'image (voir Figure 2.28 (b)). Cela a des répercussions sur l'estimation des cartes d'abondance ; la Figure 2.29 représente les « meilleures » cartes d'abondance estimées par la version en-ligne et hors-ligne, comparées à la vérité terrain. Nous constatons, sur la carte d'abondance de l'eau générée par BMDC-ADMM, que certains détails ne sont pas visibles, contrairement à celle de OMDC-ADMM. Outre le fait que OMDC-ADMM permet de suivre la variabilité spectrale, elle nécessite également moins d'itérations que

la version *batch*, ce qui réduit considérablement le temps de traitement (2.68 s pour OMDC-ADMM *vs.* 4.98 s pour BMDC-ADMM).

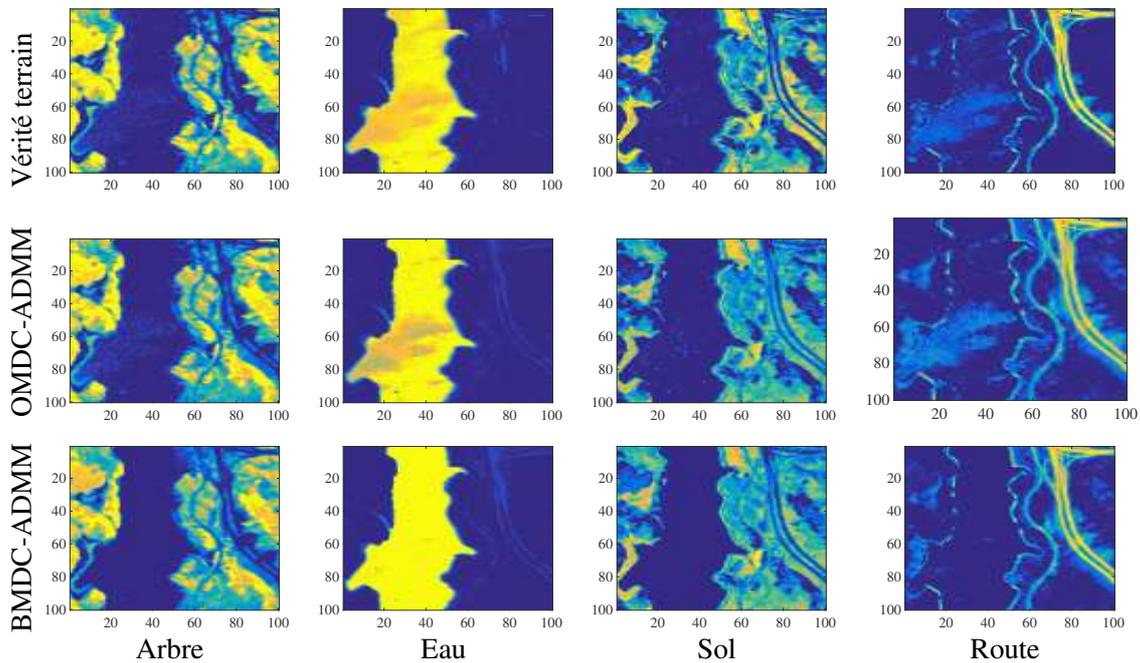


FIGURE 2.29 : Cartes d'abondance vraies et estimées par OMDC et BMDC-ADMM, pour l'image *Jasper Ridge*

2.6 Estimation du paramètre de régularisation

L'ensemble de nos méthodes introduisent un paramètre de régularisation μ relatif à MDC ou à MVC qui doit être déterminé. La mise en œuvre efficace de nos algorithmes dépend fortement du choix de la valeur de l'hyperparamètre. L'approche commune pour choisir les paramètres de régularisation consiste à utiliser des valeurs fixes basées sur des connaissances *a priori* ou sur des expériences antérieures concernant les données. Il est donc difficile de dire que les valeurs choisies sont optimales dans un certain sens. Il existe quelques approches classiques pour estimer le paramètre de régularisation telles que la méthode GCV (*Generalized Cross Validation*) [Golub 79] ou encore la courbe en L [Hansen 92, Hansen 99]. Puisque le problème NMF est bilinéaire, la méthode GCV n'est pas applicable. La courbe en L est un tracé sur une échelle log-log du coût de régularisation en fonction de l'adéquation aux données ; la valeur optimale de l'hyperparamètre correspond à la courbure maximale (MCC) de la courbe en L. Son lien avec une optimisation multi-objectif a été mentionné pour la première fois dans [Kaufman 97]. Cependant, l'approche courbe en L présente certaines propriétés indésirables, décrites dans [Hanke 96, Vogel 96] : elle est non-convexe et la courbure maximale n'est pas garantie d'être unique. Pour surmonter le caractère non-unique de la courbure maximale, [Belge 02] a proposé le critère de distance minimale (MDistC) appliqué à la courbe en L, dans un cas d'optimisation bi-objectif et multi-objectif. MDistC fournit une solution unique à condition que la courbe de réponse soit convexe ; ainsi, cela nécessite de limiter la plage de l'hyperparamètre pour s'assurer que la partie correspondante de la courbe en L est convexe. Enfin, il a été proposé dans [Song 16] d'estimer l'hyperparamètre par le biais de MDistC, en déterminant

la courbe de réponse (cas bi-objectif) ou la surface de réponse (cas multi-objectif), qui contrairement à la courbe en L, est définie comme un tracé linéaire de la pénalité en fonction de l'adéquation aux données. Cette méthode a été utilisée dans le cadre de la déconvolution d'image avec une régularisation de type Tikhonov.

Ainsi, cette section a pour objectif d'étudier l'intérêt de ces approches pour l'estimation de l'hyperparamètre μ . Nous proposons de limiter nos résultats à la contrainte de dispersion minimale dans le cadre du démelange en-ligne d'images hyperspectrales ; la procédure à mener est identique pour la contrainte de volume minimal. Pour ce faire, nous reformulons notre problème en-ligne (2.8) sous la forme d'un problème de minimisation bi-objectif. À notre connaissance, cette reformulation n'est pas très commune ; notons que dans [Mirzal 14], une approche utilisant la courbe en L a été proposée pour estimer une pénalité de type Tikhonov dans le modèle NMF multiplicatif classique. Nous proposons d'exploiter deux approches : MCC et MDistC appliquées directement sur la courbe de réponse linéaire. Les résultats obtenus sur des données simulées et réelles nous permettront de juger de l'approche la plus adaptée à notre problématique.

2.6.1 Le démelange hyperspectral comme un problème d'optimisation bi-objectif : le choix de l'hyperparamètre μ

Le point de départ de notre problème est le démelange hyperspectral en-ligne (2.3). Considérons une image hyperspectrale composée de K tranches $\tilde{\mathbf{X}}^{(k)}$, $k = 1, \dots, K$, qui est utilisée pour apprendre les valeurs de l'hyperparamètre μ . Nous supposons que les valeurs de α et N_{iter} sont fixes. Soit ${}_{(\mu)}\tilde{\mathbf{S}}^{(K)}$ et ${}_{(\mu)}\tilde{\mathbf{A}}^{(K)}$ les *endmembers* et les abondances estimés pour une valeur de μ donnée. Selon [Song 16], la courbe de réponse (dans le cas bi-objectif) est le tracé linéaire du coût de régularisation en fonction de l'adéquation aux données pour $\mu \in [0, +\infty)$. Les deux fonctions objectif sont définies respectivement de la manière suivante :

$$\mathcal{J}_1(\mu) = \mathcal{J}_1 \left({}_{(\mu)}\tilde{\mathbf{S}}^{(K)}, {}_{(\mu)}\tilde{\mathbf{A}}^{(K)} \right) = \frac{1}{K} \sum_{k=1}^K \left\| \tilde{\mathbf{X}}^{(k)} - {}_{(\mu)}\tilde{\mathbf{S}}^{(k)} {}_{(\mu)}\tilde{\mathbf{A}}^{(k)} \right\|_F^2, \quad (2.25)$$

$$\mathcal{J}_2(\mu) = \mathcal{J}_2 \left({}_{(\mu)}\tilde{\mathbf{S}}^{(K)} \right) = \frac{1}{K} \sum_{k=1}^K \text{trace} \left({}_{(\mu)}\tilde{\mathbf{S}}^{(k)} \mathbf{D}_{(\mu)} \tilde{\mathbf{S}}^{(k)T} \right). \quad (2.26)$$

Si la valeur de μ est très grande, le terme de dispersion minimale sera minimisé. Lorsque μ approche zéro, la solution tend à minimiser le terme d'adéquation aux données. Quand μ est très faible, le résultat du démelange est généralement insatisfaisant car de multiples solutions existent. De manière similaire, lorsque μ est très grand, le volume tend à être nul ce qui induit la colinéarité des *endmembers*. Cela signifie que nous ne pouvons pas améliorer un objectif sans en détériorer le deuxième. En énonçant le problème (2.3) comme un problème d'optimisation bi-objectif, il est possible d'estimer la courbe de réponse qui offre une caractérisation de l'ensemble des solutions obtenues en faisant varier les valeurs de μ . Un problème d'optimisation bi-objectif peut être formulé comme suit:

$$\underset{\mu}{\text{minimiser}} \mathcal{J}(\mu),$$

avec $\mathcal{J}(\mu) = (\mathcal{J}_1(\mu), \mathcal{J}_2(\mu))$. $\mathcal{J}_1(\mu)$ et $\mathcal{J}_2(\mu)$ définissent ainsi un espace bidimensionnel.

La minimisation de $\mathcal{J}(\mu)$ utilise la notion de domination [Deb 01]. La notion de domination définit un ordre dans l'espace des solutions, permettant de dire qu'une solution est meilleure qu'une autre. Une solution domine une autre solution, si la première n'est pas pire que la seconde pour tous les objectifs, et que la première solution est strictement meilleure que la seconde pour au moins un objectif. La solution est dite non dominée ou Pareto optimale si toutes les autres solutions de l'ensemble réalisable ont une valeur supérieure dans au moins un des objectifs. L'image de toutes les solutions non dominées s'appelle le Front de Pareto. La forme du Front de Pareto représente l'ensemble des meilleurs compromis possibles entre les objectifs.

Dans notre cas, nous estimons les paramètres $\tilde{\mathbf{S}}$ et $\tilde{\mathbf{A}}$ pour différentes valeurs de μ , et représentons le critère (2.25) en fonction de (2.26) pour l'ensemble des valeurs de μ considérées, ce qui définit une courbe de réponse. Le but est de choisir parmi cet ensemble une solution particulière qui consiste à estimer le paramètre de régularisation μ . Comme nous l'avons mentionné précédemment, deux principales méthodes vont être utilisées : le critère de courbure maximale ainsi que le critère de distance minimale.

2.6.1.1 Critère de courbure maximale

Afin de calculer la courbure, nous avons besoin d'estimer la première et la seconde dérivée de chaque fonction objectif \mathcal{J}_1 et \mathcal{J}_2 par rapport à μ . Si g est une fonction de la variable μ , nous pouvons, par exemple, estimer la première dérivée partielle de g , par rapport à μ comme suit :

$$\dot{g}_\mu(i) \approx \frac{g(\mu_{i+1}) - g(\mu_i)}{\mu_{i+1} - \mu_i}.$$

De manière similaire, la seconde dérivée de g est estimée comme suit :

$$\ddot{g}_\mu(i) \approx \frac{g(\mu_{i+1}) + g(\mu_{i-1}) - 2g(\mu_i)}{\mu_{i+1} - \mu_{i-1}}.$$

La courbure se définit de la manière suivante [Hansen 99] : soit $b = \mathcal{J}_1(\mu)$ et $c = \mathcal{J}_2(\mu)$ et κ la courbure le long de μ , alors :

$$\kappa = \frac{\dot{b}\ddot{c} - \ddot{b}\dot{c}}{(\dot{b}^2 + \dot{c}^2)^{3/2}}.$$

Ensuite, le critère de courbure maximale peut se définir ainsi :

$$\hat{\mu} = \arg \max_{\mu} \kappa(\mu). \quad (2.27)$$

Un inconvénient du critère (2.27) est la non-unicité de sa solution. Pour illustrer ce point, considérons l'exemple donné sur la Figure 2.30. Cela correspond à une courbe de réponse convexe dont la courbure présente deux maxima indiqués par les deux points. Pour surmonter ce problème de solutions multiples, nous suivons les recommandations de [Hansen 99]. En commençant par une faible valeur du paramètre de régularisation, la première courbure maximale est choisie pour estimer l'hyperparamètre.

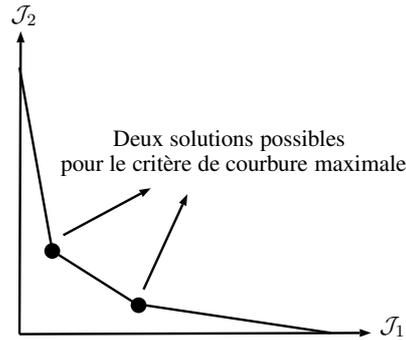


FIGURE 2.30 : Exemple d'une courbe de réponse présentant une courbure maximale non-unique

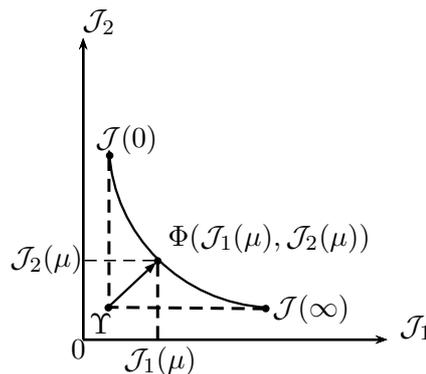
2.6.1.2 Critère de distance minimale

Le critère de distance minimale [Song 16] consiste à rechercher le point de la courbe de réponse situé à la distance minimale du *point idéal*. Le point idéal Υ correspond au point dont les coordonnées sont les minima des deux fonctions objectif (Figure 2.31). Il peut être considéré comme un point de référence et le *point optimal* Φ de la courbe de réponse sera celui ayant la distance minimale de ce point idéal. Soit $\Upsilon = (\Upsilon_1, \Upsilon_2)$ désigne les coordonnées du point idéal. La fonction $\Gamma(\mu)$ est la distance au carré du point idéal Υ au point $\Phi(\mu) = (\mathcal{J}_1(\mu), \mathcal{J}_2(\mu))$ sur la courbe de réponse :

$$\Gamma(\mu) = \sum_{i=1}^2 (\mathcal{J}_i(\mu) - \Upsilon_i)^2.$$

MDistC est défini de la manière suivante :

$$\hat{\mu} = \arg \min_{\mu} \Gamma(\mu).$$

FIGURE 2.31 : Représentation de la courbe de réponse. Le point idéal est désigné par Υ et le point optimal par Φ

D'après [Song 16], le critère de distance minimal est unique à condition que la courbe de réponse soit convexe. Étant donné que nous sommes confrontés à un problème bilinéaire, nous n'avons aucune garantie en ce qui concerne la convexité de la courbe de réponse. Dans la suite, le choix de la méthode la plus adaptée (*i.e.* MCC ou MDistC) pour l'estimation de l'hyperparamètre μ sera justifié à travers des expériences sur des données simulées et réelles.

2.6.2 Évaluation des critères MCC et MDist sur des données réelles et simulées

Nous proposons d'illustrer les performances de MCC et MDistC pour l'estimation du paramètre de régularisation à travers quelques résultats expérimentaux et numériques. Afin de quantifier les performances du démélange en fonction des valeurs de μ estimées par MCC et MDistC, les critères (2.23) et (2.24) ont été utilisés.

2.6.2.1 Forme d'une courbe de réponse

Ce point est abordé numériquement à l'aide d'une image hyperspectrale simulée, où les cartes d'abondance sont présentées en Figure 2.6. Pour toutes les simulations qui vont suivre, nous utilisons l'algorithme OMDC-ADMM qui traite l'image séquentiellement par tranche de dimension 119×36 . Les différents paramètres ont été définis de la manière suivante : $R = 3$, $\alpha = 0.99$, $\rho = 0.001$ et $N_{iter} = 200$. Les matrices ont été initialisées comme présentées à l'Algorithme 2. La courbe de réponse a été évaluée à partir de 54 valeurs de $\mu \in [0, 50]$. La Figure 2.32 correspond à la courbe de réponse obtenue. Par ailleurs, les simplexes \mathcal{S} définis par les *endmembers* estimés (à la dernière tranche de l'image), pour différentes valeurs de μ , sont affichés sur la Figure 2.33. Il sont comparés aux simplexes vrais notés \mathcal{D} .

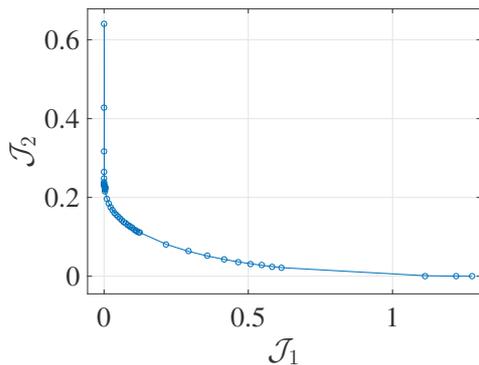


FIGURE 2.32 : Courbe de réponse estimée pour le mélange de la Figure 2.6

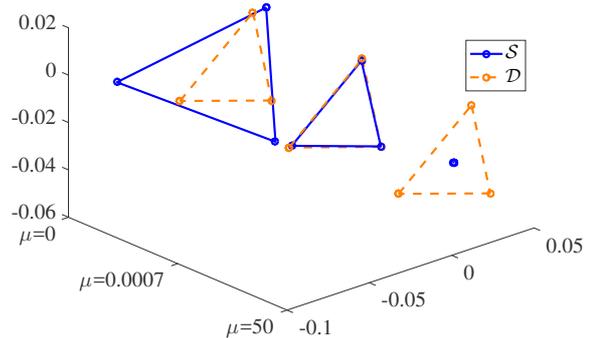


FIGURE 2.33 : Évolution du simplexe estimé pour différentes valeurs de μ , pour le mélange de la Figure 2.6

La courbe de réponse comprend trois principales parties : la première région correspond à des faibles valeurs de μ . Le vrai simplexe \mathcal{D} englobant les données est inclus dans le simplexe estimé final \mathcal{S} , *i.e.* $\mathcal{D} \subset \mathcal{S}$. Cette situation est observée sur la Figure 2.33 pour $\mu = 0$. La deuxième région correspond à de grandes valeurs de μ ; le simplexe estimé tend vers zéro, ce qui indique que les *endmembers* tendent vers la colinéarité. Dans ce cas, $\mathcal{D} \not\subset \mathcal{S}$. Cette situation est observée sur la Figure 2.33 pour $\mu = 50$. Dans ce cas limite, les *endmembers* sont situés au centre de gravité du simplexe. La troisième région est le coude de la courbe de réponse qui correspond aux valeurs de μ pour lesquelles $\mathcal{D} \approx \mathcal{S}$. La valeur optimale de μ se situe dans cette partie de la courbe. Cette situation est observée sur la Figure 2.33 pour $\mu = 0.0007$. La Figure 2.34 illustre l'influence de μ sur l'estimation des *endmembers*. Lorsque $\mu = 0$, les *endmembers* estimés s'éloignent de la vérité terrain et une multitude de solutions existe. *A contrario* lorsque la valeur de μ est élevée, les *endmembers* tendent vers la colinéarité. Enfin, dans le cas où la valeur de μ est judicieusement choisie, l'estimation des *endmembers* est pertinente.

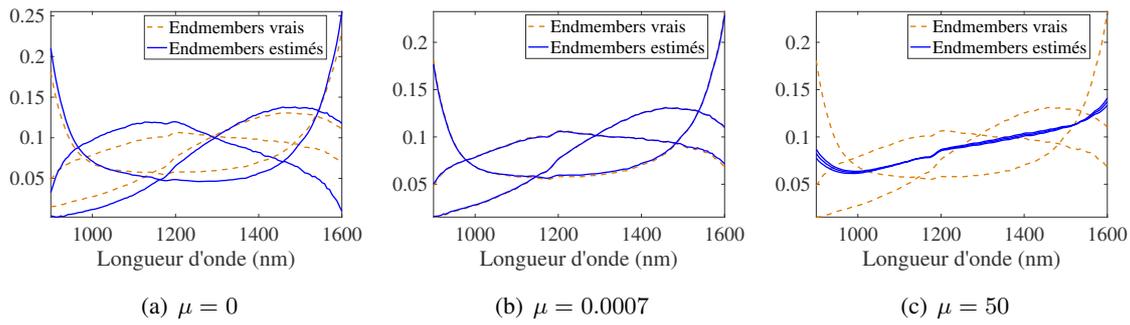


FIGURE 2.34 : *Endmembers* estimés pour différentes valeurs de μ , pour le mélange de la Figure 2.6

2.6.2.2 Influence des conditions initiales

Comme mentionné précédemment, la NMF est un problème non-convexe et peut être sensible à l'initialisation. Cela signifie que la courbe de réponse dépend des conditions initiales. Pour évaluer l'influence de ces valeurs initiales, nous avons étudié les courbes de réponse obtenues pour 50 valeurs initiales différentes $\tilde{\mathbf{S}}^{(0)}$ générées aléatoirement d'une distribution uniforme sur l'intervalle $[0, 1]$. Il apparaît (Figure 2.35) que la plupart des courbes de réponse sont très similaires. C'est un aspect très important car cela ouvre la possibilité d'utiliser une seule réalisation de la courbe de réponse pour estimer la valeur optimale de μ .

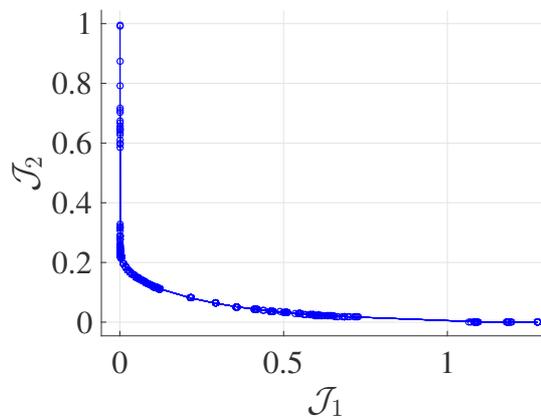


FIGURE 2.35 : Courbes de réponse estimées pour différentes valeurs de μ et différentes initialisations, pour le mélange de la Figure 2.6

2.6.2.3 Évaluation des critères MCC et MDistC sur des données simulées

Nous considérons, dans cette section, deux images hyperspectrales simulées dont les cartes d'abondance sont présentées sur les Figures 2.6 et 2.10 et les *endmembers* sur la Figure 2.7. Pour commencer, nous avons appliqué MCC et MDistC au problème de démixage hyperspectral en-ligne pour le mélange de la Figure 2.6, à l'aide de l'algorithme OMDC-ADMM. Un faible bruit a été ajouté aux données (SNR = 45 dB). Les paramètres de OMDC-ADMM ont été fixés à $R = 3$, $\alpha = 0.99$, $\rho = 0.001$ et $N_{iter} = 200$. L'algorithme a traité l'image séquentiellement, par tranche de dimension 119×36 . La courbe de réponse a été réalisée suivant une seule initialisation générée aléatoirement, et pour des valeurs

de μ comprises entre 0 et 50. La courbe de réponse et le point correspondant à MCC sont représentés sur la Figure 2.36. Le résultat du démélange hyperspectral avec comme paramètre $\mu = 0.0006$ estimé par la méthode est montré sur la Figure 2.37. Notons que \mathcal{S} caractérise le simplexe estimé à la dernière tranche de l'image hyperspectrale. La Figure 2.38 présente la courbe de réponse et le point correspondant à MDistC. Le résultat du démélange est présenté sur la Figure 2.39, pour $\mu = 0.1$. Nous constatons, pour ce problème de démélange hyperspectral, que MCC fonctionne mieux que MDistC. Les faibles performances de MDistC résultent du choix de la gamme de l'hyperparamètre. En effet, les coordonnées du point idéal Υ évoluent à mesure que μ augmente, modifiant ainsi l'estimation du point optimal Φ .

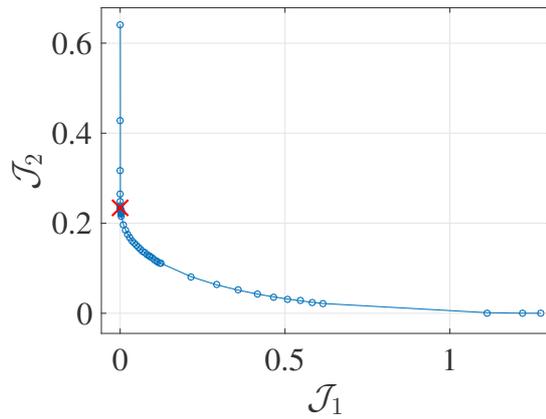


FIGURE 2.36 : Courbe de réponse estimée pour le mélange de la Figure 2.6 avec le point de courbure maximale (croix rouge)

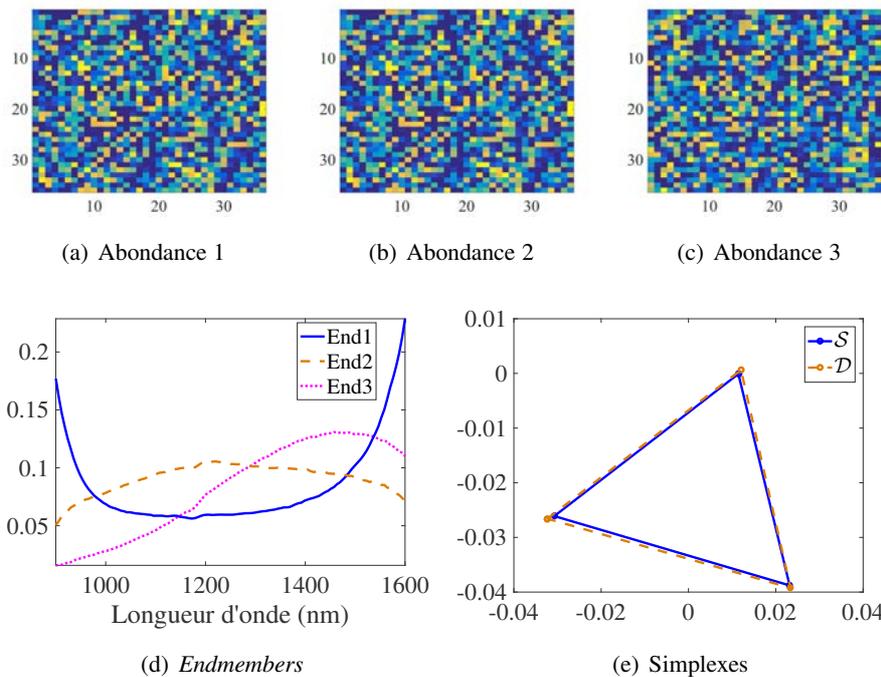


FIGURE 2.37 : Résultat du démélange hyperspectral pour le mélange de la Figure 2.6, pour $\mu = 0.0006$ (estimé par MCC)

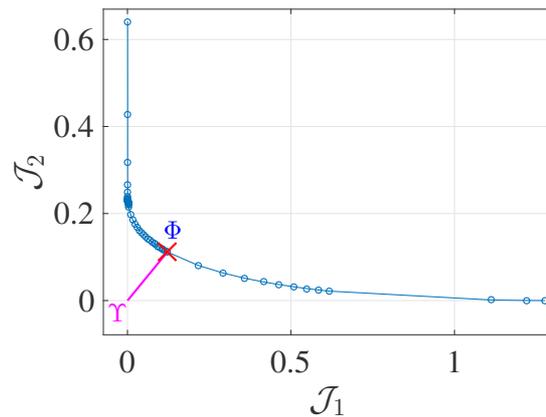


FIGURE 2.38 : Courbe de réponse estimée pour le mélange de la Figure 2.6 avec le point Φ (croix rouge) ayant la distance minimale du point idéal Υ

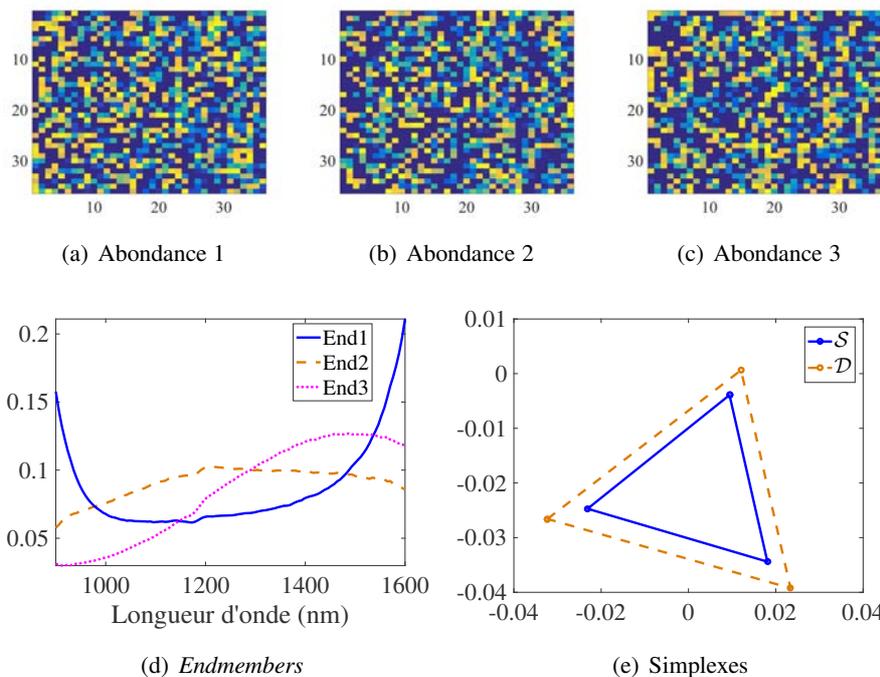
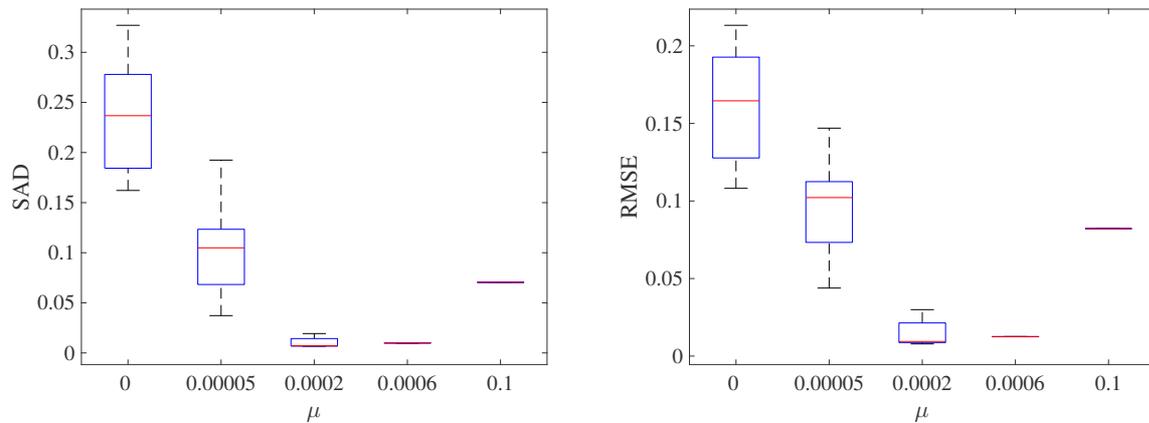
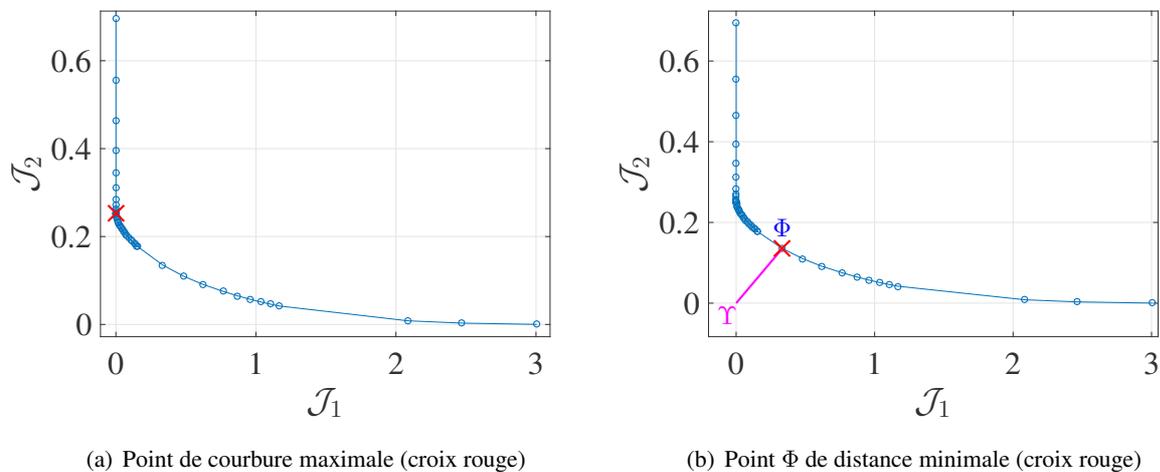


FIGURE 2.39 : Résultat du démixage hyperspectral pour le mélange de la Figure 2.6, pour $\mu = 0.1$ (estimé par MDistC)

La Figure 2.40 représente respectivement les valeurs du SAD (2.23) (calculés à partir des *endmembers* estimés à la dernière tranche) et du RMSE (2.24), sous forme de *boxplot*, obtenues à partir de 20 différentes initialisations générées aléatoirement et pour différentes valeurs de μ , dont celles estimées par MCC et MDistC. Lorsque la valeur de μ est faible, nous constatons une grande dispersion du SAD et du RMSE, révélant une multitude de solutions. À mesure que μ augmente, la variance et la médiane diminue, jusqu'à stabiliser totalement la solution lorsque μ est judicieusement choisi *e.g.* 0.0006 ; cela montre que la contrainte de dispersion minimale a un fort pouvoir de régularisation. En revanche, à partir d'une certaine valeur de μ , *e.g.* 0.1, bien que la dispersion soit nulle, les valeurs du SAD et du RMSE augmentent, indiquant que les *endmembers* tendent vers la colinéarité.

FIGURE 2.40 : SAD et RMSE en fonction de μ , pour le mélange de la Figure 2.6

Nous avons réitéré cette procédure pour l'image hyperspectrale dont les cartes d'abondance sont affichées sur la Figure 2.10, où un faible bruit a été ajouté aux données (SNR = 45 dB). Les paramètres de OMDC-ADMM sont les mêmes que précédemment. L'algorithme a traité l'image séquentiellement, par tranche de dimension 119×40 . La Figure 2.41 représente les courbes de réponse obtenues pour une seule initialisation aléatoire, ainsi que les valeurs de μ estimées par MCC ($\mu = 0.001$) et MDistC ($\mu = 0.2$).



(a) Point de courbure maximale (croix rouge)

(b) Point Φ de distance minimale (croix rouge)

FIGURE 2.41 : Courbe de réponse estimée pour le mélange de la Figure 2.10

Les cartes d'abondance relatives à $\mu = 0.001$ et 0.2 sont présentées sur les Figures 2.42 et 2.43. Nous constatons que la valeur de μ générée par MDistC ne permet pas d'obtenir des résultats pertinents en ce qui concerne le démélange hyperspectral. À titre de comparaison, la Figure 2.44 représente un résultat de démélange lorsque $\mu = 0$.

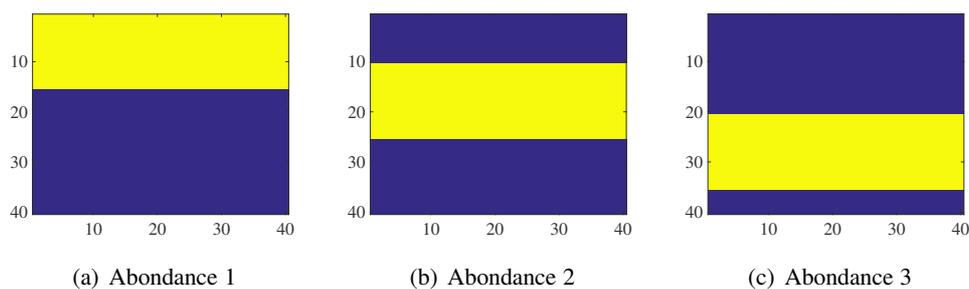


FIGURE 2.42 : Cartes d'abondance estimées pour le mélange de la Figure 2.10, pour $\mu = 0.001$ (estimé par MCC)

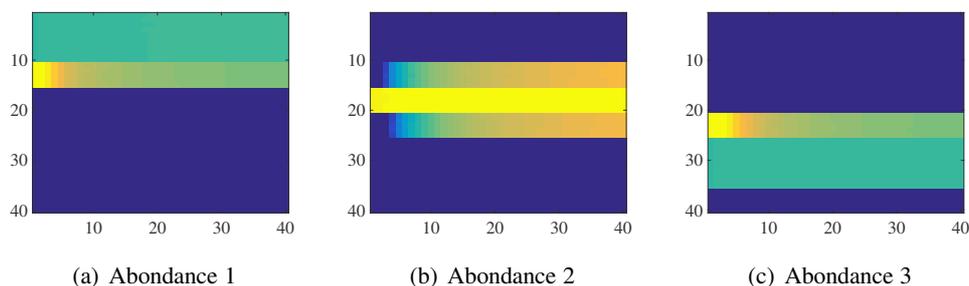


FIGURE 2.43 : Cartes d'abondance estimées pour le mélange de la Figure 2.10, pour $\mu = 0.2$ (estimé par MDistC)

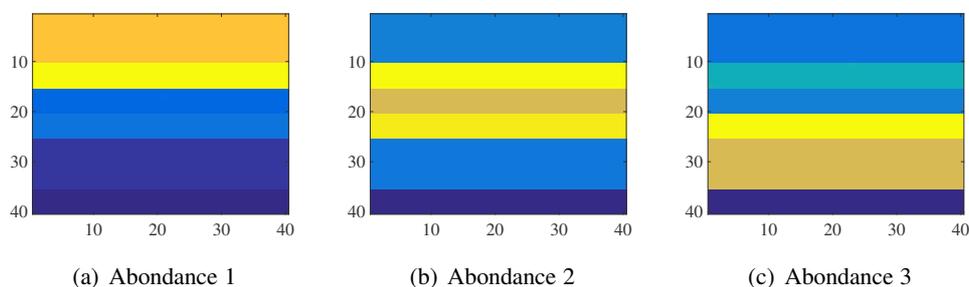
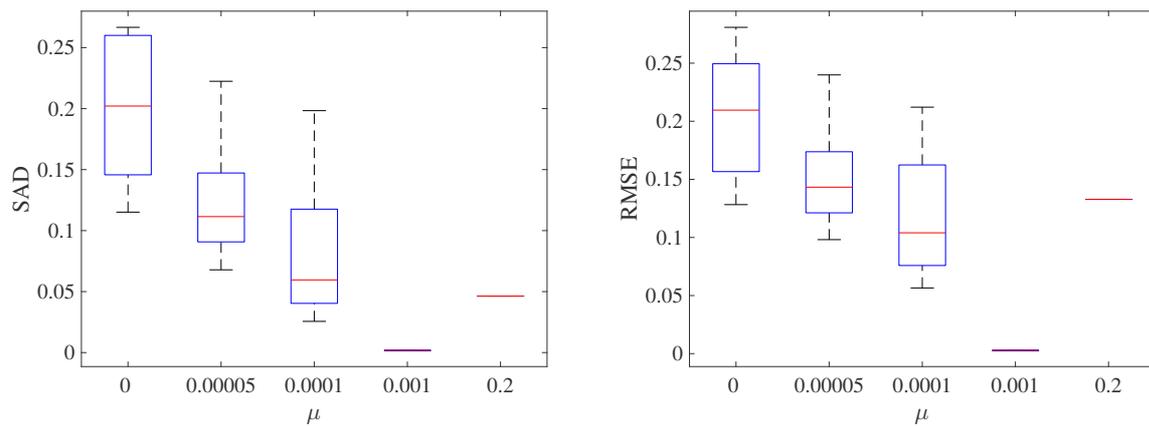


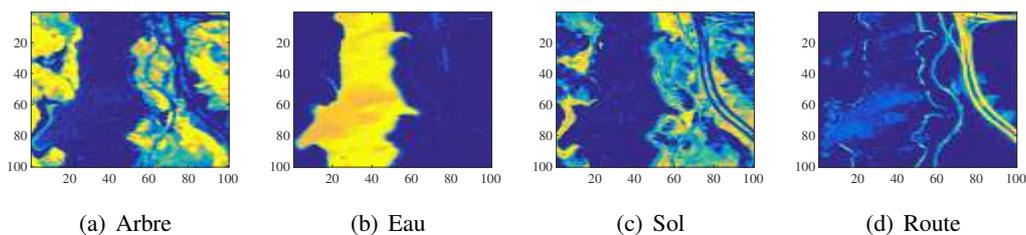
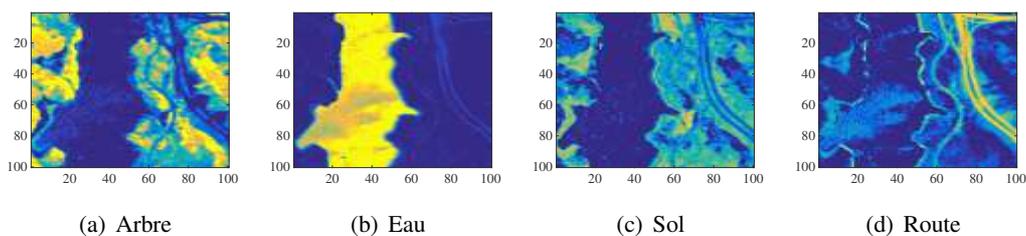
FIGURE 2.44 : Cartes d'abondance estimées pour le mélange de la Figure 2.10, pour $\mu = 0$

Les performances en termes de SAD et de RMSE sont présentées sur la Figure 2.45 pour 20 différentes initialisations (générées aléatoirement) et en fonction de plusieurs valeurs de μ . Nous constatons que la valeur de μ estimée par MCC permet à la fois de stabiliser et de minimiser les valeurs du SAD et du RMSE. *A contrario*, pour $\mu = 0.2$, les valeurs du SAD et du RMSE augmentent significativement, indiquant que, pour ce problème de démixage, MDistC n'est pas adapté.

FIGURE 2.45 : SAD et RMSE en fonction de μ , pour le mélange de la Figure 2.10

2.6.2.4 Évaluation des critères MCC et MDistC sur une image hyperspectrale réelle

Afin de tester notre approche sur des données réelles, nous avons utilisé l'image hyperspectrale satellitaire *Jasper Ridge*. Les paramètres ont été fixés de la manière suivante : $R = 4$, $\alpha = 0.99$, $\rho = 0.001$ et $N_{iter} = 300$. Les valeurs de μ ont été comprises entre 0 et 100. La Figure 2.46 montre les cartes d'abondance réelles. Les Figures 2.47 et 2.48 affichent respectivement les cartes d'abondance estimées avec le paramètre de régularisation généré par MCC et MDistC pour une initialisation aléatoire. De la même manière que dans les cas simulés, MDistC surestime la valeur de l'hyperparamètre, ce qui génère des cartes d'abondance et des *endmembers* (voir Figure 2.49) qui n'ont pas de sens physique ; les résultats deviennent donc difficilement interprétables. Notons que les *endmembers* sont ceux estimés à la dernière tranche de l'image hyperspectrale.

FIGURE 2.46 : Cartes d'abondance vraies (expertisées) pour l'image *Jasper Ridge*FIGURE 2.47 : Cartes d'abondance estimées pour l'image *Jasper Ridge*, pour $\mu = 0.018$ (estimé par MCC)

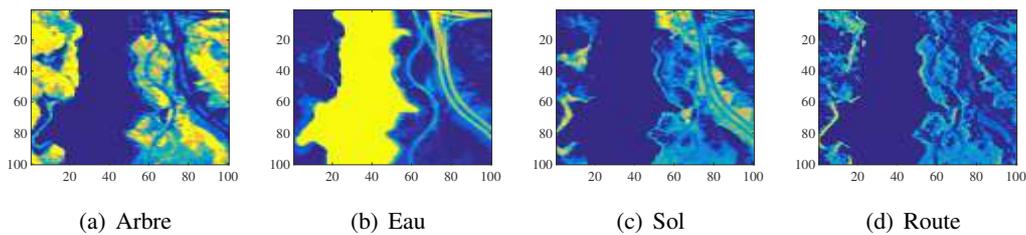


FIGURE 2.48 : Cartes d'abondance estimées pour l'image *Jasper Ridge*, pour $\mu = 0.2$ (estimé par MDistC)

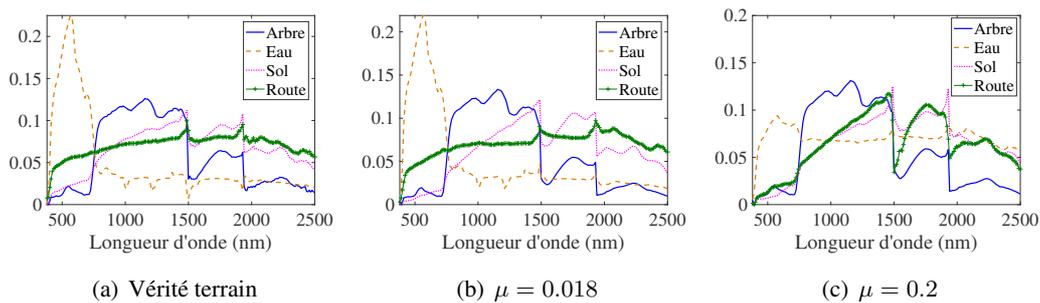


FIGURE 2.49 : Spectres des *endmembers* expertisés et estimés pour l'image *Jasper Ridge*

Pour étudier la robustesse des résultats face à 20 différentes initialisations aléatoires, les valeurs du SAD et du RMSE ont été calculées pour chacune d'entre elles et représentées sur la Figure 2.50. Nous remarquons que pour une valeur de $\mu = 0.018$, la solution est parfaitement régularisée et les estimations sont de bonne qualité.

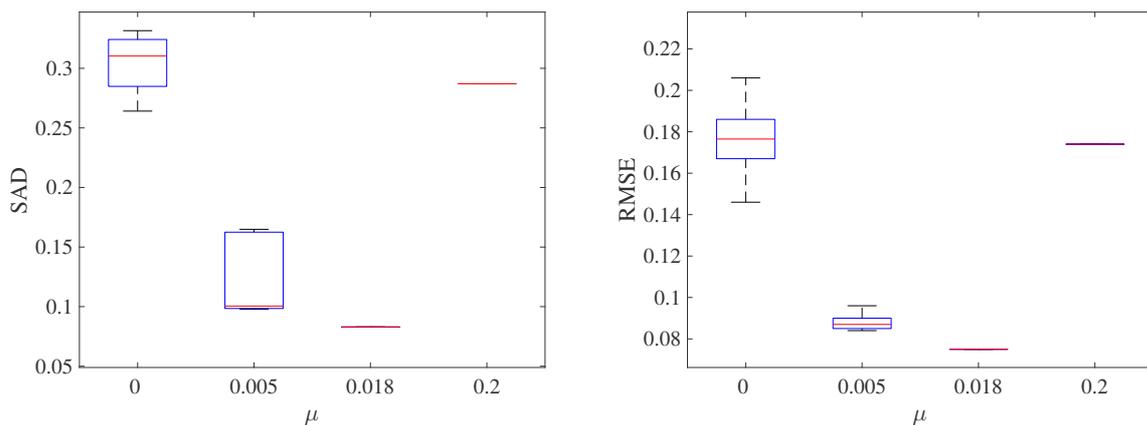


FIGURE 2.50 : SAD et RMSE en fonction de μ , pour l'image *Jasper Ridge*

D'après ces divers résultats, nous avons constaté que la mise en oeuvre efficace de l'algorithme OMDC-ADMM est conditionnée par le choix du paramètre de régularisation relatif à la contrainte de dispersion minimale ; cela est évidemment valable pour les algorithmes OMDC-MU, OMVC-MU, OMDC-Ne ainsi que OMVC-Ne. Les simulations qui ont été réalisées pour évaluer les performances de MCC et MDistC ont montré que MCC constitue une approche efficace pour estimer le paramètre de régularisation dans le cadre du démelange en-ligne d'images hyperspectrales. Notons, tout de même, que cette approche

n'offre qu'une aide pour l'estimation de l'hyperparamètre et que, dans certain cas, la valeur de μ doit être affinée manuellement pour obtenir de meilleurs résultats. Étant donné que l'application qui nous concerne est le traitement des pièces de bois, qui sont toutes de la même essence, une première idée pourrait être d'apprendre le paramètre optimal μ sur un échantillon typique et d'utiliser ce paramètre fixe pour le traitement en-ligne.

2.7 Conclusions

Dans ce chapitre, le problème de démélange en-ligne d'images hyperspectrales collectées par les systèmes industriels d'imagerie *pushbroom* a été traité. La contribution principale de ce travail est la dérivation de trois algorithmes : OMDC-ADMM, OMDC-Ne et OMVC-Ne, qui permettent le traitement rapide d'une image hyperspectrale tranche par tranche, tout en gardant une complexité de calcul constante. La principale caractéristique des approches proposées est la possibilité d'inclure un terme de régularisation de type dispersion/volume minimal, qui permet de régulariser la solution sans connaissances *a priori* des signatures spectrales. La mise en œuvre efficace de ces méthodes dépend du choix du paramètre de régularisation. C'est pourquoi, nous avons proposé une approche fondée sur le critère de courbure maximale, qui fournit une bonne indication quant à la valeur optimale de l'hyperparamètre. De plus, ces méthodes reposent sur une fonction objectif pondérée qui apporte une capacité de suivi des *endmembers* au cours de l'image. Des tests sur des données simulées ont permis de montrer que les trois approches proposées surpassent les méthodes basées sur les règles de mise à jour multiplicatives (*i.e.* OMDC-MU et OMVC-MU) en ce qui concerne la vitesse de convergence. Par ailleurs l'influence des divers paramètres a été étudiée. Enfin, des expériences sur des images hyperspectrales réelles ont révélé que OMDC-ADMM permet de réduire significativement le temps de traitement d'une image, comparé à OMDC-Ne et OMVC-Ne et à son homologue *batch* BMDC-ADMM. De plus, OMDC-ADMM est capable de suivre la variabilité spectrale des *endmembers* entre chaque tranche de l'image, ce qui est un élément essentiel pour l'analyse des données en temps réel.

Chapitre 3

Démélange hyperspectral semi-supervisé en-ligne avec suivi du nombre de *endmembers*

3.1 Introduction

Les algorithmes de démelange aveugle en-ligne proposés dans le Chapitre 2 fonctionnent bien dans la pratique si le rang de la décomposition, c'est-à-dire le nombre de *endmembers* ne varie pas entre les différents échantillons. En effet, la condition d'identifiabilité impose que pour chaque tranche de l'image $\text{rang}(\tilde{\mathbf{S}}) = \text{rang}(\tilde{\mathbf{A}}) = R$. C'est une hypothèse forte qui n'est pas toujours satisfaite dans certaines applications où une ou plusieurs sources peuvent disparaître et/ou réapparaître au cours du temps. Ainsi, l'objectif de ce chapitre est d'étudier le cas où le rang de la décomposition n'est pas strictement préservé d'une tranche à l'autre, *i.e.* $\text{rang}(\tilde{\mathbf{S}}) \leq R$. La Figure 3.1 illustre deux scénarios possibles pour l'évolution du rang au cours des échantillons k d'une image, dans le cas où $R = 3$.

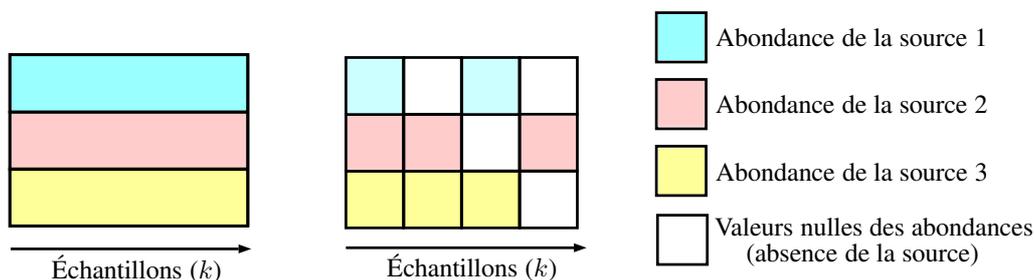


FIGURE 3.1 : Illustration des matrices d'abondance. Deux scénarios sont considérés : à gauche, rang constant au cours du traitement *i.e.* $\text{rang}(\tilde{\mathbf{S}}^{(k)}) = R$. À droite : rang variable au cours du traitement *i.e.* $\text{rang}(\tilde{\mathbf{S}}^{(k)}) \leq R$.

Dans ce contexte, le modèle de mélange en-ligne pour une tranche k de l'image peut s'écrire $\tilde{\mathbf{X}}^{(k)} = \mathbf{B}\tilde{\mathbf{A}}^{(k)}$, où la matrice \mathbf{B} regroupe sur ses colonnes tous les *endmembers* présents au cours du traitement de l'image. Si on suppose la matrice \mathbf{B} connue, le problème de poursuite de rang est donc équivalent à

$$\begin{aligned} & \underset{\tilde{\mathbf{A}}^{(k)} \geq \mathbf{0}}{\text{minimiser}} \frac{1}{2} \left\| \tilde{\mathbf{X}}^{(k)} - \mathbf{B}\tilde{\mathbf{A}}^{(k)} \right\|, \\ & \text{sous contrainte} \left\| \tilde{\mathbf{A}}^{(k)} \right\|_{2,0} \leq R. \end{aligned} \quad (3.1)$$

Dans le cas hors-ligne, les caractéristiques ou les connaissances *a priori* relatives aux données hyperspectrales sont souvent intégrées dans le modèle. Récemment, des approches semi-supervisées, utilisant une bibliothèque hyperspectrale comme connaissance *a priori*, ont été proposées pour traiter le problème de génération de *endmembers* « virtuels », causé par l'absence de pixels purs dans l'image. Elles visent à n'utiliser que quelques spectres dans une bibliothèque hyperspectrale donnée (potentiellement très volumineuse) pour modéliser chaque pixel mélangé de l'image hyperspectrale [Iordache 11]. Étant donné que le nombre de *endmembers* dans la scène hyperspectrale est généralement beaucoup plus petit que le nombre de spectres dans la bibliothèque, ces approches entraînent souvent une solution parcimonieuse [Bioucas-Dias 10, Iordache 13, Tang 14]. En d'autres termes, chaque pixel d'une image s'exprime par une combinaison linéaire d'un petit nombre de signatures hyperspectrales de la bibliothèque et les abondances sont obtenues en minimisant une fonction objectif, contenant un terme d'attache aux données et un terme de régularisation imposant la parcimonie, généralement une norme $\ell_{1,1}$ ou $\ell_{2,1}$ [Nie 10, Tang 14, Wang 16]. La norme $\ell_{2,1}$ est une relaxation convexe de la norme $\ell_{2,0}$, qui intervient dans l'expression (3.1). Ainsi, ces méthodes ne nécessitent pas d'estimer les *endmembers* mais uniquement de les sélectionner et d'estimer les abondances. Dans [Li 16], il a été proposé d'estimer conjointement les *endmembers* et les abondances en utilisant deux termes de régularisation : un premier terme qui pénalise la distance entre les *endmembers* à estimer et une référence obtenue en amont par un algorithme de démélange aveugle (*e.g.* VCA), et une norme $\ell_{2,1}$ imposée sur les abondances permettant de sélectionner uniquement les *endmembers* actifs.

Ce chapitre vise à étudier l'intérêt de ce type d'approches dans un contexte en-ligne. Ainsi, nous proposons un nouvel algorithme de démélange en-ligne, fondé sur l'algorithme ADMM, spécialement conçu pour suivre l'évolution du nombre de *endmembers* entre les tranches acquises par l'imageur *pushbroom*. Les contributions originales de cet algorithme par rapport à nos précédents travaux (notamment OMDC-ADMM) et aux approches en-ligne de l'état de l'art sont les suivantes :

1. Prise en compte d'une évolution lente des *endmembers* $\tilde{\mathbf{S}}^{(k)}$ dans un voisinage de \mathbf{B} . En d'autres termes, la contrainte de dispersion minimale (utilisée par OMDC-ADMM) est remplacée par un terme de régularisation qui consiste à « pousser » les signatures spectrales estimées à chaque tranche à être proches d'une bibliothèque hyperspectrale (ou référence) \mathbf{B} . Ainsi, le démélange hyperspectral en-ligne est considéré comme semi-supervisé, puisque des connaissances *a priori* sur les *endmembers* sont intégrées au modèle.
2. L'utilisation conjointe des normes $\ell_{2,1}$ et $\ell_{1,1}$ appliquée sur la matrice des abondances qui favorise respectivement la parcimonie par ligne (rang faible) et la parcimonie spatiale sur les lignes actives (pour la reconstruction de supports constants par morceaux sur un fond nul).

Dans ce qui suit, nous appellerons ce nouvel algorithme OSS-ADMM (pour *On-line Semi-Supervised ADMM*). Ce chapitre est organisé de la manière suivante : la Section 2 décrit la méthode en-ligne pro-

posée : sa fonction objectif, les mises à jour des paramètres d'intérêt ainsi que son implantation algorithmique. Des expériences sur des images hyperspectrales simulées sont données dans la Section 3 et permettent d'illustrer et d'évaluer les performances de OSS-ADMM. Enfin, OSS-ADMM est validé sur une application nécessitant le démixage en-ligne d'une suite d'images hyperspectrales réelles. Nous concluons le chapitre dans la Section 4.

3.2 Algorithme OSS-ADMM

3.2.1 Fonction objectif

La fonction objectif sur laquelle repose l'algorithme OSS-ADMM peut s'écrire de la manière suivante :

$$\begin{aligned} \mathcal{J}_{sparse}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}) &= \alpha \frac{1}{2} \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2 + (1 - \alpha) \frac{1}{2} \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2 \\ &+ v \left\| \tilde{\mathbf{A}}^{(k+1)} \right\|_{2,1} + \gamma \left\| \tilde{\mathbf{A}}^{(k+1)} \right\|_{1,1} + \frac{\omega}{2} \left\| \mathbf{B} - \tilde{\mathbf{S}}^{(k+1)} \right\|_F^2, \end{aligned} \quad (3.2)$$

où α symbolise le facteur d'oubli. La fonction objectif (3.2) présente cinq termes : les deux premiers termes caractérisent l'adéquation aux données des k premiers échantillons et du nouvel échantillon respectivement ; le troisième terme est la norme $\ell_{2,1}$ qui favorise la parcimonie sur les lignes de la matrice $\tilde{\mathbf{A}}^{(k+1)}$ et permet de sélectionner uniquement les *endmembers* actifs. La norme $\ell_{1,1}$ utilisée dans le quatrième terme renforce la parcimonie spatiale sur les lignes actives de $\tilde{\mathbf{A}}^{(k+1)}$. Les normes $\ell_{1,1}$ et $\ell_{2,1}$ sont des cas particuliers de la norme $\ell_{q,p}$ définie par :

$$\left\| \tilde{\mathbf{A}} \right\|_{q,p} = \left(\sum_{r=1}^R \left\| \tilde{\mathbf{a}}^r \right\|_q^p \right)^{\frac{1}{p}},$$

où $\tilde{\mathbf{a}}^r$ représente la $r^{\text{ème}}$ ligne de $\tilde{\mathbf{A}}$. Le dernier terme force les *endmembers* à être proches d'une bibliothèque hyperspectrale \mathbf{B} . La matrice \mathbf{B} permet également de lever l'indétermination d'ordre des *endmembers* entre chaque tranche, qui peut intervenir lors de l'apparition ou la disparition d'une ou plusieurs sources au cours de l'image. Elle regroupe l'ensemble des R *endmembers* de référence. Elle peut être générée de différentes façons : par apprentissage ou grâce à la connaissance d'une vérité terrain. Dans notre contexte, *i.e.* le traitement à la volée de produits appartenant à la même famille, \mathbf{B} peut être appris sur un échantillon typique et utilisé comme référence pour l'ensemble de la chaîne de traitement. Dans le cas où les spectres regroupés dans la matrice \mathbf{B} ne sont pas présents simultanément dans une tranche, les contraintes de parcimonie imposées sur la matrice des abondances forcent à zéro les lignes qui correspondent aux *endmembers* non-actifs. v , γ et ω sont les hyperparamètres liés aux différentes pénalités. Afin de minimiser la fonction objectif (3.2), nous adoptons une approche ADMM ; nous introduisons deux variables auxiliaires $\tilde{\mathbf{U}}$ et $\tilde{\mathbf{V}}$ et considérons le problème d'optimisation équivalent suivant :

$$\begin{aligned} \min_{\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}, \tilde{\mathbf{V}}^{(k+1)}, \tilde{\mathbf{U}}^{(k+1)}} \mathcal{J}_{sparse}^{(k+1)}(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}) + \mathbb{I}_{R_+}(\tilde{\mathbf{V}}^{(k+1)}) + \mathbb{I}_{R_+}(\tilde{\mathbf{U}}^{(k+1)}), \\ \text{sous contraintes } \tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{U}}^{(k+1)} \text{ et } \tilde{\mathbf{A}}^{(k+1)} = \tilde{\mathbf{V}}^{(k+1)}. \end{aligned} \quad (3.3)$$

Le Lagrangien augmenté \mathcal{L} correspondant au problème d'optimisation (3.3) est donné par :

$$\begin{aligned} \mathcal{L} \left(\tilde{\mathbf{A}}^{(k+1)}, \tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{V}}^{(k+1)}, \tilde{\mathbf{U}}^{(k+1)}, \tilde{\mathbf{\Pi}}^{(k+1)}, \tilde{\mathbf{\Lambda}}^{(k+1)} \right) \\ = \mathcal{J}_{sparse}^{(k+1)} \left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} \right) + \frac{\rho}{2} \left\| \tilde{\mathbf{A}}^{(k+1)} - \tilde{\mathbf{V}}^{(k+1)} + \tilde{\mathbf{\Pi}}^{(k+1)} \right\|_F^2 - \frac{\rho}{2} \left\| \tilde{\mathbf{\Pi}}^{(k+1)} \right\|_F^2 \\ + \frac{\rho}{2} \left\| \tilde{\mathbf{S}}^{(k+1)} - \tilde{\mathbf{U}}^{(k+1)} + \tilde{\mathbf{\Lambda}}^{(k+1)} \right\|_F^2 - \frac{\rho}{2} \left\| \tilde{\mathbf{\Lambda}}^{(k+1)} \right\|_F^2 + \mathbb{I}_{\mathbb{R}_+} \left(\tilde{\mathbf{V}}^{(k+1)} \right) + \mathbb{I}_{\mathbb{R}_+} \left(\tilde{\mathbf{U}}^{(k+1)} \right). \end{aligned} \quad (3.4)$$

De la même manière que pour OMDC-ADMM, le paramètre ρ contrôle la vitesse de convergence de la méthode. $\tilde{\mathbf{\Pi}}^{(k+1)}$ et $\tilde{\mathbf{\Lambda}}^{(k+1)}$ sont les versions mises à l'échelle des variables duales correspondant aux contraintes d'égalité $\tilde{\mathbf{A}}^{(k+1)} = \tilde{\mathbf{V}}^{(k+1)}$ et $\tilde{\mathbf{S}}^{(k+1)} = \tilde{\mathbf{U}}^{(k+1)}$ respectivement.

3.2.2 Mises à jour des paramètres d'intérêt

Nous minimisons d'abord le Lagrangien augmenté (3.4) par rapport aux paramètres $\left(\tilde{\mathbf{A}}^{(k+1)}, \tilde{\mathbf{V}}^{(k+1)} \right)$ et $\left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{U}}^{(k+1)} \right)$, puis nous mettons à jour les variables duales $\tilde{\mathbf{\Pi}}^{(k+1)}$ et $\tilde{\mathbf{\Lambda}}^{(k+1)}$. Les mises à jour des paramètres à la tranche $k+1$ s'expriment comme suit :

$$\begin{aligned} \tilde{\mathbf{A}}^{(k+1)} &= \left((1-\alpha)\tilde{\mathbf{S}}^{(k)T}\tilde{\mathbf{S}}^{(k)} + \rho\mathbf{I} + 2v\tilde{\mathbf{H}}^{(k)} \right)^{-1} \left((1-\alpha)\tilde{\mathbf{S}}^{(k)T}\tilde{\mathbf{X}}^{(k+1)} + \rho \left(\tilde{\mathbf{V}}^{(k)} - \tilde{\mathbf{\Pi}}^{(k)} \right) - \gamma\mathbf{1} \right), \\ \tilde{\mathbf{V}}^{(k+1)} &= \max \left(\mathbf{0}, \tilde{\mathbf{A}}^{(k+1)} + \tilde{\mathbf{\Pi}}^{(k)} \right), \\ \tilde{\mathbf{\Pi}}^{(k+1)} &= \tilde{\mathbf{\Pi}}^{(k)} + \tilde{\mathbf{A}}^{(k+1)} - \tilde{\mathbf{V}}^{(k+1)}, \\ \tilde{\mathbf{S}}^{(k+1)} &= \left(\mathbf{N}^{(k+1)} + \rho \left(\tilde{\mathbf{U}}^{(k)} - \tilde{\mathbf{\Lambda}}^{(k)} \right) + \omega\mathbf{B} \right) \left(\mathbf{M}^{(k+1)} + \rho\mathbf{I} + \omega\mathbf{I} \right)^{-1}, \\ \tilde{\mathbf{U}}^{(k+1)} &= \max \left(\mathbf{0}, \tilde{\mathbf{S}}^{(k+1)} + \tilde{\mathbf{\Lambda}}^{(k)} \right), \\ \tilde{\mathbf{\Lambda}}^{(k+1)} &= \tilde{\mathbf{\Lambda}}^{(k)} + \tilde{\mathbf{S}}^{(k+1)} - \tilde{\mathbf{U}}^{(k+1)}, \end{aligned}$$

où,

$$\begin{aligned} \mathbf{N}^{(k+1)} &= \alpha\mathbf{N}^{(k)} + (1-\alpha)\tilde{\mathbf{X}}^{(k+1)}\tilde{\mathbf{A}}^{(k+1)T}, \\ \mathbf{M}^{(k+1)} &= \alpha\mathbf{M}^{(k)} + (1-\alpha)\tilde{\mathbf{A}}^{(k+1)}\tilde{\mathbf{A}}^{(k+1)T}. \end{aligned}$$

La norme $\ell_{2,1}$ peut s'écrire sous la forme matricielle suivante : $\left\| \tilde{\mathbf{A}} \right\|_{2,1} = 2 \text{trace} \left(\tilde{\mathbf{A}}^T \tilde{\mathbf{H}} \tilde{\mathbf{A}} \right)$ [Han 17]. $\tilde{\mathbf{H}}$ est une matrice diagonale ayant comme élément diagonal

$$\tilde{\mathbf{H}}(r, r) = \frac{1}{\|\tilde{\mathbf{a}}^r\|_2 + \delta}, \quad (3.5)$$

où δ est un nombre positif et proche de zéro, permettant d'éviter les divisions par zéro.

3.2.3 Implantation de l'algorithme OSS-ADMM et complexité de calcul

L'Algorithme 4 résume OSS-ADMM. Sa structure est identique à l'algorithme OMDC-ADMM, *i.e.* il présente deux boucles principales : une boucle externe qui produit des estimations de l'ensemble des

paramètres pour chaque nouvel échantillon et une boucle interne qui permet d'affiner itérativement les estimations en utilisant un nombre d'itérations fixe (N_{iter}).

Algorithme 4 OSS-ADMM

Entrées : $\mathbf{X}, \mathbf{B}, R, \alpha, \nu, \gamma, \omega, \rho, \delta, N_{iter}$;

Initialisation: $k = 0, \mathbf{N} = \text{zeros}(L, R), \mathbf{M} = \text{zeros}(R, R), \tilde{\mathbf{S}} = \text{rand}(L, R), \tilde{\mathbf{V}} = \text{zeros}(R, P), \tilde{\mathbf{U}} = \text{zeros}(L, R), \tilde{\mathbf{\Pi}} = \text{zeros}(R, P), \tilde{\mathbf{\Lambda}} = \text{zeros}(L, R), \tilde{\mathbf{H}} = \text{eye}(R, R), \mathbf{A} = [], \mathbf{S} = []$;

Sorties : \mathbf{A}, \mathbf{S} ;

Tant que Nouvel échantillon $k + 1$ disponible **faire**

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

$$t = 1;$$

Tant que $t < N_{iter}$ **faire**

$$\tilde{\mathbf{A}} = \left((1 - \alpha) \tilde{\mathbf{S}}^T \tilde{\mathbf{S}} + \rho \mathbf{I} + 2\nu \tilde{\mathbf{H}} \right)^{-1} \left((1 - \alpha) \tilde{\mathbf{S}}^T \tilde{\mathbf{X}} + \rho (\tilde{\mathbf{V}} - \tilde{\mathbf{\Pi}}) - \gamma \mathbf{1} \right);$$

$$\tilde{\mathbf{H}} = \begin{bmatrix} \frac{1}{\|\tilde{\mathbf{a}}^1\|_2 + \delta} & & \\ & \ddots & \\ & & \frac{1}{\|\tilde{\mathbf{a}}^R\|_2 + \delta} \end{bmatrix};$$

$$\tilde{\mathbf{V}} = \max(\mathbf{0}, \tilde{\mathbf{A}} + \tilde{\mathbf{\Pi}});$$

$$\tilde{\mathbf{\Pi}} \leftarrow \tilde{\mathbf{\Pi}} + \tilde{\mathbf{A}} - \tilde{\mathbf{V}};$$

$$\tilde{\mathbf{N}} = \alpha \mathbf{N} + (1 - \alpha) (\tilde{\mathbf{X}} \tilde{\mathbf{A}}^T);$$

$$\tilde{\mathbf{M}} = \alpha \mathbf{M} + (1 - \alpha) (\tilde{\mathbf{A}} \tilde{\mathbf{A}}^T);$$

$$\tilde{\mathbf{S}} = \left(\tilde{\mathbf{N}} + \rho (\tilde{\mathbf{U}} - \tilde{\mathbf{\Lambda}}) + \omega \mathbf{B} \right) (\tilde{\mathbf{M}} + \rho \mathbf{I} + \omega \mathbf{I})^{-1};$$

$$\tilde{\mathbf{U}} = \max(\mathbf{0}, \tilde{\mathbf{S}} + \tilde{\mathbf{\Lambda}});$$

$$\tilde{\mathbf{\Lambda}} \leftarrow \tilde{\mathbf{\Lambda}} + \tilde{\mathbf{S}} - \tilde{\mathbf{U}};$$

$$t \leftarrow t + 1;$$

Fin tant que

$$\mathbf{N} = \tilde{\mathbf{N}}; \mathbf{M} = \tilde{\mathbf{M}};$$

$$\mathbf{A} \leftarrow [\mathbf{A}, \tilde{\mathbf{A}}]; \mathbf{S} \leftarrow [\mathbf{S}, \tilde{\mathbf{S}}];$$

Fin tant que

La complexité de calcul est du même ordre que celle de la méthode OMDC-ADMM (on néglige les opérations de multiplication qui interviennent dans les produits $2\nu \tilde{\mathbf{H}}$, $\omega \mathbf{B}$ et $\omega \mathbf{I}$). Ainsi, pour une tranche k de l'image hyperspectrale de dimension $L \times P$, et pour un nombre total d'itérations N_{iter} , la complexité de calcul de OSS-ADMM est de l'ordre de : $2N_{iter}(RPL + (L + P)R^2 + R^3)$ où R est le nombre total

de *endmembers*.

3.3 Résultats expérimentaux

Dans cette section, nous avons mené plusieurs expériences sur des images hyperspectrales simulées et réelles. Toutes ces expériences ont été réalisées à l'aide du logiciel Matlab (R2016a) sur un Macbook Pro avec processeur quadricœur à 2,7 GHz et 16 Go de RAM. Ces expériences ont pour objectifs de :

1. Évaluer l'influence des hyperparamètres v , γ et ω .
2. Étudier l'efficacité de OSS-ADMM pour le démélange hyperspectral en-ligne et le suivi du rang (nombre de sources) au cours du traitement sur plusieurs images hyperspectrales simulées.
3. Valider les performances de OSS-ADMM dans le cadre d'une application industrielle dont le but est de cartographier la quantité d'eau libre et liée au sein de produits agro-alimentaires, au cours du temps.

3.3.1 Influence des hyperparamètres v , γ et ω

Afin d'étudier l'influence des paramètres v et γ , des expériences numériques ont été menées sur une image hyperspectrale simulée de dimension $119 \times 40 \times 380$, consistant en un mélange non-négatif de trois *endmembers*, représentés sur la Figure 3.2. Ces derniers ne varient pas au cours de l'image et le rang de la décomposition est identique pour chaque tranche ($= 3$). Le premier objectif a été d'étudier la capacité des contraintes de parcimonie à désactiver les *endmembers* non présents dans l'image, dans le cas où l'on surestime la valeur de R . La première étape a été de construire la bibliothèque hyperspectrale \mathbf{B} ; pour ce faire, nous avons utilisé les vrais *endmembers* de l'image, corrompus par un bruit avec un SNR = 30 dB (afin de s'éloigner légèrement de la vérité terrain) et une quatrième source (issue de l'image *Jasper Ridge*), comme illustré sur la Figure 3.3.

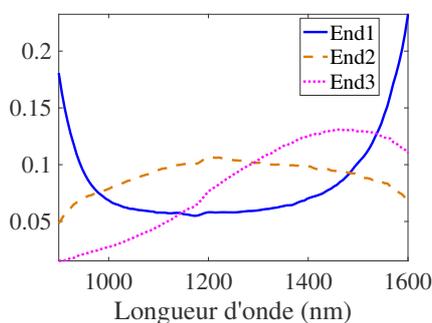


FIGURE 3.2 : Vérité terrain : spectres des trois *endmembers* (End)

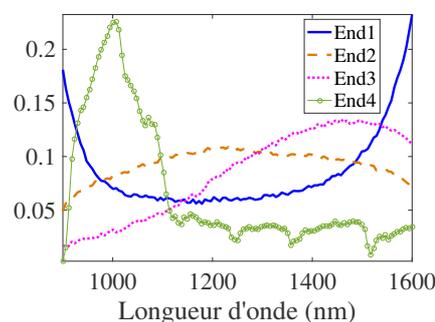


FIGURE 3.3 : Bibliothèque hyperspectrale \mathbf{B} composée de quatre *endmembers*

Les *endmembers* notés End1, End2 et End3 de la Figure 3.2 sont présents simultanément à chaque tranche de l'image hyperspectrale, tandis que End4 de la Figure 3.3 correspond à la source inactive. Pour commencer, nous avons étudié l'influence du couple (v, γ) (hyperparamètres relatifs à la norme $\ell_{2,1}$ et $\ell_{1,1}$,

respectivement) pour la désactivation de End4. Afin de représenter cette désactivation, nous avons calculé pour chaque échantillon de l'image, la quantité $\sum_{p=1}^P \tilde{a}_{rp}$, avec $r = 4$; cela revient à sommer les éléments correspondant à la quatrième ligne de la matrice des abondances (la ligne relative à End4). La valeur de v a été fixée à 0.002, de manière à ce que, elle seule, puisse désactiver End4. Les paramètres de OSS-ADMM ont été fixés de la manière suivante : $R = 4$, $\alpha = 0.9$, $\omega = 1$, $\rho = 0.001$, $\delta = 10^{-15}$ et $N_{iter} = 200$. La Figure 3.4 affiche les résultats obtenus pour les couples (v, γ) suivants : $(0.002, 0)$, $(0.002, 0.0001)$, $(0.002, 0.0002)$, $(0.002, 0.0005)$ et $(0.002, 0.001)$. Nous constatons clairement qu'à mesure que γ augmente, la désactivation de End4 s'effectue beaucoup plus rapidement, et lorsque que $\gamma = 0.001$, la désactivation est presque instantanée. Précisons que si l'on augmente v tout en fixant $\gamma = 0$, nous observons un comportement similaire à la Figure 3.4. Néanmoins, les résultats du démixage hyperspectral se dégradent significativement lorsque la valeur de v est trop élevée. Cela s'explique principalement par le fait que l'estimateur de la trace relative à la norme $\ell_{2,1}$ est biaisé par le paramètre δ , qui intervient dans l'expression (3.5). Il semble donc que l'utilisation de la norme $\ell_{1,1}$ permet d'accélérer la mise à zéro des éléments sur toute une ligne et donc d'accélérer la convergence de l'algorithme, tout en préservant de bonnes performances.

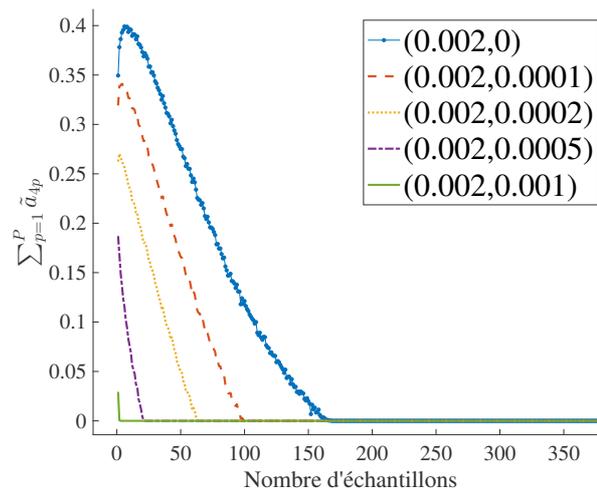


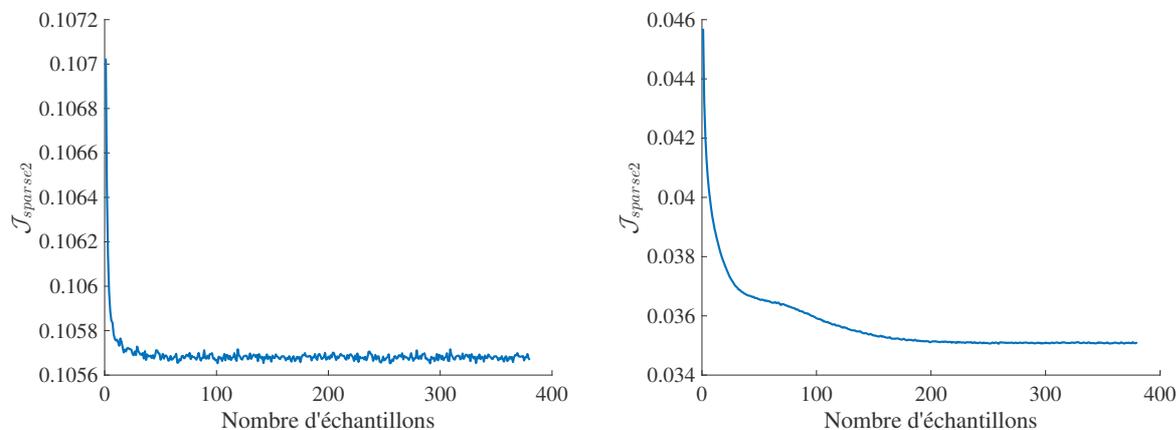
FIGURE 3.4 : Évolution de la désactivation de End4 en fonction du nombre d'échantillons, pour différents couples (v, γ)

La seconde expérience a consisté à analyser la vitesse de convergence du critère

$$\mathcal{J}_{sparse2}(\tilde{\mathbf{S}}^{(k)}, \tilde{\mathbf{A}}^{(k)}) = \left\| \tilde{\mathbf{X}}^{(k)} - \tilde{\mathbf{S}}^{(k)} \tilde{\mathbf{A}}^{(k)} \right\|_F^2 + v \left\| \tilde{\mathbf{A}}^{(k)} \right\|_{2,1} + \gamma \left\| \tilde{\mathbf{A}}^{(k)} \right\|_{1,1}$$

pour chaque échantillon k de l'image et pour deux couples (v, γ) , le premier où $v > \gamma$ et le second où $v < \gamma$. La Figure 3.5 illustre la vitesse de convergence du critère pour les couples $(0.0001, 0.002)$ et $(0.002, 0.0001)$. D'après la Figure 3.5, il est intéressant de souligner que l'utilisation d'une norme $\ell_{1,1}$ plus forte qu'une norme $\ell_{2,1}$ permet à l'algorithme OSS-ADMM de converger plus rapidement ; c'est un phénomène pour lequel nous n'avons pas encore trouvé une explication complètement satisfaisante,

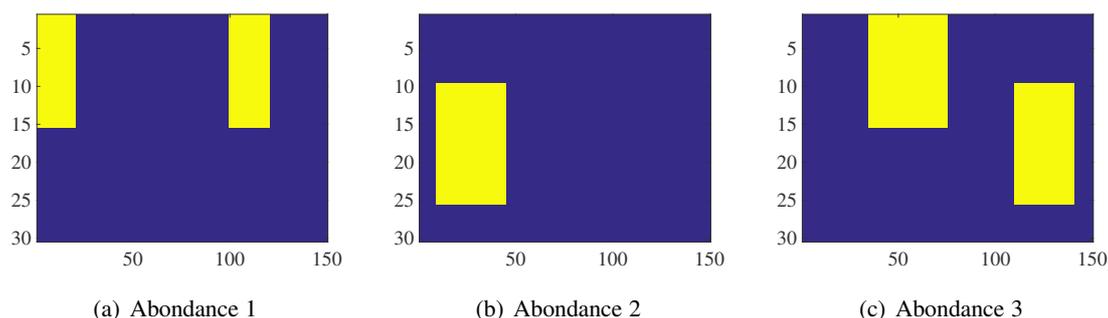
mais cette analyse, bien que sommaire, nous pousse à choisir par la suite une valeur de γ plus élevée que la valeur de ν .



(a) Vitesse de convergence pour $\nu = 0.0001$ et $\gamma = 0.002$ (b) Vitesse de convergence pour $\nu = 0.002$ et $\gamma = 0.0001$

FIGURE 3.5 : Vitesse de convergence du critère $\mathcal{J}_{sparse2}$ en fonction du nombre d'échantillons, pour différents couples (ν, γ)

L'utilisation de la norme $\ell_{1,1}$ permet d'accélérer la mise à zéro des coefficients d'abondance relatifs à la source inactive, mais également d'imposer de la parcimonie spatiale sur les lignes actives de la matrice des abondances ; cela est un point essentiel pour la reconstruction de supports constants par morceaux sur un fond nul. Pour illustrer nos propos, nous avons simulé une image hyperspectrale de dimension $119 \times 30 \times 150$ dont les cartes d'abondance, représentées sur la Figure 3.6, sont composées de trois *endmembers* (identiques à ceux représentés sur la Figure 3.2). Cette fois ci, le rang de la décomposition n'est plus le même entre les tranche de l'image, et il évolue comme illustré sur la Figure 3.7.



(a) Abondance 1

(b) Abondance 2

(c) Abondance 3

FIGURE 3.6 : Cartes d'abondance simulées ayant des supports constant par morceaux

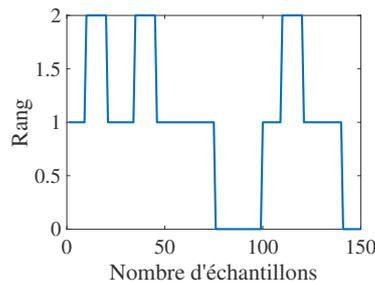
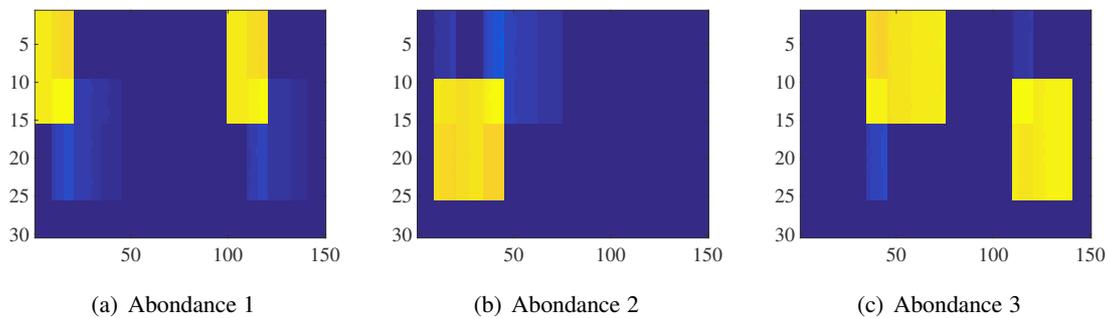
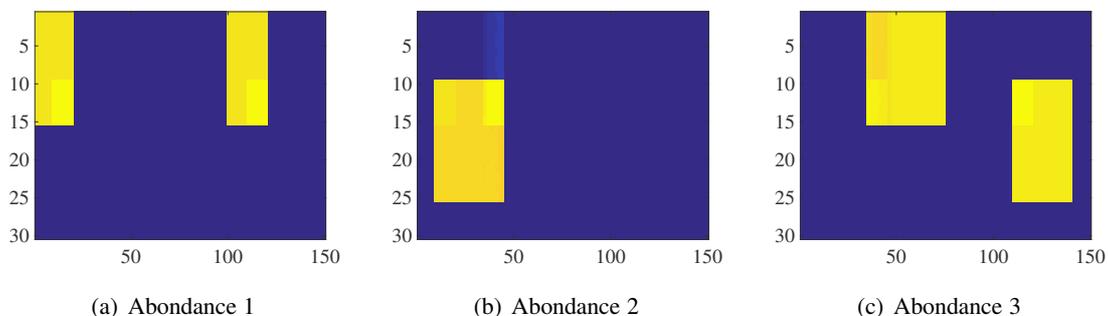


FIGURE 3.7 : Rang de la décomposition relatif au mélange de la Figure 3.6

En considérant le cas où les *endmembers* ne sont pas surestimés, *i.e.* $R = 3$, et en imposant $\nu = 0.002$ et $\gamma = 0$, les cartes d'abondance générées par OSS-ADMM sont représentées sur la Figure 3.8.

FIGURE 3.8 : Cartes d'abondance générées par OSS-ADMM pour le mélange de la Figure 3.6, avec $\nu = 0.002$ et $\gamma = 0$

On constate que l'utilisation de la pénalité de norme $\ell_{2,1}$ à elle seule, induit des artefacts, notamment lorsque les supports se chevauchent. Cela s'explique par le fait que la norme $\ell_{2,1}$ ne permet pas d'imposer de la parcimonie spatiale et structurée sur les lignes actives de la matrice des abondances. La Figure 3.9 illustre les cartes d'abondance générées par OSS-ADMM pour $\nu = 0.002$ et pour une valeur de γ plus élevée (0.009). Nous remarquons que la norme $\ell_{1,1}$ permet d'améliorer significativement les résultats. En revanche, l'utilisation conjointe des paramètres ν et γ doit se faire avec précaution : d'expérience, une valeur faible de ν permet d'obtenir de meilleurs résultats qu'une valeur forte.

FIGURE 3.9 : Cartes d'abondance générées par OSS-ADMM pour le mélange de la Figure 3.6, avec $\nu = 0.002$ et $\gamma = 0.009$

Pour finir, nous avons étudié l'influence de l'hyperparamètre ω , relatif à la contrainte de distance à une

bibliothèque hyperspectrale. Comme nous l'avons mentionné précédemment, l'utilisation de ω a deux principaux rôles : le premier est de forcer les *endmembers* estimés à chaque tranche de l'image par OSS-ADMM à être proches des éléments de la bibliothèque et le second, d'éviter les permutations d'ordre des *endmembers* au cours du traitement de l'image. Nous avons fait varier ω de 0 à 10 avec un pas variable, tout en fixant $\nu = 0.0001$ et $\gamma = 0.002$. Afin d'analyser l'influence de ω , nous avons calculé, pour chacune de ses valeurs, le RMSE entre les vraies abondances et les estimées. Les résultats sont affichés sur la Figure 3.10 sous forme de *bargraph*.

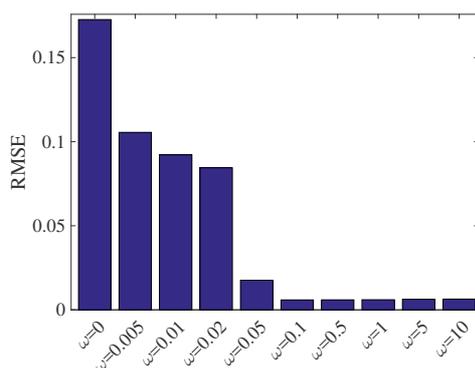


FIGURE 3.10 : Évolution du RMSE en fonction de ω , pour le mélange de la Figure 3.6

À mesure que la valeur de ω augmente, le RMSE diminue, signe que les abondances sont de mieux en mieux estimées par l'algorithme. En effet, lorsque $\omega = 0$, les termes de parcimonie à eux seuls ne permettent pas de régulariser suffisamment la solution et c'est d'ailleurs ce que l'on peut observer sur la Figure 3.11 qui illustre un des résultats de démélange hyperspectral réalisé par OSS-ADMM. Nous pouvons constater, sur les cartes d'abondance, de nombreuses permutations des *endmembers* au cours du traitement. D'après la Figure 3.10, il semble judicieux d'adopter une valeur de $\omega > 0.1$.

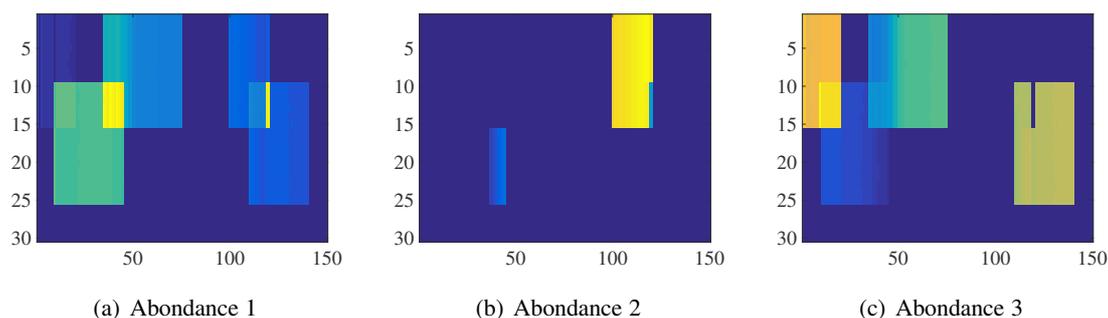


FIGURE 3.11 : Cartes d'abondance générées par OSS-ADMM pour le mélange de la Figure 3.6, avec $\omega = 0$

3.3.2 OSS-ADMM appliqué aux images hyperspectrales simulées

Pour commencer, nous avons testé les performances de notre algorithme sur quatre images hyperspectrales simulées, où le nombre de *endmembers* varie d'une tranche à l'autre. L'objectif est d'évaluer la capacité de l'approche proposée à estimer correctement les *endmembers* et les abondances. Toutes les images ont été corrompues par un faible bruit. La Figure 3.12 représente les cartes d'abondance ainsi

que les *endmembers* associés pour les trois premières images. Les caractéristiques de la quatrième image sont affichées sur la Figure 3.13. Les cartes d'abondance ont des supports constants par morceaux sur un fond nul.

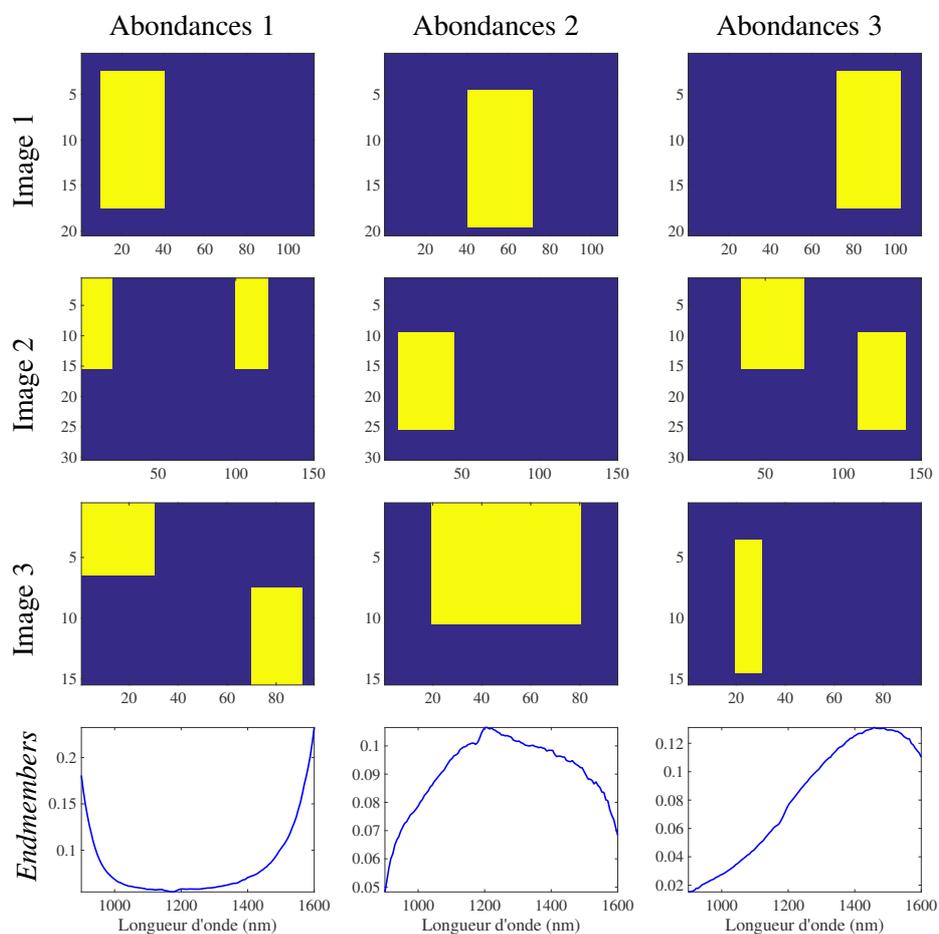


FIGURE 3.12 : Cartes d'abondance et signatures spectrales associées pour les images simulées 1, 2 et 3

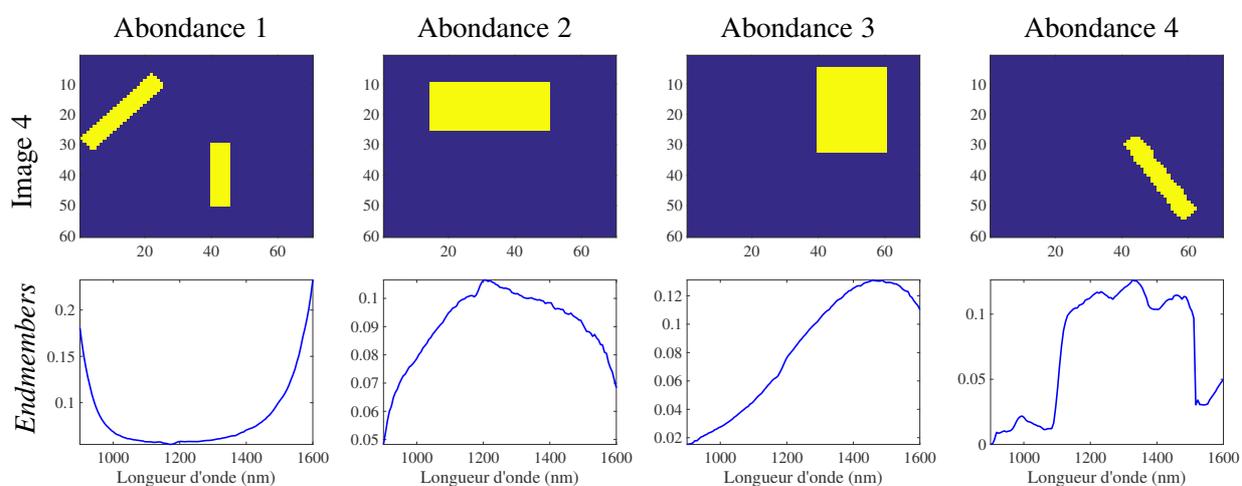


FIGURE 3.13 : Cartes d'abondance et signatures spectrales associées pour l'image simulée 4

L'image 1 a pour dimension $119 \times 20 \times 112$ où 119 caractérise la dimension spectrale (pour toutes les images) et 20×112 la dimension spatiale. Cette image est composée de trois *endmembers* juxtaposés et ne présente donc pas de mélange effectif. Les images 2 et 3, de dimensions respectives $119 \times 30 \times 150$ et $119 \times 15 \times 95$ sont également composées de trois *endmembers* mais nous pouvons observer, à partir des cartes d'abondance, le chevauchement de 2 ou 3 *endmembers* modifiant ainsi localement le rang de la décomposition. Enfin, l'image 4 a pour dimension $119 \times 60 \times 70$ et est composée de quatre *endmembers*.

La première expérience a pour but de montrer que l'algorithme OMDC-ADMM n'est pas adapté à des situations d'apparition ou de disparition de sources au cours de l'image. Pour ce faire, nous l'avons appliqué sur l'image 1 dont les cartes d'abondance et les *endmembers* sont représentés sur la Figure 3.12. Les paramètres ont été fixés de la manière suivante : $N_{iter} = 200$, $\alpha = 0.99$, $\mu = 0.01$, $R = 3$ et $\rho = 0.001$. Le traitement a été effectué suivant l'axe horizontal, séquentiellement, par tranche de dimension 119×20 . Les cartes d'abondance estimées sont illustrées sur la Figure 3.14. Nous constatons que le résultat du démélange hyperspectral est loin d'être satisfaisant ; en effet, pour chaque tranche, OMDC-ADMM met à jour trois *endmembers* et les fractions d'abondance associées, ce qui ne reflète pas la vérité terrain puisqu'une seule source est active à chaque échantillon. Notons que la Figure 3.14 (d) représente le rang de la décomposition calculé à partir des *endmembers* et des abondances générés à chaque tranche par OMDC-ADMM. Les *endmembers* non-actifs sont donc estimés et propagés au cours du traitement (par récursivité), dégradant considérablement les résultats. Ainsi, cette simple simulation montre la motivation principale du suivi de rang durant le traitement et donc de OSS-ADMM.

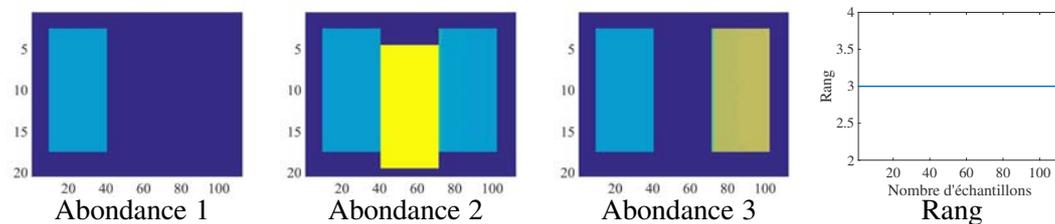


FIGURE 3.14 : Cartes d'abondance générées par OMDC-ADMM et rang de la décomposition, pour le mélange de l'image 1

La seconde expérience a consisté à tester les performances de OSS-ADMM pour le suivi du rang sur les quatre images hyperspectrales simulées présentées ci-dessus. Avant toute chose, une étape essentielle est la création de la bibliothèque hyperspectrale \mathbf{B} . Comme nous l'avons mentionné précédemment, \mathbf{B} peut être établie par apprentissage, *e.g.* à l'aide d'un algorithme de démélange hors-ligne, qui traite l'image dans sa totalité en une seule fois, et n'est donc pas sensible aux variations locales du nombre de sources. \mathbf{B} peut être aussi construit *a priori*, *e.g.* par la connaissance préalable des composants présents dans l'image et de leurs signatures spectrales. Dans le cas de nos images simulées, où nous connaissons parfaitement les spectres présents, nous avons élaboré une bibliothèque hyperspectrale composée des vrais *endmembers* ainsi que des *endmembers* fictifs (extraits d'autres images ou synthétisés). L'objectif est de surestimer le nombre de sources afin d'observer au mieux l'intérêt des contraintes de parcimonie pour l'activation et la désactivation des abondances. De plus, un bruit a été ajouté sur la matrice \mathbf{B} pour obtenir un $\text{SNR} = 30$ dB afin de s'éloigner légèrement de la vérité terrain. La bibliothèque est visible sur la

Figure 3.15, et regroupe un total de huit *endmembers*.

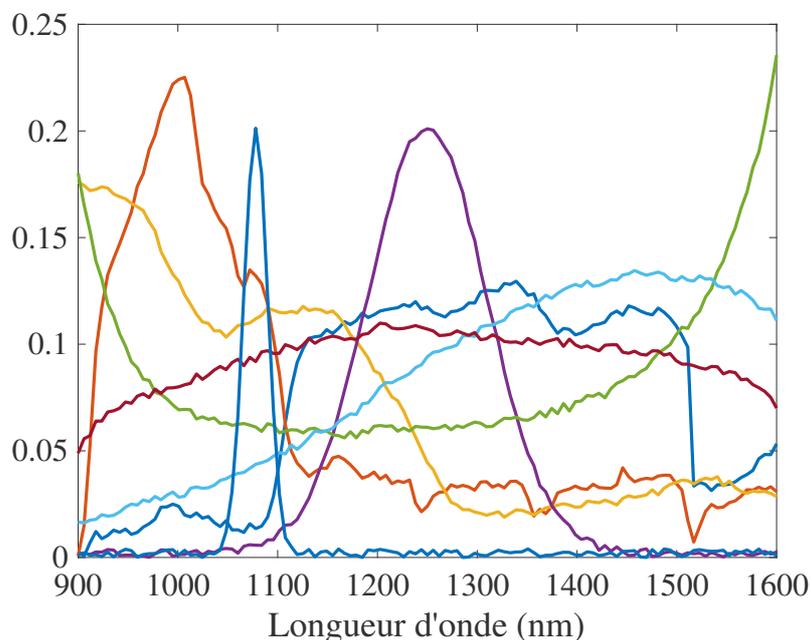


FIGURE 3.15 : Composition de la bibliothèque hyperspectrale \mathbf{B} pour le traitement des images 1 à 4

Pour l'ensemble des simulations, les paramètres relatifs à OSS-ADMM ont été fixés à : $N_{iter} = 50$, $\alpha = 0.9$, $\nu = 10^{-5}$, $\gamma = 0.0015$, $\omega = 1$, $\delta = 10^{-15}$, $R = 8$ et $\rho = 0.001$. Les résultats sont affichés sur la Figure 3.16, où l'on peut observer les cartes d'abondance obtenues ainsi que le rang de la décomposition calculé en fonction des paramètres estimés à chaque tranche. Notons que les cartes d'abondance affichées sont caractéristiques uniquement des *endmembers* actifs. Nous constatons que OSS-ADMM permet une très bonne reconstruction des cartes d'abondance, bien que le nombre de *endmembers* soit surestimé. Les coefficients d'abondance des sources inactives sont mis à zéro, ce qui permet de suivre parfaitement le rang de la décomposition. Précisons qu'une des principales limitations de cette méthode concerne la construction de la matrice \mathbf{B} dont les colonnes doivent être peu corrélées, pour un bon fonctionnement de l'algorithme.

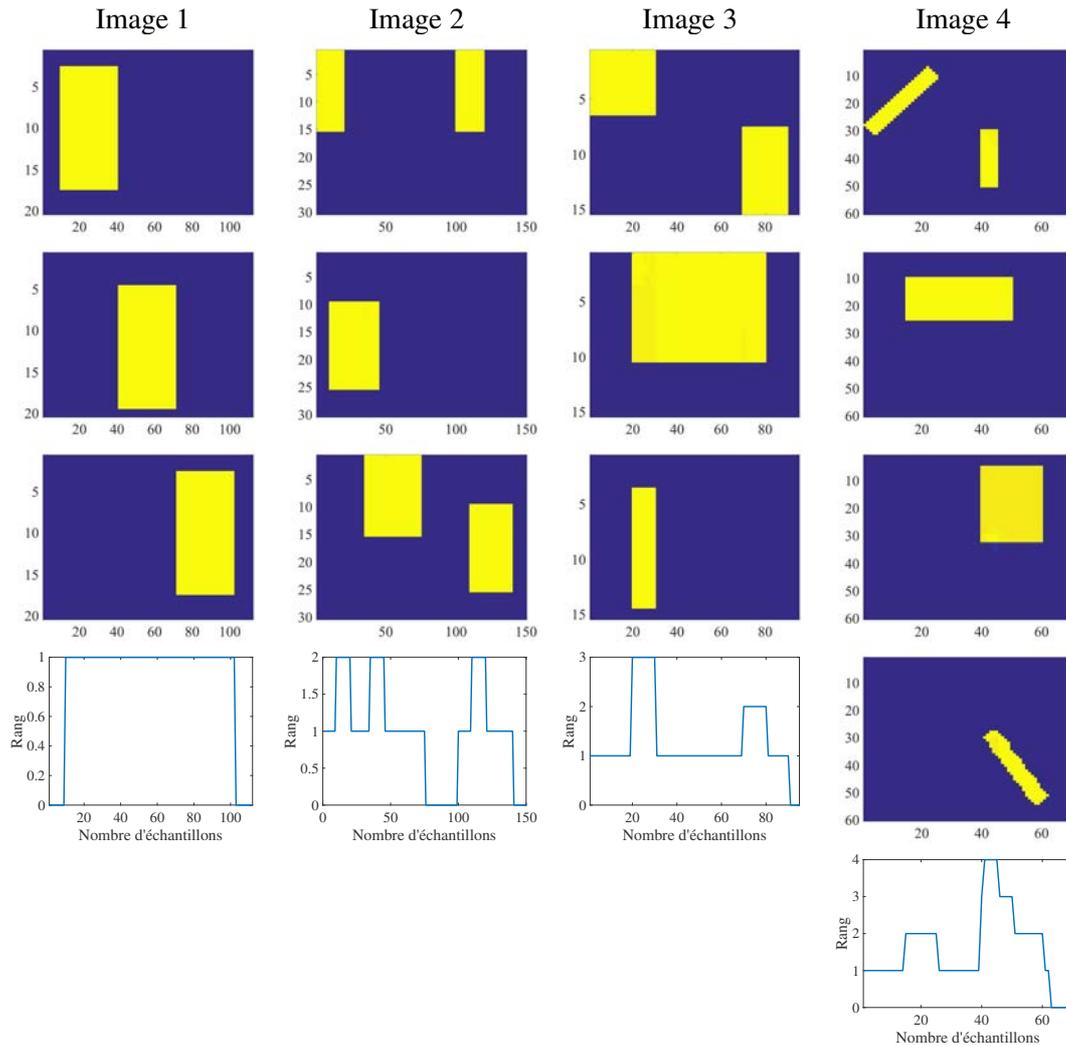


FIGURE 3.16 : Cartes d'abondance et rang de la décomposition estimés par OSS-ADMM , pour les mélanges des images 1 à 4

3.3.3 OSS-ADMM pour l'analyse de la diffusion de l'eau au sein d'un produit céréalier

L'objectif de cette expérience a été de suivre la diffusion de l'eau au sein de produits céréaliers (biscuits de la marque *Cracotte*). En effet, les produits alimentaires sont hygroscopiques et poreux et contiennent deux types d'eau, à savoir l'eau libre (FW pour *Free Water*) et l'eau liée (BW pour *Bound Water*). FW est présent dans les espaces inter-cellulaires tandis que l'eau dans les espaces intra-cellulaires est appelée BW. Au cours de la phase de séchage (pour un produit hydraté en amont), les voies de migration de FW et de BW peuvent entraîner un retrait (ou rupture) et ainsi dégrader la qualité des attributs des produits alimentaires. Par conséquent, nous visons à étudier les mécanismes de transfert de FW et de BW pendant la phase de séchage du biscuit en utilisant OSS-ADMM. Des expériences ont été effectuées sur un biscuit où deux gouttes ont été ajoutées en surface (voir Figure 3.17), de manière à simuler une phase d'absorption.



FIGURE 3.17 : Ajout de deux gouttes d'eau au centre du biscuit

Le séchage a été effectué naturellement, à l'air ambiant. Durant le séchage, toutes les minutes, une image hyperspectrale du biscuit entier a été acquise par un imageur *pushbroom* PIR situé à l'INRA de Nantes, travaillant dans la gamme spectrale 900-2500 nm. La gamme spectrale est plus grande que celle fixée par notre imageur FX17 (900-1700 nm) et cela est d'un grand intérêt pour la détection de l'eau, puisque cette dernière présente un pic d'absorption significatif à 1900 nm. L'acquisition des images a été réalisée pendant 73 minutes. Ainsi, nous avons obtenu une séquence de 73 cubes hyperspectraux. Le traitement de série temporelle d'images hyperspectrales est un sujet d'actualité et suscite un vif intérêt dans la communauté du traitement du signal et des images (voir *e.g.* [Henrot 16, Thouvenin 16]). Suite aux acquisitions, chaque image a été traitée séquentiellement par OSS-ADMM. Elles ont toutes pour dimension $231 \times 224 \times 461$ où 231 représente le nombre de longueurs d'onde et 224×461 les dimensions spatiales. N'ayant aucune connaissance *a priori* des données, la bibliothèque hyperspectrale \mathbf{B} a été estimée par le biais de l'algorithme hors-ligne BMDC-ADMM, sur l'image hyperspectrale correspondant au temps $t = 1$ min. Le nombre de *endmembers* à extraire a été déterminé expérimentalement et il a été fixé à 4. Les spectres de référence sont affichés sur la Figure 3.18, à la fois en réflectance et en absorbance.

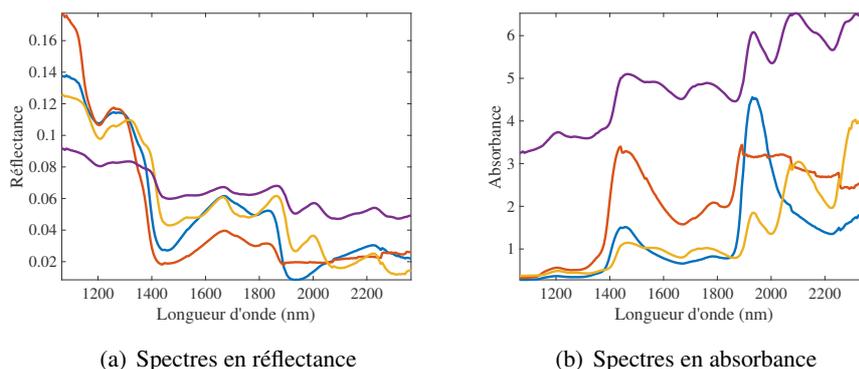


FIGURE 3.18 : Bibliothèque hyperspectrale estimée par BMDC-ADMM pour le traitement des images hyperspectrales de biscuit

Les paramètres de OSS-ADMM ont été définis tels que $N_{iter} = 10$, $\alpha = 0.99$, $v = 10^{-5}$, $\gamma = 0.001$, $\omega = 50$, $\delta = 10^{-15}$, $R = 4$ et $\rho = 0.001$. L'algorithme traite chaque image hyperspectrale suivant l'axe horizontal, par échantillon de dimension 231×224 (précisons que le temps de traitement d'une image est de l'ordre de 0.7 s). Nous pouvons facilement observer, par exemple sur la Figure 3.17, que le nombre de *endmembers* est susceptible d'évoluer dès lors que la tranche à traiter se situe au niveau de la goutte d'eau ; en effet, l'eau est caractérisée par un spectre bien précis, qui la différencie du spectre

de l'amidon, par exemple, qui est présent sur tout le biscuit. La Figure 3.20 représente essentiellement les cartes d'abondance et les spectres d'absorbance associés, estimés par OSS-ADMM, relatifs aux deux états de l'eau, pour les temps intermédiaires suivants $t = 1, 15, 25, 35, 45,$ et 55 min. Précisons que les deux autres cartes d'abondance et les *endmembers* estimés sont communs à tous les temps t et sont représentés sur la Figure 3.19. Par expertise, l'un des spectres caractérise l'amidon, que l'on retrouve sur tout le biscuit, et le deuxième, très ressemblant, semble être celui de fibres cellulosiques caractéristiques des céréales.

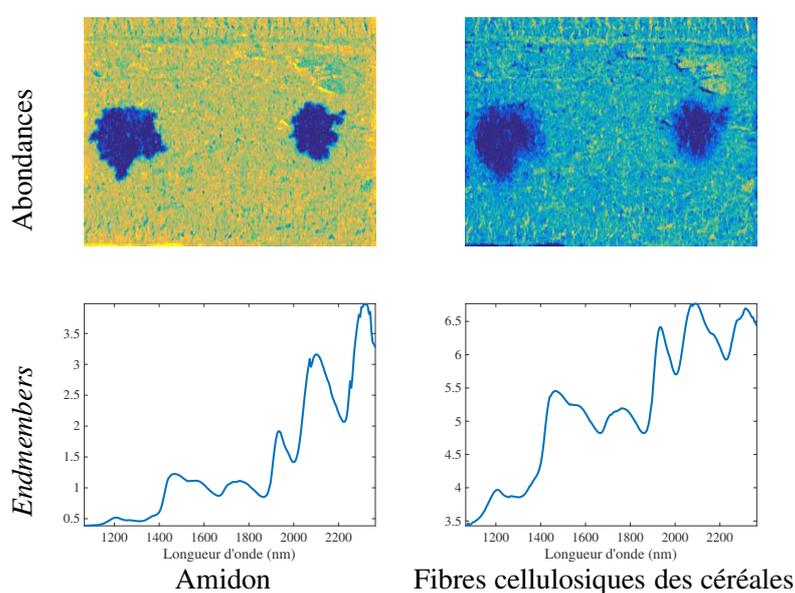


FIGURE 3.19 : Cartes d'abondance et spectres de l'amidon et de fibres cellulosiques estimés par OSS-ADMM pour les images hyperspectrales de biscuit

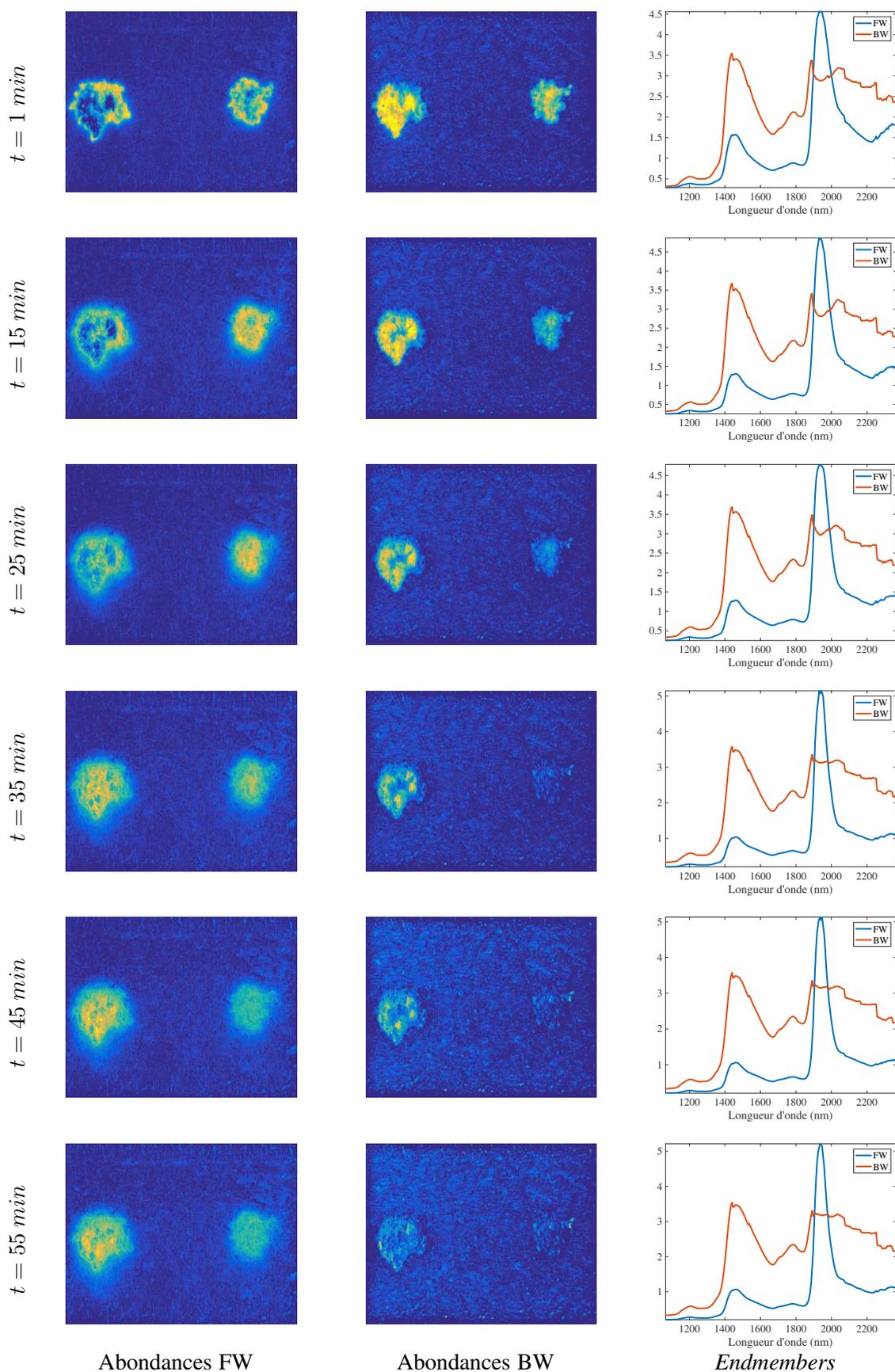


FIGURE 3.20 : Résultats du démelange hyperspectral par OSS-ADMM pour les images hyperspectrales de biscuit acquises aux temps $t = 1, 15, 25, 35, 45,$ et 55 min

Les résultats observés sur la Figure 3.20 sont très intéressants car nous sommes capables de fournir une interprétation physico-chimique des spectres obtenus. Le spectre représenté en bleu caractérise l'eau libre (pics significatifs et connus à 1440 et 1940 nm). Le second spectre tracé en rouge présente un élargissement important du pic vers 1440 lié aux liaisons hydrogènes inter-moléculaires (1530-1610 nm) et intra-moléculaires (1430-1520 nm), ainsi qu'un pic beaucoup plus fin situé à 1885 nm (un zoom de cette région est représenté sur la Figure 3.21). De plus, le décalage de la ligne de base par rapport au spectre de l'eau libre est un comportement spécifique à l'eau liée. Le pic, situé à 1885 nm, tend à disparaître à mesure que t augmente, indiquant une évolution de la proportion d'eau liée dans le biscuit. Nous signalons que les résultats de cette analyse et notamment les valeurs des pics caractéristiques sont en accord avec les analyses présentées, *e.g.*, dans [Carteret 99].

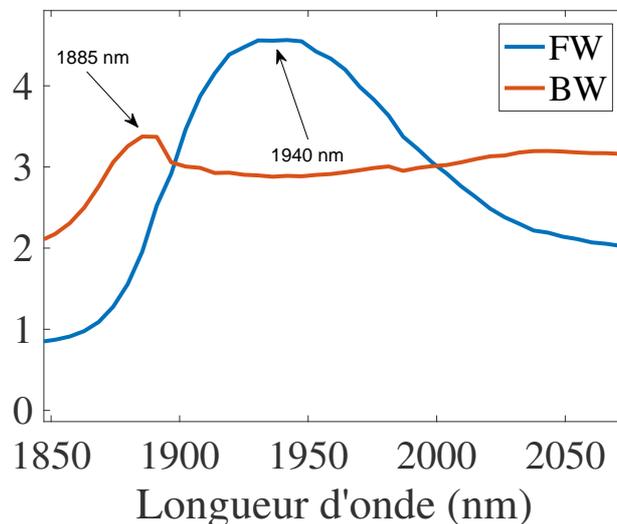


FIGURE 3.21 : Zoom sur un des pics caractéristiques de l'eau liée et de l'eau libre, à $t = 1$ min

Les cartes d'abondance apportent également de nombreuses informations : à $t = 1$ min, nous pouvons observer que l'eau libre se situe plutôt sur les bords de la zone humide tandis que l'eau liée au centre. Sur la Figure 3.22 est représentée la même goutte d'eau, à $t = 1$ s, et à quelques secondes d'intervalle et il apparaît clairement que l'eau pénètre plus rapidement au milieu que sur les côtés. De plus, comme nous pouvons le voir sur la Figure 3.23, les tissus mouillés ne sont pas identiques : certains semblent plus gris que d'autres. Cette différence de structure explique les résultats obtenus. À mesure que t augmente, *i.e* au cours de la phase d'absorption, nous pouvons observer que la diffusion de l'eau libre est synchronisée avec la diffusion de l'eau liée dans les parois du biscuit ; c'est d'ailleurs ce que l'on observe à partir des cartes d'abondance à $t = 15$ min, où un contraste apparaît au niveau de la goutte d'eau située à droite pour FW et BW. En d'autres termes, l'eau libre n'est observée qu'après que les parois des cellules sont remplies d'eau liée. Ces résultats sont cohérents avec ceux présentés *e.g.* dans [Gezici-Koç 17], où l'eau liée est la première espèce qui apparaît majoritairement après rajout de l'eau, tandis que l'eau libre est majoritaire en fin d'hydratation (saturation en eau).

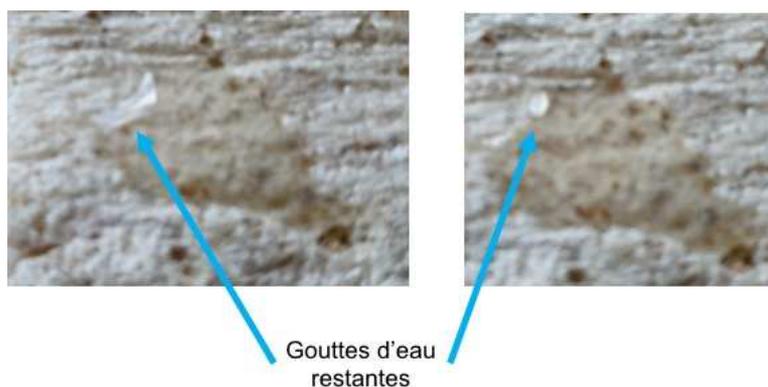


FIGURE 3.22 : Pénétration de l'eau au sein du biscuit. À gauche : $t = 1$ s, à droite $t = 1 +$ quelques secondes



FIGURE 3.23 : Tissus hydratés non homogènes du biscuit

Bien que ces résultats nous informent principalement sur le processus d'absorption de FW et BW dans le produit, ils peuvent être d'une aide capitale pour améliorer le processus industriel de séchage. En effet, nous constatons à $t = 55$ min, que le biscuit est encore hydraté et il l'est encore à $t = 73$ min (non montré ici). L'eau libre est la première à être évacuée lors de la phase de séchage. Son évacuation (par évaporation, par exemple) est facile et se réalise sans variation dimensionnelle du produit. Lorsque toute l'eau libre a disparu et que l'eau liée demeure intégralement, on atteint le point de saturation des fibres (qui varie en fonction des matériaux). C'est à ce stade que le séchage doit être méticuleusement mené, car une perte de l'eau liée trop rapide peut dégrader considérablement le produit. Ainsi, être en mesure de cartographier la diffusion de l'eau en temps réel sur une ligne de production est un enjeu essentiel dans l'industrie agro-alimentaire, et non seulement.

Pour finir, afin d'étudier la capacité de suivi de rang de l'algorithme OSS-ADMM, nous nous sommes limités au traitement de l'image hyperspectrale au temps $t = 1$ min. Afin d'observer ce comportement, il est nécessaire d'augmenter significativement la contrainte de parcimonie et de modifier les autres paramètres en conséquence ; ainsi, pour cette expérience, les paramètres de OSS-ADMM ont été fixés tels

que $N_{iter} = 40$, $\alpha = 0.99$, $v = 10^{-5}$, $\gamma = 0.045$, $\omega = 30$, $\delta = 10^{-15}$, $R = 4$ et $\rho = 0.001$. Les cartes d'abondance obtenues pour les 4 *endmembers* sont représentées sur la Figure 3.24.

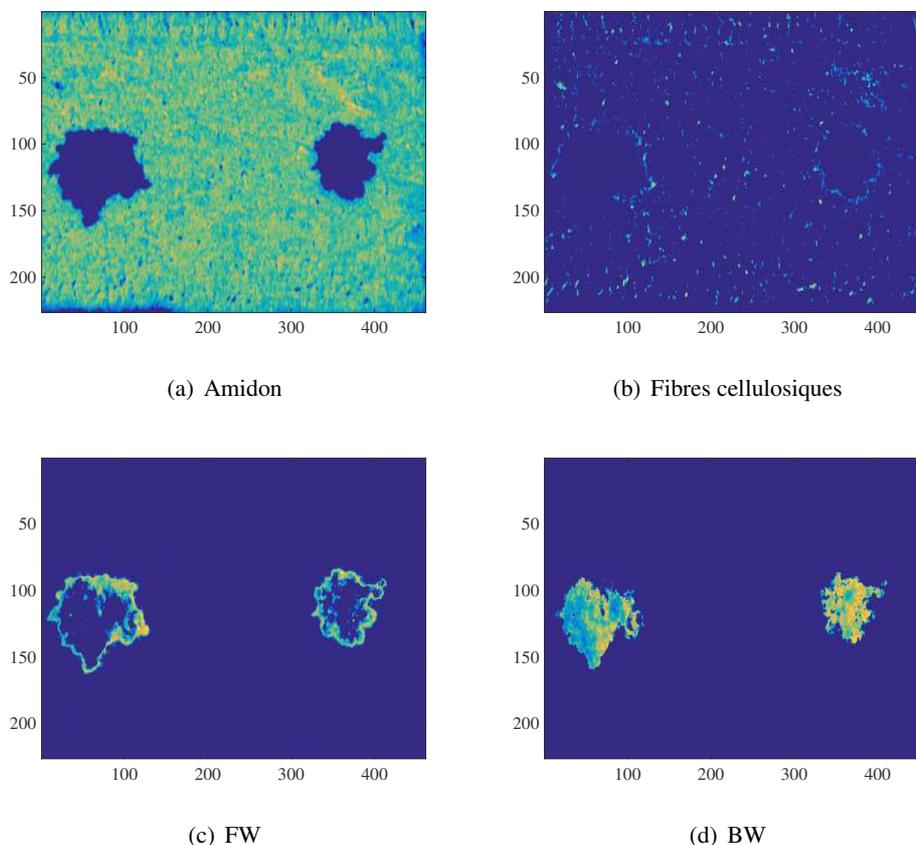


FIGURE 3.24 : Cartes d'abondance estimées par OSS-ADMM pour l'image hyperspectrale de biscuit à $t = 1$ min, pour $v = 10^{-5}$ et $\gamma = 0.045$

On constate que la parcimonie permet d'isoler parfaitement les gouttes d'eau du reste du biscuit, ce qui n'est pas le cas pour les expériences précédentes lorsque $\gamma = 0.001$. La Figure 3.25, illustre le rang de la décomposition, calculé à chaque tranche de l'image à partir des *endmembers* et des abondances estimés par OSS-ADMM. Nous constatons que le rang évolue au cours du traitement de l'image et on identifie trois principales régions : les deux premières, situées entre les tranches 15-122 et 331-413 affichent un rang constant et égal à 4 indiquant que les 4 *endmembers* sont présents simultanément au sein de chaque tranche de l'image. Cela est cohérent puisque la position des deux gouttes correspond exactement à ces deux régions de l'image. Par ailleurs, on observe de part et d'autre de ces deux régions, des petits effets en forme d'escalier qui illustrent la présence de FW en périphérie de chaque goutte d'eau. Dans la troisième région, située entre les tranches 132- 325, le rang est constant et égal à 2, caractérisant uniquement la présence du cellulose et des fibres cellulosiques. En effet, dans cette zone, les deux états de l'eau ne sont plus présents. Ce sont des résultats très intéressants, qui nous offre la possibilité de suivre au cours du traitement, l'activation et la désactivation des sources présentes dans une image.

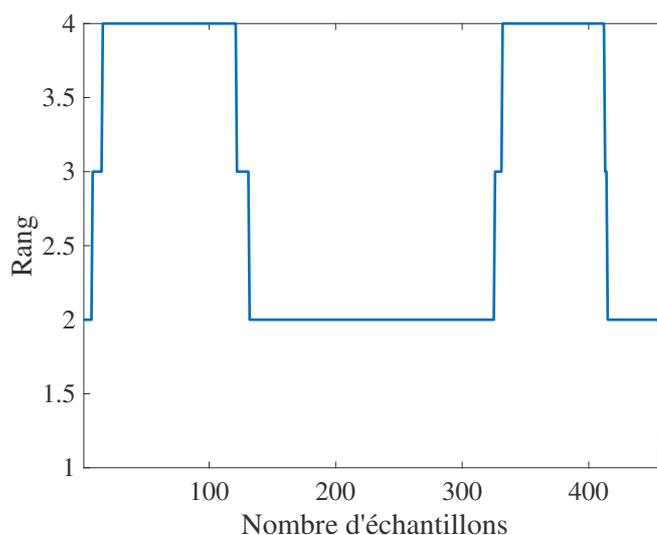


FIGURE 3.25 : Rang de la décomposition estimé par OSS-ADMM en fonction du nombre d'échantillons pour l'image hyperspectrale de biscuit à $t = 1$ min

3.4 Conclusions

Dans ce chapitre nous avons traité le problème du suivi du nombre de sources dans une image hyperspectrale, qui intervient dans certaines applications en-ligne. L'approche OSS-ADMM proposée inclut un terme de régularisation qui consiste à minimiser la distance entre les *endmembers* estimés à chaque tranche et une référence (générée soit par connaissance *a priori* soit par apprentissage) ; les normes $\ell_{1,1}$ et $\ell_{2,1}$ favorisent conjointement la parcimonie par ligne et spatiale de la matrice des abondances, permettant ainsi de sélectionner les sources actives de la bibliothèque hyperspectrale. Plusieurs expériences sur des images hyperspectrales simulées ont montré que OSS-ADMM parvient à suivre le rang de la décomposition en-ligne, tout en générant des estimations de bonne qualité. Par ailleurs, OSS-ADMM s'est montré robuste à la surestimation du nombre de *endmembers*, à condition que ces derniers ne soient pas trop corrélés. Nous avons testé OSS-ADMM sur une application originale, consistant à suivre la diffusion de l'eau dans des produits agro-alimentaires. Il s'agit d'un problème récurrent et de grand intérêt pratique dans l'industrie, puisque les mécanismes de transport de l'eau sont encore mal connus et peuvent altérer la qualité des produits. L'imagerie hyperspectrale PIR combinée à notre approche nous a permis de générer en-ligne des spectres dont l'interprétation physico-chimique est conforme à la littérature ainsi que des cartes d'abondance qui offrent une très bonne indication quant au stade d'hydratation du matériau. Enfin, OSS-ADMM s'est révélé être un algorithme rapide ; ainsi, il serait judicieux de s'orienter vers cette méthode pour le traitement des pièces de bois.

Chapitre 4

Analyse de planches de chêne

4.1 Introduction

Les experts du fabricant de parquet Berry Wood, l'un des partenaires industriels de l'ANR OPTIFIN, ont annoté et fourni trente échantillons de bois présentant de nombreuses singularités, dont certaines sont difficilement visibles à l'œil nu. Ces singularités posent problème pour la prédiction de la qualité et du rendu du revêtement. Chaque échantillon a été scanné par l'imageur *pushbroom* FX17 de chez Specim afin d'obtenir des images hyperspectrales. La fréquence d'acquisition (nombre d'images par seconde) ainsi que le temps d'exposition ont été respectivement de 145 Hz et 2.1 ms ; la vitesse de déplacement de l'échantillon était de 90 mm/s. Pour chaque mesure, nous avons utilisé une référence pour le noir et une pour le blanc, utilisées pour normaliser les données selon :

$$\text{Réflectance} = \frac{\text{mesures brutes} - \text{référence noire}}{\text{référence blanche} - \text{référence noire}}. \quad (4.1)$$

Les différentes étapes de traitement sont illustrées sur la Figure 4.1.

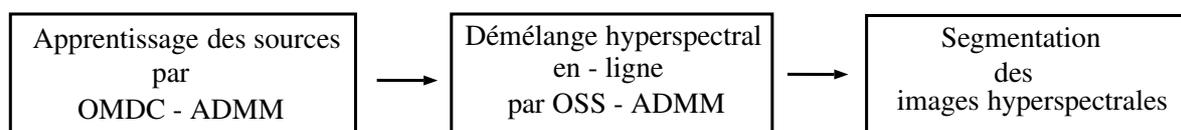


FIGURE 4.1 : Étapes de traitement des pièces de bois

La première étape (une fois les données acquises et normalisées) consiste en l'utilisation de l'algorithme OMDC-ADMM sur un échantillon de bois typique. L'objectif est d'utiliser cette méthode dans une phase d'apprentissage, permettant d'estimer et d'apprendre les sources d'intérêt. Le démélange hyperspectral aveugle est effectué suivant le rythme d'acquisition de l'imageur *pushbroom*, *i.e.* tranche par tranche séquentiellement dans le temps. La seconde étape consiste à traiter l'ensemble des pièces de bois à l'aide de l'algorithme OSS-ADMM ; la bibliothèque hyperspectrale pour OSS-ADMM est construite lors de l'étape 1. Enfin, la troisième étape a pour objectif la segmentation (la classification) des images hyperspectrales par la méthode des *k-means* [Hartigan 79] ; cela permet le regroupement des pixels ayant les

mêmes propriétés spectrales. Cette étape est essentielle dans la prise de décision, notamment lorsqu'il s'agit du tri des pièces de bois en temps réel.

Ce chapitre est organisé de la manière suivante : la Section 2 présente les divers échantillons de bois annotés par expertise ainsi que la présentation des singularités observées. La Section 3 regroupe les résultats du démélange obtenus avec les algorithmes OMDC-ADMM et OSS-ADMM. La segmentation des images hyperspectrales est présentée ainsi qu'une analyse des anomalies localisées. Nous concluons le chapitre dans la Section 4.

4.2 Présentation des pièces de bois

4.2.1 Échantillons à traiter

Trente pièces de bois (chêne), non conformes, ont été mises à notre disposition, toutes annotées par des experts. Précisons que ces pièces de bois, présentées ici à l'état brut, sont ensuite transformées pour confectionner du parquet. Nous proposons de présenter dans ce chapitre neuf d'entre elles, qui nous semblent les plus pertinentes en ce qui concerne les singularités. Elles sont représentées sur les Figures 4.2 à 4.10.

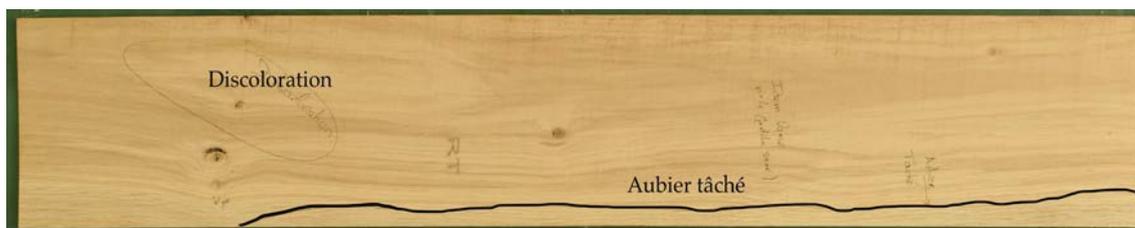


FIGURE 4.2 : Échantillon 1

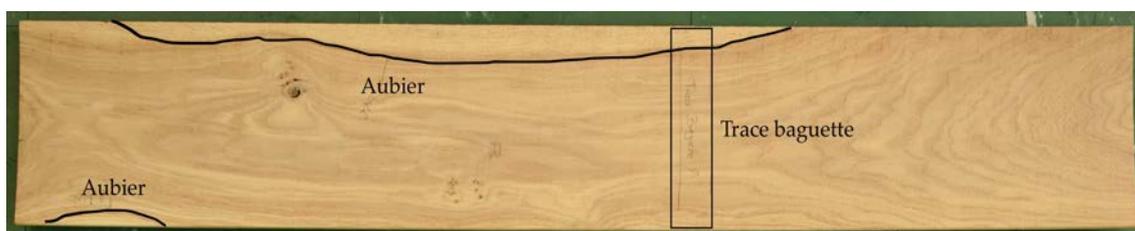


FIGURE 4.3 : Échantillon 2



FIGURE 4.4 : Échantillon 3

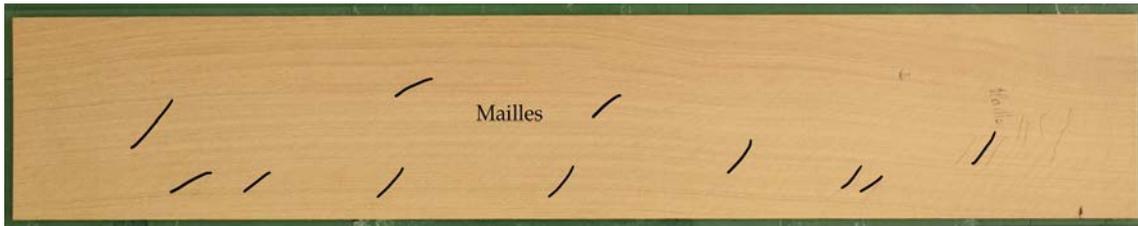


FIGURE 4.5 : Échantillon 4

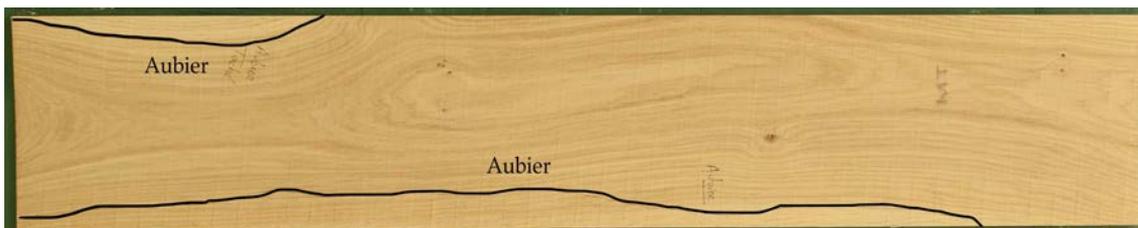


FIGURE 4.6 : Échantillon 5



FIGURE 4.7 : Échantillon 6



FIGURE 4.8 : Échantillon 7



FIGURE 4.9 : Échantillon 8

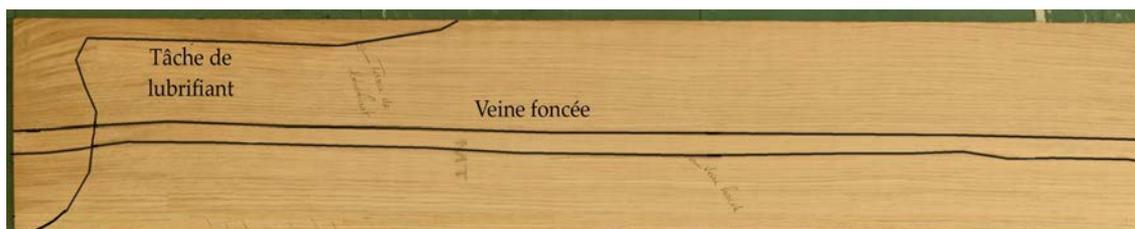


FIGURE 4.10 : Échantillon 9

4.2.2 Définition des singularités

Notons que le rendu esthétique de certaines singularités est plus ou moins apprécié ; cela dépend du choix ou du grade attribué au parquet. Il existe trois grades caractérisant l'esthétique du revêtement :

- Le grade A : le bois ne présente aucun nœud ni aubier et son rendu est homogène.
- Le grade B : quelques irrégularités naturelles du bois sont visibles dont les aubiers.
- Le grade C : le bois présente une quantité importante de défauts naturels en surface.

Nous avons relevé, sur les neuf échantillons, un total de douze singularités. Nous pouvons citer :

- Discoloration : elle est due à une réaction chimique qui s'opère généralement lors du séchage du bois (suite à un vernissage ou à l'application d'une peinture). Les tannins (substances végétales présentes dans le bois), migrent vers la surface, se transforment chimiquement et brunissent.
- Aubier : tâché ou non, c'est la partie poreuse et blanchâtre située entre le bois dur et l'écorce d'un arbre. Il est courant que l'aubier réagisse avec la finition appliquée sur le bois, générant en surface des lignes blanches plus ou moins importantes. Il n'est pas toujours détectable à l'œil (voir par exemple l'échantillon 1, Figure 4.2)
- Trace baguette : générée par le support où le bois est empilé et stocké.
- Veine foncée : le veinage est plus ou moins foncé en fonction des essences. Il correspond aux « lignes » que l'on peut observer sur le parquet. Pour un parquet en chaîne clair, ce qui est le cas ici, les veines foncées ne sont pas désirées.

- Nœud : le nœud est une modification de la structure du bois liée à l'existence d'une branche. On distingue plusieurs sortes de nœuds (voir Figure 4.11) :
 - le nœud vivant (ou sain) : il est issu d'une branche encore vivante au moment de l'abattage. Le nœud est parfaitement englobé dans le reste de la surface, auquel il a une bonne adhérence.
 - le nœud mort (ou noir) : il est issu d'une branche morte. Il est en discontinuité avec les tissus du bois et risque de se détacher au cours du séchage ou d'un usinage.
 - le nœud vicieux (ou pourri) : on parle de nœud vicieux lorsque ses tissus sont affectés par la pourriture. Le plus souvent, ce dernier apparaît lors d'une infiltration d'eau de pluie après le cassage accidentel d'une branche.
 - le nœud «patte de chat» : regroupement de petits nœuds sains en grappes. Ce type de nœud est souvent recherché pour son aspect décoratif.

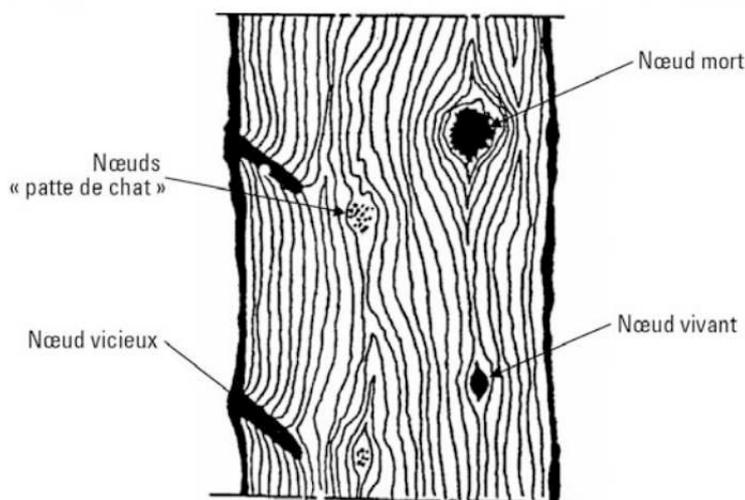


FIGURE 4.11 : Différents types de nœuds

Image extraite de : <http://lapins.menuisier.free.fr>

- Maille : la maille est le nom donné aux petits « motifs », d'aspect nacré, caractéristiques du veinage de certaines essences de bois (le chêne étant l'exemple le plus connu). Ces motifs ne sont présents que sur les planches débitées aux plus près du diamètre de la bille de bois (les planches dites « de quartier »).
- Fente : elle peut être causée par le vieillissement du bois, la sécheresse, le froid ou encore lors d'un choc.
- Tâche de lubrifiant : résidus de machine.

4.3 Démélange hyperspectral en-ligne

Dans cette section, nous avons mené plusieurs expériences sur les neuf images hyperspectrales de bois. L'objectif principal est d'étudier la capacité de nos algorithmes à détecter les singularités présentées ci-dessus. Nous proposons d'organiser la phase de démélange hyperspectral en-ligne et d'analyse des résultats en trois étapes :

- La première étape consiste en l'utilisation de l'algorithme OMDC-ADMM en tant que phase d'apprentissage. Le but est d'apprendre les signatures spectrales et de constituer une bibliothèque, regroupant les sources permettant de discriminer au mieux les singularités.
- La deuxième étape est l'utilisation de l'algorithme OSS-ADMM avec les sources préalablement apprises sur l'ensemble des échantillons.
- La dernière étape présente la segmentation des cartes d'abondance obtenues par OSS-ADMM, par la méthode des *k-means*, afin d'observer le regroupement des pixels ayant les mêmes propriétés spectrales.

4.3.1 Traitement des pièces de bois à l'aide de OMDC-ADMM

Dans un but de validation, OMDC-ADMM a été appliqué sur l'ensemble des images hyperspectrales de bois. Ceci étant, la bibliothèque hyperspectrale retenue est issue du traitement d'un seul échantillon de bois. Pour l'ensemble des données, les paramètres relatifs à OMDC-ADMM ont été fixés de la manière suivante : $N_{iter} = 30$, $\alpha = 0.99$, $\mu = 2$ et $\rho = 0.001$. Le nombre de *endmembers* à extraire a été fixé à $R = 3$; divers essais en amont n'ont pas montré l'intérêt d'en estimer davantage. Les cartes d'abondance estimées par OMDC-ADMM sont présentés sur les Figures 4.12 à 4.20.

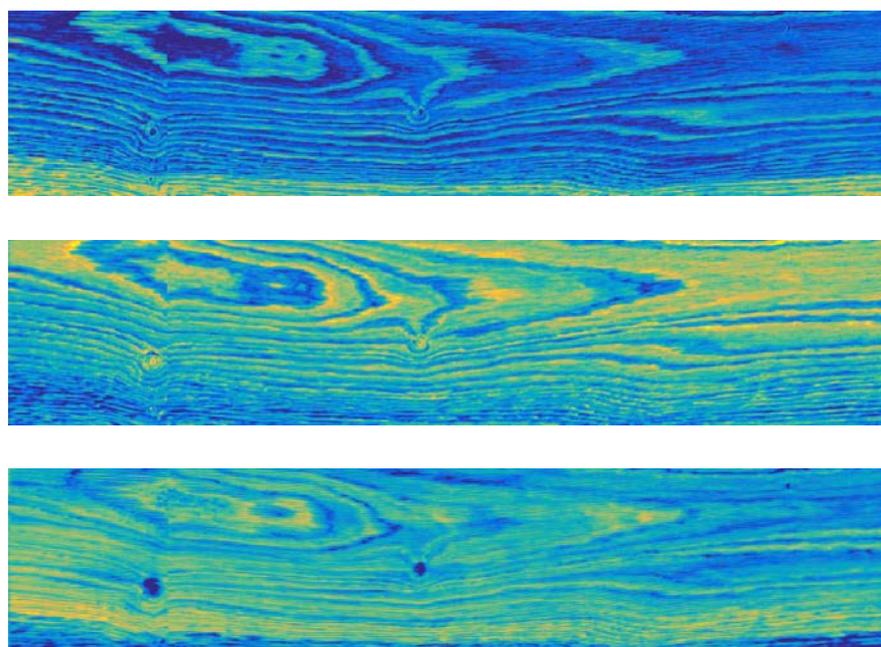


FIGURE 4.12 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 1

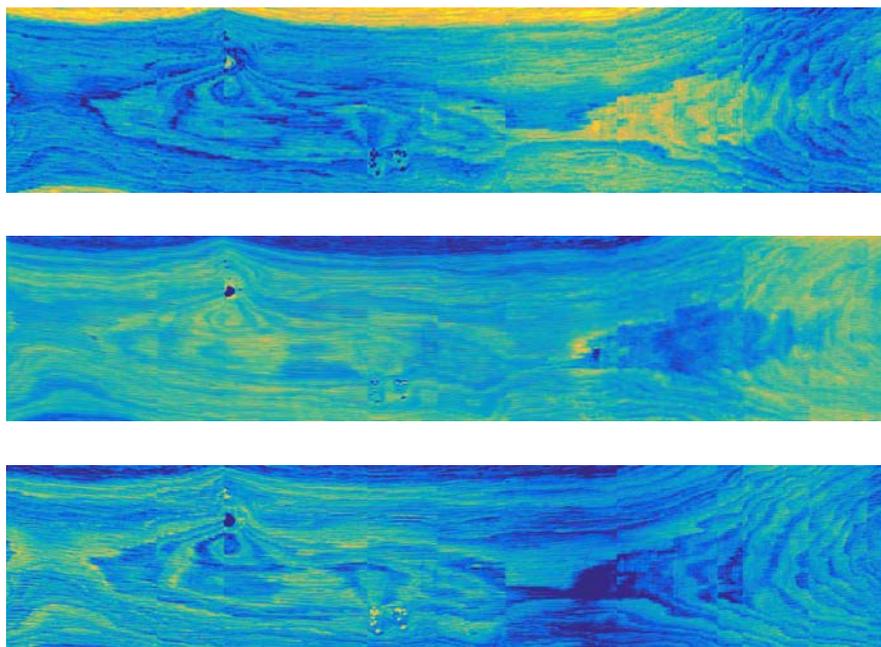


FIGURE 4.13 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 2

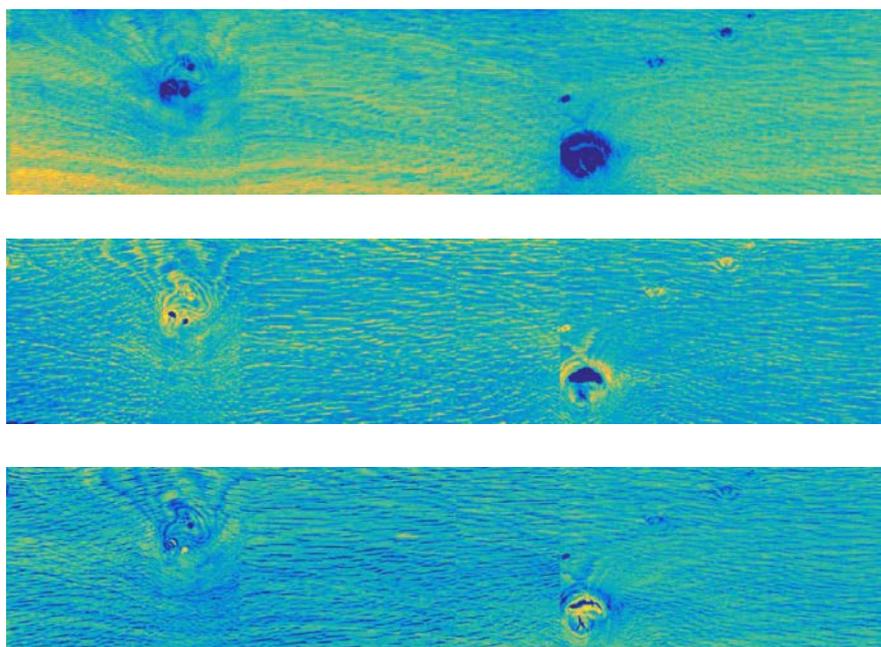


FIGURE 4.14 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 3

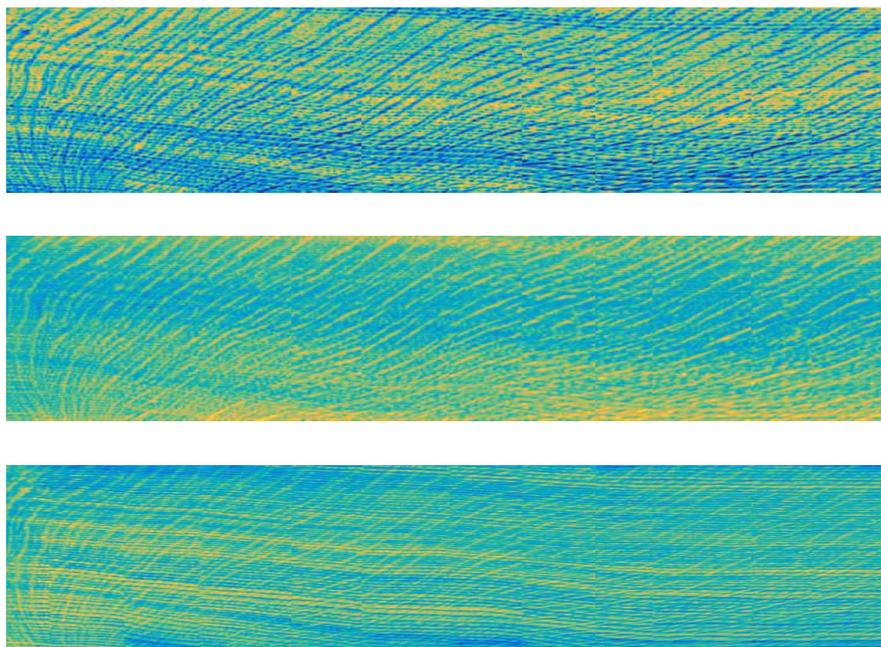


FIGURE 4.15 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 4

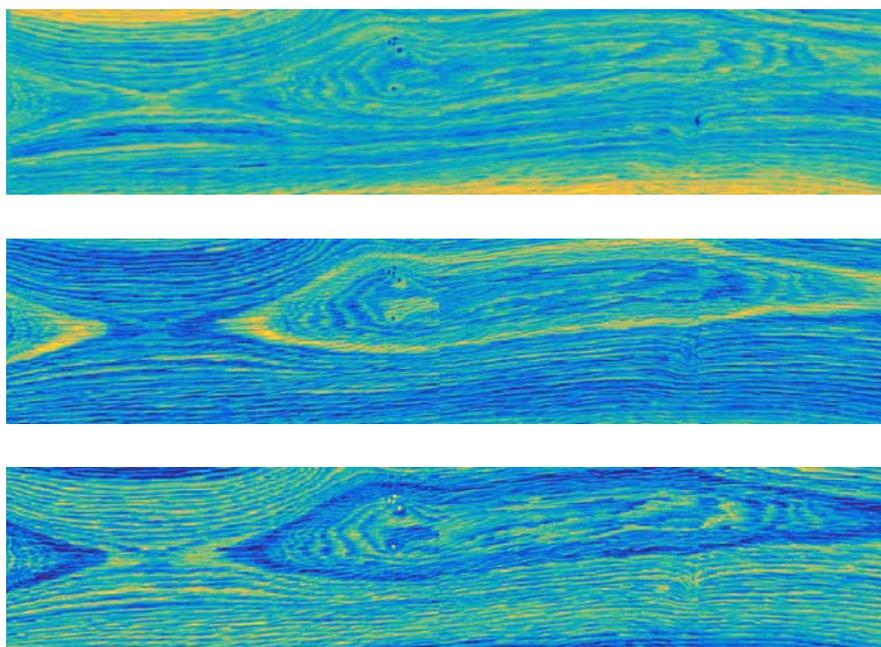


FIGURE 4.16 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 5

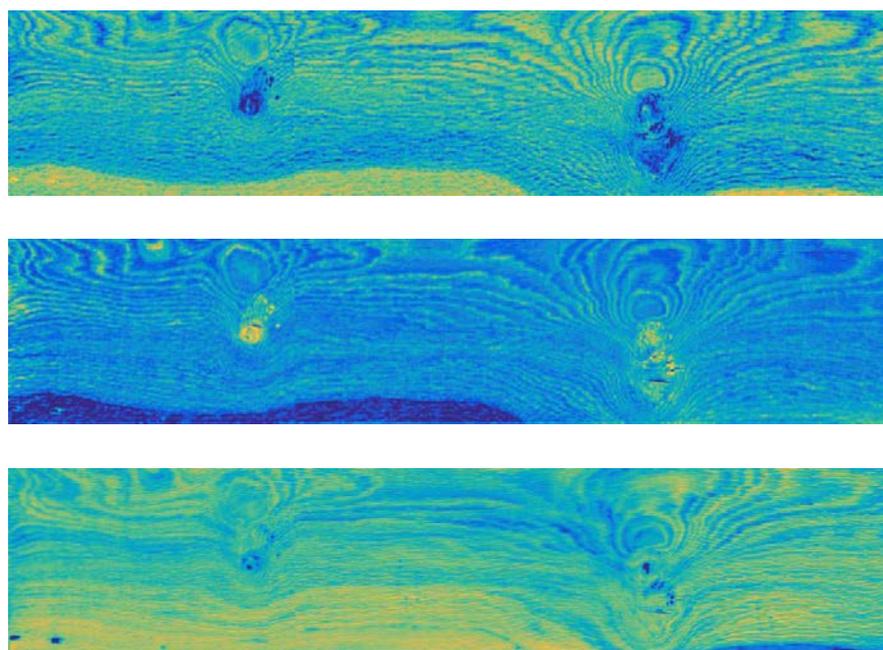


FIGURE 4.17 : RCartes d'abondance générées par OMDC-ADMM pour l'échantillon 6

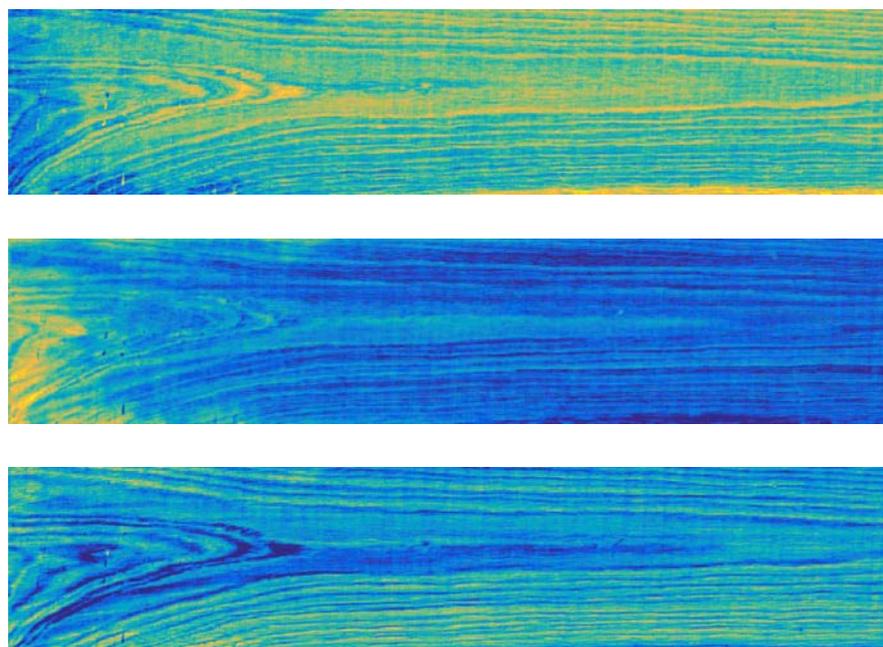


FIGURE 4.18 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 7

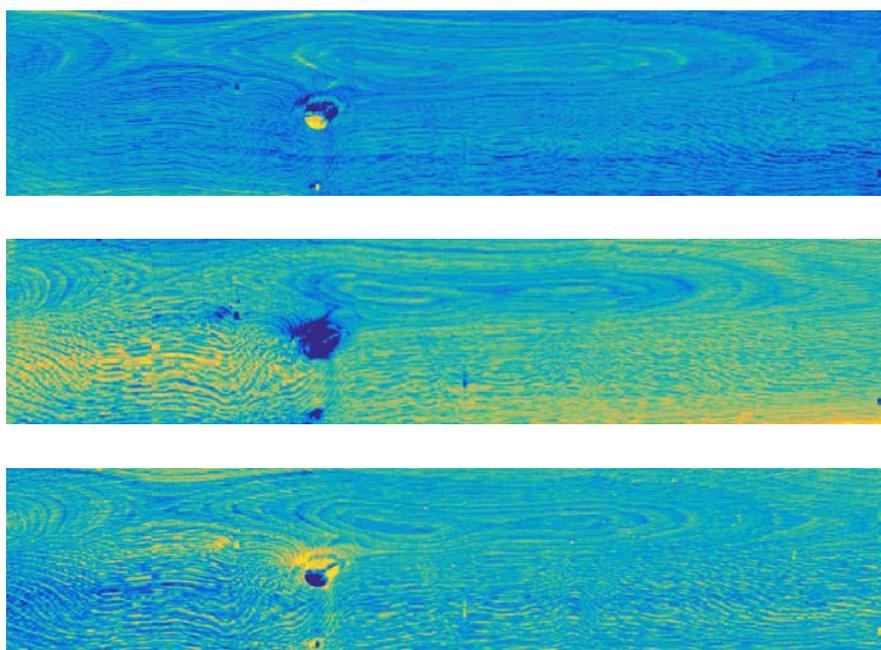


FIGURE 4.19 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 8

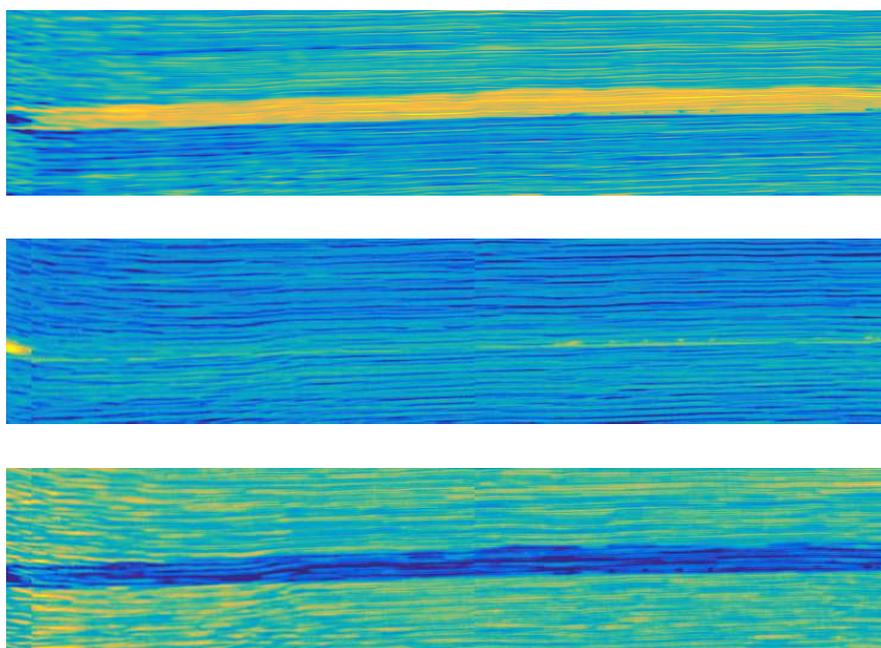


FIGURE 4.20 : Cartes d'abondance générées par OMDC-ADMM pour l'échantillon 9

Les cartes d'abondance nous fournissent un aspect visuel très intéressant : sur chacune d'entre elles, nous pouvons observer de nombreuses structures en surface du bois, qui ne sont pas toutes visibles à l'œil nu. De plus, la grande majorité des singularités, relevées par les experts, sont détectées par l'algorithme. Un récapitulatif est donné dans le Tableau 4.1. Les cellules du tableau représentées en vert indiquent les singularités visibles sur les cartes d'abondance, tandis que les singularités non détectées par l'algorithme

sont représentées par des cellules rouges. Ce sont des résultats prometteurs car les solutions techniques industrialisées à ce jour se cantonnent à détecter des singularités très marquées comme les nœuds par exemple. D'ailleurs, nous constatons avec succès que notre méthode est capable de distinguer les nœuds sains aux nœuds noirs (voir échantillon 3) et les nœuds sains aux nœuds pourris (voir échantillon 8). Le Tableau 4.2 regroupe les dimensions de chaque image ainsi que le temps de traitement relatif à OMDC-ADMM.

Échantillons	1	2	3	4	5	6	7	8	9
Discoloration									
Aubier									
Aubier tâché									
Trace baguette									
Nœud sain									
Nœud noir									
Nœud pourri									
Nœud patte de chat									
Veine foncée									
Maille									
Fente									
Tâche de lubrifiant									

TABLE 4.1 : Récapitulatif des singularités détectées par l'algorithme OMDC-ADMM

	Dimension	Temps (s)
Échantillon 1	224 × 256 × 1491	7.74
Échantillon 2	224 × 273 × 1216	6.45
Échantillon 3	224 × 264 × 1430	7.76
Échantillon 4	224 × 241 × 745	3.78
Échantillon 5	224 × 264 × 1253	6.52
Échantillon 6	224 × 270 × 1468	7.64
Échantillon 7	224 × 271 × 1493	7.96
Échantillon 8	224 × 262 × 1243	6.98
Échantillon 9	224 × 177 × 737	3.28

TABLE 4.2 : Dimension de chaque image hyperspectrale de bois et temps de traitement par l'algorithme OMDC-ADMM

Lors de cette phase d'apprentissage, nous avons constaté que les signatures spectrales n'évoluent que très peu d'un échantillon à l'autre ; c'est un point essentiel puisque cela facilite grandement la création de la bibliothèque hyperspectrale. Les trois spectres que nous avons désigné comme référence sont illustrés sur la Figure 4.21 ; ce sont les spectres estimés par OMDC-ADMM suite au traitement de l'échantillon 5. Ces derniers, représentés en réflectance, sont très corrélés ; c'est un problème donc difficile pour le démixage hyperspectral. Une étude approfondie des spectres mérite d'être réalisée pour une

compréhension précise des phénomènes physico-chimiques impactant le rendu et la qualité des pièces de bois.

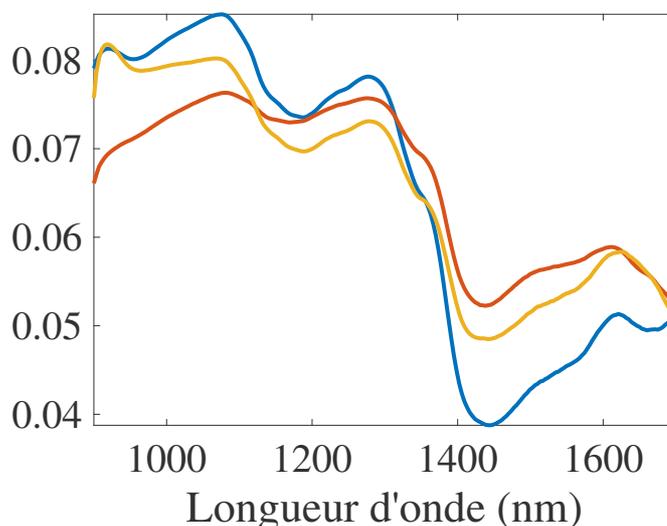


FIGURE 4.21 : Composition de la bibliothèque hyperspectrale **B** pour le traitement des images de pièces de bois

4.3.2 Traitement des pièces de bois à l'aide de OSS-ADMM

Comme nous l'avons montré dans le Chapitre 3, pour la même image hyperspectrale, le traitement réalisé à l'aide de l'algorithme OSS-ADMM est plus rapide que celui réalisé par OMDC-ADMM. Bien que OSS-ADMM met à jour conjointement les abondances et les *endmembers*, ces derniers sont approximativement connus grâce à l'étape d'apprentissage et par conséquent, la convergence de l'algorithme est plus rapide, ce qui le désigne comme un bon candidat pour le traitement en-ligne. Nous avons appliqué OSS-ADMM sur les neuf échantillons présentés précédemment, en utilisant la bibliothèque apprise à l'aide de l'algorithme OMDC-ADMM. Les paramètres relatifs à l'algorithme ont été fixés de la manière suivante : $N_{iter} = 10$, $\alpha = 0.9$, $\nu = 10^{-5}$, $\gamma = 0.001$, $\omega = 50$, $\delta = 10^{-15}$, $R = 3$ et $\rho = 0.001$. Dans le cas de notre application bois, le nombre de sources n'évolue pas au cours de l'image, elles sont toutes présentes simultanément dans des proportions plus ou moins importantes. Par conséquent, la parcimonie imposée par OSS-ADMM n'a pas réellement d'intérêt, ne serait-ce qu'elle régularise tout de même la solution. Les cartes d'abondance estimées par OSS-ADMM sont présentées sur les Figures 4.22 à 4.30.

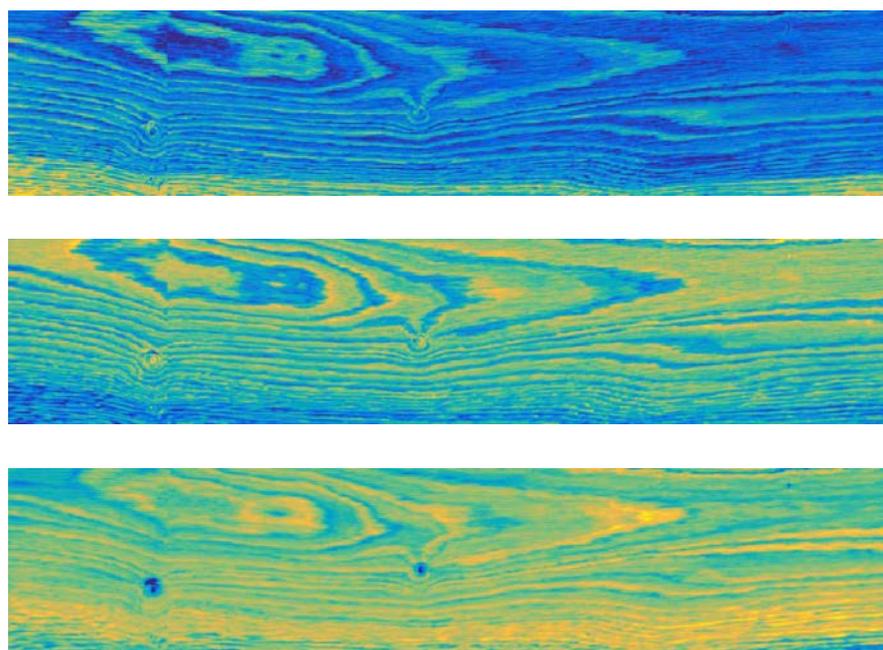


FIGURE 4.22 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 1

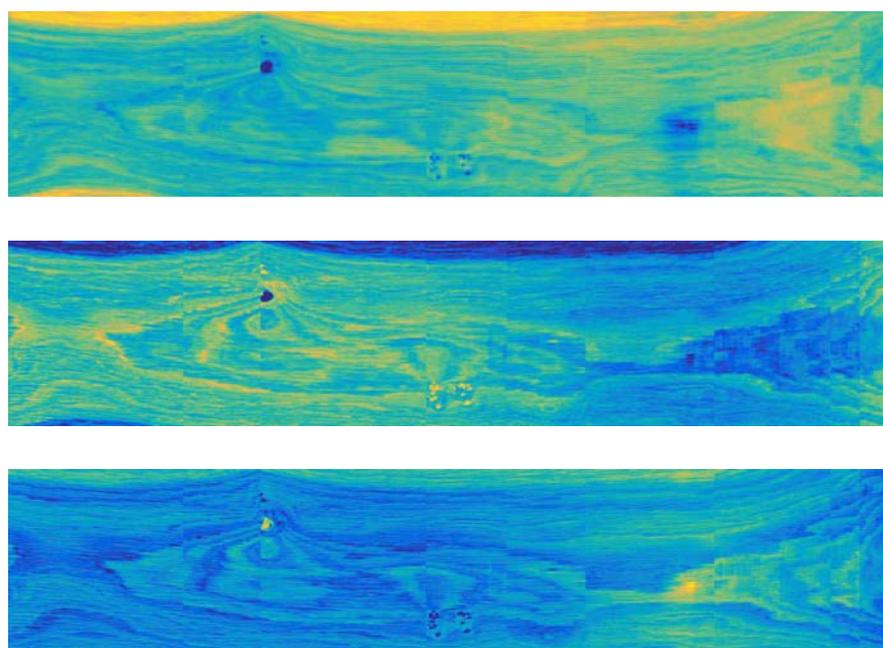


FIGURE 4.23 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 2

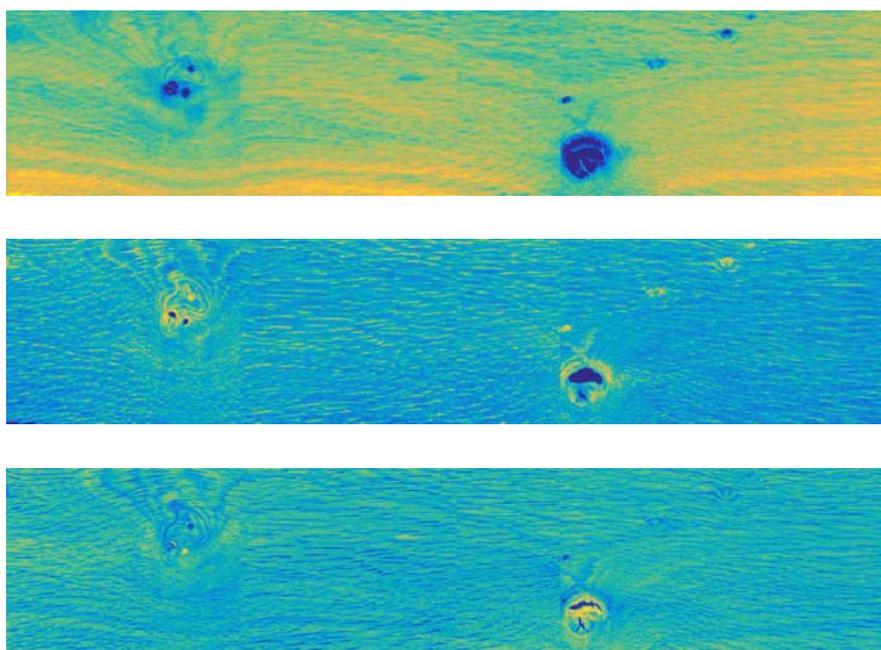


FIGURE 4.24 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 3

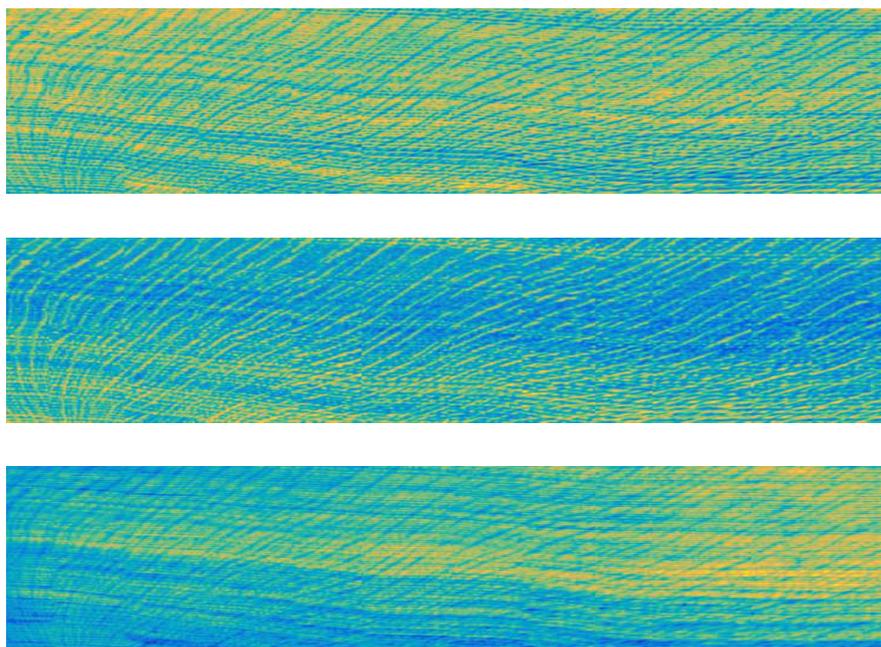


FIGURE 4.25 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 4

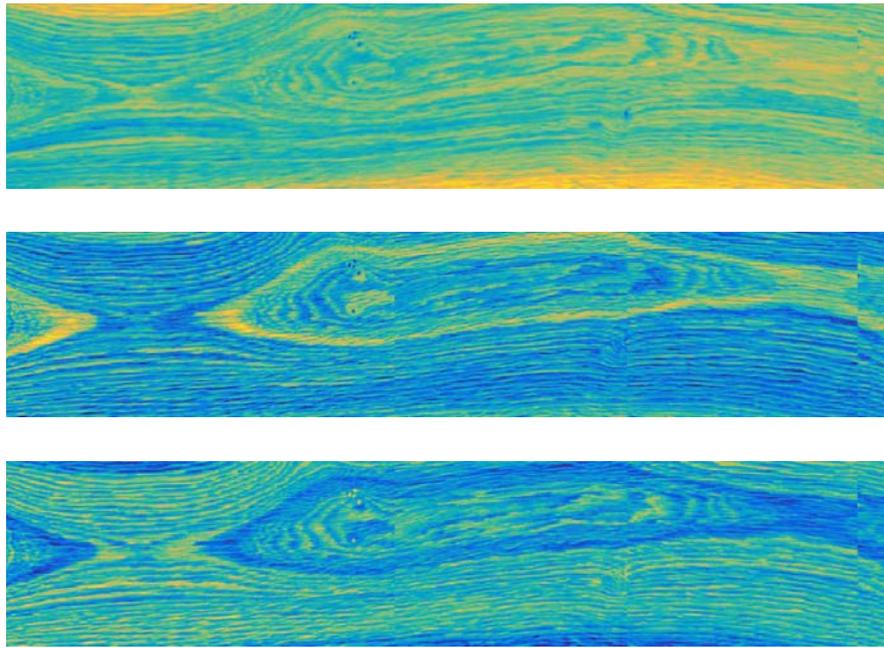


FIGURE 4.26 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 5

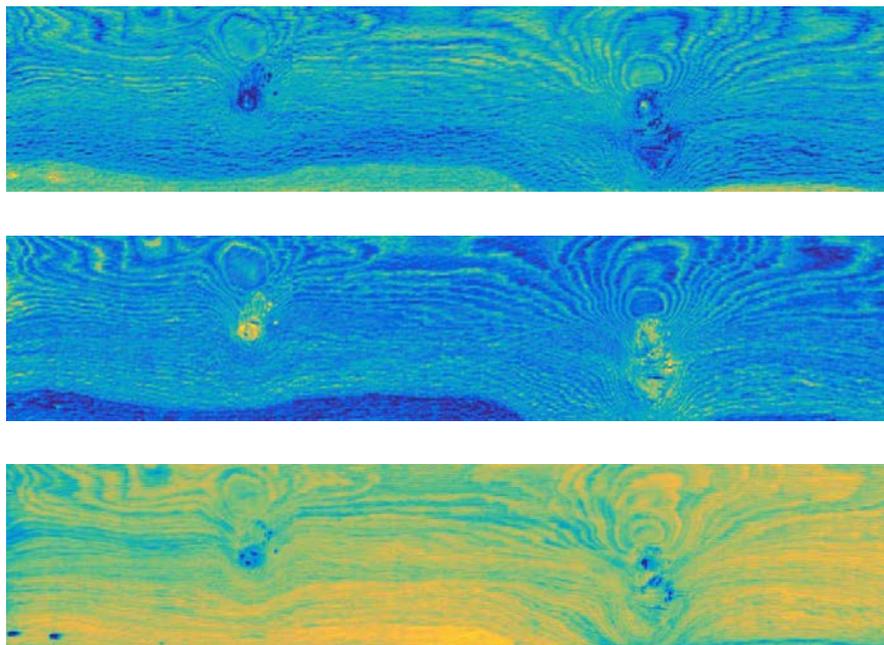


FIGURE 4.27 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 6

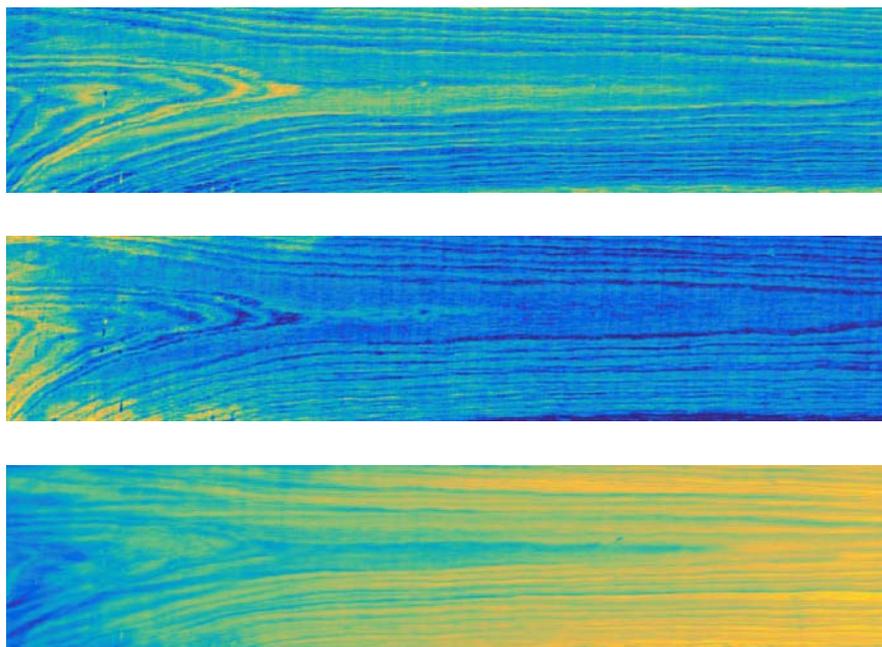


FIGURE 4.28 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 7

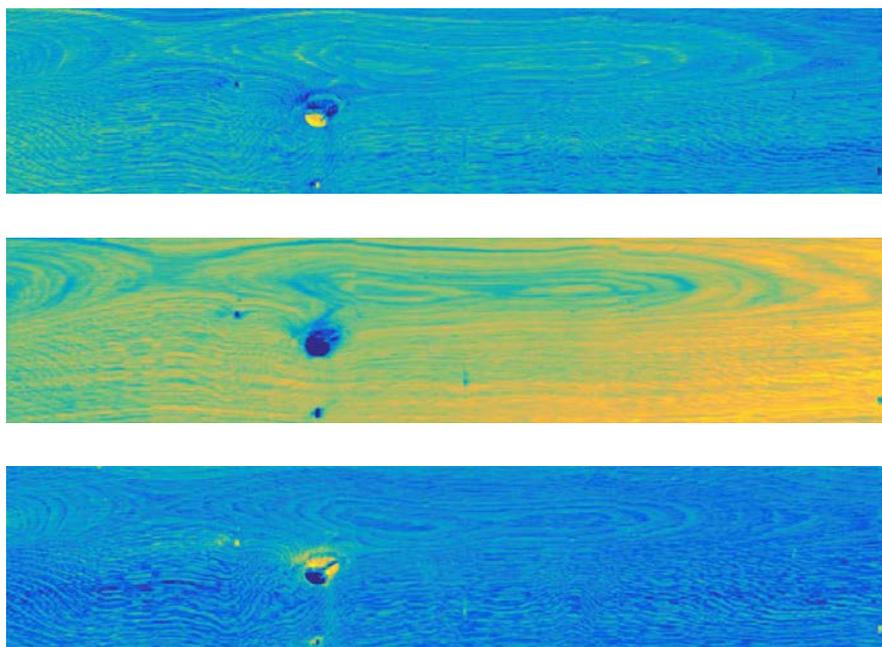


FIGURE 4.29 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 8

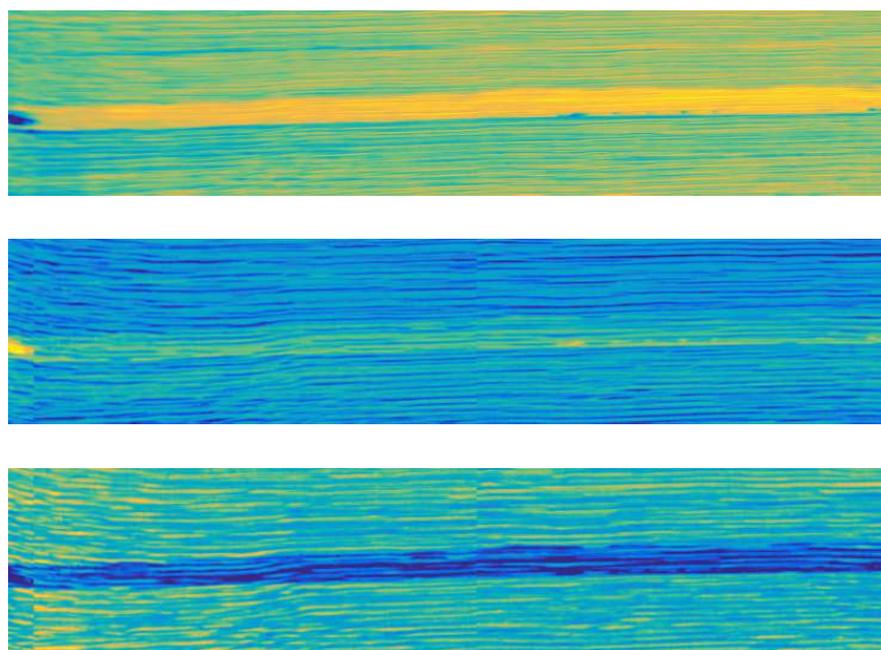


FIGURE 4.30 : Cartes d'abondance générées par OSS-ADMM pour l'échantillon 9

Nous constatons que OSS-ADMM génère des cartes d'abondance similaires à celles obtenues avec OMDC-ADMM. Les performances, en ce qui concerne la détection des singularités sont identiques au Tableau 4.1. Il semble donc que réaliser le démélange hyperspectral aveugle ou en semi-supervisé permet d'aboutir aux mêmes observations. Il y a tout de même un véritable intérêt à utiliser OSS-ADMM même dans le cas où le nombre de sources est constant tout au long du traitement. En effet, comme nous pouvons le constater dans le Tableau 4.3, les temps de traitement sont réduits en moyenne par trois, comparés aux temps de traitement de l'algorithme OMDC-ADMM (Tableau 4.2). Cela représente un point essentiel, notamment pour des applications industrielles en-ligne, où la rapidité est un élément clé.

	Temps (s)
Échantillon 1	2.18
Échantillon 2	1.56
Échantillon 3	2.06
Échantillon 4	1.04
Échantillon 5	2.03
Échantillon 6	2.09
Échantillon 7	2.20
Échantillon 8	1.91
Échantillon 9	0.91

TABLE 4.3 : Temps de traitement de chaque échantillon par OSS-ADMM

4.3.3 Segmentation des images hyperspectrales

Dans cette section, nous présentons les résultats obtenus quant à la segmentation des cartes d'abondance générées par OSS-ADMM. Précisons que la segmentation a été effectuée hors-ligne, une fois les images hyperspectrales acquises en totalité. L'implantation en-ligne du processus de segmentation fera l'objet des futures investigations. Au préalable, les cartes d'abondance ont été lissées par un filtre passe bas de taille 10×10 , puis segmentées par la méthode des *k-means* [Hartigan 79]. Le nombre de classes a été fixé expérimentalement pour chaque échantillon. Afin de comparer le degré de similarité entre les différentes observations, la distance Euclidienne a été utilisée. De plus, pour chaque carte d'abondance, la méthode des *k-means* a été réitérée cinq fois, dans le cas d'un mauvais choix pour le centre des classes. Les résultats sont illustrés sur les Figures 4.31 à 4.39.

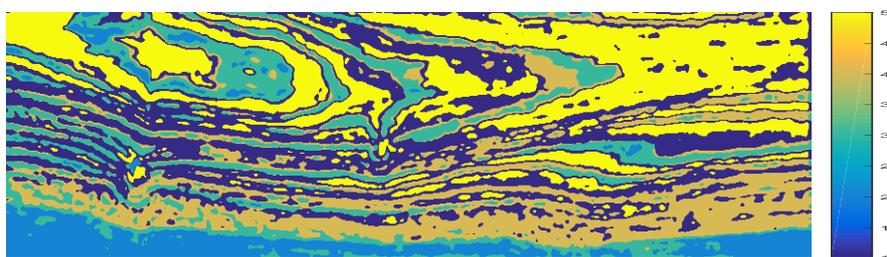


FIGURE 4.31 : Segmentation des cartes d'abondance de l'échantillon 1 : 5 classes

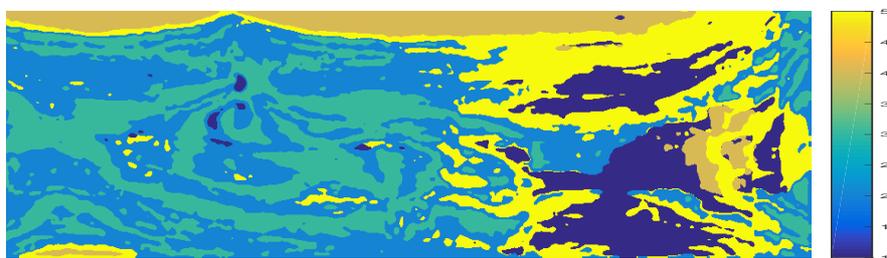


FIGURE 4.32 : Segmentation des cartes d'abondance de l'échantillon 2 : 5 classes

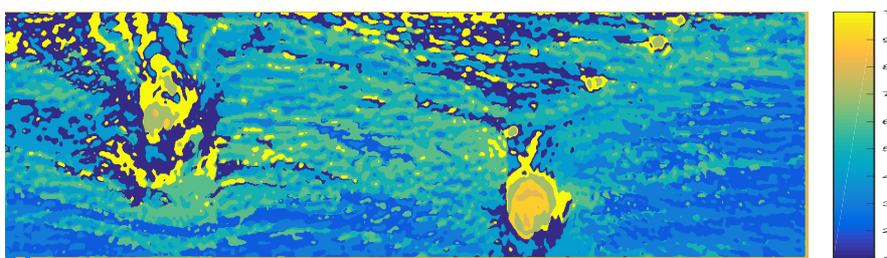


FIGURE 4.33 : Segmentation des cartes d'abondance de l'échantillon 3 : 10 classes

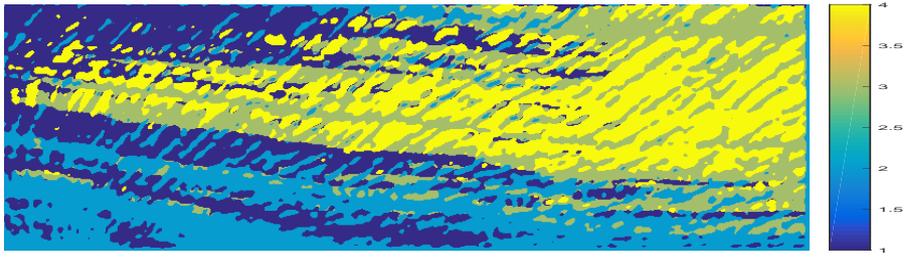


FIGURE 4.34 : Segmentation des cartes d'abondance de l'échantillon 4 : 4 classes

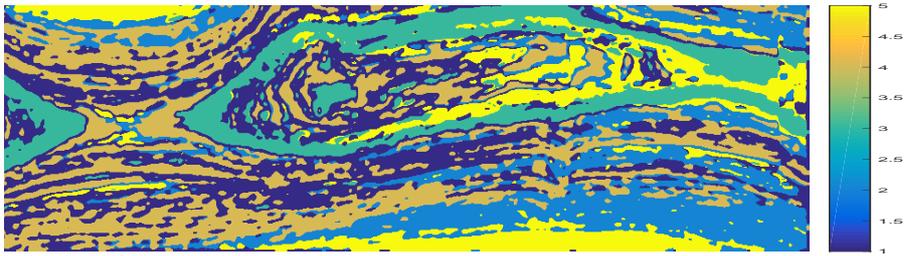


FIGURE 4.35 : Segmentation des cartes d'abondance de l'échantillon 5 : 5 classes

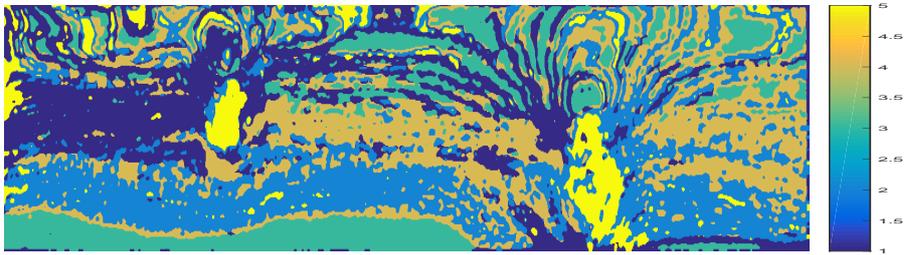


FIGURE 4.36 : Segmentation des cartes d'abondance de l'échantillon 6 : 5 classes

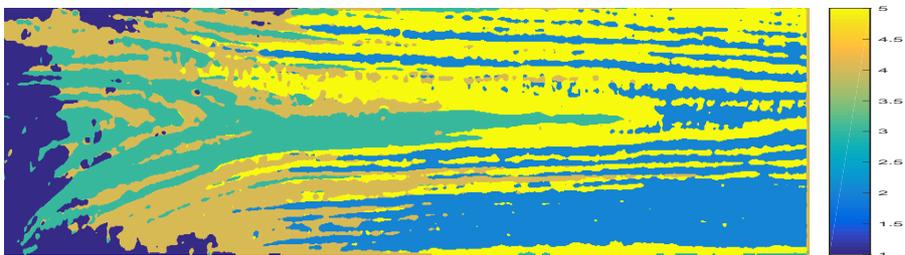


FIGURE 4.37 : Segmentation des cartes d'abondance de l'échantillon 7 : 5 classes

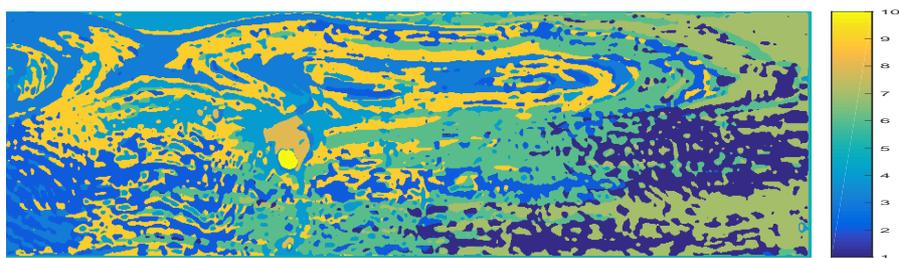


FIGURE 4.38 : Segmentation des cartes d'abondance de l'échantillon 8 : 10 classes

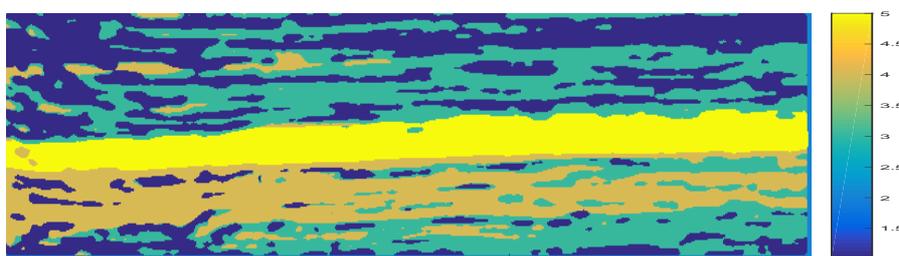


FIGURE 4.39 : Segmentation des cartes d'abondance de l'échantillon 9 : 5 classes

Nous observons que certaines classes correspondent parfaitement aux zones défectueuses recherchées, ce qui valide les performances de nos algorithmes de démélange hyperspectral en-ligne. Ci-dessous, nous avons répertorié, pour chaque échantillon de bois, les classes générées par la méthode des *k-means*, annotant les singularités recherchées :

- Échantillon 1. Classe 2 : aubier tâché. Classe 5 : discoloration.
- Échantillon 2. Classe 4 : aubier. Classes 5 et 1 : trace baguette.
- Échantillon 3. Classes 2 et 3 : aubier et veine foncée. Classe 7 : nœud sain. Classes 8 et 9 : nœud noir.
- Échantillon 4. Classes 2 et 3 : maille.
- Échantillon 5. Classe 5 : aubier.
- Échantillon 6. Classe 3 : aubier tâché. Classe 5 : fente et nœud patte de chat.
- Échantillon 7. Classe 1 : tâche de lubrifiant. Classe 5 : aubier.
- Échantillon 8. Classe 10 : nœud pourri.
- Échantillon 9. Classe 5 : veine foncée.

4.4 Conclusions

Divers échantillons de bois annotés par des experts ont été traités en deux étapes distinctes : la première étape a consisté en un démixage hyperspectral aveugle, en utilisant l'algorithme OMDC-ADMM. Les résultats observés à travers les cartes d'abondance ont révélé que l'imagerie hyperspectrale proche infrarouge combinée à notre algorithme de démixage en-ligne est un duo puissant pour la détection de nombreuses singularités, difficilement visibles à l'œil nu. De plus, n'ayant aucune connaissance *a priori* sur les signatures spectrales à extraire, cette étape nous a permis de constituer une bibliothèque de référence des *endmembers*. Dans la seconde étape, nous avons exploité cette bibliothèque afin de réaliser un démixage hyperspectral semi-supervisé, par le biais de l'algorithme OSS-ADMM. Ce dernier est également très performant en ce qui concerne la localisation des diverses singularités. Par ailleurs, il permet de réduire considérablement le temps de traitement de chaque image comparé à la méthode OMDC-ADMM ; ainsi, il semble judicieux de s'orienter vers une telle méthode pour des applications industrielles en temps réel. Enfin, la dernière étape a été réservée à la segmentation des images hyperspectrales par la méthode des *k-means*, en utilisant les cartes d'abondance estimées à l'étape précédente. Nous avons remarqué que la répartition des différentes classes, pour chaque échantillon, est tout à fait cohérente avec les zones défectueuses localisées au préalable par expertise. Ce sont des résultats encourageants et les perspectives ouvertes par ces travaux sont nombreuses. L'une d'entre elles est d'intégrer nos méthodes dans des classifieurs plus complexes, pour l'analyse rapide des données d'un système de production de grande dimension, en temps réel. Ce travail fait l'objet de la thèse de Radouan Dahbi, également réalisée dans le cadre du projet ANR-OPTIFIN.

Chapitre 5

Conclusions et perspectives

Dans cette thèse, nous nous sommes intéressés au démélange séquentiel d'images hyperspectrales *pushbroom* pour des applications industrielles. La première contribution de cette thèse a été axée sur le développement de méthodes en-ligne inspirées de l'approche INMF [Bucak 09]. Fondée sur une fonction objectif avec pondération entre les nouveaux et les anciens échantillons, INMF permet la prise en compte de la variabilité spectrale et garantit une complexité de calcul constante par échantillon. Cependant, elle présente aussi un certain nombre d'inconvénients : les mises à jour multiplicatives peuvent altérer l'estimation des paramètres et souffrir d'une convergence lente qui n'est pas adaptée au suivi temps-réel des processus industriels rapides. Par ailleurs, la solution de la NMF (hors-ligne ou en-ligne) est généralement non-unique, ce qui engendre des difficultés quant à l'interprétabilité des résultats. Pour pallier le problème de la non-unicité, deux contraintes ont été proposées : la contrainte de dispersion minimale (MDC) et la contrainte de volume minimal (MVC), qui permettent la préservation du rang entre les tranches de l'image hyperspectrale. De plus, afin de générer un traitement rapide de l'image, en accord avec les contraintes industrielles, nous avons utilisé deux méthodes d'optimisation, basées sur le gradient optimal de Nesterov et ADMM, ce qui nous a permis de proposer trois algorithmes en-ligne : OMVC-Ne, OMDC-Ne et OMDC-ADMM. Des simulations sur des images synthétiques et réelles ont montré que OMDC-ADMM surpasse les algorithmes fondés sur les règles de mise à jour multiplicatives et le gradient optimal de Nesterov, en ce qui concerne la vitesse de convergence et le temps de traitement.

La seconde contribution de cette thèse a porté sur l'estimation de l'hyperparamètre relatif à la contrainte de dispersion ou de volume minimal. En formulant le problème de démélange en tant que problème d'optimisation bi-objectif, deux méthodes ont été proposées pour estimer automatiquement le paramètre de régularisation : le critère de distance minimale (MDistC) et le critère de courbure maximale (MCC). En se basant sur des images hyperspectrales simulées, les approches proposées ont été comparées, confirmant ainsi les meilleures performances de MCC par rapport à MDistC. Les performances de MDistC dépendent fortement du choix du point de référence pour la mesure de distance. La supériorité de l'approche MCC sur MDistC a été validée ensuite sur une image hyperspectrale réelle avec vérité terrain. Le critère d'estimation de l'hyperparamètre proposé peut être appliqué au démélange hyperspectral en-ligne, en estimant d'abord le paramètre sur un échantillon typique puis en l'utilisant pour l'ensemble des échantillons suivants.

La troisième contribution de ces travaux a traité du problème de l'évolution possible du nombre de

sources au cours du traitement. En utilisant une bibliothèque hyperspectrale et des termes de régularisation imposant la parcimonie, nous avons proposé l'algorithme en-ligne semi-supervisé, OSS-ADMM, capable d'estimer pour chaque nouvel échantillon les *endmembers* actifs uniquement. Des tests sur des données simulées et réelles ont permis d'illustrer sa capacité de suivi du rang au cours du traitement d'une image. Enfin, son utilisation sur une application industrielle originale a confirmé son avantage quant à l'estimation des paramètres d'intérêt lorsque le nombre de sources varie dans le temps (entre les échantillons).

Enfin, la quatrième contribution a été réservée à la caractérisation du matériau bois. Divers échantillons, annotés par des experts, ont été traités par OMDC-ADMM et OSS-ADMM. Les résultats obtenus par les deux algorithmes en ce qui concerne la détection des singularités sont prometteurs. De plus, OSS-ADMM s'est révélé être un algorithme de traitement très rapide, représentant ainsi un excellent candidat pour le traitement temps-réel sur les lignes de production industrielles.

La première perspective de ces travaux est l'analyse théorique du comportement transitoire des algorithmes en-ligne proposés. En s'inspirant des travaux menés par [Zhao 16], l'objectif est d'analyser la vitesse de convergence vers un état stationnaire ainsi que de la poursuite des variations spectrales.

La deuxième perspective est d'intégrer la diversité de polarisation à nos images hyperspectrales. L'imagerie hyperspectrale polarisée suscite un intérêt croissant en imagerie optique de par sa capacité à fournir des images à fort contraste (voir *e.g.* [Usenik 12, Qu 16]). En particulier, elle permet de discriminer les matériaux en fonction de leurs propriétés d'absorption de la lumière. Dans le contexte de l'imagerie hyperspectrale PIR, l'information de polarisation permet d'accéder à des informations relatives à la conformation moléculaire, c'est-à-dire à la façon dont les molécules sont arrangées spatialement. C'est donc une technique qui présente un réel intérêt pour l'analyse des matériaux organiques de type polymère tels que le bois. La diversité de polarisation peut être utilisée comme une dimension supplémentaire des données acquises ; cela permet d'aboutir à des modèles tensoriels de type CANDECOMP/PARAFAC non-négatifs (voir *e.g.* [Miron 11, Guo 12a, Guo 12b]). Néanmoins, ces approches présentent certaines limitations pour le traitement de données polarisées puisque la polarisation ne varie pas avec les longueurs d'onde. Ainsi, l'objectif est de s'inspirer de la méthode proposée par [Flamant 19] qui permet de palier ce problème.

La troisième perspective consiste en le développement de nouveaux modèles tensoriels pour le traitement de séries temporelles d'images hyperspectrales, en s'inspirant du modèle PARAFAC2 [Henrot 16]. Les conditions d'identifiabilité devront être étudiées afin de définir un protocole expérimental permettant d'identifier sans ambiguïté les paramètres d'intérêt.

Enfin, une quatrième direction pour les futurs travaux porte sur la complétion de matrices pour la prédiction de données manquantes au sein des images hyperspectrales (voir *e.g.* [Xu 12]). L'implantation en-ligne de ces méthodes pourrait permettre de diminuer considérablement le temps d'acquisition des images hyperspectrales et d'accélérer ainsi la vitesse des processus industriels.

Bibliographie

- [Abd-Elrahman 11] Amr Abd-Elrahman, Matthew Croxton, Roshan Pande-Chettri, Gурpal S Toor, Scot Smith & Jeffrey Hill. *In situ estimation of water quality parameters in freshwater aquaculture ponds using hyperspectral imaging system*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 4, pages 463–472, 2011.
- [Amari 98] Shun-Ichi Amari. *Natural gradient works efficiently in learning*. Neural Computation, vol. 10, no. 2, pages 251–276, 1998.
- [Ang 18] Andersen M S Ang & Nicolas Gillis. *Volume regularized non-negative matrix factorizations*. In 9th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, pages 1–5, 2018.
- [Baillères 02] Henri Baillères, Fabrice Davrieux & Frédérique Ham-Pichavant. *Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program*. Annals of Forest Science, vol. 59, no. 5-6, pages 479–490, 2002.
- [Belge 02] Murat Belge, Misha E Kilmer & Eric L Miller. *Efficient determination of multiple regularization parameters in a generalized L-curve framework*. Inverse Problems, vol. 18, no. 4, page 1161, 2002.
- [Belmerhnia 17] Leila Belmerhnia. *Approches parcimonieuses pour la sélection de variables et la classification : application à la spectroscopie IR de déchets de bois*. PhD thesis, Université de Lorraine, 2017.
- [Berman 04] Mark Berman, Harri Kiiveri, Ryan Lagerstrom, Andreas Ernst, Rob Dunne & Jonathan F Huntington. *ICE: a statistical approach to identifying endmembers in hyperspectral images*. IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 10, pages 2085–2095, 2004.
- [Berné 10] Olivier Berné, Alexander Tielens, Paolo Pilleri & Christine Joblin. *Non-negative matrix factorization pansharpening of hyperspectral data: an application to mid infrared astronomy*. In 2nd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, pages 1–4, 2010.
- [Bertaud 04] Frédérique Bertaud & Bjarne Holmbom. *Chemical composition of earlywood and latewood in Norway spruce heartwood, sapwood and transition zone wood*. Wood Science and Technology, vol. 38, no. 4, pages 245–256, 2004.

- [Berthold 98] Jesper Berthold, Ralf J O Olsson & Lennart Salmén. *Water sorption to hydroxyl and carboxylic acid groups in carboxymethylcellulose (CMC) studied with nir spectroscopy*. Cellulose, vol. 5, no. 4, pages 281–298, 1998.
- [Bertsekas 97] Dimitri P Bertsekas. *Non-linear programming*. Journal of the Operational Research Society, vol. 48, no. 3, pages 334–334, 1997.
- [Bioucas-Dias 09] José M Bioucas-Dias. *A variable splitting augmented Lagrangian approach to linear spectral unmixing*. In 1st IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, pages 1–4, 2009.
- [Bioucas-Dias 10] José M Bioucas-Dias & Mário A T Figueiredo. *Alternating direction algorithms for constrained sparse regression: application to hyperspectral unmixing*. In 2nd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, pages 1–4, 2010.
- [Bioucas-Dias 11] José M Bioucas-Dias & Antonio Plaza. *An overview on hyperspectral unmixing: geometrical, statistical, and sparse regression based approaches*. In IEEE International Geoscience and Remote Sensing Symposium, pages 1135–1138, 2011.
- [Bioucas-Dias 12] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader & Jocelyn Chanussot. *Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 2, pages 354–379, 2012.
- [Boardman 95] Joseph W Boardman, Fred A Kruse & Robert O Green. *Mapping target signatures via partial unmixing of AVIRIS data*. 1995.
- [Boulais 17] Axel Boulais. *Méthodes de séparation aveugle de sources et application à l'imagerie hyperspectrale en astrophysique*. PhD thesis, Université Paul Sabatier-Toulouse III, 2017.
- [Boyd 04] Stephen Boyd & Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [Boyd 11] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato & Jonathan Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Foundations and Trends® in Machine Learning, vol. 3, no. 1, pages 1–122, 2011.
- [Bucak 09] Serhat S Bucak & Bilge Günsel. *Incremental subspace learning via non-negative matrix factorization*. Pattern Recognition, vol. 42, no. 5, pages 788–797, 2009.
- [Cao 07] Bin Cao, Dou Shen, Jian-Tao Sun, Xuanhui Wang, Qiang Yang & Zheng Chen. *Detect and track latent factors with on-line non-negative matrix factorization*. In 12th Proceedings of the International Joint Conferences on Artificial Intelligence, volume 7, pages 2689–2694, 2007.

- [Carteret 99] Cédric Carteret. *Étude, par spectroscopie dans le proche infrarouge, et modélisation des structures de surface et de l'hydratation de silices amorphes*. PhD thesis, Nancy 1, 1999.
- [Chan 09] Tsung-Han Chan, Chong-Yung Chi, Yu-Min Huang & Wing-Kin Ma. *A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing*. IEEE Transactions on Signal Processing, vol. 57, no. 11, pages 4418–4432, 2009.
- [Chang 00] Chein-I Chang & Daniel C Heinz. *Constrained subpixel target detection for remotely sensed imagery*. IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 3, pages 1144–1159, 2000.
- [Chen 96] Yud-Ren Chen, Roy W Huffman, Bosoon Park & Minh Nguyen. *Transportable spectrophotometer system for on-line classification of poultry carcasses*. Applied Spectroscopy, vol. 50, no. 7, pages 910–916, 1996.
- [Chen 07] Qing-Min Chen, Zhoujian Hu, Hou-Min Chang & Bailian Li. *Micro analytical methods for determination of compression wood content in loblolly pine*. Journal of Wood Chemistry and Technology, vol. 27, no. 3-4, pages 169–178, 2007.
- [Cichocki 09] Andrzej Cichocki, Rafal Zdunek, Anh H Phan & Shun-ichi Amari. *Non-negative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [Colthup 12] Norman Colthup. *Introduction to infrared and raman spectroscopy*. Elsevier, 2012.
- [Craig 94] Maurice D Craig. *Minimum-volume transforms for remotely sensed data*. IEEE Transactions on Geoscience and Remote Sensing, vol. 32, no. 3, pages 542–552, 1994.
- [De Viguerie 16] Laurence De Viguerie, Matthias Alfeld & Philippe Walter. *La lumière pour une imagerie chimique des peintures*. Reflets de la physique, no. 47-48, pages 106–111, 2016.
- [Deb 01] Kalyanmoy Deb. *Multi-objective optimization using evolutionary algorithms*, volume 16. John Wiley & Sons, 2001.
- [Dobigeon 08] Nicolas Dobigeon, Jean-Yves Tourneret & Chein-I Chang. *Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery*. IEEE Transactions on Signal Processing, vol. 56, no. 7, pages 2684–2695, 2008.
- [Donoho 04] David Donoho & Victoria Stodden. *When does non-negative matrix factorization give a correct decomposition into parts?* In 16th International Conference on Neural Information Processing Systems, pages 1141–1148, 2004.
- [Drumetz 16] Lucas Drumetz. *Endmember Variability in hyperspectral image unmixing*. PhD thesis, Université Grenoble Alpes, 2016.

- [Eches 10] Olivier Eches, Nicolas Dobigeon, Corinne Mailhes & Jean-Yves Tourneret. *Bayesian estimation of linear mixtures using the normal compositional model. Application to hyperspectral imagery*. IEEE Transactions on Image Processing, vol. 19, no. 6, pages 1403–1413, 2010.
- [Fazel 03] Maryam Fazel, Haitham Hindi & Stephen P Boyd. *Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices*. In Proceedings of the American Control Conference, volume 3, pages 2156–2162, 2003.
- [Feng 12] Yao-Ze Feng & Da-Wen Sun. *Application of hyperspectral imaging in food safety inspection and control: a review*. Critical Reviews in Food Science and Nutrition, vol. 52, no. 11, pages 1039–1058, 2012.
- [Févotte 11] Cédric Févotte & Jérôme Idier. *Algorithms for non-negative matrix factorization with the β -divergence*. Neural Computation, vol. 23, no. 9, pages 2421–2456, 2011.
- [Flamant 19] Julien Flamant, Sebastian Miron & David Brie. *Quaternion non-negative matrix factorization: definition, uniqueness and algorithm*. arXiv preprint arXiv:1903.10593, 2019.
- [Fu 15] Xiao Fu, Wing-Kin Ma, Kejun Huang & Nicholas D Sidiropoulos. *Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain*. IEEE Transactions on Signal Processing, vol. 63, no. 9, pages 2306–2320, 2015.
- [Fu 16] Xiao Fu, Kejun Huang, Bo Yang, Wing-Kin Ma & Nicholas D Sidiropoulos. *Robust volume minimization-based matrix factorization for remote sensing and document clustering*. IEEE Transactions on Signal Processing, vol. 64, no. 23, pages 6254–6268, 2016.
- [Fu 18a] Xiao Fu, Kejun Huang & Nicholas D Sidiropoulos. *On identifiability of non-negative matrix factorization*. IEEE Signal Processing Letters, vol. 25, no. 3, pages 328–332, 2018.
- [Fu 18b] Xiao Fu, Kejun Huang, Nicholas D Sidiropoulos & Wing-Kin Ma. *Non-negative matrix factorization for signal and data analytics: identifiability, algorithms, and applications*. arXiv preprint arXiv:1803.01257, 2018.
- [Genot 14] Valérie Genot, Laurent Bock, Pierre Dardenne & Gilles Colinet. *L'intérêt de la spectroscopie proche infrarouge en analyse de terre (synthèse bibliographique)*. Biotechnologie, Agronomie, Société et Environnement, vol. 18, no. 2, pages 247–261, 2014.
- [Gezici-Koç 17] Özlem Gezici-Koç, Sebastiaan J F Erich, Hendrik P Huinink, Leendert G J van der Ven & Olaf C G Adan. *Bound and free water distribution in wood during water uptake and drying as measured by 1D magnetic resonance imaging*. Cellulose, vol. 24, no. 2, pages 535–553, 2017.

- [Gierlinger 02] Notburga Gierlinger, Manfred Schwanninger, Barbara Hinterstoisser & Rupert Wimmer. *Rapid determination of heartwood extractives in Larix sp. by means of Fourier transform near infrared spectroscopy*. Journal of Near Infrared Spectroscopy, vol. 10, no. 3, pages 203–214, 2002.
- [Gillis 12] Nicolas Gillis & François Glineur. *Accelerated multiplicative updates and hierarchical ALS algorithms for non-negative matrix factorization*. Neural Computation, vol. 24, no. 4, pages 1085–1105, 2012.
- [Girod 08] Denis Girod. *Détermination de la maturité des avocats hass par imagerie hyperspectrale*. PhD thesis, École de technologie supérieure, 2008.
- [Goetz 09] Alexander F H Goetz. *Three decades of hyperspectral remote sensing of the Earth: a personal view*. Remote Sensing of Environment, vol. 113, pages S5–S16, 2009.
- [Goldstein 14] Tom Goldstein, Brendan O’Donoghue, Simon Setzer & Richard Baraniuk. *Fast alternating direction optimization methods*. SIAM Journal on Imaging Sciences, vol. 7, no. 3, pages 1588–1623, 2014.
- [Golub 79] Gene H Golub, Michael Heath & Grace Wahba. *Generalized cross-validation as a method for choosing a good ridge parameter*. Technometrics, vol. 21, no. 2, pages 215–223, 1979.
- [Gonzalez 05] Edward F Gonzalez & Yin Zhang. *Accelerating the Lee-Seung algorithm for non-negative matrix factorization*. Rapport technique, 2005.
- [Gowen 07] Aoife A Gowen, Colm P O’Donnell, Patrick J Cullen, Gérard Downey & Jesus M Frias. *Hyperspectral imaging—an emerging process analytical tool for food quality and safety control*. Trends in Food Science & Technology, vol. 18, no. 12, pages 590–598, 2007.
- [Green 88] Andrew A Green, Mark Berman, Paul Switzer & Maurice D Craig. *A transformation for ordering multispectral data in terms of image quality with implications for noise removal*. IEEE Transactions on Geoscience and Remote Sensing, vol. 26, no. 1, pages 65–74, 1988.
- [Grippo 00] Luigi Grippo & Marco Sciandrone. *On the convergence of the block non-linear Gauss–Seidel method under convex constraints*. Operations Research Letters, vol. 26, no. 3, pages 127–136, 2000.
- [Guan 12a] Naiyang Guan, Dacheng Tao, Zhigang Luo & Bo Yuan. *NeNMF: an optimal gradient method for non-negative matrix factorization*. IEEE Transactions on Signal Processing, vol. 60, no. 6, pages 2882–2898, 2012.
- [Guan 12b] Naiyang Guan, Dacheng Tao, Zhigang Luo & Bo Yuan. *On-line non-negative matrix factorization with robust stochastic approximation*. IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 7, pages 1087–1099, 2012.

- [Guo 12a] Xijing Guo, Sebastian Miron & David Brie. *The effect of polarization separation on the performance of CANDECOMP/PARAFAC-based vector sensor array processing*. *Physical Communication*, vol. 5, no. 4, pages 289–295, 2012.
- [Guo 12b] Xijing Guo, Sebastian Miron, David Brie & Alwin Stegeman. *Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings*. *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 1, pages 111–129, 2012.
- [Hagen 12] Nathan A Hagen, Liang S Gao, Tomasz S Tkaczyk & Robert T Kester. *Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems*. *Optical Engineering*, vol. 51, no. 11, page 111702, 2012.
- [Hajinezhad 16] Davood Hajinezhad, Tsung-Hui Chang, Xiangfeng Wang, Qingjiang Shi & Mingyi Hong. *Non-negative matrix factorization using ADMM: algorithm and convergence analysis*. In 41th IEEE International Conference on Acoustics, Speech and Signal Processing, pages 4742–4746, 2016.
- [Halimi 15] Abderrahim Halimi, Nicolas Dobigeon & Jean-Yves Tourneret. *Unsupervised unmixing of hyperspectral images accounting for endmember variability*. *IEEE Transactions on Image Processing*, vol. 24, no. 12, pages 4904–4917, 2015.
- [Han 09] Lixing Han, Michael Neumann & Upendra Prasad. *Alternating projected Barzilai-Borwein methods for non-negative matrix factorization*. *Electronic Transactions on Numerical Analysis*, vol. 36, no. 6, pages 54–82, 2009.
- [Han 17] Xu Han, Laurent Albera, Amar Kachenoura, Lotfi Senhadji & Huazhong Shu. *Low rank canonical polyadic decomposition of tensors based on group sparsity*. In 25th IEEE European Signal Processing Conference, pages 668–672, 2017.
- [Hanke 96] Martin Hanke. *Limitations of the L-curve method in ill-posed problems*. *BIT Numerical Mathematics*, vol. 36, no. 2, pages 287–301, 1996.
- [Hansen 92] Per C Hansen. *Analysis of discrete ill-posed problems by means of the L-curve*. *SIAM Review*, vol. 34, no. 4, pages 561–580, 1992.
- [Hansen 99] Per Christian Hansen. *The L-curve and its use in the numerical treatment of inverse problems*. Department of Mathematical Modelling, Technical University of Denmark, 1999.
- [Harsanyi 94] Joseph C Harsanyi & C-I Chang. *Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach*. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pages 779–785, 1994.
- [Hartigan 79] John A Hartigan & Manchek A Wong. *Algorithm AS 136: a k-means clustering algorithm*. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pages 100–108, 1979.

- [Hege 04] E Keith Hege, Dan O’Connell, William Johnson, Shridhar Basty & Eustace L Dereniak. *Hyperspectral imaging for astronomy and space surveillance*. In *Imaging Spectrometry IX*, volume 5159, pages 380–392. International Society for Optics and Photonics, 2004.
- [Heinz 01] Daniel C Heinz & Chein-I Chang. *Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery*. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pages 529–545, 2001.
- [Henrot 16] Simon Henrot, Jocelyn Chanussot & Christian Jutten. *Dynamical spectral unmixing of multitemporal hyperspectral images*. *IEEE Transactions on Image Processing*, vol. 25, no. 7, pages 3219–3232, 2016.
- [Hoffmeyer 95] Preben Hoffmeyer & JG Pedersen. *Evaluation of density and strength of Norway spruce wood by near infrared reflectance spectroscopy*. *Holz als Roh-und werkstoff*, vol. 53, no. 3, pages 165–170, 1995.
- [Hoyer 02] Patrik O Hoyer. *Non-negative sparse coding*. In *2th IEEE Workshop on Neural Networks for Signal Processing*, pages 557–565, 2002.
- [Huang 14] Kejun Huang, Nicholas D Sidiropoulos & Ananthram Swami. *Non-negative matrix factorization revisited: uniqueness and algorithm for symmetric decomposition*. *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pages 211–224, 2014.
- [Huang 16] Kejun Huang, Nicholas D Sidiropoulos & Athanasios P Liavas. *A flexible and efficient algorithmic framework for constrained matrix and tensor factorization*. *IEEE Transaction Signal Processing*, vol. 64, no. 19, pages 5052–5065, 2016.
- [Huck 10] Alexis Huck, Mireille Guillaume & Jacques Blanc-Talon. *Minimum dispersion constrained non-negative matrix factorization to unmix hyperspectral data*. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pages 2590–2602, 2010.
- [Hyvärinen 00] Aapo Hyvärinen & Erkki Oja. *Independent component analysis: algorithms and applications*. *Neural Networks*, vol. 13, no. 4-5, pages 411–430, 2000.
- [Ifarraguerri 99] Agustin Ifarraguerri & Chein-I Chang. *Multispectral and hyperspectral image analysis with convex cones*. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pages 756–770, 1999.
- [Iordache 11] Marian-Daniel Iordache, José M Bioucas-Dias & Antonio Plaza. *Sparse unmixing of hyperspectral data*. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pages 2014–2039, 2011.
- [Iordache 13] Marian-Daniel Iordache, José M Bioucas-Dias & Antonio Plaza. *Collaborative sparse regression for hyperspectral unmixing*. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pages 341–354, 2013.

- [Jia 09] Sen Jia & Yuntao Qian. *Constrained non-negative matrix factorization for hyperspectral unmixing*. IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 1, pages 161–173, 2009.
- [Jones 05] Paul D Jones, Laurence R Schimleck, Gary F Peter, Richard F Daniels & Alexander Clark. *Non-destructive estimation of Pinus taeda L tracheid morphological characteristics for samples from a wide range of sites in Georgia*. Wood Science and Technology, vol. 39, no. 7, page 529, 2005.
- [Kaufman 97] Linda Kaufman & Arnold Neumaier. *Regularization of ill-posed problems by envelope guided conjugate gradients*. Journal of Computational and Graphical Statistics, vol. 6, no. 4, pages 451–463, 1997.
- [Kelley 04] Stephen S Kelley, Timothy G Rials, Rebecca Snell, Leslie H Groom & Amie Sluiter. *Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood*. Wood Science and Technology, vol. 38, no. 4, pages 257–276, 2004.
- [Keshava 02] Nirmal Keshava & John F Mustard. *Spectral unmixing*. IEEE Signal Processing Magazine, vol. 19, no. 1, pages 44–57, 2002.
- [Kim 07] Dongmin Kim, Suvrit Sra & Inderjit S Dhillon. *Fast Newton-type methods for the least squares non-negative matrix approximation problem*. In 7th Proceedings of the SIAM International Conference on Data Mining, pages 343–354, 2007.
- [Kohan 12] Neil J Kohan, Brian K Via & Steven E Taylor. *Prediction of strand feedstock mechanical properties with near infrared spectroscopy*. Bioresources, vol. 7, no. 3, pages 2996–3007, 2012.
- [Laurberg 08] Hans Laurberg, Mads G Christensen, Mark D Plumbley, Lars K Hansen & Søren H Jensen. *Theorems on positive data: on the uniqueness of NMF*. Computational Intelligence and Neuroscience, vol. 2008, 2008.
- [Leblon 13] Brigitte Leblon, Oluwatosin Adedipe, Guillaume Hans, Ataollah Haddadi, Satoru Tsuchikawa, James Burger, Rod Stirling, Zarin Pirouz, Kevin Groves, Joseph Nader & Armand LaRocque. *A review of near infrared spectroscopy for monitoring moisture content and density of solid wood*. The Forestry Chronicle, vol. 89, no. 5, pages 595–606, 2013.
- [Lee 99] Daniel D Lee & Hyunjune S Seung. *Learning the parts of objects by non-negative matrix factorization*. Nature, vol. 401, no. 6755, page 788, 1999.
- [Lee 01] Daniel D Lee & Hyunjune S Seung. *Algorithms for non-negative matrix factorization*. In 13th Proceedings of Advances in Neural Information Processing Systems, pages 556–562, 2001.
- [Lefevre 11] Augustin Lefevre, Francis Bach & Cédric Févotte. *On-line algorithms for non-negative matrix factorization with the Itakura-Saito divergence*. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 313–316, 2011.

- [Leplat 19] Valentin Leplat, Andersen M S Ang & Nicolas Gillis. *Minimum-volume rank-deficient non-negative matrix factorizations*. In 44th IEEE International Conference on Acoustics, Speech and Signal Processing, pages 3402–3406, 2019.
- [Li 08] Jun Li & José M Bioucas-Dias. *Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data*. In IEEE International Geoscience and Remote Sensing Symposium, volume 3, pages III–250, 2008.
- [Li 12] Jun Li, José M Bioucas-Dias & Antonio Plaza. *Collaborative non-negative matrix factorization for remotely sensed hyperspectral unmixing*. In IEEE International Geoscience and Remote Sensing Symposium, pages 3078–3081, 2012.
- [Li 13] Qingli Li, Xiaofu He, Yiting Wang, Hongying Liu, Dongrong Xu & Fangmin Guo. *Review of spectral imaging technology in biomedical engineering: achievements and challenges*. Journal of Biomedical Optics, vol. 18, no. 10, page 100901, 2013.
- [Li 16] Jun Li, José M Bioucas-Dias, Antonio Plaza & Lin Liu. *Robust collaborative non-negative matrix factorization for hyperspectral unmixing*. IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pages 6076–6090, 2016.
- [Lin 07a] Chih-Jen Lin. *On the convergence of multiplicative update algorithms for non-negative matrix factorization*. IEEE Transactions on Neural Networks, vol. 18, no. 6, pages 1589–1596, 2007.
- [Lin 07b] Chih-Jen Lin. *Projected gradient methods for non-negative matrix factorization*. Neural Computation, vol. 19, no. 10, pages 2756–2779, 2007.
- [Lin 15] Chia-Hsiang Lin, Wing-Kin Ma, Wei-Chiang Li, Chong-Yung Chi & ArulMurugan Ambikapathi. *Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: the no-pure-pixel case*. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 10, pages 5530–5546, 2015.
- [Lleó 11] Lourdes Lleó, Jean-Michel Roger, Ana Herrero-Langreo, Belen Diezma-Iglesias & Pilar Barreiro. *Comparison of multispectral indexes extracted from hyperspectral images for the assessment of fruit ripening*. Journal of Food Engineering, vol. 104, no. 4, pages 612–620, 2011.
- [Lu 14] Guolan Lu & Baowei Fei. *Medical hyperspectral imaging: a review*. Journal of Biomedical Optics, vol. 19, no. 1, page 010901, 2014.
- [Mairal 10] Julien Mairal, Francis Bach, Jean Ponce & Guillermo Sapiro. *On-line learning for matrix factorization and sparse coding*. Journal of Machine Learning Research, vol. 11, no. Jan, pages 19–60, 2010.
- [Mazet 05] Vincent Mazet. *Développement de méthodes de traitement de signaux spectroscopiques : estimation de la ligne de base et du spectre de raies*. PhD thesis, Université Henri Poincaré-Nancy 1, 2005.

- [McLellan 91] Toni M McLellan, John D Aber, Mary E Martin, Jerry M Melillo & Knute J Nadelhoffer. *Determination of nitrogen, lignin, and cellulose content of decomposing leaf material by near infrared reflectance spectroscopy*. Canadian Journal of Forest Research, vol. 21, no. 11, pages 1684–1688, 1991.
- [Miao 07] Lidan Miao & Hairong Qi. *Endmember extraction from highly mixed data using minimum volume constrained non-negative matrix factorization*. IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 3, pages 765–777, 2007.
- [Miron 11] Sebastian Miron, Manuel Dossot, Cédric Carteret, Samuel Margueron & David Brie. *Joint processing of the parallel and crossed polarized Raman spectra and uniqueness in blind non-negative source separation*. Chemometrics and Intelligent Laboratory Systems, vol. 105, no. 1, pages 7–18, 2011.
- [Mirzal 14] Andri Mirzal. *Non-parametric tikhonov regularized NMF and its application in cancer clustering*. IEEE Transactions on Computational Biology and Bioinformatics, vol. 11, no. 6, pages 1208–1217, 2014.
- [Mora 11] Christian R Mora, Laurence R Schimleck, Seung-Chul Yoon & Chi N Thai. *Determination of basic density and moisture content of loblolly pine wood disks using a near infrared hyperspectral imaging system*. Journal of Near Infrared Spectroscopy, vol. 19, no. 5, pages 401–409, 2011.
- [Nascimento 05] José M P Nascimento & José MB Dias. *Vertex component analysis: a fast algorithm to unmix hyperspectral data*. IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 4, pages 898–910, 2005.
- [Nesterov 83] Yu Nesterov. *A method of solving a convex programming problem with convergence rate $O(1/k^2)$* . Soviet Mathematics Doklady, vol. 27, no. 2, pages 372–367, 1983.
- [Nie 10] Feiping Nie, Heng Huang, Xiao Cai & Chris H Ding. *Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization*. In 23rd International Conference on Neural Information Processing Systems, pages 1813–1821, 2010.
- [Ono 03] Kenji Ono, Masakazu Hiraide & Masahiro Amari. *Determination of lignin, holocellulose, and organic solvent extractives in fresh leaf, litterfall, and organic material on forest floor using near infrared reflectance spectroscopy*. Journal of Forest Research, vol. 8, no. 3, pages 191–198, 2003.
- [Osborne 06] Brian G Osborne. *Near-infrared spectroscopy in food analysis*. Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, 2006.
- [Paatero 94] Pentti Paatero & Unto Tapper. *Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values*. Environmetrics, vol. 5, no. 2, pages 111–126, 1994.
- [Parente 10] Mario Parente & Antonio Plaza. *Survey of geometric and statistical unmixing algorithms for hyperspectral images*. In 2nd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, pages 1–4, 2010.

- [Pauca 06] Victor P Pauca, Jon Piper & Robert J Plemmons. *Non-negative matrix factorization for spectral data analysis*. Linear Algebra and Its Applications, vol. 416, no. 1, pages 29–47, 2006.
- [Peng 07] Bo Peng. *The determinant: A means to calculate volume*. Recall, vol. 21, page a22, 2007.
- [Pettersen 84] Roger C Pettersen. *The chemical composition of wood*. The Chemistry of Solid Wood, vol. 207, pages 57–126, 1984.
- [Plaza 12] Javier Plaza, Eligijs M T Hendrix, Inmaculada García, Gabriel Martín & Antonio Plaza. *On endmember identification in hyperspectral images without pure pixels: a comparison of algorithms*. Journal of Mathematical Imaging and Vision, vol. 42, no. 2-3, pages 163–175, 2012.
- [Powell 73] Michael J D Powell. *On search directions for minimization algorithms*. Mathematical Programming, vol. 4, no. 1, pages 193–201, 1973.
- [Qian 11] Yuntao Qian, Sen Jia, Jun Zhou & Antonio Robles-Kelly. *Hyperspectral unmixing via $\ell_{1/2}$ sparsity-constrained non-negative matrix factorization*. IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 11, pages 4282–4297, 2011.
- [Qin 17] Jianwei Qin, Moon Kim, Kuanglin Chao, Diane Chan, Stephen Delwiche & Byoung-Kwan Cho. *Line-scan hyperspectral imaging techniques for food safety and quality applications*. Applied Sciences, vol. 7, no. 2, page 125, 2017.
- [Qu 16] Yingjie Qu, Wenqi Ren, Songde Liu, Peng Liu, Lan Xie, Xiaoyuan Zhang, Shiwu Zhang, Shufang Chang & Ronald Xu. *Polarized hyperspectral imaging system for in vivo detection of vulvar lichen sclerosus*. In Optical Diagnostics and Sensing XVI: Toward Point-of-Care Diagnostics, volume 9715, page 97150Q. International Society for Optics and Photonics, 2016.
- [Rakotovao 12] Georges Rakotovao, Andrianasolo R Rabevohitra, Philippe C de Chatelperron, Daniel Guibal & Jean Gérard. Atlas des bois de madagascar. Editions Quae, 2012.
- [Ravel 17] Sylvain Ravel. *Démixage d'images hyperspectrales en présence d'objets de petite taille*. PhD thesis, École Centrale Marseille, 2017.
- [Sandak 16] Jakub Sandak, Anna Sandak & Roger Meder. *Assessing trees, wood and derived products with near infrared spectroscopy: hints and tips*. Journal of Near Infrared Spectroscopy, vol. 24, no. 6, pages 485–505, 2016.
- [Sandberg 09] Karin Sandberg & Magdalena Sterley. *Separating Norway spruce heartwood and sapwood in dried condition with near infrared spectroscopy and multivariate data analysis*. European Journal of Forest Research, vol. 128, no. 5, pages 475–481, 2009.

- [Schachtner 09] Reinhard Schachtner, Gerhard Pöppel, Ana M Tomé & Elmar W Lang. *Minimum determinant constraint for non-negative matrix factorization*. In 8th International Conference on Independent Component Analysis and Signal Separation, pages 106–113. Springer, 2009.
- [Schimleck 02] Laurence R Schimleck & Robert Evans. *Estimation of microfibril angle of increment cores by near infrared spectroscopy*. IAWA Journal, vol. 23, no. 3, pages 225–234, 2002.
- [Schwanninger 11] Manfred Schwanninger, José C Rodrigues & Karin Fackler. *A review of band assignments in near infrared spectra of wood and wood components*. Journal of Near Infrared Spectroscopy, vol. 19, no. 5, pages 287–308, 2011.
- [Song 16] Yingying Song, David Brie, El-Hadi Djermoune & Simon Henrot. *Regularization parameter estimation for non-negative hyperspectral image deconvolution*. IEEE Transactions on Image Processing, vol. 25, no. 11, pages 5316–5330, 2016.
- [Stirling 13] Rod Stirling. *Near infrared spectroscopy as a potential quality assurance tool for the wood preservation industry*. The Forestry Chronicle, vol. 89, no. 5, pages 654–658, 2013.
- [Strang 06] Gilbert Strang. *Linear algebra and its applications*. Cengage Learning, 2006.
- [Sun 14] Dennis L Sun & Cedric Fevotte. *Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence*. In 39th IEEE International Conference on Acoustics, Speech and Signal Processing, pages 6201–6205, 2014.
- [Svedas 00] Vitas Svedas. *Transformation of the near infrared bands of cellulose surface hydroxyls under the influence of adsorbed water molecules*. Applied Spectroscopy, vol. 54, no. 3, pages 420–425, 2000.
- [Tang 14] Wei Tang, Zhenwei Shi, Ying Wu & Changshui Zhang. *Sparse unmixing of hyperspectral data using spectral a priori information*. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 2, pages 770–783, 2014.
- [Tatzer 05] Petra Tatzer, Markus Wolf & Thomas Panner. *Industrial application for inline material sorting using hyperspectral imaging in the nir range*. Real-Time Imaging, vol. 11, no. 2, pages 99–107, 2005.
- [Theys 09] Céline Theys, Nicolas Dobigeon, Jean-Yves Tourneret & Henri Lantéri. *Linear unmixing of hyperspectral images using a scaled gradient method*. In 15th IEEE Workshop on Statistical Signal Processing, pages 729–732, 2009.
- [Thouvenin 16] Pierre-Antoine Thouvenin, Nicolas Dobigeon & Jean-Yves Tourneret. *On-line unmixing of multitemporal hyperspectral images accounting for spectral variability*. IEEE Transactions on Image Processing, vol. 25, no. 9, pages 3979–3990, 2016.
- [Usenik 12] Peter Usenik, Miran Bürmen, Aleš Fidler, Franjo Pernuš & Boštjan Likar. *Evaluation of cross-polarized near infrared hyperspectral imaging for early detection of*

- dental caries*. In Lasers in Dentistry XVIII, volume 8208, page 82080G. International Society for Optics and Photonics, 2012.
- [Via 13] Brian K Via & Weilin Jiang. *Non-linear multivariate modeling of strand mechanical properties with near infrared spectroscopy*. The Forestry Chronicle, vol. 89, no. 5, pages 621–630, 2013.
- [Vogel 96] Curtis R Vogel. *Non-convergence of the L-curve regularization parameter selection method*. Inverse Problems, vol. 12, no. 4, page 535, 1996.
- [Wang 11] Fei Wang, Chenhao Tan, Ping Li & Arnd Christian König. *Efficient document clustering via on-line non-negative matrix factorizations*. In 11th Proceedings of the SIAM International Conference on Data Mining, pages 908–919, 2011.
- [Wang 13a] Dong Wang & Huchuan Lu. *On-line learning parts-based representation via incremental orthogonal projective non-negative matrix factorization*. Signal Processing, vol. 93, no. 6, pages 1608–1623, 2013.
- [Wang 13b] Naiyan Wang, Jingdong Wang & Dit-Yan Yeung. *On-line robust non-negative dictionary learning for visual tracking*. In 14th Proceedings of the IEEE International Conference on Computer Vision, pages 657–664, 2013.
- [Wang 13c] Nan Wang, Bo Du & Liangpei Zhang. *An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 6, no. 2, pages 554–569, 2013.
- [Wang 16] Dong Wang, Jin-Xing Liu, Ying-Lian Gao, Jiguo Yu, Chun-Hou Zheng & Yong Xu. *An NMF- $\ell_{2,1}$ -norm constraint method for characteristic gene selection*. PLOS One, vol. 11, no. 7, page e0158494, 2016.
- [Willett 14] Rebecca M Willett, Marco F Duarte, Mark A Davenport & Richard G Baraniuk. *Sparsity and structure in hyperspectral imaging: sensing, reconstruction, and target detection*. IEEE Signal Processing Magazine, vol. 31, no. 1, pages 116–126, 2014.
- [Winter 99] Michael E Winter. *N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data*. In Imaging Spectrometry V, volume 3753, pages 266–276. International Society for Optics and Photonics, 1999.
- [Wolfe 97] William L Wolfe. Introduction to imaging spectrometers, volume 25. SPIE Press, 1997.
- [Wu 14] Yi Wu, Bin Shen & Haibin Ling. *Visual tracking via on-line non-negative matrix factorization*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 3, pages 374–383, 2014.
- [Xu 12] Yangyang Xu, Wotao Yin, Zaiwen Wen & Yin Zhang. *An alternating direction algorithm for matrix completion with non-negative factors*. Frontiers of Mathematics in China, vol. 7, no. 2, pages 365–384, 2012.

- [Yang 11] Zuyuan Yang, Guoxu Zhou, Shengli Xie, Shuxue Ding, Jun-Mei Yang & Jun Zhang. *Blind spectral unmixing based on sparse non-negative matrix factorization*. IEEE Transactions on Image Processing, vol. 20, no. 4, pages 1112–1125, 2011.
- [Yang 14] Sang-Yun Yang, Chang-Deuk Eom, Yeonjung Han, Yoonseong Chang, Yonggun Park, Jun-Jae Lee, Joon-Weon Choi & Hwanmyeong Yeo. *Near infrared spectroscopic analysis for classification of water molecules in wood by a theory of water mixtures*. Wood and Fiber Science, vol. 46, no. 2, pages 138–147, 2014.
- [Yao 10] Sheng Yao, Guo-Feng Wu, Yi-Fei Jiang, Xiao-Dong Fu, Hong-Kun Lü, Mei Su & Jun-We Pu. *Extending hemicelluloses content calibration of acacia spp using NIR to new sites*. Spectroscopy and Spectral Analysis, vol. 30, no. 5, pages 1206–1209, 2010.
- [Yu 07] Yue Yu & Weidong Sun. *Minimum distance constrained non-negative matrix factorization for the endmember extraction of hyperspectral images*. In MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications, volume 6790, page 679015. International Society for Optics and Photonics, 2007.
- [Zare 13] Alina Zare & Dominic KC Ho. *Endmember variability in hyperspectral analysis: addressing spectral variability during spectral unmixing*. IEEE Signal Processing Magazine, vol. 31, no. 1, pages 95–104, 2013.
- [Zdunek 06] Rafal Zdunek & Andrzej Cichocki. *Non-negative matrix factorization with quasi-newton optimization*. In 8th International Conference on Artificial Intelligence and Soft Computing, pages 870–879. Springer, 2006.
- [Zhang 14] Lefei Zhang, Liangpei Zhang, Dacheng Tao, Xin Huang & Bo Du. *Hyperspectral remote sensing image subpixel target detection based on supervised metric learning*. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 8, pages 4955–4965, 2014.
- [Zhao 16] Renbo Zhao & Vincent Y F Tan. *On-line non-negative matrix factorization with outliers*. IEEE Transactions on Signal Processing, vol. 65, no. 3, pages 555–570, 2016.
- [Zhou 11a] Guoxu Zhou, Shengli Xie, Zuyuan Yang, Jun-Mei Yang & Zhaoshui He. *Minimum-volume-constrained non-negative matrix factorization: enhanced ability of learning parts*. IEEE Transactions on Neural Networks, vol. 22, no. 10, pages 1626–1637, 2011.
- [Zhou 11b] Guoxu Zhou, Zuyuan Yang, Shengli Xie & Jun-Mei Yang. *On-line blind source separation using incremental non-negative matrix factorization with volume constraint*. IEEE Transactions on Neural Networks, vol. 22, no. 4, pages 550–560, 2011.

- [Zhu 17] Feiyun Zhu. *Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey*. arXiv preprint arXiv:1708.05125, 2017.
- [Zimmermann 03] Timo Zimmermann, Jens Rietdorf & Rainer Pepperkok. *Spectral imaging and its applications in live cell microscopy*. FEBS letters, vol. 546, no. 1, pages 87–92, 2003.