# Traitement statistique d'images hyperspectrales pour la détection d'objets diffus : application aux données astronomiques du spectro-imageur MUSE

Jean-Baptiste Courbot

Université de Strasbourg

École doctorale MSII 269
Université de Strasbourg

msii École doctorale

# THÈSE

présentée par

**Jean-Baptiste COURBOT**

le 13 octobre 2017

pour obtenir le grade de **docteur de l'Université de Strasbourg**

discipline : traitement du signal et des images

# Traitement statistique d'images hyperspectrales pour la détection d'objets diffus : application aux données astronomiques du spectro-imageur MUSE

**Thèse dirigée par :**

| | |
|---|---|
| **M. COLLET Christophe** | Professeur ICube, Université de Strasbourg. |
| **M. BACON Roland** | Directeur de recherches CRAL, Observatoire de Lyon. |

**Rapporteurs :**

| | |
|---|---|
| **M. TOURNERET Jean-Yves** | Professeur IRIT, INP Toulouse. |
| **M. PIECZYNSKI Wojciech** | Professeur IMT - Telecom SudParis. |

**Autres membres du jury :**

| | |
|---|---|
| **M. MICHEL Olivier** | Professeur GIPSA-lab, Université Grenoble Alpes. |
| **M. THIEBAUT Éric** | Astronome adjoint CRAL, Observatoire de Lyon. |

**Encadrants :**

| | |
|---|---|
| **M. MAZET Vincent** | Maître de conférences ICube, Université de Strasbourg. |
| **M. MONFRINI Emmanuel** | Maître de conférences IMT - Telecom SudParis. |

*À mon père*

# *Remerciements*

Après trois ans de thèse, ces remerciements signent la fin du voyage. Faire cette comparaison n'est sans doute pas faire preuve d'originalité, mais elle me semble pourtant adaptée. Si le chemin a été long, je ne l'ai pas parcouru seul et je voudrais remercier celles et ceux sans qui cette aventure n'aurait – de loin – pas été si enrichissante, tant scientifiquement que personnellement.

Je tiens à remercier les membres du jury, qui m'ont accueilli au terme du voyage : Jean-Yves Tourneret et Wojciech Pieczynski, pour avoir accepté d'être rapporteurs de ces travaux, ainsi que Éric Thiébaut et Olivier Michel pour avoir accepté d'en être examinateurs. Tous quatre ont permis des échanges très enrichissants en soulevant des questions très pertinentes lors de nos diverses discussions.

Mes directeurs m'ont été d'une aide précieuse durant ces trois années : Christophe, qui m'a permis de m'initier à la recherche puis de continuer sur cette voie, et Roland, qui m'a plongé dans un milieu de recherche bien différent, avec ses propres problématiques et enjeux.

Je remercie tout particulièrement mes encadrants pour leurs discussions et apports, scientifiques et non-scientifiques : Emmanuel, qui me suit depuis « tout petit » et m'a accompagné, en gardant sa jovialité, sur les chemins de la markovianité, et Vincent, qui a su poser les bonnes questions – tant scientifiques que typographiques – avec une pertinence qui m'a plusieurs fois frappé. Interagir avec Vincent et Emmanuel m'a été très précieux et j'espère que la fin de cette thèse ne signifiera, avec les années, que le début de nos échanges !

Dans le cadre de ces travaux, nous nous sommes réunis régulièrement avec les membres du projet ERC MUSICOS au CRAL, à l'Observatoire de Lyon. Cela a pour moi été l'occasion de côtoyer un milieu de recherche différent, et d'y faire des rencontres humainement et scientifiquement enrichissantes. Je pense en particulier à Benjamin, Jérémy, Alyssa, David, Sylvie et Mylène pour leur bonne humeur et leur convivialité lors de nos rencontres à Saint-Genis. Merci aux membre du projet MUSICOS : Simon et Laure, Carole, Floriane et Raphaël : que dire de plus ? J'ai vraiment aimé travailler avec vous.

Pendant trois ans et demi, j'ai côtoyé les membres de l'équipe MIV à Strasbourg. Je les remercie pour l'ambiance de travail exceptionnellement bonne, les sorties, soirées, petit-déjeuners, pauses café et tous les autres bons moments passés ensemble. Merci donc à Florent, Ola, Anthony, Alix, Marine et Pierre-Henri, qui sont partis sous de nouveaux horizons ; à Vincent, Sylvain, Étienne, Loïc, à Hassan et Yves, et aux « nouveaux » : Adrien, Anastasia, Cagdas, Argheesh et Hugo, Florian, Odyssée et Céline. Tous me semblent armés pour animer le laboratoire entier !

Je remercie enfin ma famille et ma belle-famille pour leur soutien et leur accueil toujours chaleureux. Une mention toute spéciale pour mon père, qui m'a encouragé depuis le début et aurait aimé le faire jusqu'à la fin, et pour ma mère qui m'a soutenu tout du long et a eu le courage de relire intégralement le présent manuscrit.

Enfin, un grand merci à Lauriane, qui m'a supporté au quotidien le long de ce voyage, et sans laquelle rien de tout ceci n'aurait eu lieu.

# *Table des matières*

# *Table des figures*

# *Acronymes et notations*

## Acronymes

| | |
|---|---|
| ACE | Adaptive cosine/coherence estimator |
| AMC | Arbre de Markov caché à bruit indépendant |
| AMC | Arbre de Markov couple |
| AMF | Adaptive matched filter |
| AMT | Arbre de Markov triplet |
| AMTS | Arbre de Markov triplet spatial |
| AUC | Area under curve |
| CGM | Circum-galactic medium |
| CMC | Champ de Markov caché à bruit indépendant |
| CMco | Champ de Markov couple convolutif |
| CMTO | Champ de Markov triplet orienté |
| EM | Expectation-maximization |
| FDR | False discovery rate |
| FSF | Field spread function |
| GLR | Generalized likelihood ratio |
| GLR-1s | GLR 1-parcimonieux |
| HC | Higher Criticism |
| HDFS | Hubble deep field south |
| HUDF | Hubble ultra-deep field |
| ICE | Iterative conditional estimation |
| ICM | Iterated conditional modes |
| IFU | Integral Field Unit |
| IGM | Inter-galactic medium |
| LR | Likelihood ratio |
| LRMAP | Likelihood ratio avec le MAP |
| MAP | Maximum *a posteriori* |
| MCEM | Monte Carlo expectation-maximization |
| MCMC | Monte Carlo par chaînes de Markov |
| MPM | Maximum posterior mode |
| MUSE | Multi-Unit Spectroscopic Explorer |
| MV | Maximum de vraisemblance |
| OSP | Orthogonal subspace projector |
| PDR | Posterior density ratio |
| PP | Projection pursuit |
| ROC | Receiver operating characteristic |
| RSB | Rapport signal-sur-bruit |
| RX | Reed-Xiaoli |
| SACE | Squared adaptive cosine/coherence estimator |
| SAEM | Stochastic approximation expectation-maximization |
| SEM | Stochastic expectation-maximization |

| | |
|---|---|
| SS-GRF | Spatio-spectral gaussian random fields |
| SVM | Support vector machine |

## Notations

| | |
|---|---|
| $\|\cdot\|_2$ | Norme euclidienne |
| $\|\cdot\|$ | Cardinal d'un ensemble |
| $\bar{\cdot}$ | Moyenne d'un vecteur ou d'une matrice |
| $\setminus$ | Différence ensembliste |
| $\cup$ | Union ensembliste |
| $\langle\cdot,\cdot\rangle$ | Produit scalaire |
| $\hat{\cdot}$ | Estimateur d'une valeur |
| Å | Angström ($0{,}1$ nm) |
| $A$ | Variable aléatoire scalaire |
| $a$ | Réalisation de $A$ |
| $\mathbf{A}$ | Variable aléatoire vectorielle |
| $\mathbf{a}$ | Réalisation de $\mathbf{A}$ |
| $\delta$ | Fonction de Kronecker |
| $\mathbf{I}_K$ | Matrice identité de taille $K \times K$ |
| $\mathbb{1}$ | Fonction indicatrice |
| $\Lambda$ | Nombre de bandes spectrales d'un spectre, d'une image hyperspectrale |
| $\lambda$ | Longueur d'onde |
| $\boldsymbol{\mu}$ | Spectre non nul, non bruité |
| $s$ | Site d'une image |
| $N_s$ | Voisinage local du site $s$ |
| $p$ | Première dimension spatiale d'une image |
| $P_{\mathrm{D}}$ | Probabilité de détection |
| $P_{\mathrm{FA}}$ | Probabilité de fausse alarme |
| $\otimes$ | Produit de Kronecker |
| $q$ | Seconde dimension spatiale d'une image |
| $\rho$ | Corrélation |
| $r$ | Corrélation mesurée |
| $\mathcal{S}$ | Ensemble des sites $s$ d'une image, d'une image hyperspectrale |
| $\boldsymbol{\Sigma}$ | Matrice de covariance |
| $\sigma$ | Écart-type |
| $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ | Processus de classe |
| $\mathbf{Y} = (\mathbf{Y}_s)_{s\in\mathcal{S}}$ | Processus d'observation |
| $\mathbf{y}$ | Observation vectorielle |
| $y$ | Observation réelle |
| $\mathbf{f}$ | Fonction d'étalement spatial (FSF) |
| $z$ | Redshift |
| $\mathbf{0}_K$ | Vecteur nul comportant $K$ composantes |

## Détection par tests d'hypothèses

| | |
|---|---|
| $\alpha$ | Niveau de signification d'un test statistique |
| $\mathcal{B}$ | Ensemble de spectres brillants |
| $\mathcal{O}$ | Ensemble des observations d'une même scène |
| $\mathbf{C}$ | Catalogue de raies spectrales |
| $\mathcal{F}$ | Ensemble de spectres non brillants |
| $\mathcal{H}$ | Hypothèse dans un test statistique |
| $\mathcal{T}$ | Dénomination générique d'un test GLR |
| $T$ | Statistique de test |
| $\xi$ | Seuil utilisé dans un test statistique |
| $\alpha_s$, $\beta_{f,b}$ | Coefficients d'atténuation |

## Modèles de Markov

| | |
|---|---|
| $\alpha$, $\beta$ | Paramètres d'un champ de Markov ou d'un arbre de Markov |
| $\theta$ | Ensemble de paramètres |
| $\mathcal{C}$ | Ensemble des cliques $c$ d'une image |
| $\gamma$ | Constante de normalisation |
| $G = (\mathcal{S}, \Gamma)$ | Graphe formé de l'ensemble des noeuds $\mathcal{S}$ et de l'ensemble des arêtes $\Gamma$ |
| $\Omega_{\mathbf{v}}$ | Ensemble dans lequel les $\mathbf{V}_s$ multivariés prennent valeurs |
| $\Omega_v$ | Ensemble dans lequel les $V_s$ scalaires prennent valeurs |
| $\Omega_x$ | Ensemble dans lequel les $X_s$ prennent valeurs |
| $\phi$ | Fonction de potentiel |
| $\mathbf{u}^v = (u_s^v)_{s\in\mathcal{S}}$ | Carte des incertitudes associées à une segmentation de $\mathbf{V}$ |
| $\mathbf{u}^x = (u_s^x)_{s\in\mathcal{S}}$ | Carte des incertitudes associées à une segmentation de $\mathbf{X}$ |
| $\mathbf{V} = (V_s)_{s\in\mathcal{S}}$ | Processus auxiliaire |
| $\mathbf{T}$ | Notation générale pour le triplet $(\mathbf{X}, \mathbf{Y}, \mathbf{V})$ |
| $\mathbf{Z}$ | Notation générale pour le couple $(\mathbf{X}, \mathbf{Y})$ |

# Introduction générale

Aucune image n'est réelle : en effet, elles forment une représentation partielle de la réalité physique, résultant de nombreuses transformations. Certaines images sont cependant perçues comme plus fidèles à la réalité que d'autres. Cette distinction s'opère notamment par leur dégradation : lorsque l'information d'intérêt dans l'image peut être facilement appréhendée à l'œil nu, celle-ci sera jugée « de bonne qualité ». Cela ne signifie pas pour autant que dans le cas contraire, une image « de mauvaise qualité » soit dépourvue d'intérêt. La question qui se pose alors est « comment accéder aux informations d'intérêt ». Ce problème se décline dans de très nombreuses situations. Par exemple, le photographe maladroit souhaitera corriger le flou de ses clichés, le médecin voudra estimer des caractéristiques physiologiques dans une image de scanner suffisamment bien résolue, et l'astronome s'intéressera aux propriétés des corps célestes à différentes longueurs d'onde.

La discipline du traitement des images a pour but de répondre à ce type de questions. Les solutions dépendent naturellement du problème posé, et reposent sur une modélisation prenant en compte l'incertitude des données observées. L'adéquation d'un modèle ne correspond pas nécessairement à une modélisation des nombreux processus intervenant dans la formation de l'image : un grand nombre d'entre eux restent et resteront en pratique imprécis ou très partiellement connus. Ce constat est à l'origine des approches statistiques de traitement des signaux, et plus généralement du traitement des données : ces phénomènes sont modélisés comme des processus aléatoires, et le problème devient celui de la recherche d'information « la plus probable » et aboutit par exemple à une prise de décision basée sur l'inférence bayésienne.

Le problème de l'accès à des informations d'intérêt dans des images est un cas particulier de la catégorie des *problèmes inverses*. Cette dénomination n'est pas le fruit du hasard : en effet, l'objectif est d'inverser autant que possible les phénomènes de dégradation afin de retrouver les informations d'intérêt, aussi peu dégradées que possible. Le problème est devenu plus complexe avec l'apparition d'images dont la dynamique dépasse largement la capacité de résolution de la vision humaine (images multispectrales et hyperspectrales). Outre le problème de la visualisation de données toujours plus complexes, se pose le problème de l'étendue des données disponibles par rapport à l'information d'intérêt recherchée.

Les travaux abordés dans ce manuscrit s'inscrivent dans le cadre des approches stochastiques du traitement des signaux et des images. Le premier problème abordé est un problème de *détection*, dont le but est d'infirmer ou non la présence d'objets dans des images. Le problème de détection peut être vu comme un cas particulier du problème de *segmentation*, que nous abordons dans un second temps. Dans ce cadre, l'objectif est la catégorisation de chaque élément de l'image en classes. Nous nous intéressons donc également, de manière plus générale, au problème de la segmentation des images.

## Thématiques abordées

Dans le cadre de ces travaux de thèse, la problématique générale est la détection d'objets dans des images hyperspectrales très bruitées, et plus généralement la segmentation d'images en zones d'intérêt. D'un point de vue applicatif, l'objectif est la détection de structures astrophysiques très lointaines formant des *halos* et des *filaments* au sein des données issues du *multi-unit spectroscopic explorer* (MUSE). Cela se traduit par la formulation de plusieurs problématiques méthodologiques.

### Détection statistique d'objets ténus

Cette approche consiste à décider la présence d'un objet à partir de la formulation de deux hypothèses. Cette décision se base sur un *test d'hypothèses* dont le but est la bonne discrimination entre les alternatives envisagées. Le test doit donc, pour être efficace, renvoyer des valeurs très différentes si la structure recherchée est présente ou absente.

### Modèles markoviens, détection, segmentation

La problématique de détection d'objets peut être formulée comme un cas particulier de la segmentation d'images. Nous nous plaçons dans le cadre bayésien, et étudions des modèles de Markov pour permettre la détection d'objets dans des données extrêmement bruitées, où la perception par l'œil est mise en défaut. Les modèles markoviens proposent en effet une grande richesse de modélisation, qui permet lors de la segmentation de prendre au mieux en compte les phénomènes intrinsèques aux images.

Les travaux sur les modèles de Markov se sont traduits par les études suivantes :

— la modélisation d'images bruitées et convoluées. Ce modèle permet en particulier la formulation du problème de détection comme un problème de segmentation ;
— la modélisation conjointe des classes et orientations dans les images. Certains types d'images présentent en effet des structures orientées, il est pertinent de les modéliser afin de permettre la segmentation conjointe des classes et des orientations ;
— la modélisation de structures hiérarchiques homogènes. De nombreuses images sont composées de régions cohérentes, qu'il n'est pas souhaitable de séparer même en présence de fortes dégradations. Une modélisation en conséquence permet une bonne segmentation d'images excessivement bruitées.

## Organisation du manuscrit

Une introduction au traitement d'images hyperspectrales est tout d'abord présentée dans le chapitre 1. Ensuite, le manuscrit est partagé en trois parties :

I. La première partie porte sur la détection statistique par tests d'hypothèses. Le chapitre 2 présente un état de l'art sur la détection en imagerie hyperspectrale, et le chapitre 3 introduit la méthode de détection de sources ténues et étendues correspondant aux travaux publiés dans [Courbot et al., 2016a] et [Courbot et al., 2017a]. Les contributions de cette partie sont les suivantes :
   — la mise au point de tests prenant en compte la similarité d'un spectre par rapport à une référence et par rapport à un ensemble de références, ainsi que l'étude des propriétés statistiques de ces tests ;
   — la formulation de ces tests prenant en compte la forme spatiale et spectrale des intensités lumineuses recherchées, ainsi que l'étude de la détection sur un ensemble d'observations de la même scène ;

— l'introduction d'une stratégie de détection en deux temps, permettant la détection successive des spectres les plus brillants et les plus ténus de l'image.

II. Dans la deuxième partie, nous nous intéressons à la segmentation bayésienne des images. Le chapitre 4 présente une synthèse sur les modèles de Markov employés dans ce contexte. Ensuite, le chapitre 5 décrit un modèle de champ de Markov couple dans le même contexte applicatif que dans le chapitre 3. Nous nous intéressons plus généralement, dans le chapitre 6, à la modélisation d'images présentant des orientations marquées, par le biais d'un modèle de champs de Markov triplets. Enfin, le chapitre 7 présente un modèle d'arbres de Markov triplets permettant la segmentation bayésienne reposant sur un *a priori* hiérarchique et d'homogénéité spatiale. Les contributions de cette partie peuvent se résumer comme suit :
   — l'introduction d'un modèle de champ Markov couple original permettant la segmentation d'images convoluées, et plus particulièrement la détection non supervisée d'objets dans des données extrêmement bruitées ;
   — l'introduction d'un nouveau modèle de champ de Markov triplet, qui permet la modélisation des orientations dans les images et la segmentation bayésienne conjointe des orientations et des classes, en contexte non supervisé. Ces travaux correspondent aux publications [Courbot et al., 2016b] et [Courbot et al., 2017c] ;
   — un modèle d'arbre de Markov triplet original est également introduit, et permet lors de la segmentation la prise en compte simultanée de l'information hiérarchique inter-résolution et d'homogénéité au sein d'une même résolution. Ces travaux ont été publiés dans [Courbot et al., 2017b].

III. La troisième partie concerne l'application aux images hyperspectrales astronomiques issues de l'instrument MUSE. Nous présentons tout d'abord le contexte astronomique et instrumental dans le chapitre 8. Une analyse du bruit affectant les images MUSE est ensuite présentée dans le chapitre 9. Le chapitre 10 présente enfin les résultats comparatifs des méthodes développées sur des données synthétiques et réelles. Les contributions de cette partie sont :
   — la mise en évidence de structures de corrélations dans les images MUSE, et de distributions de Student (observations individuelles) et normale (observations fusionnées) pour les intensités ;
   — la comparaison des 4 méthodes proposées dans les chapitres 3, 5, 6 et 7 dans le cadre de la détection de sources ténues, sur des données synthétiques élaborées à partir d'images MUSE
   — l'obtention des premières cartographies de halos circum-galactiques présents autour de galaxies lointaines.

Ces trois parties ont des thématiques relativement distinctes, c'est pourquoi elles peuvent être lues de manière indépendante. Ainsi, le lecteur intéressé par les modèles markoviens pourra débuter par la partie II. De même, pour un accès rapide aux résultats applicatifs, il est possible de débuter la lecture par la partie III.

## Contexte

Les travaux de thèse présentés dans ce manuscrit ont été réalisés à partir d'octobre 2014 dans l'équipe MIV du laboratoire ICube [1] à Strasbourg, en collaboration avec le CRAL [2] à Lyon. Ils s'inscrivent dans le cadre du projet MUSE, initié en 2005 et aboutissant en 2014 aux

1. ICube, Université de Strasbourg, CNRS (UMR 7357).
2. Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, CRAL UMR 5574.

premières observations de l'instrument MUSE sur la plate-forme du *Very Large Telescope* européen à Paranal, au Chili [3]. À ce titre, ces travaux participent également aux recherches menées par le consortium MUSE [4] formé des laboratoires européens ayant participé à la construction de l'instrument.

La thèse est financée par le projet ERC MUSICOS débuté en 2014 [5], dont l'objectif est l'analyse de données MUSE pour la détection de galaxies et de structures caractérisant l'Univers jeune. Ce projet a vu le démarrage simultané de trois thèses en traitement d'images et d'une thèse en astronomie, et a notamment permis des échanges réguliers avec le laboratoire GIPSA-lab [6]. Le projet ANR DSIM [7], débuté en 2014, a de plus fourni un support et des échanges scientifiques sur les problématiques de décomposition spectrale de données hyperspectrales.

Ce travail fait donc suite aux projets initiés avec le CRAL et l'OCA (Observatoire de Nice) il y a une dizaine d'années, au travers des projets ANR SpaceFusion [8] et ANR Dahlia [9] au sein desquels l'équipe MIV fut active.

## Contributions

### Articles en journal

Les travaux présentés dans ce manuscrit font l'objet de deux principales publications en revue.

[Courbot et al., 2017a] : Courbot, J.-B., Mazet, V., Monfrini, E. et Collet, C. (2017). Extended Faint Source Detection in Astronomical Hyperspectral Images. *Signal Processing*, 135 :274–283.
Cet article présente la méthode de détection par tests d'hypothèses développée dans ce manuscrit (chapitre 3), en présentant des tests statistiques adaptés au problème de détection de halos dans les données MUSE.

[Courbot et al., 2017c] : Courbot, J.-B., Monfrini, E., Mazet, V., et Collet, C. (2017). Oriented Triplet Markov Fields. Soumis aux *Pattern Recognition Letters*.
Cet article présente un modèle original de champs de Markov triplets permettant la segmentation conjointe non supervisée des classes et des orientations dans les images. Ce modèle est décrit dans le chapitre 6.

De plus, les collaborations au sein du consortium MUSE ont permis la participation à deux autres articles en revue.

[Wisotzki et al., 2016] : Wisotzki, L. et al. (2016). Extended Lyman $\alpha$ haloes around individual high-redshift galaxies revealed by MUSE. *Astronomy & Astrophysics*, 587 :A98.
Cet article présente les première détection de halos effectuées dans les données MUSE. Il démontre la faisabilité de telles détections, et illustre quelques propriétés physiques de ces objets.

[Bacon et al., 2017] : Bacon, R., et al (2017). The MUSE Hubble Ultra Deep Field Survey : I. Survey description, data reduction and source detection. Accepté à *Astronomy & Astrophysics*.

3. http://www.eso.org/public/teles-instr/paranal-observatory/vlt/
4. http://muse.univ-lyon1.fr/spip.php?article97
5. Projet ERC 339659-MUSICOS.
6. GIPSA-lab, Université Grenoble Alpes, CNRS UMR 5216.
7. Projet ANR-14-CE27-0005.
8. 2005-2008, projet ANR-05-JCJC-0002.
9. 2008-2013, projet ANR-08-BLAN-0253.

Cet article présente les données MUSE d'observations de champs profonds, en détaillant le processus de formation des images hyperspectrales, les propriétés de ces images, et un outil de détection de sources ponctuelles. Il s'agit du premier d'une série de 10 articles portant sur ces données et leur analyse.

### Conférences internationales et francophone

[Courbot et al., 2016a] : Courbot, J.-B., Mazet, V., Monfrini, E., and Collet, C. (2016). Detection of faint extended sources in hyperspectral data and application to HDF-S MUSE observations. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1891–1895. IEEE.
Dans cet article, nous présentons les premiers travaux menés sur la détection de halos par une méthode de détection par tests d'hypothèses. Cet article forme la version préliminaire des travaux publiés dans [Courbot et al., 2017a].

[Courbot et al., 2016b] : Courbot, J.-B., Monfrini, E., Mazet, V., and Collet, C. (2016). Oriented triplet Markov fields for hyperspectral image segmentation. In *2016 8th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS).* IEEE.
Cet article présente un modèle de champs de Markov triplets pour la segmentation d'images hyperspectrales, et forme la version préliminaire des travaux présentés dans [Courbot et al., 2017c].

[Courbot et al., 2017b] : Courbot, J.-B., Monfrini, E., Mazet, V., and Collet, C. (2017). Arbres de Markov triplets pour la segmentation d'images. *Colloque GRETSI.*
Cet article présente des travaux portant sur la segmentation d'images par un modèle d'arbres de Markov triplets qui permet la prise en compte d'informations hiérarchiques et d'homogénéité spatiale. Ces travaux sont décrits dans le chapitre 7.

# 1

# Introduction



Les problématiques abordées dans ce manuscrit relèvent du traitement d'images hyperspectrales. Bien qu'elles soient similaires à celles du traitement d'images au sens général, plusieurs spécificités existent. Après avoir présenté, dans la section 1.1, l'imagerie hyperspectrale et quelques domaines d'application, nous décrivons les grandes catégories de traitements d'images hyperspectrales en section 1.2. Enfin, la section 1.3 nous permet de préciser les applications étudiées dans le cadre de ce manuscrit et les démarches entreprises pour les résoudre.

## 1.1 Images hyperspectrales et applications

### 1.1.1 Définition

Une image hyperspectrale est une image contenant, en chaque pixel, une information sur le spectre lumineux de la scène observée. Il s'agit d'une généralisation de l'image tricolore, qui est elle-même une généralisation de l'image monochrome ne contenant qu'une information d'intensité lumineuse en chaque pixel. Une image hyperspectrale repose sur une « granularité » spatiale et spectrale, c'est pourquoi elle dépend d'une *résolution* spatiale et spectrale. Plus cette résolution est fine, plus spécifique est l'information que l'on peut extraire de l'image : une image tricolore permet de distinguer une encre verte d'une encre bleue, une image hyperspectrale peut permettre d'identifier les éléments chimiques les constituant. La figure 1.1 illustre la différence entre l'image hyperspectrale et tricolore d'une même scène.

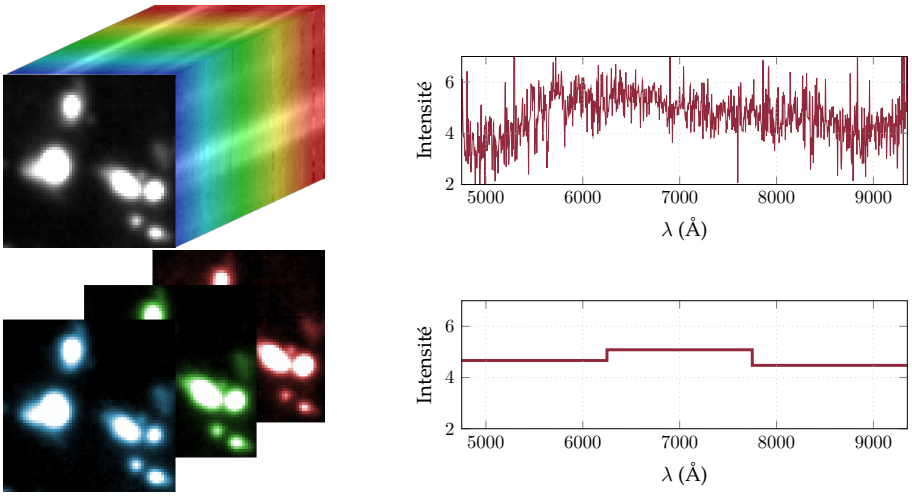



Figure 1.1 – Informations fournies et représentation d'une images hyperspectrale (première ligne) et d'une image tricolore (seconde ligne) d'une même observation. Les spectres présentés à droite sont les informations contenues en un pixel dans les deux cas.

Par analogie avec une image standard, une image hyperspectrale est souvent représentée sous la forme d'un *cube de données*, qui permet de regrouper les dimensions spatiales et spectrales. En ce sens chaque bande spectrale correspond à une image monochrome.

Il existe par ailleurs un type d'image intermédiaire, dite *multispectrale*. La distinction avec l'imagerie hyperspectrale est souvent floue, et porte généralement sur le nombre de bandes spectrales considérées : quelques dizaines de bandes spectrales pour une image multispectrale, quelques dizaines à quelques centaines de bandes pour une image hyperspectrale. Le nombre de canaux spectraux ne caractérise pas en lui-même l'image hyperspectrale. C'est pourquoi nous optons pour une définition plus précise : nous dirons d'une image qu'elle est hyperspectrale lorsqu'elle permet de résoudre spectralement les phénomènes physiques observés, c'est-à-dire lorsque les bandes spectrales couvrent des canaux très fins et contigus.

Formellement, une image hyperspectrale est un ensemble de spectres $\mathbf{y}_s$, comportant un nombre $\Lambda$ de composantes réelles correspondant aux longueurs d'onde. $\mathbf{y}_s$ est spatialement localisé au site $s$ de l'image, et nous notons l'ensemble de ces sites $\mathcal{S}$. Cela permet de noter l'image hyperspectrale entière, $\mathbf{y}$, sous la forme $\mathbf{y} = (\mathbf{y}_s)_{s\in\mathcal{S}}$. Nous avons $\mathbf{y}_s \in \mathbb{R}^\Lambda$, ce qui donne $\mathbf{y} \in \mathbb{R}^{|S|\times\Lambda}$. La figure 1.2 reprend les notations employées.

Ces notations illustrent en particulier l'aspect massif des données hyperspectrales, qui en fait l'intérêt majeur de ce type d'image : la combinaison des dimensions spatiales et spectrales les rend très riches en informations. C'est aussi un frein au traitement de ces données : très volumineuses, elles requièrent des capacités de stockage et de calcul non négligeables. De plus, les informations d'intérêt sont très dispersées [1], ce qui complique leur recherche et rend difficile, sinon impossible, l'extension de méthodes de traitement d'images conventionnelles. Cela justifie le développement de méthodes aptes à résoudre les problèmes applicatifs apparaissant avec l'imagerie hyperspectrale, que nous présentons dans la section 1.2.

1. Il s'agit d'une instance de la *malédiction de la dimension*, terme attribué à R. Bellman [Bellman, 1956].

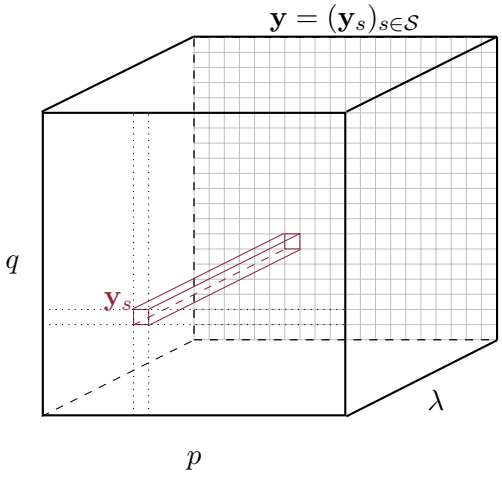


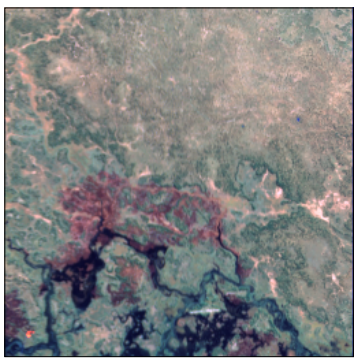

Figure 1.2 – Notations employées pour l'image hyperspectrale $\mathbf{y}$. $p$ et $q$ repèrent les deux dimensions spatiales, et $\lambda$ repère la dimension spectrale.

### 1.1.2 Domaines d'application

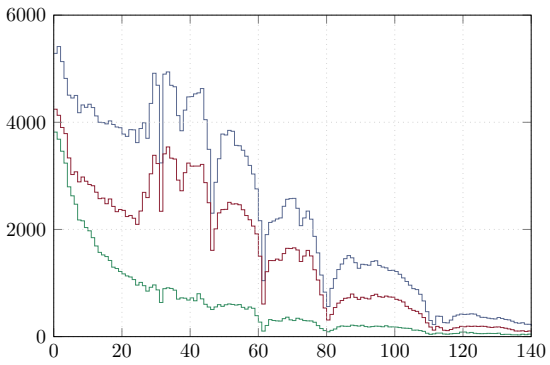
L'imagerie hyperspectrale est d'un intérêt pratique conséquent pour de nombreux domaines applicatifs. Le plus important est l'imagerie de télédétection [Varshney et Arora, 2004], qui a vu apparaître les premiers imageurs hyperspectraux dans les années 1970. Ce domaine couvre de très nombreuses applications et est moteur au sein de la communauté du traitement d'images hyperspectrales. Les autres domaines couverts par l'imagerie hyperspectrale se sont développés progressivement. Ainsi, l'imagerie hyperspectrale astronomique [Soulez et al., 2011; Petremand et al., 2012] a fait son apparition dans les années 1990, l'analyse d'aliments [Wu et Sun, 2013] dans les années 2000, tout comme la caractérisation de tissus [Lu et al., 2014; Pike et al., 2016] ou encore l'analyse légale [Brewer et al., 2008; Edelman et al., 2012]. Nous détaillons ici les spécificités des applications :

— de télédétection, car ce domaine motive la plupart des méthodes développées dans la littérature ;
— à l'astronomie, qui est le domaine applicatif de ce manuscrit.

L'application à la télédétection consiste en l'observation de la terre, à l'aide d'instruments aéroportés ou satellitaires. Ce type d'observation peut avoir des retombées civiles, avec la

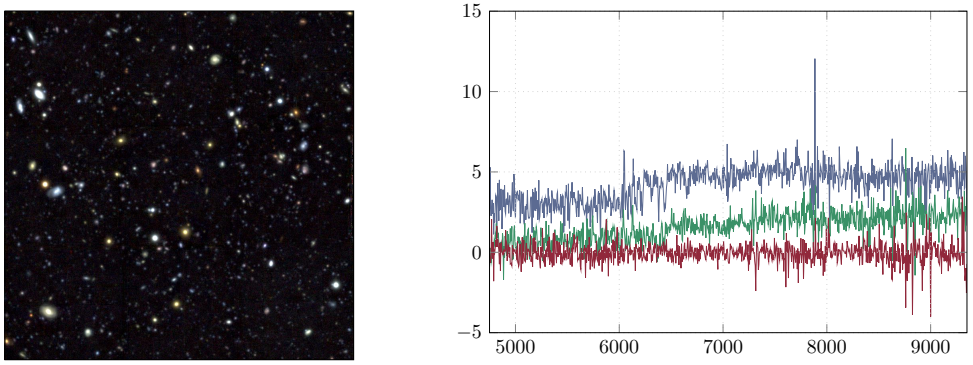


(a) Reconstitution en fausses couleurs.




(b) Trois spectres de cette image.

Figure 1.3 – Image hyperspectrale de télédétection issue de l'instrument AVIRIS [avi, 2017]. Source : `http://www.ehu.eus/`.




(a) Reconstitution en fausses couleurs. (b) Trois spectres de cette image.

Figure 1.4 – Image hyperspectrale astronomique issue de l'instrument MUSE [mus, 2017; Bacon et al., 2017].

caractérisation urbaine ou de régions naturelles par exemple. L'autre motivation importante de la télédétection est militaire, pour une observation avancée de régions inaccessibles. Le rayonnement observé dans une image hyperspectrale de télédétection est majoritairement celui de la lumière du soleil réfléchie. Les spectres observés mettent en évidence les éléments chimiques qui ont interagi avec la lumière sur son trajet. Cela permet de distinguer les régions selon leur contenu (région urbaines, aquatiques, forestières, etc.). La lumière a aussi interagi avec les différents composants de l'atmosphère, ce qui peut permettre de traduire la présence de composants volatils par exemple. La portée et la résolution spatiales varient selon les outils d'acquisition. La résolution spatiale est souvent trop grossière (rarement inférieure au mètre) pour observer de petits objets, de sorte à produire un mélange, en chaque site $s$, de plusieurs spectres *élémentaires*. Un exemple d'image hyperspectrale de télédétection est donné en figure 1.3.

L'application à l'astronomie est une branche relativement récente de l'imagerie hyperspectrale. Les spectres observés sont majoritairement des spectres d'émission : la lumière observée a été émise par les objets eux-mêmes. L'instrumentation hyperspectrale astronomique permet la synthèse de deux catégories d'instruments existants :

— les imageurs, qui couvrent de larges régions du ciel mais seulement quelques canaux spectraux produits à l'aide de filtres ;
— les spectrographes, qui permettent d'obtenir un spectre résolu d'une région ponctuelle du ciel, localisée au préalable.

Les éléments observés sont les mêmes que ceux observés par les instruments « standards » : planètes, nébuleuses, étoiles, galaxies, etc. Un exemple d'image hyperspectrale astronomique est donné en figure 1.4.

Les travaux de ce manuscrit ont pour application principale les images hyperspectrales astronomiques. Nous en présentons ici quelques spécificités par rapport aux images de télédétection. La première d'entre elles est le caractère parcimonieux des observations : les sources sont très localisées spatialement, et leurs spectres peuvent être décomposés en une combinaison de raies d'émission, d'absorption, et d'une composante continue à variation douce. Cela signifie également qu'une partie non négligeable de l'image hyperspectrale ne contient aucun signal d'intérêt. Cette partie forme ce que l'on appellera « l'arrière-plan ». Celui-ci est uniforme sur toute l'image dans le cas d'un instrument idéal, ne laissant aucune « signature » instrumentale.

Nous avons par ailleurs noté, pour la télédétection, l'existence d'objets trop petits pour être résolus spatialement dans l'image. C'est le phénomène inverse qui se produit en

imagerie astronomique : une source idéalement ponctuelle, comme peut l'être une étoile ou une galaxie lointaine, formera une sorte de « tache » couvrant plusieurs sites. Plus précisément, cela signifie que la réponse impulsionnelle spatiale de l'instrument (FSF pour *field spread function*) n'est pas négligeable par rapport à la taille des pixels. Cette FSF est produite par la dispersion atmosphérique dans le cas d'observations au sol, et provient dans une moindre mesure des caractéristiques de l'instrument.

La comparaison des figures 1.3b et 1.4b illustre un autre phénomène important : les images astronomiques peuvent se montrer extrêmement bruitées. En effet, l'observation d'objets très lointains requiert de longs temps de pose pour pouvoir collecter une lumière très ténue, rendant l'image très sensible aux imperfections du système.

Nous avons vu qu'une image hyperspectrale forme une masse riche et complexe de données. Les informations qu'elle contient sont rarement exploitables directement pour l'application souhaitée, c'est pourquoi il est souvent nécessaire de recourir à des méthodes de traitement d'images pour accéder aux informations les plus pertinentes.

## 1.2 Le traitement de données hyperspectrales

Nous présentons dans cette partie une vue synthétique du traitement d'images hyperspectrales. Les méthodes sont présentées en deux familles. Dans la première (section 1.2.1), l'objectif est l'amélioration des images, en inversant les différents processus de dégradation qui ont pu intervenir lors de leur formation. Dans la seconde (section 1.2.2), l'objectif est l'extraction d'informations spatiales ou spectrales dans les images.

### 1.2.1 Reconstruction, fusion et démélange

La reconstruction d'images hyperspectrales a pour objectif de reconstituer une image « source », telle qu'elle serait sans les processus de convolution, diffusion, bruit et distorsion intervenant lors de sa formation. Les méthodes de débruitage appartiennent à cette catégorie, et reposent sur la formulation judicieuse d'un problème d'optimisation sous contrainte [Rasti et al., 2014; Lin et Bourennane, 2013; Yuan et al., 2012].

Nous avons vu que la formation des images astronomiques fait intervenir un processus de convolution, par le biais de la FSF. Plusieurs méthodes [Villeneuve et Carfantan, 2014; Soulez et al., 2013] proposent d'inverser ce processus, pour accéder à une image aussi proche que possible d'une version « non convoluée ».

Lorsqu'une même scène est observée plusieurs fois par un ou plusieurs instruments, il peut être souhaitable d'exploiter au mieux toutes les données observées par le biais d'une méthode de fusion. C'est à cette catégorie qu'appartiennent les méthodes dites de *super-résolution* ou *pansharpening* [Loncan et al., 2015; Zhang et al., 2017]. Dans le cas de la télédétection, ce type de méthode utilise une image bien résolue spatialement et mal résolue spectralement conjointement avec une image hyperspectrale de moindre résolution spatiale pour produire une image « super-résolue ». Notons que la problématique de fusion a aussi été abordée dans le cadre d'images astronomiques [Petremand et al., 2012], pour des observations répétées d'une même région du ciel.

Nous avons vu que pour les images de télédétection, un spectre observé peut contenir un mélange de spectres élémentaires. L'objectif des méthodes de démélange est de retrouver ces composantes élémentaires (*endmembers*) et leurs abondances respectives. Les méthodes de démélange font l'objet d'une riche littérature, parmi laquelle nous pouvons citer les méthodes géométriques [Parente et Plaza, 2010], bayésiennes [Mittelman et al., 2012], et plus généralement non-linéaires [Dobigeon et al., 2014].

### 1.2.2 Classification et détection

La classification d'images hyperspectrales est un cas particulier de la problématique générale de classification, qui peut porter sur des images simples, des séquences, ou des données complexes [Bishop, 2006]. L'objectif est, dans le cas général, le regroupement de données *similaires* : dans le cas de l'imagerie hyperspectrale, il s'agit le plus souvent de regrouper les spectres observés par catégories [Camps-Valls et al., 2014], de sorte à produire une carte dite de *classification* ou de *segmentation*.

Les méthodes de classification d'images hyperspectrales sont une adaptation ou une extension de méthodes existantes, par exemple dans le cas d'images 2D. Nous pouvons par exemple citer des méthodes d'apprentissage supervisé comme les *support vector machines* (SVM) [Mountrakis et al., 2011] et plus généralement les méthodes à noyau [Camps-Valls et Bruzzone, 2005], ou les modélisations bayésiennes [Bali et Mohammad-Djafari, 2008; Borges et al., 2011]. D'autres auteurs [Benediktsson et al., 2005; Dalla Mura et al., 2011] utilisent des méthodes morphologiques pour gérer les caractéristiques spatiales dans l'image. Enfin, les méthodes basées sur la réduction de dimension linéaire [Rodarmel et Shan, 2002] ou sur des variétés [Lunga et al., 2014] permettent une description des images hyperspectrales dans des espaces adaptés, afin de faciliter les classifications. Les travaux sur la classification abordés dans ce manuscrit se placent dans le cadre des méthodes bayésiennes de segmentation : ce cadre est détaillé dans le chapitre 4.

Les problématiques de détection peuvent être vues comme un cas particulier de classification. En effet, la détection dans une image hyperspectrale a pour objectif d'isoler des spectres particuliers, présentant une forme d'*anomalie*. Une synthèse sur les méthodes de détection est présenté dans le chapitre 2.

## 1.3 Cadre applicatif et démarches adoptées

### 1.3.1 Problématiques applicatives

La problématique astronomique abordée dans ce manuscrit est double :

— la cartographie de *halos* autour de galaxies jeunes et lointaines (cf. figure 1.5). Ces halos sont majoritairement constitués de gaz, et traduisent l'histoire et l'interaction d'une galaxie avec son environnement.

— le repérage d'éventuels *filaments* formant une « toile cosmique » (cf. figure 1.5). L'existence de ces structures est prédite par les modèles astrophysiques. Elles couvrent de très grandes échelles, englobant plusieurs galaxies, voire plusieurs ensembles de galaxies. Ces structures sont elles aussi composées de gaz, et sont des résidus de l'accrétion gravitationnelle ayant conduit à la formation des galaxies.

Ces deux types d'objets sont repérables par le biais du rayonnement dit de *Lyman-alpha* illustré en figure 1.6.

Le détail des problématiques et motivations astronomiques est donné en section 8.1. Nous utilisons des données issues de l'instrument MUSE [mus, 2017]. Celui-ci fournit des images hyperspectrales astronomiques du ciel, à une résolution suffisamment fine pour accéder au rayonnement de Lyman-alpha, qui est inaccessible avec les techniques d'imagerie standard. L'instrument est décrit plus en détail dans la section 8.2.

D'un point de vue méthodologique, les problématiques astronomiques se traduisent sous la forme suivante :

— les objets sont localisés spectralement (une raie d'émission) et étendus spatialement ;

— leur forme spatiale est inconnue, et il n'existe pas d'*a priori* les concernant ;

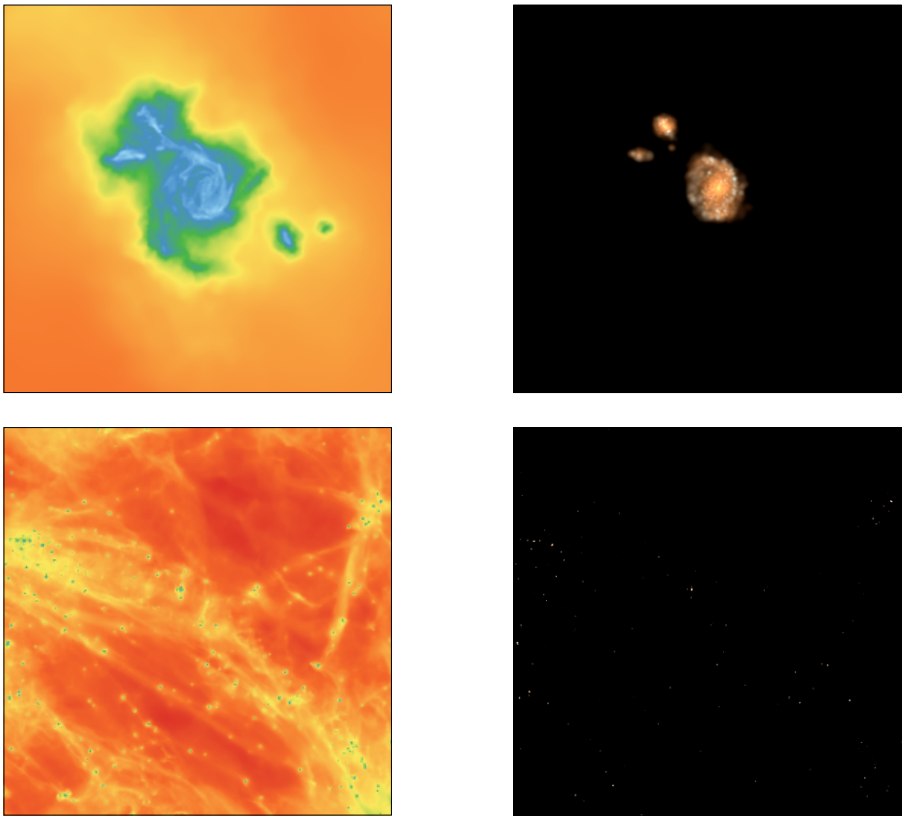


Figure 1.5 – Première ligne : halo entourant une galaxie lointaine, seconde ligne : filaments regroupant et reliant plusieurs galaxies. Ces images sont issues de simulations astrophysiques, permettant de visualiser la densité des gaz présents (à gauche, bleu-vert : importante ; orange : faible), et les intensités lumineuses effectivement visibles (à droite). Source : http://www.illustris-project.org/ [Vogelsberger et al., 2014].

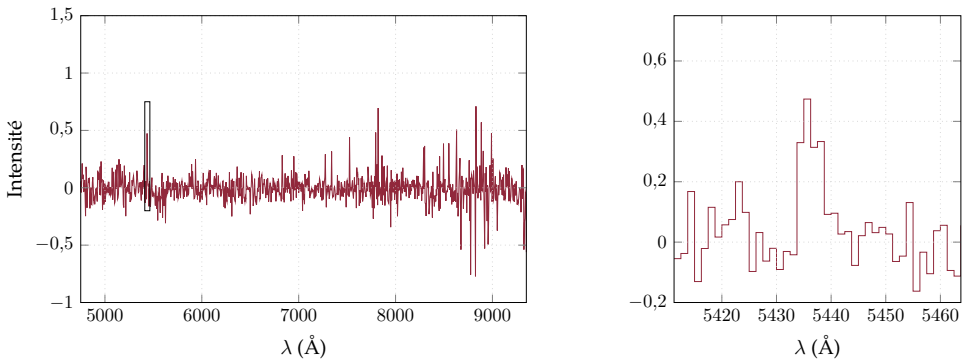



Figure 1.6 – Exemple de spectre présentant un rayonnement de Lyman-alpha, caractéristique des halos et filaments. Cet exemple est tiré d'une observation MUSE, avec à gauche un spectre sur toute la portée spectrale de l'instrument, et à droite un affichage restreint sur le rayonnement de Lyman-alpha.

— ils sont très ténus, et le bruit est d'une amplitude très importante par rapport au signal.

De plus, concernant la cartographie des halos :

— la localisation spectrale est connue : il s'agit de celle de la galaxie ;
— spatialement, ils sont situés à la périphérie de galaxies.

Cela permettra d'isoler un « sous-cube » de données pour ces traitements.

Enfin, concernant la recherche de filaments :

— spectralement, ils sont localisés dans la même bande spectrale qu'un ensemble de galaxies de décalage spectral similaire ;
— leur position dans l'espace est inconnue, mais doit traduire la distribution des galaxies.

Cela permettra d'isoler des régions susceptibles d'héberger des filaments dans les données MUSE.

### 1.3.2 Démarche et contributions

La première problématique abordée est celle de la détection de halos. Dans un premier temps, nous présentons l'objectif sous la forme d'un problème de détection, qui nous permet de proposer, dans le chapitre 3, une méthode de détection de halos par tests d'hypothèses. Ces travaux prennent notamment en compte la spécificité du problème, *via* les notions d'étalement spatial, de localisation spectrale, de similarité et d'observations multiples.

Bien que la méthode prenne explicitement en compte les paramètres de formation des images, les résultats montrent une disproportion des détections de type « faux positifs » et un manque de régularité des détections. Cela nous conduit à proposer dans le chapitre 5 une méthode alternative basée sur une modélisation bayésienne de la détection sous la forme d'un cas particulier de segmentation par champs de Markov couples.

Dans un second temps, les travaux développés visent une application à la détection de filaments. Ceux-ci étant caractérisés par une structure orientée et à grande échelle, deux approches ont été adoptées. Nous proposons d'abord dans le chapitre 6 une méthode de segmentation conjointe des classes et des orientations au sein d'une image, par le biais d'une modélisation par champs de Markov triplets. Cette méthode fournit des résultats de segmentation satisfaisants, mais ne permet pas de prendre en compte les phénomènes se produisant à grande échelle au sein d'une image.

C'est pourquoi nous proposons ensuite, dans le chapitre 7, une approche hiérarchique à la segmentation d'image, par le biais d'un modèle par arbre de Markov triplet. La modélisation proposée permet, dans le cadre de la segmentation d'image, d'éviter certains effets introduits par l'utilisation de la structure en arbre.

Enfin, une analyse comparative de ces quatre méthode est conduite dans le chapitre 10. Nous y présentons également les résultats obtenus sur les images issues de l'instrument MUSE.

# I

# Détection par tests d'hypothèses dans les images hyperspectrales

**Résumé.** Dans cette partie, nous nous plaçons dans le cadre de la détection statistique par tests d'hypothèses, avec pour contexte applicatif la détection de halos dans les données hyperspectrales astronomiques. Le chapitre 2 présente tout d'abord une synthèse sur la détection par tests dans le contexte de l'imagerie hyperspectrale. Ensuite, nous introduisons dans le chapitre 3 une méthode originale pour la détection de signaux ténus, que nous évaluons sur des données synthétiques et issues de l'instrument MUSE.

# 2

# État de l'art sur la détection par tests



Ce chapitre présente une synthèse sur la détection de sources dans des images hyperspectrales par tests d'hypothèses, avec pour objectif l'introduction des éléments importants pour la suite du manuscrit. Dans un premier temps, nous présentons en section 2.1 le cas général de la formulation par tests d'hypothèses, et l'application à la problématique de détection. Ce problème, et les méthodes existantes, sont déclinés selon la connaissance de la signature spectrale recherchée. Lorsqu'elle est parfaitement connue, nous parlerons de détection supervisée et les méthodes correspondantes sont présentées en section 2.2. Lorsqu'elle est inconnue, la détection est non-supervisée : ce type de méthode est décrit en section 2.3. Enfin, la signature spectrale peut être contrainte, ou partiellement connue. Nous parlerons dans ce cas de détection semi-supervisée et présentons ce type de méthode en section 2.4.

## 2.1 Introduction à la détection par tests d'hypothèses

### 2.1.1 Décision par tests d'hypothèses

L'observation d'un signal ou d'une scène est généralement motivée par l'acquisition d'une certaine information, permettant une prise de décision ultérieure. La question qui se pose alors est celle du lien entre les données observées et la décision à prendre. La décision dite par *tests d'hypothèses* est basée sur une famille de modèles qui permet de gérer cette prise de décision.

La formulation par tests d'hypothèses propose une description des alternatives envisagées, sous forme d'*hypothèses* que l'on notera $\mathcal{H}$. La formulation la plus courante repose sur la description d'une *hypothèse nulle*, $\mathcal{H}_0$. Dans le cas binaire, on se donne une alternative que

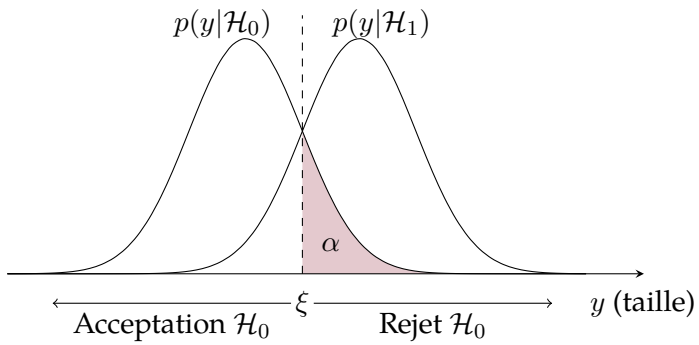



Figure 2.1 – Illustration du problème de décision statistique lors de l'observation de $y$.

l'on notera $\mathcal{H}_1$. Le cas général correspond à la formulation d'hypothèses multiples [Kay, 1998], que nous n'approfondirons pas dans ce manuscrit.

Il est souhaitable que la décision soit motivée par un certain *niveau de signification*, que l'on note $\alpha$. Ce degré de confiance permet de lier la décision à une probabilité d'erreur. Celle-ci n'est pas toujours connue, car elle dépend de l'échantillon examiné et de ses propriétés, qui ne sont pas toujours déterminées.

**Exemple 2.1.1.** *Dans une population mixte hommes/femmes, nous souhaitons déterminer le sexe d'un individu de la population à partir de sa taille uniquement. Le problème se traduira sous la forme de deux hypothèses $\mathcal{H}_0$ : "l'individu sélectionné est une femme" et $\mathcal{H}_1$ : "l'individu sélectionné est un homme". La conclusion du test sera soit le rejet soit l'impossibilité de rejeter $\mathcal{H}_0$ avec un niveau de signification $\alpha$.*

La prise de décision requiert le calcul d'une grandeur $t$ que l'on appelle *statistique de test* qui est la réalisation d'une variable aléatoire $T$. Cette valeur est réelle, et la décision résulte en général de la comparaison de $t$ avec un seuil $\xi$ : si $t < \xi$, nous retenons $\mathcal{H}_0$, sinon nous retenons $\mathcal{H}_1$ [1]. Cela s'écrit :

$$t(y) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi. \tag{2.1}$$

Le seuil $\xi$ peut, selon les situations, être lié à des grandeurs d'intérêt comme le taux d'erreur. Il peut alors être choisi pour remplir des conditions pré-définies.

**Remarque 2.1.1.** *Précisons que la décision ne consiste pas à retenir $\mathcal{H}_0$ ou $\mathcal{H}_1$, mais à rejeter ou ne pas rejeter $\mathcal{H}_0$. Cet amalgame est très fréquent dans la littérature. Dans la suite du manuscrit, nous utilisons indifféremment les deux versions, étant entendu que « accepter $\mathcal{H}_1$ » signifiera « rejeter $\mathcal{H}_0$ ».*

**Exemple 2.1.2** (suite)**.** *Dans notre exemple, l'observation $y$ est une valeur (la taille des individus) ; on peut proposer pour $T$ la fonction identité. La décision sera alors : accepter $\mathcal{H}_0$ avec un risque d'erreur $\alpha$ lorsque $t < \xi$, et ne pas pouvoir affirmer $\mathcal{H}_0$ avec ce risque si $t > \xi$. Une illustration de cette situation est donnée en figure 2.1.*

1. De manière plus générale, il est possible de comparer la statistique de test à plusieurs valeurs, par exemple pour décider $\mathcal{H}_0$ pour $t(y) \in [\xi_1, \xi_2]$. Ce type de situation n'est pas abordé dans ce manuscrit.

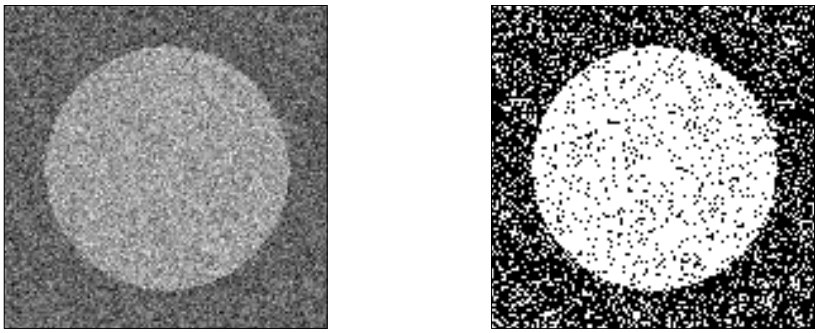


Figure 2.2 – Exemple de test LR : image bruitée à gauche et ensemble des décisions (noir : $\mathcal{H}_0$, blanc : $\mathcal{H}_1$) à droite.

### 2.1.2 Le problème de détection

Nous avons vu comment une hypothèse peut être retenue ou rejetée à l'aide d'une formulation par tests. Dans le cas de la détection, les hypothèses prennent la forme suivante :

$$\left|\begin{array}{ll} \mathcal{H}_0 & : \text{ le signal est absent de l'observation} \\ \mathcal{H}_1 & : \text{ le signal est présent dans l'observation} \end{array}\right. \tag{2.2}$$

Ces deux alternatives se déclinent selon les situations, en particulier lorsque le bruit ou le signal font l'objet d'hypothèses particulières : signal connu ou partiellement connu, distribution du bruit, etc. Le problème de détection consiste alors à trouver le test permettant la meilleure discrimination entre les deux alternatives. Le sens à attribuer à cette *meilleure discrimination* peut dépendre des situations, et est généralement lié aux erreurs susceptibles d'être commises. Nous verrons dans la section 2.1.3 que plusieurs mesures permettent de qualifier les erreurs commises lors de la détection.

Au sens du lemme de Neyman-Pearson [Kay, 1998; Neyman et Pearson, 1933], la meilleure discrimination est celle qui maximise la probabilité de détection, à probabilité de fausse alarme fixée. Il s'agit dans ce cas du test *le plus puissant*.

**Lemme 2.1.1** (Neyman-Pearson)**.** *Soient $p(y, \mathcal{H}_0)$ et $p(y, \mathcal{H}_1)$ les distributions continues de $y$ sous $\mathcal{H}_0$ et $\mathcal{H}_1$ respectivement. Pour maximiser la probabilité de détection, $P_{\mathrm{D}}$, pour une probabilité de fausse alarme, $P_{\mathrm{FA}} = \alpha$, fixée, il faut refuser $\mathcal{H}_0$ si :*

$$L(y) = \frac{p(y, \mathcal{H}_1)}{p(y, \mathcal{H}_0)} > \xi, \tag{2.3}$$

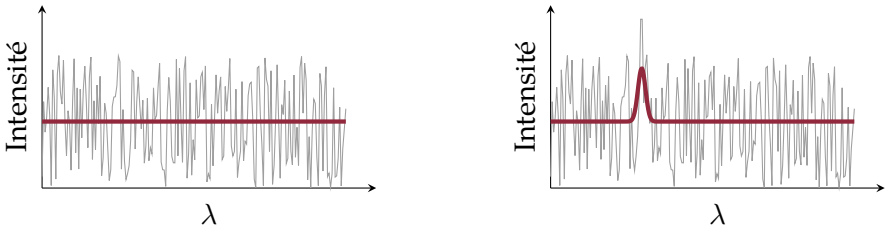
*où le seuil $\xi$ doit vérifier :*

$$\int_{\{y:L(y)>\xi\}} p(y, \mathcal{H}_0) dy = \alpha \tag{2.4}$$

*L'ensemble de la procédure est le test du* rapport de vraisemblance *(ou LR pour* likelihood ratio*).*

La preuve de ce lemme peut être trouvée dans [Kay, 1998].

**Exemple 2.1.3** (Détection dans une image bruitée)**.** *Soit une image comportant deux catégories de pixels. Dans la première, sous $\mathcal{H}_0$, les intensités des pixels ont une distribution gaussienne, de moyenne $0$ et d'écart-type $0{,}5$. Dans la seconde, vérifiant $\mathcal{H}_1$, la distribution des intensités est normale, de moyenne $1$ et de même écart-type. Les distributions des intensités sous $\mathcal{H}_0$ et $\mathcal{H}_1$ sont représentées en figure 2.1. Un exemple d'image bruitée et des résultats de tests LR (2.3), avec $\xi = 0{,}5$, est donné en figure 2.2.*




(a) $\mathcal{H}_0$ (b) $\mathcal{H}_1$

Figure 2.3 – Illustration de spectre $\mathbf{y}_s$ en présence de bruit (grisé) et sans bruit (rouge), sous les deux hypothèses alternatives.

La détection statistique peut concerner des signaux d'une ou de plusieurs dimensions. Les travaux présentés dans ce manuscrit traitent des données hyperspectrales. Dans ce cadre, la détection s'applique à chacun des spectres $\mathbf{y}_s$ de l'image hyperspectrale $(\mathbf{y}_s)_{s\in\mathcal{S}}$. La figure 2.3 donne un exemple de spectres observés sous les deux hypothèses.

Par ailleurs, le fait d'effectuer plusieurs décisions simultanément, par exemple avec l'ensemble des pixels d'une image, peut mériter une attention particulière. En effet, il peut être souhaitable de contrôler le nombre d'erreurs global, plutôt que la probabilité d'erreur de chaque décision. Il s'agit alors d'une approche de type *false discovery rate* (FDR). Le lecteur pourra se référer à [Meillier, 2015, Annexe A.2] pour une introduction aux tests multiples, qui ne seront pas développés dans ce manuscrit.

Plusieurs éléments lexicaux, souvent rencontrés dans la littérature, méritent d'être précisés pour une meilleure compréhension. Tout d'abord, le terme de *détection d'anomalies* est fréquemment rencontré. Ce vocabulaire est en particulier largement utilisé dans un contexte de télédétection, dans lequel la présence d'objets *anormaux* dans une image est significative. C'est le cas, par exemple, de caractéristiques propres à des objets artificiels au sein d'une scène ne présentant que des éléments naturels. Dans ce cadre, les méthodes reposent souvent sur une description fine de ce qui est *normal* afin de mieux repérer les anomalies (cf. section 2.3). Ce repérage repose, comme cela a déjà été mentionné, sur une discrimination efficace entre les alternatives. Il s'agit de formuler un test qui permet de séparer au mieux les deux alternatives. Comme la séparation est faite par un seuillage, il est souhaitable que la statistique associée $T$ offre un contraste pertinent entre les deux alternatives. C'est pourquoi la littérature mentionne parfois les tests de détection comme des *fonctions de contraste* [Huck et Guillaume, 2010]. Enfin, lorsque l'on effectue de la détection en imagerie hyperspectrale, le spectre caractéristique de $\mathcal{H}_1$ est indifféremment appelé *signature spectrale* ou *spectre cible*.

### 2.1.3 Évaluer les détections

L'évaluation des détections est nécessaire pour déterminer le type d'erreur susceptible d'être commise lors de la détection. Cette qualification peut aussi être utilisée dans la conception des méthodes de détection, pour minimiser la fréquence de certaines erreurs par exemple.

La détection peut résulter en une réussite ou une erreur, et chaque situation recouvre deux cas de figures qu'il faut distinguer. Une erreur peut en effet être un *faux positif* ou un *faux négatif*, et une réussite peut être un *vrai positif* ou un *vrai négatif* (cf. tableau 2.1)

Ces cas de figures sont illustrés en figure 2.4. Lorsque les calculs théoriques le permettent, il est possible d'associer une probabilité à chaque situation. On parle par exemple de

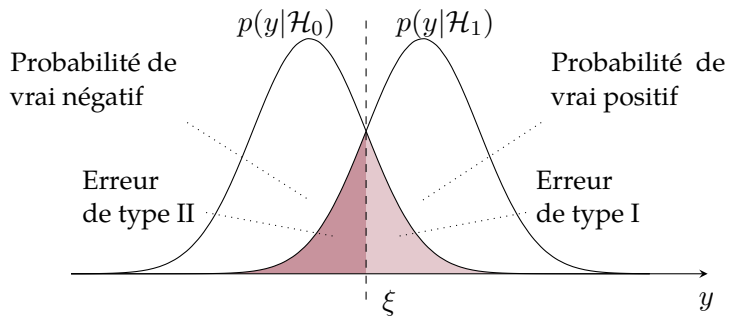



Figure 2.4 – Résultats de détection et distributions associées.

probabilités de fausse alarme ($P_{\mathrm{FA}}$) et de détection correcte ($P_{\mathrm{D}}$). Si une série de détections est effectuée et que les vérités terrain sont connues, il est possible d'évaluer empiriquement ces valeurs.

**Remarque 2.1.2.** *Plusieurs termes partagent le même sens :*
- *la probabilité d'erreur de type I ($P_{\mathrm{FA}}$) est le niveau de signification $\alpha$ ;*
- *la probabilité d'erreur de type II est notée $\beta$ ;*
- *la probabilité de détection ($P_{\mathrm{D}}$) est la puissance du test, valant $1-\beta$ ;*
- *la description d'une statistique de test $T$ associé à un seuil $\xi$ est également appelée* détecteur.

Toutes les détections dépendent d'un seuil $\xi$ à fixer. Les résultats pouvant différer selon cette valeur, il est intéressant de visualiser les performances lorsque $\xi$ varie. C'est ce que proposent les courbes ROC (*receiver operating characteristic*), à travers un affichage de $P_{\mathrm{D}}$ en fonction de $P_{\mathrm{FA}}$ pour un seuil $\xi$ variable. Un exemple de courbe ROC est donné en figure 2.5. Elle est jugée selon sa proximité avec l'axe des ordonnées ($P_{\mathrm{FA}}$ faible) et sa distance avec l'axe des abscisses ($P_{\mathrm{D}}$ élevée). Cet outil est en particulier pertinent pour comparer rapidement deux tests différents : lorsque la courbe ROC d'un premier test majore celle d'un second test, le premier est considéré de meilleure qualité. Cela se traduit aussi grâce aux mesures de type *area under curve* (AUC) sur la courbe ROC. En revanche, la courbe ROC ne contient que les informations sur la partie « positive » de la détection (décision correcte ou incorrecte de $\mathcal{H}_1$), et n'illustre pas les vrais négatifs ni les faux négatifs. Notons enfin que la courbe ROC est moins adaptée lorsqu'il s'agit d'évaluer une méthode qui n'a pas de paramètre variable : la courbe résultante n'est alors qu'un point.

Nous avons présenté le cadre général de la décision par tests d'hypothèses, puis le problème spécifique de la détection, avec le cas particulier de l'imagerie hyperspectrale. Jusqu'ici, nous avons présenté le test d'un spectre seul, en supposant implicitement la connaissance des paramètres du modèle. Nous opérons dans la suite la distinction suivante :
- la *détection supervisée*, supposant la connaissance complète du signal recherché ;
- la *détection non supervisée*, lorsque le signal à détecter est inconnu ;
- la *détection semi-supervisée*, dans le cadre d'une connaissance partielle du signal recherché.

Les parties suivantes présentent et illustrent ces catégories de détection.

Table 2.1 – Résultats possibles d'une détection par tests d'hypothèse.

| Réalité | | $\mathcal{H}_0$ | $\mathcal{H}_1$ |
|---|---|---|---|
| Décision | $\mathcal{H}_0$ | Vrai négatif | Faux négatif (type II) |
| | $\mathcal{H}_1$ | Faux positif (type I) | Vrai positif |

## 2.2 Détection supervisée

Nous présentons tout d'abord plusieurs méthodes de détection statistique que l'on considère supervisées, c'est-à-dire qu'elles reposent sur la connaissance *a priori* du spectre recherché $\boldsymbol{\mu}$. Les méthodes présentées dans cette partie font référence pour l'application à l'imagerie hyperspectrale de télédétection, et sont envisageables pour l'application à l'astronomie.

Nous nous intéressons d'abord au cas du bruit additif gaussien, multivarié spectralement. Cela se traduit par la formulation, en chaque site $s$ de l'image, des deux hypothèses suivantes :

$$\left|\begin{array}{ll} \mathcal{H}_0 & : \mathbf{y}_s = \boldsymbol{\epsilon}_s \\ \mathcal{H}_1 & : \mathbf{y}_s = \boldsymbol{\mu} + \boldsymbol{\epsilon}_s \end{array}\right. \quad \text{avec } \boldsymbol{\epsilon}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \text{ et } \boldsymbol{\mu} \neq \mathbf{0} \text{ connu} ; \tag{2.5}$$

où $\boldsymbol{\epsilon}_s$ représente une réalisation du bruit.

Un test très répandu pour la détection au sein d'images hyperspectrales est le détecteur *adaptive matched filter* (AMF) [Reed et al., 1974]. Il s'agit d'une adaptation du LR (2.3) au cas où le bruit est additif et gaussien multivarié spectralement ; il bénéficie donc des mêmes propriétés. Il repose sur la connaissance de la signature $\boldsymbol{\mu}$ recherchée ainsi que de la matrice de covariance spectrale $\boldsymbol{\Sigma}$. Lorsque cette dernière n'est pas connue, elle peut être estimée au sens du maximum de vraisemblance (MV). En remplaçant $\boldsymbol{\Sigma}$ par cette estimation, nous obtenons un test généralisant le LR, qui est de la catégorie des *generalized likelihood ratio* (GLR) [2].

Une autre approche pour fournir une discrimination entre la cible et l'arrière-plan utilise une mesure d'angle entre le spectre $\mathbf{y}_s$ traité et la signature recherchée $\boldsymbol{\mu}$. Cette mesure peut être obtenue, notamment, en calculant le cosinus entre $\mathbf{y}_s$ et $\boldsymbol{\mu}$, dans l'espace multivarié dans lequel ces vecteurs vivent. Le test qui en résulte est le *adaptive cosine/coherence estimator* (ACE) [Kraut et al., 2001; Pieper et al., 2011]. Ce test est dit cohérent car il retourne une valeur négative si $\mathbf{y}_s$ et $\boldsymbol{\mu}$ ne sont pas de même direction, et permet donc de rejeter facilement ces

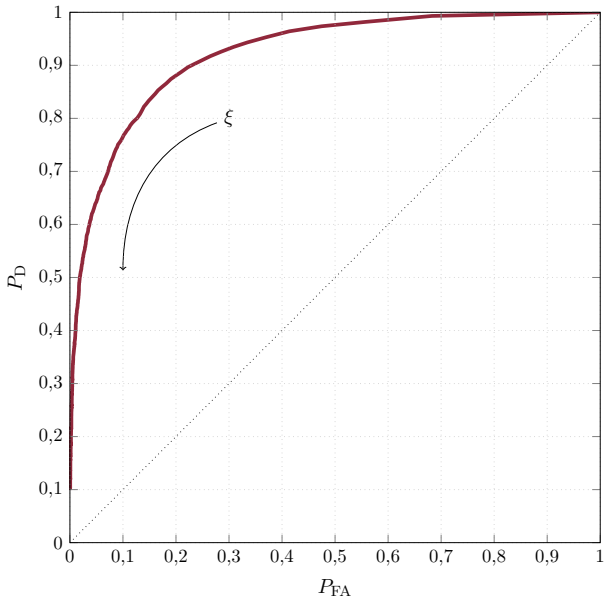



Figure 2.5 – Exemple de courbe ROC, mesurée sur l'exemple donné en figure 2.2. La flèche indique les valeurs croissantes de $\xi$.

2. Plusieurs tests de la catégorie GLR, avec $\boldsymbol{\Sigma}$ inconnu et $\boldsymbol{\mu}$ inconnu ou partiellement connu, sont présentés dans la partie suivante.

spectres. Il existe également une alternative non-cohérente du test ACE, qui est le *squared ACE* (SACE) [Kraut et al., 2001], parfois dénommé $\beta$-test.

Les tests AMF et ACE bénéficient tous deux de la propriété dite CFAR pour *constant false alarm rate* : cela signifie que la probabilité de fausse alarme est indépendante des paramètres du bruit [Kay, 1998]. Cette propriété est particulièrement attrayante pour les implémentations pratiques de ces méthodes. Elle n'est en revanche pas valide pour tous les tests, en raison notamment des étapes intermédiaires nécessaires à l'estimation d'un ou de plusieurs paramètres [3].

Parmi les méthodes plus générales, nous pouvons citer les détecteurs comme OSP (*orthogonal subspace projector*) [Chang, 2005]. La détection opère par le biais de la projection de l'observation sur un sous-espace représentant l'arrière-plan. Si la valeur projetée résultante est faible, alors le sous-espace décrit mal le spectre : il n'appartient vraisemblablement pas à l'arrière-plan. Le point délicat de ce type de modèle est la description appropriée du fond, qui doit être à la fois pertinente pour modéliser l'arrière-plan et pour construire une projection discriminante.

Dans les méthodes décrites, le bruit est supposé additif et gaussien, multivarié spectralement. Plusieurs auteurs [Manolakis et al., 2009; Schweizer et Moura, 2000] soulignent que pour l'application aux images hyperspectrales « réelles », ces hypothèses sont erronées. Ces remarques s'appliquent en particulier aux images de télédétection et ne concernent pas les

Table 2.2 – Formulation de quelques détecteurs supervisés, appliqués à une observation $\mathbf{y}_s$, dans le cadre d'un bruit additif gaussien multivarié de moyenne $\boldsymbol{\mu}$ et de covariance spectrale $\boldsymbol{\Sigma}$.

| Nom | Formulation du test | Commentaire | Références |
|---|---|---|---|
| AMF (*Adaptive Matched Filter*) | $\frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_s}{\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}}$ | Ce test est une formulation particulière de test LR (2.3). Le numérateur correspond à un filtrage adapté. | [Manolakis et al., 2009; Reed et al., 1974] |
| GLR (*Generalized Likelihood Ratio*) | $\frac{\boldsymbol{\mu}^\top (\hat{\boldsymbol{\Sigma}}^{\mathrm{MV}})^{-1} \mathbf{y}_s}{\sqrt{\boldsymbol{\mu}^\top (\hat{\boldsymbol{\Sigma}}^{\mathrm{MV}})^{-1} \boldsymbol{\mu}}}$ | Généralisation du LR (2.3) au cas où $\boldsymbol{\mu}$ est connu et $\boldsymbol{\Sigma}$ inconnu. $\hat{\boldsymbol{\Sigma}}^{\mathrm{MV}}$ est l'estimation au sens du MV de $\boldsymbol{\Sigma}$. | [Kay, 1998] |
| ACE (*Adaptive Cosine Estimator*) | $\frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_s}{\sqrt{(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})(\mathbf{y}_s^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_s)}}$ | Propriétés d'invariance en échelle et en rotation. | [Kraut et al., 2001; Manolakis et al., 2009; Pieper et al., 2011] |
| SACE (*Squared ACE*) | $\frac{(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_s)^2}{(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})(\mathbf{y}_s^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_s)}$ | | [Kraut et al., 2001] |
| SP (*Subspace Projector*) | $\frac{\boldsymbol{\mu}^\top \mathbf{P} \mathbf{y}_s}{\sqrt{\boldsymbol{\mu}^\top \mathbf{P} \boldsymbol{\mu}}}$ | Si le dénominateur vaut 1, il s'agit du détecteur OSP. | [Chang, 2005] |

3. Se reporter au chapitre 3 pour un exemple de test CFAR et des conditions nécessaires.

images hyperspectrales astronomiques que nous utilisons [4].

Mentionnons néanmoins quelques méthodes proposant une modélisation plus riche du bruit. Dans certains travaux [Frontera-Pons et al., 2016; Richmond, 2000], les auteurs modélisent en effet des distributions de bruit elliptiques. Celles-ci généralisent les distributions normales multivariées, et permettent de gérer notamment des queues de distribution plus lourdes. De plus, les détecteurs AMF, ACE et le GLR restent valables dans ce cadre [Richmond, 2000]. D'autres méthodes [Erdinç et Aksoy, 2015] représentent l'arrière-plan comme un mélange de lois normales, ce qui offre là aussi une plus grande richesse de modélisation.

L'aspect spectral de la détection est très largement exploité dans les méthodes que nous avons évoquées. En revanche, l'aspect spatial est en comparaison très peu utilisé en contexte supervisé. En effet, la plupart des algorithmes de la littérature traitent des images hyperspectrales de télédétection, dans lesquelles les cibles occupent une faible surface par rapport à la taille de l'image hyperspectrale : d'une fraction de pixel à quelques pixels [5]. Dès lors, les considérations liées au voisinage spatial des pixels traités peuvent être moins pertinentes dans le cadre de la détection.

L'expression des tests mentionnés figure dans le tableau 2.2, pour le cas d'un bruit normal multivarié. Pour approfondir le sujet de la détection supervisée, le lecteur pourra par exemple se reporter à [Manolakis et al., 2009; Chang, 2003; Kay, 1998].

## 2.3 Détection non supervisée ou détection d'anomalies

Nous nous plaçons dans le cadre suivant :

$$\left|\begin{array}{ll} \mathcal{H}_0 & : \mathbf{y}_s = \boldsymbol{\epsilon}_s \\ \mathcal{H}_1 & : \mathbf{y}_s = \boldsymbol{\mu} + \boldsymbol{\epsilon}_s \end{array}\right. \quad \text{avec } \boldsymbol{\epsilon}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \text{ et } \boldsymbol{\mu} \neq \mathbf{0} \text{ inconnu.} \tag{2.6}$$

Contrairement au cadre présenté en (2.5), le spectre du signal recherché, $\boldsymbol{\mu}$, est inconnu. Ainsi, le problème est double : il faut, en présence de bruit, conjointement détecter et reconnaître le signal d'intérêt. L'hypothèse $\mathcal{H}_0$ correspond au cas « normal » de signal bruité : l'observation $\mathbf{y}_s$ ne contient aucun élément autre que le bruit. En ce sens, la recherche de $\boldsymbol{\mu}$ correspond à de la détection d'anomalies.

Le GLR peut permettre de détecter une anomalie $\boldsymbol{\mu}$ de forme spectrale inconnue. Ce test propose de remplacer, dans le test LR (2.3), $\boldsymbol{\mu}$ par une estimation au sens du maximum de vraisemblance, $\hat{\boldsymbol{\mu}}^{\mathrm{MV}}$. Sans information supplémentaire, nous avons en réalité $\hat{\boldsymbol{\mu}}^{\mathrm{MV}} = \mathbf{y}_s$. Cela permet de formuler le GLR comme un détecteur d'énergie dans le cadre d'un bruit gaussien.

Un autre détecteur très utilisé en tant que référence (et outil de comparaison) est le détecteur de Reed-Xiaoli (RX) [Reed et Xiaoli, 1990]. Cette méthode mesure à quel point le spectre en un site est anormal, étant donné un fond et sa matrice de covariance estimée. La particularité de cette méthode est l'estimation de paramètres du bruit localement, au sein d'une fenêtre spatialement centrée sur le spectre testé. Ce test est dérivé du test GLR, et bénéficie également de la propriété CFAR [Reed et Xiaoli, 1990]. Le détecteur RX fait l'hypothèse d'un modèle de mélange linéaire local : sous $\mathcal{H}_0$, le spectre observé est de loi normale. Des situations plus complexes peuvent être modélisées en faisant l'hypothèse d'un mélange de lois normales, ce qui est pertinent pour la détection d'anomalies dans un environnement non homogène. Deux types de méthodes se placent dans ce cadre. Les

4. Se reporter au chapitre 9 pour une analyse des données réelles.

5. Les sources occupant une fraction de pixel sont gérées par la famille méthodes de détection dite *sous-pixélique*, que nous ne développons pas dans ce manuscrit. Le lecteur intéressé pourra se reporter à [Manolakis et al., 2001].

premières [Ashton, 1998; Carlotto, 2005] opèrent une classification des spectres en amont des détections, qui sont réalisées par classes ensuite. Les secondes [Chen et Reed, 1987; Stein et al., 2002] correspondent à un GLR par composante du mélange.

Parmi les travaux plus récents, nous pouvons mentionner les méthodes dites de *projection pursuit* (PP) [Chiang et al., 2001; Du et Kopriva, 2008; Huck et Guillaume, 2010; Malpica et al., 2008]. Le but de ce type de méthode est de projeter les observations sur un espace d'une dimension, avec comme objectif l'optimisation d'un critère particulier, comme l'estimation du kurtosis qui est sensible aux valeurs aberrantes. Il s'agit de méthodes itératives, qui n'ont pas de formulation en une étape comme les tests mentionnés précédemment.

D'autres méthodes reposent sur la théorie des valeurs extrêmes [Broadwater et Chellappa, 2010; Zhang et al., 2013] et gèrent explicitement la présence de valeurs aberrantes. Ce type de méthode est utile lorsque la distribution du bruit est inconnue ou difficile à estimer à partir de l'observation. Elles sont également efficaces lorsque le spectre recherché est rarement rencontré dans l'image.

Nous donnons dans le tableau 2.3 l'expression du RX et du GLR, les autres méthodes mentionnées ici n'ayant pas d'expression analytique directe. Le lecteur pourra notamment se référer à [Matteoli et al., 2010] pour approfondir le sujet de la détection non supervisée.

## 2.4 Détection semi-supervisée

Le cadre de la détection semi-supervisée est le suivant :

$$\left| \begin{array}{ll} \mathcal{H}_0 & : \mathbf{y}_s = \boldsymbol{\epsilon}_s \\ \mathcal{H}_1 & : \mathbf{y}_s = \boldsymbol{\mu} + \boldsymbol{\epsilon}_s \end{array} \right. \quad \text{avec } \boldsymbol{\epsilon}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \text{ et } \boldsymbol{\mu} \neq \mathbf{0} \text{ est contraint.} \tag{2.7}$$

La contrainte sur le spectre $\boldsymbol{\mu}$ est une information permettant de le caractériser, comme sa distribution ou sa forme. Le problème est facilité par rapport à celui de la détection non supervisée et constitue son extension naturelle permettant l'utilisation d'informations sur $\boldsymbol{\mu}$. Cette partie détaille plus particulièrement les méthodes pertinentes vis-à-vis de l'application et des développements effectués dans la suite du manuscrit.

La méthode LRMAP [Paris et al., 2011] repose sur l'estimation de $\boldsymbol{\mu}$ au sens du MAP dans le test LR, au lieu de l'estimation au sens du MV menant au test GLR. Ce type d'estimation requiert un *a priori* sur ce spectre : dans [Paris et al., 2011] il s'agit d'un *a priori* laplacien sur les intensités des spectres qui est utilisé. Cet *a priori* a vocation à favoriser les estimations parcimonieuses de $\boldsymbol{\mu}$, comportant peu de valeurs élevées. La méthode *posterior density ratio*

Table 2.3 – Formulation de deux détecteurs non supervisés dans le cadre d'un bruit additif gaussien multivarié de moyenne $\boldsymbol{\mu}$ et de covariance spectrale $\boldsymbol{\Sigma}$.

| Nom | Formulation | Commentaire | Références |
|---|---|---|---|
| GLR (*Generalized Likelihood Ratio*) | $\dfrac{\max_{\boldsymbol{\mu}} p(\mathbf{y}_s \vert \boldsymbol{\mu}, \mathcal{H}_1)}{p(\mathbf{y}_s \vert \mathcal{H}_0)}$ | Le numérateur peut être remplacé par $p(\mathbf{y}_s \vert \hat{\boldsymbol{\mu}}^{\mathrm{MV}}, \mathcal{H}_1)$, où $\hat{\boldsymbol{\mu}}^{\mathrm{MV}}$ est l'estimateur de $\boldsymbol{\mu}$ au sens du maximum de vraisemblance. | [Kay, 1998] |
| RX (Reed-Xiaoli) | $\mathbf{y}_s^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_s$ | Il s'agit de la distance de Mahalanobis entre l'observation et un fond de moyenne nulle. | [Reed et Xiaoli, 1990] |

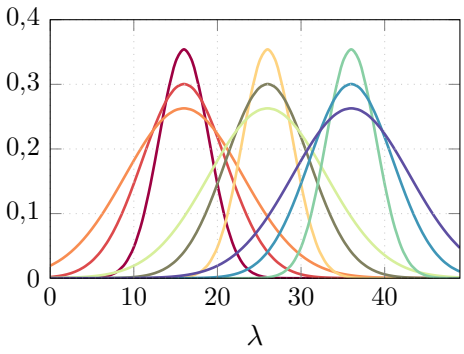



FIGURE 2.6 – Exemple de catalogue $\mathbf{C}$ de raies gaussiennes, pour trois positions et trois largeurs différentes.

(PDR), utilisée également dans [Paris et al., 2011] et proposée dans [Basu, 1996], repose aussi sur des *a priori* faits sur $\boldsymbol{\mu}$. En revanche, le PDR repose sur un rapport de vraisemblances *a posteriori*, de type $p(\mathcal{H}|\mathbf{y}_s)$, contrairement aux rapports de vraisemblances de type $p(\mathbf{y}_s|\mathcal{H})$ utilisés dans les tests dérivés du LR.

Le PDR et le LRMAP ont tous deux été développés dans le but de favoriser la parcimonie de $\boldsymbol{\mu}$. Les hypothèses considérées sont toujours identiques, et seules les formulations des tests et des estimées y intervenant changent.

La parcimonie peut également être traduite sous la forme de tests par éléments du vecteur $\mathbf{y}_s$. C'est l'approche du test dit *de signification de niveau 2*, ou du *higher criticism* (HC) [Donoho et Jin, 2004]. Le HC formule le test d'un spectre comme la résultante des tests de toutes ses composantes, dans le cadre de tests multiples.

Pour renforcer l'incitation à la parcimonie, il est possible d'expliciter les formes spectrales attendues de manière déterministe, dès la formulation des hypothèses. Supposons par exemple que nous souhaitions contraindre la forme spectrale de $\boldsymbol{\mu}$ à appartenir, à un coefficient d'atténuation près, à un catalogue de $J$ raies spectrales $\mathbf{c}$ de la forme $(\mathbf{c}_j)_{1\leq j\leq J}$. Nous pouvons écrire :

$$\left|\begin{array}{ll} \mathcal{H}_0 & : \mathbf{y}_s = \boldsymbol{\epsilon}_s \\ \mathcal{H}_1 & : \mathbf{y}_s = \boldsymbol{\mu} + \boldsymbol{\epsilon}_s = \mathbf{c}_j\alpha + \boldsymbol{\epsilon}_s \end{array}\right. \quad \text{et } \boldsymbol{\epsilon}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}); \tag{2.8}$$

où l'indice $j$ et le coefficient d'atténuation $\alpha \neq 0$ sont inconnus. Le catalogue $\mathbf{c}$ peut prendre

TABLE 2.4 – Formulation de quelques détecteurs semi-supervisés.

| Acronyme | Formulation | Commentaire | Références |
|---|---|---|---|
| LRMAP (*Likelihood Ratio*/MAP) | $\frac{p(\mathbf{y}_s\|\hat{\boldsymbol{\mu}}^{\text{MAP}}, \mathcal{H}_1)}{p(\mathbf{y}_s\|\mathcal{H}_0)}$ | Ce test requiert la définition d'un *a priori* sur $\boldsymbol{\mu}$. | [Paris et al., 2011; Paris, 2013] |
| PDR (*Posterior Density Ratio*) | $\frac{\max_{\boldsymbol{\mu}} p(\boldsymbol{\mu}, \mathcal{H}_1\|\mathbf{y}_s)}{p(\mathcal{H}_0\|\mathbf{y}_s)}$ | Le numérateur peut être remplacé par $p(\hat{\boldsymbol{\mu}}^{\text{MAP}}, \mathcal{H}_1\|\mathbf{y}_s)$. Un *a priori* sur $\boldsymbol{\mu}$ est aussi nécessaire. | [Paris et al., 2011; Paris, 2013] |
| GLR-1s (GLR 1-parc.) | $\frac{\max_{j,\alpha} p(\mathbf{y}_s\|\mathbf{c}_j\alpha, \mathcal{H}_1)}{p(\mathbf{y}_s\|\mathcal{H}_0)}$ | Ce test requiert la connaissance d'un catalogue $\mathbf{C}$ de raies spectrales. | [Paris et al., 2013b; Paris, 2013] |

une forme quelconque. Dans le cas de la 1-parcimonie de [Paris et al., 2013b], il prend la forme d'une série de raies gaussiennes, de largeurs et de positions variables, comme illustré en figure 2.6. Ainsi, dans le cadre des hypothèses (2.8), il est possible de formuler un GLR *1-parcimonieux* (GLR-1s), tel qu'il est décrit dans [Paris et al., 2013b]. Il est aussi possible de contraindre la forme spatiale du signal recherché, dans le cas où il couvrirait plusieurs pixels. Ce type d'approche est décrit dans le chapitre 3. Les expressions des tests LRMAP, PDR et GLR-1s sont reportées dans le tableau 2.4.

## 2.5 Bilan

Nous avons présenté dans cette partie les principes de la détection par tests d'hypothèses dans des images hyperspectrales. Nous avons ensuite classé les méthodes existantes, selon leur utilisation du spectre ciblé $\boldsymbol{\mu}$. Nous avons ainsi pu détailler les méthodes les plus connues, ainsi que celles pertinentes vis-à-vis de la suite du manuscrit. Remarquons également que la plupart des méthodes récentes de détection traitent des images hyperspectrales de télédétection, et résolvent des problèmes qui ne sont pas ceux rencontrés en imagerie hyperspectrale astronomique, comme par exemple la modélisation complexe du bruit.

Dans la partie suivante, nous nous placerons dans la continuité des tests semi-supervisés présentés dans la section 2.4. En effet, l'application astronomique permet d'utiliser plusieurs informations sur la forme spectrale et spatiale des objets recherchés.

# 3

# Détection par tests d'hypothèses de sources ténues et étendues



## 3.1 Introduction

La problématique applicative est celle de la détection de halos dans les données MUSE, décrite dans le chapitre 1. Dans ce contexte, les halos sont localisés spectralement en une unique raie d'émission, et localisés spatialement en périphérie d'une galaxie. Nous supposerons qu'un objet (le halo et la galaxie) est caractérisé par un ensemble de spectres, que nous appellerons *sources* en l'absence de bruit. Le spectre des sources présente seulement quelques coefficients non nuls, correspondant à une raie d'émission. Spatialement, l'intensité lumineuse s'atténue en s'éloignant du centre : cela signifie que quelques spectres centraux peuvent être considérés comme *brillants* et que tous les autres sont *ténus*. Cette distinction est reprise dans la construction de la stratégie de détection proposée.

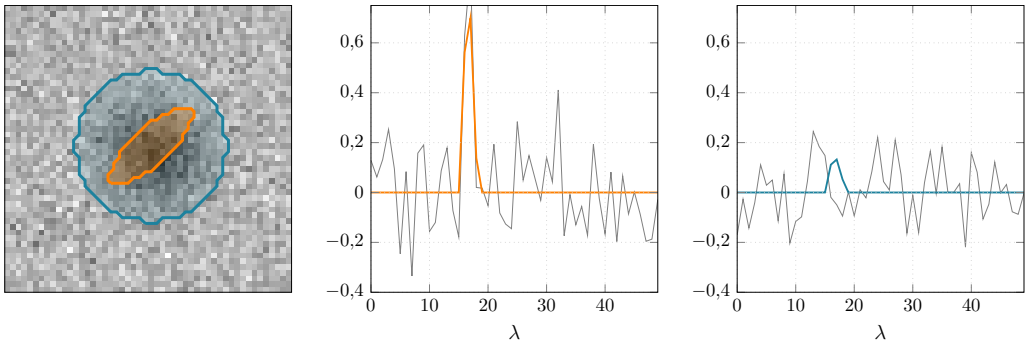



Figure 3.1 – Illustration du problème de détection posé dans ce chapitre à l'aide de données de synthèse. À gauche : moyenne spectrale d'une image hyperspectrale (inverse vidéo), à laquelle sont superposées une région brillante (orange, au centre) et une région ténue (bleue, en périphérie). Les spectres correspondants sont illustrés à droite : en couleur les spectres d'origine, en gris un spectre observé.

Enfin, nous supposerons que les spectres brillants et ténus sont similaires à un facteur d'atténuation près. Cette problématique est illustrée en figure 3.1.

Nous retenons également trois éléments concernant l'instrument d'observations :

— nous supposons la PSF de l'instrument séparable en une fonction d'étalement spectral (LSF) et une fonction d'étalement spatial (FSF), cette dernière n'étant pas négligeable par rapport à la taille des pixels ;
— plusieurs observations d'une même région du ciel sont disponibles, ainsi que leur fusion ;
— nous nous plaçons pour cette fusion dans le cadre d'un bruit gaussien, multivarié spectralement [1].

Les principales contributions méthodologiques de ce chapitre sont les suivantes :

— une stratégie de détection en deux temps est proposée, consistant d'abord en une *détection de sources brillantes* puis en une *détection de sources ténues*. La première étape permet de repérer, de manière non supervisée, les spectres les plus brillants. La seconde étape s'appuie sur ces premiers résultats pour rechercher les spectres similaires et plus faibles au sein du reste de l'image hyperspectrale ;
— nous introduisons un test GLR avec une contrainte de *similarité* avec un autre spectre de l'image. Nous montrons également que ce test peut avoir la propriété CFAR ;
— par extension, nous introduisons un test composite, qui gère la similarité avec un ensemble de spectres. Il est utilisé dans la seconde étape de la stratégie de détection. Nous montrons que son taux de fausse alarme a une limite supérieure qui est indépendante des paramètres du bruit ;
— les modèles sont étendus pour l'utilisation des relations au sein d'un voisinage local, induites par la FSF ;
— nous explorons également l'utilisation d'observations multiples d'une même scène afin d'exploiter au mieux les informations disponibles. Expérimentalement, cette approche ne permet pas d'amélioration notable des résultats (section 3.5.2).

Du point de vue de la validation expérimentale, nous présentons :

— des résultats numériques sur des données de synthèse, qui nous permettent une comparaison avec d'autres méthodes non supervisées [Ahmad et al., 2015; Reed et Xiaoli, 1990] par le biais d'une analyse des courbes ROC ;

1. Se reporter au chapitre 9 pour une analyse plus détaillée du bruit dans les données MUSE.

Table 3.1 – Résumé des tests de détection employés. Les lettres entre guillemets représentent les indices des tests $\mathcal{T}$ et des seuils utilisés $\xi$ dans ce contexte. "✓","✗", "–" repèrent lorsque le modèle correspondant est utilisé, non utilisé ou non pertinent.

| | | | | Voisinage | Observations | Composite | Similarité |
|---|---|---|---|---|---|---|---|
| | Nom | Section | Équation | « n » | « o » | « c » | « s » |
| Brillant | $\mathcal{T}$ | 3.2.1 | (3.3) | ✗ | ✗ | – | – |
| | $\mathcal{T}_{\text{n}}$ | 3.3.1 | (3.25) | ✓ | ✗ | – | – |
| | $\mathcal{T}_{\text{no}}$ | 3.3.2 | (3.34) | ✓ | ✓ | – | – |
| Ténu | $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}$ | 3.2.2 | (3.12) | ✗ | ✗ | ✗ | ✓ |
| | $\mathcal{T}_{\text{ns}}^{\hat{\mathbf{x}}_b^{\text{N}}}$ | 3.2.2 | (3.26) | ✓ | ✗ | ✗ | ✓ |
| | $\mathcal{T}_{\text{nos}}^{\hat{\mathbf{x}}_b^{\text{NO}}}$ | 3.3.2 | (3.35) | ✓ | ✓ | ✗ | ✓ |
| | $\mathcal{T}_{\text{cs}}^{\mathcal{B}}$ | 3.2.2 | (3.17) | ✗ | ✗ | ✓ | ✓ |
| | $\mathcal{T}_{\text{ncs}}^{\mathcal{B}}$ | 3.3.1 | (3.28) | ✓ | ✗ | ✓ | ✓ |
| | $\mathcal{T}_{\text{nocs}}^{\mathcal{B}}$ | 3.3.2 | (3.36) | ✓ | ✓ | ✓ | ✓ |

— l'application à des données réelles issues de l'instrument MUSE [2].

Nous utilisons tout au long de ce chapitre plusieurs tests GLR, la plupart étant contraints à la manière du $\text{GLR}_{\text{1s}}^{(\text{1D})}$ proposé dans [Paris et al., 2013b]. Par souci de concision, nous noterons $\mathcal{T}$ tout test GLR avec une contrainte spectrale (définie ensuite). Nous explorons différents types de variations de test, incluant ou non des contraintes spatiales, d'observations multiples, de similarité et composites. Ces tests, leur désignation et les sections dans lesquelles ils sont définis sont reportés dans le tableau 3.1.

Dans les Dans les sections 3.2 et 3.3, nous introduisons les modèles et les tests de détection associés. Les propriétés de certains des tests tests sont étudiées dans la section 3.4. Nous étudions ensuite les performances des tests introduits sur données de synthèse dans la section 3.5. Enfin, la section 3.6 présente la stratégie de détection adoptée, ses performances, ainsi que l'application aux données réelles.

## 3.2 Modèles et tests de détection

Dans cette partie, nous présentons les modèles employés pour la détection. L'approche est séparée en deux étapes : la détection de sources brillantes (section 3.2.1) et de sources ténues (section 3.2.2).

Nous notons $\mathbf{y}_s$ la réalisation de la variable aléatoire $\mathbf{Y}_s$, et $\mathcal{S}$ est l'ensemble des sites dans l'image. Par concision et lorsque cela ne gênera pas la compréhension du texte, nous noterons $p(\mathbf{y}_s)$ la densité de probabilité de $\mathbf{Y}_s$, et de même pour les autres variables aléatoires. Nous repérerons de plus l'ensemble $\mathcal{B}$ de spectres *brillants* obtenus à l'issue de la première étape et $\mathcal{F} = \mathcal{S} \setminus \mathcal{B}$ son complément dans $\mathcal{S}$. Enfin, nous ferons usage d'un *catalogue* $\mathbf{c} \in \mathbb{R}^{\Lambda \times J}$ de raies spectrales contenant $J$ colonnes notées $\mathbf{c}_j$ (formées par exemple de gaussiennes,

2. Des résultats complets sont de plus présentés dans le chapitre 10.

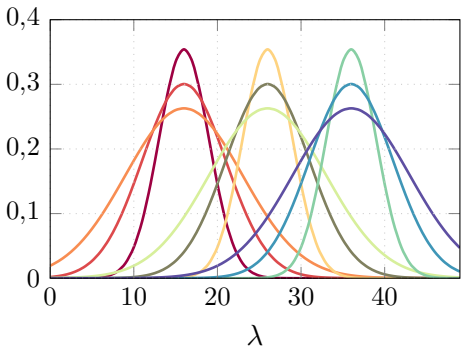



Figure 3.2 – Quelques raies d'un catalogue $\mathbf{c}$ utilisé pour la construction de tests GLR contraints.

comme illustré en figure 3.2).

### 3.2.1 Détection de sources brillantes

Cette étape a pour objectif la détection non supervisée des spectres les plus brillants de l'image hyperspectrale. Pour tout $s \in \mathcal{S}$, nous posons deux hypothèses concurrentes :

$$\left| \begin{array}{l} \mathcal{H}_0 : \mathbf{y}_s = \boldsymbol{\epsilon}_s; \\ \mathcal{H}_1 : \mathbf{y}_s = \mathbf{x}_s + \boldsymbol{\epsilon}_s. \end{array} \right. \tag{3.1}$$

$\boldsymbol{\epsilon} \in \mathbb{R}^{\Lambda}$ est une réalisation d'un bruit additif et suit une loi normale centrée en $\mathbf{0}$ et de matrice de covariance $\boldsymbol{\Sigma}$, notée $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. La recherche d'une raie d'émission nous permet de supposer que le contenu spectral $\mathbf{x}_s$ peut être décrit à partir d'une colonne $\mathbf{c}_j$ de $\mathbf{c}$ et d'un coefficient $\alpha_s \in \mathbb{R}^{*+}$ :

$$\left| \begin{array}{l} \mathcal{H}_0 : \mathbf{y}_s = \boldsymbol{\epsilon}_s; \\ \mathcal{H}_1 : \mathbf{y}_s = \alpha_s \mathbf{c}_j + \boldsymbol{\epsilon}_s. \end{array} \right. \tag{3.2}$$

Nous utilisons le test GLR contraint [Paris et al., 2013b] qui correspond à ce nouveau jeu d'hypothèses :

$$\mathcal{T}(\mathbf{y}_s) = \frac{\max\limits_{\mathbf{c}_j, \alpha_s} p(\mathbf{y}_s \,|\, \mathbf{c}_j, \alpha_s, \mathcal{H}_1)}{p(\mathbf{y}_s | \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi. \tag{3.3}$$

L'expression analytique de ce test est :

$$\frac{(\mathbf{c}_{\hat{\jmath}}^{\top} \boldsymbol{\Sigma}^{-1} \mathbf{y}_s)^2}{\mathbf{c}_{\hat{\jmath}}^{\top} \boldsymbol{\Sigma}^{-1} \mathbf{c}_{\hat{\jmath}}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\ln(\xi) \text{ où } \hat{\jmath} = \arg\max_j p(\mathbf{y}_s \,|\, \hat{\alpha}_s, \mathbf{c}_j, \mathcal{H}_1); \tag{3.4}$$

où $\hat{\alpha}_s$ est l'estimation au sens du maximum de vraisemblance de $\alpha_s$. Dans le cas d'un bruit normal, $\hat{\jmath}$ peut être obtenu par recherche exhaustive et $\hat{\alpha}_s$ est obtenu par (3.8).

*Démonstration.* Développons l'expression (3.3) dans le cas où $\boldsymbol{\epsilon}_s$ est de distribution normale,

multivariée spectralement :

$$\begin{aligned}\mathcal{T}(\mathbf{y}_s) &= \frac{\max\limits_{\mathbf{c}_j,\alpha_s} \exp\left[-\frac{1}{2}(\mathbf{y}_s-\alpha_s\mathbf{c}_j)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_s-\alpha_s\mathbf{c}_j)\right]}{\exp\left[-\frac{1}{2}\mathbf{y}_s^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s\right]} \\ &= \max_{\mathbf{c}_j,\alpha_s} \exp\left[-\frac{1}{2}\mathbf{y}_s^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s - \frac{\alpha_s^2}{2}\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j + \alpha_s\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s + \frac{1}{2}\mathbf{y}_s^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s\right] \\ &= \max_{\mathbf{c}_j,\alpha_s} \exp\left[-\frac{\alpha_s^2}{2}\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j + \alpha_s\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s\right] \end{aligned} \tag{3.5}$$

Nous estimons $\alpha_s$ au sens du maximum de vraisemblance :

$$\begin{aligned}\hat{\alpha}_s &= \arg\max_{\alpha_s} p(\mathbf{y}_s \mid \alpha_s, \mathbf{c}_j, \mathcal{H}_1) \\ &= \arg\max_{\alpha_s} \exp\left[-\frac{1}{2}(\mathbf{y}_s-\alpha_s\mathbf{c}_j)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_s-\alpha_s\mathbf{c}_j)\right]. \end{aligned} \tag{3.6}$$

En annulant la dérivée du logarithme de la vraisemblance, nous avons :

$$0 = \mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s - \hat{\alpha}_s\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j. \tag{3.7}$$

L'estimateur au sens du maximum de vraisemblance de $\alpha_s$ est donc :

$$\hat{\alpha}_s = \frac{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s}{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j}. \tag{3.8}$$

Nous pouvons donc remplacer son expression dans (3.5) :

$$\begin{aligned}\mathcal{T}(\mathbf{y}_s) &= \max_{\mathbf{c}_j} \exp\left[-\frac{1}{2}\left(\frac{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s}{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j}\right)^2\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j + \frac{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s}{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j}\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s\right]; \\ &= \max_{\mathbf{c}_j} \exp\left[\frac{1}{2}\frac{(\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_s)^2}{\mathbf{c}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}_j}\right]. \end{aligned} \tag{3.9}$$

Ainsi, si nous prenons pour $j$ l'estimation $\hat{\jmath} = \arg\max\limits_j p(\mathbf{y}_s \mid \hat{\alpha}_s, \mathbf{c}_j, \mathcal{H}_1)$, nous obtenons la formule (3.4).

□

L'expression de (3.4) permet d'exprimer le taux de fausse alarme grâce à une distribution du $\chi^2$, dans le cas où $\mathbf{c}_j$ est connu indépendamment de l'observation. Dans [Paris et al., 2013b], il a été montré que ce test est plus performant que le GLR sans contraintes pour l'application à des signaux parcimonieux. Une fois appliqué à l'ensemble $\mathcal{S}$ des spectres de l'image hyperspectrale, nous catégorisons la région dans laquelle $\mathcal{H}_1$ est retenue comme l'ensemble des spectres brillants $\mathcal{B}$. La figure 3.3 donne un exemple de détection suite au test (3.3), et montre que ce test n'est pas assez puissant pour détecter les signaux les plus ténus. En revanche, ils permettent une première caractérisation des sources d'émission, sur laquelle s'appuiera la recherche de sources ténues.

**Remarque 3.2.1.** *Le test (3.3) repose sur la connaissance de la matrice de covariance $\boldsymbol{\Sigma}$ et du catalogue $\mathbf{c}$. La matrice de covariance peut, par exemple, être estimée sur l'ensemble des spectres de l'image. Le catalogue peut quant à lui être paramétré, par exemple sous la formes de gaussiennes dont les moyenne et écart-types couvrent les éventualités (cf. section 3.5.1). Cela signifie que les imprécisions du catalogue résultent d'imprécisions des paramètres, ce qui rend le résultat final plus robuste à ces erreurs.*

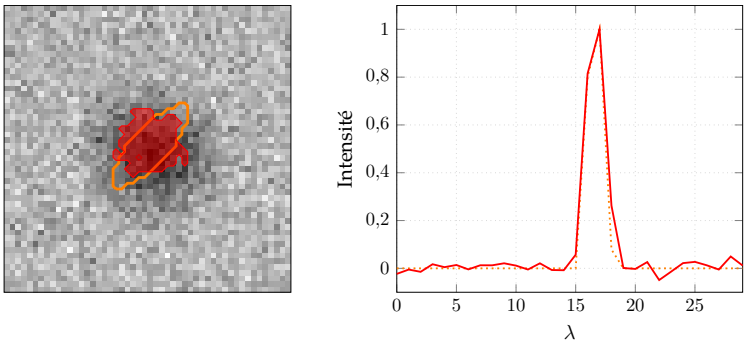



Figure 3.3 – Exemple de détection à l'aide du test (3.3) sur des données de synthèse avec RSB $= 5$ dB. À gauche : moyenne spectrale de l'image hyperspectrale (inverse vidéo) et contours réels (orange) et détectés (rouge). Les spectres normalisés réels et estimés (couleurs identiques) sont présentés dans la figure de droite.

### 3.2.2 Détection de sources ténues

Nous supposons dans la suite que l'ensemble $\mathcal{B}$ est fixé. L'objectif de cette partie est la détection de sources ténues dans les autres spectres de l'image indexés par $\mathcal{F} = \mathcal{S} \setminus \mathcal{B}$. Nous présentons dans cette partie deux tests GLR sous contrainte : le premier exploite la notion de similarité avec un spectre et le second (dit composite) la similarité avec un ensemble de spectres.

#### Test GLR avec contrainte de similarité

Soient $\mathbf{y}_b$ et $\mathbf{y}_f$ deux spectres de l'image hyperspectrale, respectivement brillant ($b \in \mathcal{B}$) et ténu ($f \in \mathcal{F}$). Le modèle que nous présentons suppose qu'en présence de source ténue ces deux spectres sont caractérisés par les vecteurs sous-jacents $\mathbf{x}_b$ et $\mathbf{x}_f$ respectivement. Ceux-ci sont liés par un coefficient d'atténuation $\beta_{f,b} \in \mathbb{R}^{*+}$. Les hypothèses émises pour le test d'un spectre $\mathbf{y}_f$ sont les suivantes :

$$\left| \begin{array}{l} \mathcal{H}_0 : \mathbf{y}_f = \boldsymbol{\epsilon}_f; \\ \mathcal{H}_1 : \mathbf{y}_f = \mathbf{x}_f + \boldsymbol{\epsilon}_f = \beta_{f,b}\mathbf{x}_b + \boldsymbol{\epsilon}_f; \end{array} \right. \tag{3.10}$$

où $\boldsymbol{\epsilon}_f$ est une réalisation d'une loi normale, multivariée spectralement et centrée en $\mathbf{0}$, de matrice de covariance $\boldsymbol{\Sigma}$. Comme $\mathbf{x}_b$ fait partie des spectres de $\mathcal{B}$, nous pouvons écrire :

$$\left| \begin{array}{l} \mathcal{H}_0 : \mathbf{y}_f = \boldsymbol{\epsilon}_f; \\ \mathcal{H}_1 : \mathbf{y}_f = \beta_{f,b}\alpha_b\mathbf{c}_j + \boldsymbol{\epsilon}_f. \end{array} \right. \tag{3.11}$$

Posons $\hat{\mathbf{x}}_b = \hat{\alpha}_b\mathbf{c}_{\hat{\jmath}}$ l'estimation au sens du maximum de vraisemblance de $\mathbf{x}_b$. Les termes $\hat{\alpha}_b$ et $\hat{\jmath}$ sont obtenus de la même manière que dans l'équation (3.4).

**Définition 3.2.1.** *Soit un spectre de référence $\hat{\mathbf{x}}_b$. Le test GLR avec contrainte de similarité $\mathcal{T}_{\mathrm{s}}^{\hat{\mathbf{x}}_b}$ est défini par :*

$$\begin{aligned} \mathcal{T}_{\mathrm{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f) &= \frac{\max\limits_{\mathbf{x}_f} p(\mathbf{y}_f \mid \mathbf{x}_f, \mathcal{H}_1)}{p(\mathbf{y}_f \mid \mathcal{H}_0)} \\ &= \frac{p(\mathbf{y}_f \mid \hat{\beta}_{f,b}\hat{\mathbf{x}}_b, \mathcal{H}_1)}{p(\mathbf{y}_f \mid \mathcal{H}_0)}; \end{aligned} \tag{3.12}$$

*où $\hat{\beta}_{f,b}$ est l'estimation de $\beta_{f,b}$ au sens du maximum de vraisemblance.*

Ce test peut être formulé de la manière suivante :

$$\mathcal{T}_{\mathrm{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi_{\mathrm{s}} \;\Leftrightarrow\; \frac{(\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f)^2}{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \hat{\mathbf{x}}_b} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\ln(\xi_{\mathrm{s}}). \tag{3.13}$$

*Démonstration.* Nous nous plaçons dans le cas d'un bruit normal, multivarié spectralement. Nous pouvons d'abord écrire l'estimation de $\beta_{f,b}$ au sens du maximum de vraisemblance, en suivant le même raisonnement que dans (3.6-3.8). Pour tout $f \in \mathcal{F}$ et $b \in \mathcal{B}$ :

$$\hat{\beta}_{f,b} = \arg\max_{\beta_{f,b}} p(\mathbf{y}_f \mid \beta_{f,b}, \hat{\mathbf{x}}_b, \mathcal{H}_1) = \frac{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f}{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \hat{\mathbf{x}}_b}. \tag{3.14}$$

En suivant un raisonnement similaire à (3.5), nous pouvons écrire le test sous la forme :

$$\mathcal{T}_{\mathrm{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f) = \exp\left[-\frac{\beta_{f,b}^2}{2} \hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \hat{\mathbf{x}}_b + \beta_{f,b} \hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f\right] \tag{3.15}$$

En remplaçant $\beta_{f,b}$ par son estimation au sens du maximum de vraisemblance (3.14), nous obtenons enfin :

$$2\ln[\mathcal{T}_{\mathrm{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)] = \frac{(\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f)^2}{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \hat{\mathbf{x}}_b}. \tag{3.16}$$

□

Cette expression nous permet en section 3.4 de déduire le taux d'erreur de type I (fausse alarme) et de montrer que ce test a la propriété CFAR.

Le modèle présenté est inspiré du test « LR-MP$\beta$ » présenté dans [Paris et al., 2011], avec les différences suivantes :

— seule la première étape de l'algorithme *Orthogonal Matching Pursuit* [Pati et al., 1993] est employée, car une seule raie d'émission est recherchée ;
— toutes les combinaisons $(b, f)$ de spectres sont envisageables, au lieu des seuls spectres contigus ;
— nous ne tirons pas les mêmes conclusions quant à l'expression du taux de fausse alarme (cf. section 3.4).

Remarquons enfin que, bien que nous utilisions une estimation au sens du maximum de vraisemblance pour $\hat{\mathbf{x}}_b$, cela n'est pas une nécessité dans le cas général : il est envisageable d'utiliser d'autres estimateurs, ou des données supplémentaires le cas échéant.

#### Test GLR composite avec contrainte de similarité

Le test proposé (3.13) permet la détection basée sur une notion de similarité par rapport à un spectre $\mathbf{y}_b$. Or nous disposons, à l'issue de la détection des spectres brillants de la section 3.2.1, d'un ensemble de spectres indexés par $\mathcal{B}$. Il est possible d'utiliser pour $\mathbf{y}_b$ un élément de $\mathcal{B}$ ou bien le spectre moyen $1/|\mathcal{B}| \sum_{b\in\mathcal{B}} \mathbf{y}_b$. Ces approches n'exploitent cependant pas toutes les informations disponibles, et en réduisent la variété statistique.

C'est pourquoi nous proposons dans cette partie d'utiliser l'ensemble des spectres $(\mathbf{y}_b)_{b\in\mathcal{B}}$, en combinant des applications du test (3.12) avec $f$ fixé et tous les $b \in \mathcal{B}$. Cette combinaison est formulée comme un produit, afin d'offrir à ce test des propriétés similaires à celles du test (3.13) (cf. section 3.4).

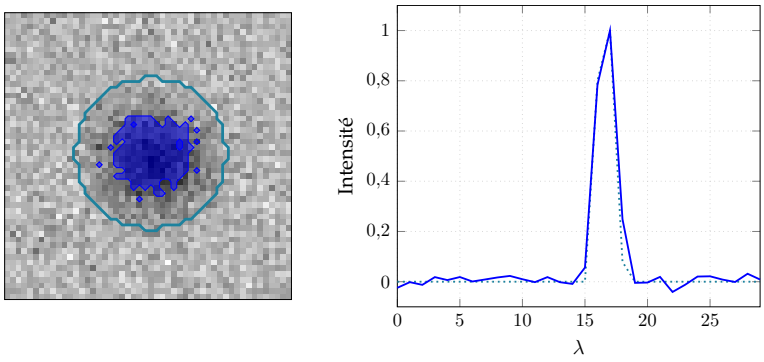



Figure 3.4 – Exemple de détection à l'aide du test (3.17). La légende est identique à celle de la figure 3.3, avec en bleu la détection et en cyan la vérité terrain.

**Définition 3.2.2.** *Soit $(\mathbf{y}_b)_{b\in\mathcal{B}}$ un ensemble de spectres de référence. Nous lui associons un ensemble d'estimations $(\hat{\mathbf{x}}_b)_{b\in\mathcal{B}}$. Le test GLR composite avec contraintes de similarité sera noté $\mathcal{T}_{\text{cs}}^{\mathcal{B}}$ et défini par :*

$$\mathcal{T}_{\text{cs}}^{\mathcal{B}}(\mathbf{y}_f) = \prod_{b\in\mathcal{B}} \mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f). \tag{3.17}$$

En utilisant l'expression (3.13), nous pouvons exprimer ce test sous la forme :

$$\mathcal{T}_{\text{cs}}^{\mathcal{B}}(\mathbf{y}_f) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi_{\text{cs}} \Leftrightarrow \sum_{b\in\mathcal{B}} \frac{(\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1}\mathbf{y}_f)^2}{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1}\hat{\mathbf{x}}_b} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\ln(\xi_{\text{cs}}). \tag{3.18}$$

Cela nous permettra de rechercher une expression du taux de fausse alarme. Nous montrons dans la section 3.4 qu'une borne supérieure sur ce taux peut effectivement être exprimée.

**Remarque 3.2.2.** *Lorsque les composantes individuelles $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)$ sont calculées pour le test $\mathcal{T}_{\text{cs}}^{\mathcal{B}}(\mathbf{y}_f)$, les estimations $\hat{\mathbf{x}}_b$ peuvent varier en forme ou en intensité. Ainsi, toutes les composantes d'une source potentiellement variable en forme spectrale ou en intensité sont prises en compte.*

La figure 3.4 illustre enfin les résultats du test de détection étendue. Nous pouvons constater que, bien que la région détectée soit plus grande que dans la première partie, elle ne couvre pas l'ensemble des spectres ténus. C'est pourquoi nous enrichissons les modèles proposés dans la section suivante.

## 3.3 Extensions des modèles

### 3.3.1 Caractéristiques spatiales

Les tests que nous avons présentés considèrent la dimension spectrale des données mais omettent les dimensions spatiales. Dans cette section, nous présentons une adaptation des tests pour permettre la prise en compte de telles composantes.

À faible RSB, les signaux les plus ténus sont invisibles dans un spectre seul. Si l'on considère en revanche l'ensemble des spectres au sein d'un voisinage local, la présence d'un signal dispersé peut rester significative. Cette dispersion est en pratique décrite par la FSF de l'instrument. Nous supposons que celle-ci diminue en s'éloignant de son centre, de sorte à pouvoir être fenêtrée au sein de quelques pixels (cf. figure 3.5).

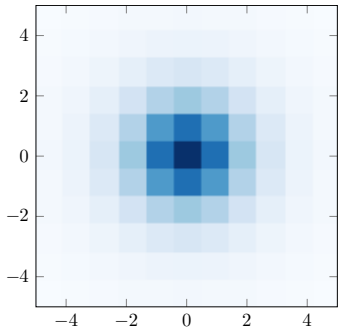



Figure 3.5 – Exemple de FSF couvrant $K = 11^2$ pixels, décrite par une fonction de Moffat (3.44).

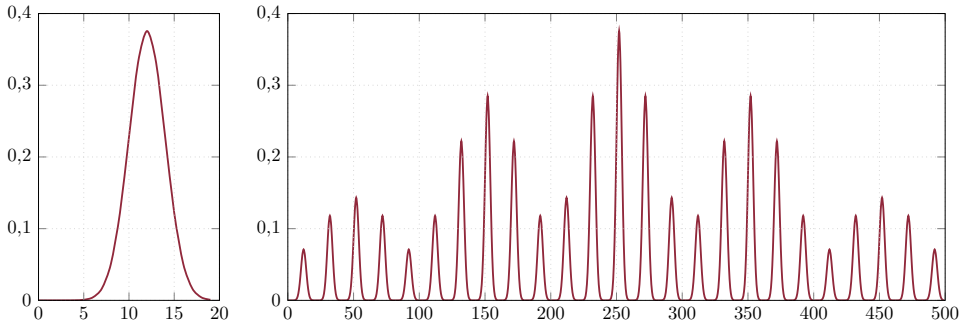



Figure 3.6 – Une raie $\mathbf{c}_j$ de taille $\Lambda = 20$ du catalogue $\mathbf{c}$ et son adaptation $\mathbf{c}_j^{\mathrm{N}}$ en utilisant une FSF fenêtrée à $K = 5^2$ pixels.

Pour pouvoir prendre en compte la FSF, nous introduisons une nouvelle variable d'observation prenant en compte en chaque site $s \in \mathcal{S}$ le spectre $\mathbf{y}_s$ et l'ensemble des spectres au sein d'un voisinage $N_s$ :

$$\forall s \in \mathcal{S},\ \mathbf{y}_s^{\mathrm{N}} = (\mathbf{y}_r)_{r \in \{s\} \cup N_s}. \tag{3.19}$$

Un nouveau catalogue $\mathbf{c}^{\mathrm{N}} \in \mathbb{R}^{K\Lambda \times J}$ de raies spectrales est également introduit. Il contient $J$ colonnes et est construit à partir de $\mathbf{c}$ de sorte à contenir les éléments du catalogue et leurs versions atténuées par la FSF :

$$\forall j \in \{1, \ldots, J\},\ \mathbf{c}_j^{\mathrm{N}} = \mathbf{f}\mathbf{c}_j; \tag{3.20}$$

où $\mathbf{f} \in \mathbb{R}^{K\Lambda \times \Lambda}$ est la matrice de représentant la FSF. En notant $[f_1, \ldots, f_K]^\top$ les $K$ coefficients de la FSF, cette matrice s'écrit comme :

$$\mathbf{f} = [f_1 \mathbf{I}_\Lambda, \ldots, f_K \mathbf{I}_\Lambda]^\top = [f_1, \ldots, f_K]^\top \otimes \mathbf{I}_\Lambda; \tag{3.21}$$

où $\mathbf{I}_\Lambda$ est une matrice identité de taille $\Lambda$, et $\otimes$ est le produit de Kronecker. La figure 3.6 illustre le passage d'une raie du catalogue $\mathbf{c}$ à son homologue dans $\mathbf{c}^{\mathrm{N}}$.

Les hypothèses concurrentes à tester (3.2) deviennent :

$$\left| \begin{array}{l} \mathcal{H}_0 : \mathbf{y}_s^{\mathrm{N}} = \boldsymbol{\epsilon}_s^{\mathrm{N}} \\ \mathcal{H}_1 : \mathbf{y}_s^{\mathrm{N}} = \mathbf{c}_j^{\mathrm{N}} \alpha_s + \boldsymbol{\epsilon}_s^{\mathrm{N}} = \mathbf{f}\mathbf{c}_j \alpha_s + \boldsymbol{\epsilon}_s^{\mathrm{N}} \end{array} \right. \tag{3.22}$$

et les hypothèses (3.11) deviennent :

$$\left| \begin{array}{l} \mathcal{H}_0 : \mathbf{y}_f^{\mathrm{N}} = \boldsymbol{\epsilon}_f^{\mathrm{N}} \\ \mathcal{H}_1 : \mathbf{y}_f^{\mathrm{N}} = \beta_{f,b} \mathbf{c}_j^{\mathrm{N}} \alpha_b + \boldsymbol{\epsilon}_f^{\mathrm{N}} = \beta_{f,b} \mathbf{f}\mathbf{c}_j \alpha_b + \boldsymbol{\epsilon}_f^{\mathrm{N}}. \end{array} \right. \tag{3.23}$$

$\boldsymbol{\epsilon}_s^{\mathrm{N}}$ et $\boldsymbol{\epsilon}_f^{\mathrm{N}}$ sont des réalisations de bruit normal multivarié, centré et de matrice de covariance $\boldsymbol{\Sigma}^{\mathrm{N}}$. Les corrélations dans $\boldsymbol{\Sigma}^{\mathrm{N}}$ sont spectrales comme dans le cas précédent, mais aussi spatiales. Nous supposons dans ce chapitre que la distribution du bruit n'est pas corrélée spatialement. En conséquence, $\boldsymbol{\Sigma}^{\mathrm{N}}$ est formulée comme une matrice bloc-diagonale :

$$\boldsymbol{\Sigma}^{\mathrm{N}} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0}_\Lambda & \cdots & \mathbf{0}_\Lambda \\ \mathbf{0}_\Lambda & \ddots & & \\ & & & \mathbf{0}_\Lambda \\ \mathbf{0}_\Lambda & \cdots & \mathbf{0}_\Lambda & \boldsymbol{\Sigma}_K \end{bmatrix}; \tag{3.24}$$

où $\boldsymbol{\Sigma}_i$ est la matrice de covariance spectrale du $i$-ième spectre dans $N_s$ et $\mathbf{0}_\Lambda$ est la matrice nulle de taille $\Lambda \times \Lambda$. En pratique, si l'image $\mathbf{y}$ est stationnaire, les $\boldsymbol{\Sigma}_i$ sont tous identiques.

**Définition 3.3.1.** *Dans le cadre des hypothèses (3.22), nous pouvons étendre le test (3.3) en un test de détection de sources brillantes utilisant les caractéristiques spatiales :*

$$\mathcal{T}_{\mathrm{n}}(\mathbf{y}_s) = \mathcal{T}(\mathbf{y}_s^{\mathrm{N}}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi_{\mathrm{n}}. \tag{3.25}$$

**Remarque 3.3.1.** *Le test « $\mathrm{GLR}_{\mathrm{1s}}^{(\mathrm{3D})}$ » de [Paris et al., 2013b] est un cas particulier du test (3.25) dans le cadre d'un bruit blanc, c'est-à-dire avec $\boldsymbol{\Sigma}^{\mathrm{N}} = \mathbf{I}_{K\Lambda}$.*

**Définition 3.3.2.** *Dans le cadre des hypothèses (3.22), la reformulation du test (3.12) pour intégrer l'aspect spatial est :*

$$\mathcal{T}_{\mathrm{ns}}^{\hat{\mathbf{x}}_b^{\mathrm{N}}}(\mathbf{y}_f) = \mathcal{T}_{\mathrm{s}}^{\hat{\mathbf{x}}_b^{\mathrm{N}}}(\mathbf{y}_f^{\mathrm{N}}); \tag{3.26}$$

*où $\hat{\mathbf{x}}_b^{\mathrm{N}} = \hat{\alpha}_b \mathbf{c}_{\hat{\jmath}}^{\mathrm{N}}$ est estimé au sens du maximum de vraisemblance par :*

$$\begin{aligned} \hat{\alpha}_b &= \arg\max_{\alpha_b} p(\mathbf{y}_s^{\mathrm{N}} \mid \alpha_s, \mathbf{c}_j^{\mathrm{N}}, \mathcal{H}_1) \\ \hat{\jmath} &= \arg\max_{j} p(\mathbf{y}_s^{\mathrm{N}} \mid \hat{\alpha}_s, \mathbf{c}_j^{\mathrm{N}}, \mathcal{H}_1) \end{aligned} \tag{3.27}$$

**Définition 3.3.3.** *L'extension du test composite $\mathcal{T}_{\mathrm{cs}}^{\mathcal{B}}$ (3.17) aux composantes spatiales est :*

$$\mathcal{T}_{\mathrm{ncs}}^{\mathcal{B}}(\mathbf{y}_f) = \prod_{b \in \mathcal{B}} \mathcal{T}_{\mathrm{ns}}^{\hat{\mathbf{x}}_b^{\mathrm{N}}}(\mathbf{y}_f) \tag{3.28}$$

**Remarque 3.3.2.** *La démarche présentée ici n'est pas équivalente à une déconvolution suivie d'un test de détection : nous considérons la convolution au sein même du test. En effet, une procédure en deux étapes rend la qualification statistique des résultats difficile en raison des corrélations induites dans les données testées.*

### 3.3.2 Observations multiples

Les images astronomiques de champs profonds, permettant d'accéder aux signaux les plus ténus, requièrent d'effectuer plusieurs observations du ciel à des temps différents en raison notamment des effets de saturation des capteurs. Nous avons utilisé jusqu'ici une information résultant de la fusion de ces observations [3]. Ce processus, bien qu'améliorant le RSB, a pour effet de diminuer la quantité d'information disponible. C'est pourquoi cette

3. La formation des images MUSE est détaillée en section 8.2, et la fusion des données est effectuée par une méthode de *sigma-clipping* [Bacon et al., 2017, section 3.2].

partie propose l'adaptation des tests précédents pour prendre en compte conjointement en compte une série d'observations au lieu de la fusion de ces observations.

Nous supposons dans ce cadre que :

— le signal d'intérêt (les sources astrophysiques) est identique entre toutes les observations ;
— les erreurs de recalages sont négligeables par rapport à la taille des pixels ;
— nous nous plaçons dans un cadre simplifié, pour lequel la FSF est identique sur toutes les observations.

Nous notons $\mathcal{O}$ l'ensemble des observations. De la même manière que dans (3.19), nous pouvons alors écrire les spectres en tenant compte de leurs composantes spatiales et d'observation :

$$\forall s \in \mathcal{S},\ \mathbf{y}_s^{\mathrm{NO}} = (\mathbf{y}_{o,s}^{\mathrm{N}})_{o\in\mathcal{O}}. \tag{3.29}$$

Le catalogue $\mathbf{c}$ peut être lui aussi adapté aux observations multiples. Comme chaque observation a une importance identique, $\mathbf{c}$ est dupliqué :

$$\forall j \in \{1,\dots,J\},\ \mathbf{c}_j^{\mathrm{NO}} = (\mathbf{c}_j^{\mathrm{N}})_{o\in\mathcal{O}} \tag{3.30}$$

Les hypothèses à tester (3.22) deviennent :

$$\left|\begin{array}{l} \mathcal{H}_0 : \mathbf{y}_s^{\mathrm{NO}} = \boldsymbol{\epsilon}_s^{\mathrm{NO}} \\ \mathcal{H}_1 : \mathbf{y}_s^{\mathrm{NO}} = \mathbf{c}_j^{\mathrm{NO}}\alpha_s + \boldsymbol{\epsilon}_s^{\mathrm{NO}} \end{array}\right. \tag{3.31}$$

et les hypothèses (3.23) deviennent :

$$\left|\begin{array}{l} \mathcal{H}_0 : \mathbf{y}_f^{\mathrm{NO}} = \boldsymbol{\epsilon}_f^{\mathrm{NO}} \\ \mathcal{H}_1 : \mathbf{y}_f^{\mathrm{NO}} = \beta_{f,b}\alpha_b\mathbf{c}_j^{\mathrm{NO}} + \boldsymbol{\epsilon}_f^{\mathrm{NO}}. \end{array}\right. \tag{3.32}$$

Ces deux jeux d'hypothèses font intervenir des réalisations de bruit normal multivarié sous la forme de $\boldsymbol{\epsilon}_s^{\mathrm{NO}}$ et $\boldsymbol{\epsilon}_f^{\mathrm{NO}}$. Dans le cas général, la définition de leur distribution fait intervenir des termes de corrélations spectrales, spatiales et par observation dans la matrice de covariance $\boldsymbol{\Sigma}^{\mathrm{NO}}$. Dans le cadre de l'application astronomique, le signal est supposé cohérent entre observations et le bruit est indépendant d'une observation à l'autre. En conséquence, $\boldsymbol{\Sigma}^{\mathrm{NO}}$ sera bloc-diagonale :

$$\boldsymbol{\Sigma}^{\mathrm{NO}} = \begin{bmatrix} \boldsymbol{\Sigma}_{(o=1)}^{\mathrm{N}} & \mathbf{0}_{K\Lambda} & \cdots & \mathbf{0}_{K\Lambda} \\ \mathbf{0}_{K\Lambda} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0}_{K\Lambda} \\ \mathbf{0}_{K\Lambda} & \cdots & \mathbf{0}_{K\Lambda} & \boldsymbol{\Sigma}_{(o=|\mathcal{O}|)}^{\mathrm{N}} \end{bmatrix}; \tag{3.33}$$

où $\boldsymbol{\Sigma}_{(o=i)}^{\mathrm{N}}$ est la matrice de covariance décrite en (3.24) pour la $i-$ième observation, et $\mathbf{0}_{K\Lambda}$ est la matrice nulle de taille $K\Lambda \times K\Lambda$. En pratique, $\boldsymbol{\Sigma}_{(o=i)}^{\mathrm{N}}$ est estimée avec la $i-$ème observation uniquement.

**Définition 3.3.4.** *Nous nous plaçons dans le cadre des hypothèses* (3.31)*. Le test de détection de sources brillantes étendu aux caractéristiques spatiales et d'observations est :*

$$\mathcal{T}_{\mathrm{no}}(\mathbf{y}_s) = \mathcal{T}(\mathbf{y}_s^{\mathrm{NO}}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi_{\mathrm{no}}. \tag{3.34}$$

**Définition 3.3.5.** *Nous nous plaçons dans le cadre des hypothèses (3.32). La reformulation du test (3.26) pour intégrer les composantes d'observation sont :*

$$\mathcal{T}_{\text{nos}}^{\hat{\mathbf{x}}_b^{\text{NO}}}(\mathbf{y}_f) = \mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b^{\text{NO}}}(\mathbf{y}_f^{\text{NO}}); \tag{3.35}$$

*où $\hat{\mathbf{x}}_b^{\text{N}} = \hat{\alpha}_b \mathbf{c}_{\hat{j}}^{\text{N}}$ est estimé au sens du maximum de vraisemblance de la même manière que dans l'équation (3.27).*

**Définition 3.3.6.** *La reformulation du test composite (3.28) avec les composantes d'observations est :*

$$\mathcal{T}_{\text{nocs}}^{\mathcal{B}}(\mathbf{y}_f) = \prod_{b \in \mathcal{B}} \mathcal{T}_{\text{nos}}^{\hat{\mathbf{x}}_b^{\text{NO}}}(\mathbf{y}_f) \tag{3.36}$$

Remarquons enfin qu'il n'existe pas, à notre connaissance, de test de détection tenant compte d'observations multiples dans des images hyperspectrales. Cet aspect est en effet propre à l'imagerie astronomique, dont les développements vers l'imagerie hyperspectrale sont relativement récents.

## 3.4 Qualification des probabilités de fausse alarme

Il est généralement souhaitable de pouvoir exprimer le taux d'erreur de type I ou $P_{\text{FA}}$ en fonction des seuils des tests statistiques. Cette partie présente les résultats obtenus sur ce point pour les tests $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}$ et $\mathcal{T}_{\text{cs}}^{\mathcal{B}}$.

**Proposition 3.4.1.** *Soit $\hat{\mathbf{x}}_b$ un spectre de référence. Le test GLR avec contrainte de similarité, décrit en (3.12) et noté $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}$, a la propriété CFAR : le taux de fausse alarme ne dépend que du seuil du test.*

*Démonstration.* Reformulons tout d'abord (3.13) :

$$\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi_{\text{s}} \Leftrightarrow \left( \frac{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f}{\|\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-\frac{1}{2}}\|_2} \right)^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\ln(\xi_{\text{s}}). \tag{3.37}$$

Nous pouvons donc déduire, sous $\mathcal{H}_0$ :

$$\begin{aligned} \mathbf{y}_f \overset{\mathcal{H}_0}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) &\Rightarrow \hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f \overset{\mathcal{H}_0}{\sim} \mathcal{N}(0, \|\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-\frac{1}{2}}\|_2^2) \\ &\Rightarrow \frac{\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}_f}{\|\hat{\mathbf{x}}_b^\top \boldsymbol{\Sigma}^{-\frac{1}{2}}\|_2} \overset{\mathcal{H}_0}{\sim} \mathcal{N}(0,1) \\ &\Rightarrow 2\ln[\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)] \overset{\mathcal{H}_0}{\sim} \chi_1^2; \end{aligned} \tag{3.38}$$

où $\chi_1^2$ est une distribution du $\chi^2$ à un degré de liberté.

En conséquence, la $P_{\text{FA}}$ est donnée par :

$$\begin{aligned} P_{\text{FA}}\left[\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)\right] &= p(\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f) > \xi_{\text{s}} \,|\, \mathcal{H}_0) \\ &= 1 - p(2\ln\left[\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)\right] < 2\ln(\xi_{\text{s}}) \,|\, \mathcal{H}_0) \\ &= 1 - \Phi_{\chi_1^2}\left[2\ln(\xi_{\text{s}})\right]; \end{aligned} \tag{3.39}$$

où $\Phi_{\chi_1^2}$ est la fonction de répartition de la distribution $\chi_1^2$.

□

Nous souhaitons dans un second temps généraliser ce résultat au test composite $\mathcal{T}_{\text{cs}}^{\mathcal{B}}$. Rappelons qu'il est construit comme le produit de plusieurs tests $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}$ dont nous connaissons la distribution grâce à (3.38).

**Proposition 3.4.2.** *Soit $(\hat{\mathbf{x}}_b)_{b\in\mathcal{B}}$ un ensemble de spectres de référence. Le test GLR composite avec contraintes de similarité $\mathcal{T}_{\text{cs}}^{\mathcal{B}}$ a un taux de fausse alarme borné . La borne supérieure de ce test ne dépend que du seuil $\xi_{\text{cs}}$.*

*Démonstration.* Supposons dans un premier temps que les vecteurs $\mathbf{y}_b$ sont indépendants. Nous pouvons, grâce au résultat (3.39), établir le résultat suivant :

$$P_{\text{FA}}\left[\mathcal{T}_{\text{cs}}^{\mathcal{B}}(\mathbf{y}_f)\right] = 1 - \Phi_{\chi^2_{|\mathcal{B}|}}\left[2\ln(\xi_{\text{cs}})\right]. \tag{3.40}$$

Le nombre de degrés de liberté de cette distribution dépend alors du nombre de spectres faisant partie de la détection initiale $\mathcal{B}$.

Cependant, supposer l'indépendance entre les différents vecteurs $\mathbf{y}_b$ n'est pas une hypothèse satisfaisante :

— le signal sous-jacent $\mathbf{x}_b$ est vraisemblablement corrélé d'un site à l'autre en raison de l'étalement spatial des sources lumineuses ;

— lorsque les composantes spatiales sont considérées, les vecteurs $\mathbf{y}_b^{\text{N}}$ ont des composantes identiques (cf. (3.19)).

En relaxant cette hypothèse d'indépendance, la distribution de $\chi^2$ a au plus $|\mathcal{B}|$ degrés de libertés. En effet, la non-indépendance des $\mathbf{y}_b$ implique la non-indépendance des tests $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)$. En conséquence, la loi du logarithme du produit des $|\mathcal{B}|$ tests $\mathcal{T}_{\text{s}}^{\hat{\mathbf{x}}_b}(\mathbf{y}_f)$ peut être décrite avec au plus $|\mathcal{B}|$ composantes, ce qui borne le nombre de degrés de libertés de la loi de $\chi^2$.

De plus, à argument fixé, la fonction de répartition de la distribution de $\chi^2$ décroît quand le degré de liberté croît :

$$\forall u \in \mathbb{R}^+, \forall a, b \in \mathbb{N}^+ : a \leq b \Rightarrow \Phi_{\chi^2_a}(u) \geq \Phi_{\chi^2_b}(u). \tag{3.41}$$

Nous pouvons donc déduire de (3.40) la limite suivante :

$$P_{\text{FA}}\left[\mathcal{T}_{\text{cs}}^{\mathcal{B}}(\mathbf{y}_f)\right] \leq 1 - \Phi_{\chi^2_{|\mathcal{B}|}}\left[2\ln(\xi_{\text{cs}})\right]. \tag{3.42}$$

□

Ainsi, pour un seuil $\xi_{\text{cs}}$ donné, un taux d'erreur maximal peut être estimé. Les résultats numériques de la section 3.5.3 apportent une vérification expérimentale des résultats de cette partie.

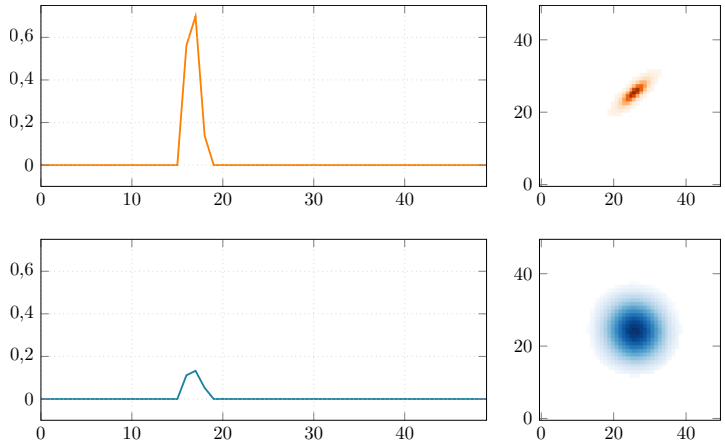
Remarquons enfin que ces propriétés sont valables lorsque les références ($\hat{\mathbf{x}}_b$ ou $(\hat{\mathbf{x}}_b)_{b\in\mathcal{B}}$) ne dépendent pas des observations. L'application pratique présentée dans ce chapitre requiert l'estimation de ces références à partir des observations, mais d'autres applications pourraient bénéficier par exemple de données supplémentaires.

## 3.5 Résultats numériques

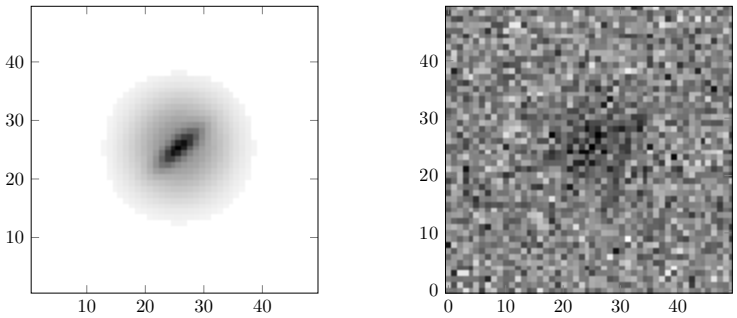
### 3.5.1 Simulation d'images hyperspectrales

Dans les travaux présentés dans ce manuscrit, les images astronomiques n'ont pas de « vérité terrain » ni de référence établie par des experts qui permette de qualifier les résultats sur des données réelles. En conséquence, le recours à des images de synthèse est nécessaire pour permettre la validation de la méthode. L'image simulée est composée de deux éléments :

— une source d'émission *brillante* elliptique occupant peu de spectres de l'image ;




(a) Composantes brillantes (orange) et ténues (bleu) intervenant dans la formation des images de synthèse. Les spectres sont représentés à gauche, et les cartes d'intensité à droite (en inverse vidéo).




(b) Moyenne spectrale de l'image hyperspectrale sans bruit (à gauche) et bruitée avec RSB $= 0$ dB à droite, en inverse vidéo.

Figure 3.7 – Formation des images de synthèse.

— une source circulaire plus étendue, ténue.

Toutes deux ont des intensités s'affaiblissant à partir de leur centre, suivant un profil gaussien tronqué : les luminosités sont nulles en-dehors des objets. L'émission spectrale de ces objets est similaire, à un facteur d'atténuation près. L'image hyperspectrale $(\mathbf{y}_s)_{s\in\mathcal{S}}$ comporte $50 \times 50$ spectres et $50$ bandes spectrales, et est centrée autour de l'objet d'intérêt. La figure 3.7 illustre la formation des images de synthèse.

Le bruit est ajouté dans un second temps, sous la forme de réalisations de lois normales multivariées spectralement. Nous définissons de plus le RSB comme :

$$\mathrm{RSB} = 10\log_{10}\left(\frac{\|\mathbf{x}_{\mathrm{halo}}\|_2^2}{\mathrm{Tr}(\mathbf{\Sigma})}\right); \tag{3.43}$$

où $\mathbf{x}_{\mathrm{halo}}$ est le spectre le plus brillant de la partie ténue avant l'ajout du bruit, et $\mathbf{\Sigma}$ est la matrice de covariance du bruit. Cette approche pour la définition du RSB est analogue à un « pic-RSB » ou PSNR, dans le sens où ce RSB est le maximum des RSB possibles au sein de l'objet.

Nous utilisons comme catalogue $\mathbf{c}$ une collection de raies gaussiennes normalisées [4] (cf. figure 3.2). Les moyennes spectrales sont échantillonnées comme les bandes spectrales de l'image ($1$ à $50$), et les écarts-types varient entre $0{,}1$ et $5$ bandes spectrales.

4. En pratique, le coefficient de normalisation est calculé sur la raie la plus centrale, de sorte à éviter des différences dans les valeurs maximales des raies.

**Remarque 3.5.1.** *Dans le contexte applicatif, nous recherchons des émissions lumineuses couvrant une fine bande spectrale. Ces raies d'émissions peuvent être asymétriques. Les expérimentations ont montré des résultats très proches lors de l'utilisation d'éléments symétriques et asymétriques dans le catalogue* **c***.*

La FSF est représentée par une distribution de Moffat 2D [Moffat, 1969], qui a des extrémités plus élevées qu'une gaussienne et qui est plus adaptée à la description des fonctions d'étalement d'instruments astrophysiques. Une loi de Moffat s'écrit :

$$m^{a,b}(p,q) = \frac{b-1}{\pi a^2}\left[1 + \frac{p^2+q^2}{a^2}\right]^{-b}; \tag{3.44}$$

où $p$ et $q$ représentent les deux coordonnées spatiales, et $a$ et $b$ sont deux paramètres de la loi. Pour les images simulées, nous considérons ces derniers constants : $a = 2{,}0$ et $b = 2{,}6$. Notons que pour les données MUSE, le paramètre $a$ varie légèrement en fonction de la longueur d'onde [Serre et al., 2010; Bacon et al., 2015].

### 3.5.2 Performances par étape

Les résultats présentés dans cette section sont évalués à l'aide de 100 simulations comportant des réalisations différentes du bruit. Ils consistent en :

— la probabilité de détection $P_{\mathrm{D}}$ ;
— la probabilité de fausse alarme $P_{\mathrm{FA}}$.

Les expériences ont de plus lieu :

— sous RSB variable, entre $-20$ et $5$ dB.
— en utilisant différents fenêtrages pour la FSF, de $1$ à $11^2$ pixels.

De plus, l'image hyperspectrale est blanchie avant traitement : cela assure que les corrélations spectrales sont négligeables, et permet de contraindre l'estimation de la matrice de covariance à être diagonale. De plus, la contamination de la matrice de covariance par les signaux est négligeable dans le contexte d'éventuels signaux ténus. Nous supposons que la FSF est connue, par exemple à l'aide d'un modèle physique de l'instrument ou d'une calibration.

**Remarque 3.5.2.** *La FSF est généralement estimée par ses paramètres : les erreurs d'estimation concernent ces paramètres et non les coefficients. Cela implique que les propriétés que l'on peut généralement attendre de la FSF, comme une évolution « douce » des coefficients et une atténuation en s'éloignant du centre, sont toujours valables. L'influence de ces erreurs sur les paramètres de la FSF sur la détection est donc supposée négligeable.*

#### Détection de sources brillantes

La figure 3.8 illustre les résultats obtenus par le test $\mathcal{T}_{\mathrm{ns}}$ (3.25). Nous y constatons en particulier que l'extension du voisinage spatial considéré permet une nette amélioration des résultats à faible SNR : la différence entre les valeurs estimées de $P_{\mathrm{D}}$ va de $23\%$ avec $-20$ dB à $75\%$ avec $-10$ dB. Pour comparaison, l'algorithme RX [Reed et Xiaoli, 1990] a été employé dans les mêmes conditions. Il est moins efficace dans ce contexte, vraisemblablement en raison de la faible intensité des « anomalies » simulées dans l'image hyperspectrale.

La seconde observation importante concerne l'utilisation d'observations multiples. La figure 3.8 montre que cette approche diminue les performances du test. Ce phénomène est explicable par la recherche de maximum dans l'équation (3.25). Cette recherche est en effet sensible au RSB : dans les observations individuelles, celui-ci est nettement plus faible

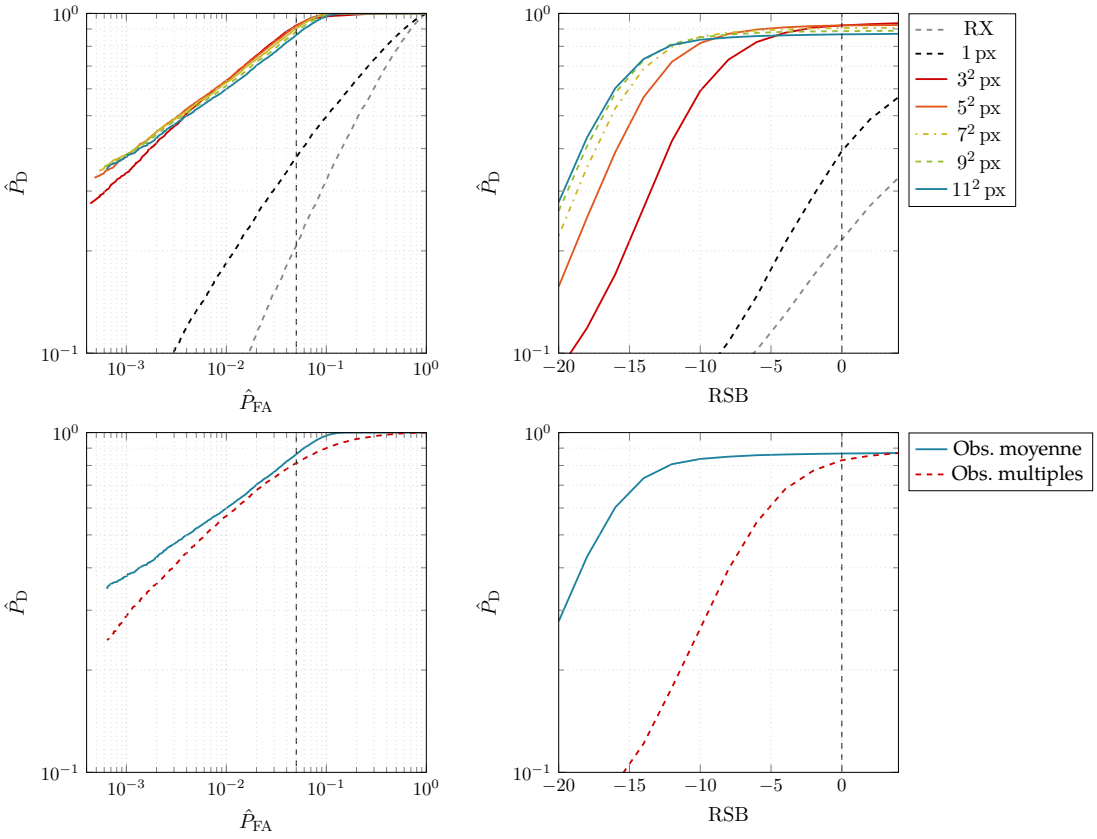



Figure 3.8 – Résultats numériques concernant la détection de sources brillantes. Les graphiques de gauche montrent les courbes ROC avec $\mathrm{SNR} = 0$ dB, et ceux de droite montrent l'évolution de $\hat{P}_{\mathrm{D}}$ en fonction du SNR lorsque $\hat{P}_{\mathrm{FA}} = 0{,}05$. Les lignes verticales en tirets sont communes aux deux graphes : $\mathrm{SNR} = 0$ dB et $\hat{P}_{\mathrm{FA}} = 0{,}05$. La première ligne illustre l'utilisation des composantes spatiales, et la seconde l'utilisation d'observations multiples en utilisant une fenêtre de FSF de $11^2$ px.

en chaque spectre. En conclusion, la recherche du maximum semble handicapante lors de l'utilisation d'observations multiples à très faible RSB.

### Détection de sources ténues

Nous étudions dans un second temps les performances du test composite $\mathcal{T}_{\mathrm{ncs}}$ (3.28) seul : nous utilisons comme référence $\mathcal{B}$ la partie correspondant au signal brillant de forme elliptique décrit en section 3.5.1. Les résultats expérimentaux sont visibles en figure 3.9. Ils illustrent le gain apporté par l'ajout de composantes spatiales additionnelles. Nous pouvons en particulier remarquer que :

— utiliser une FSF bornée à $3^2$ pixels apporte des résultats significativement meilleurs que lorsque la FSF est ignorée (« 1 pixel ») ;
— en raison de l'atténuation de la FSF, utiliser des fenêtres de plus en plus grandes pour la FSF donne des gains consécutifs de résultats de plus en plus petits.

En pratique, le test composite $\mathcal{T}_{\mathrm{ncs}}$ (3.28) requiert l'estimation successive des coefficients $\alpha_b$ (basé sur l'opérateur $\max$) et du coefficient $\beta_{f,b}$ (3.12). Nous avons vu pour le test de détection brillante que la première étape est peu efficace dans le cas d'observations multiples. C'est pourquoi nous proposons dans ce cas une approche « mixte » :

— estimer $\alpha_b$ à l'aide de la moyenne des observations ;

— estimer $\beta_{f,b}$ sur l'ensemble des observations.

Les résultats de cette approche sont illustrés dans la seconde ligne de la figure 3.9. Ils montrent que l'adoption de cette approche mixte permet l'obtention de résultats équivalents pour l'utilisation conjointe des observations ou de leur moyenne. Nous pouvons conclure de ces résultats que le gain attendu, en terme d'information contenue dans l'ensemble des informations, est compensé dans le meilleur des cas par le très faible RSB des observations individuelles.

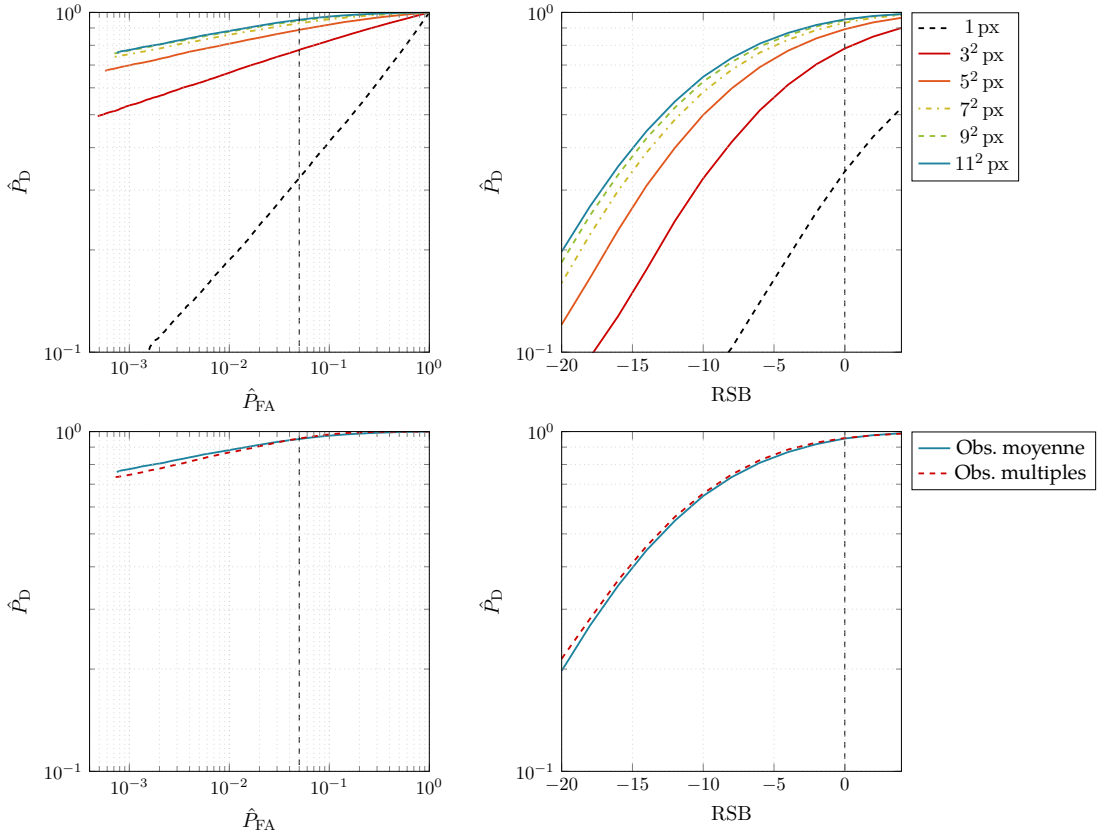



Figure 3.9 – Résultats numériques pour la détection de sources étendues. La légende est identique à celle de la figure 3.8.

### 3.5.3 Distributions des statistiques de tests

Cette section présente les résultats numériques sur les statistiques des tests $\mathcal{T}_\mathrm{s}^{\hat{\mathbf{x}}_b}$ (3.12) et $\mathcal{T}_\mathrm{cs}^{\mathcal{B}}$ (3.17) et leurs extensions respectives $\mathcal{T}_\mathrm{sn}^{\hat{\mathbf{x}}_b^\mathrm{N}}$ (3.26) et $\mathcal{T}_\mathrm{ncs}^{\mathcal{B}}(\mathbf{y}_f)$ (3.28). L'objectif est de caractériser leur distribution sous $\mathcal{H}_0$ et de confronter ces résultats empiriques aux propriétés exprimées dans la section 3.4. Les tests sont effectués sur des données de synthèse contenant uniquement du bruit, et étant donné un ensemble de spectres « brillants » correspondant à l'ensemble $\mathcal{B}$. Nous nous plaçons à $\mathrm{RSB} = -10$ dB et effectuons ces mesures pour différentes tailles de la FSF étudiées dans la section précédente. Cela nous permettra en particulier d'évaluer les propriétés de tests utilisant les caractéristiques spatiales. Nous mesurons en particulier l'adéquation à des lois du $\chi^2$, en estimant sur les données le paramètre de degré de liberté.

La figure 3.10a présente les résultats numériques pour le test $\mathcal{T}_\mathrm{s}^{\hat{\mathbf{x}}_b}$ (3.12) et son extension $\mathcal{T}_\mathrm{ns}^{\hat{\mathbf{x}}_b^\mathrm{N}}$ (3.26), en utilisant pour ce dernier différentes tailles pour la FSF. Nous pouvons constater la bonne adéquation générale de la statistique de $\mathcal{T}_\mathrm{s}^{\hat{\mathbf{x}}_b}$ sous $\mathcal{H}_0$ à la distribution du $\chi^2$ qui

lui correspond théoriquement. Remarquons également que la queue de distribution semble légèrement sous-estimée, avec un écart de probabilité maximum de $3.10^{-4}$. Constatons enfin que cette distribution semble très peu dépendante de la taille de la FSF, au point de rendre indiscernables les distributions affichées en figure 3.10a.

La figure 3.10b présente les résultats expérimentaux pour le test $\mathcal{T}_{\text{cs}}^{\mathcal{B}}$ et son extension aux composantes spatiales $\mathcal{T}_{\text{ncs}}^{\mathcal{B}}$ avec différents fenêtrages de la FSF. Nous pouvons là aussi constater la bonne adéquation des distributions empiriques sous $\mathcal{H}_0$ avec une distribution du $\chi^2$. Enfin, les variations du fenêtrage de la FSF induisent des variations dans les distributions. En effet, le paramètre de degré de liberté diminue lorsque la taille de la FSF augmente. Cela s'explique par le fait que pour décrire l'ensemble des spectres indexés par $\mathcal{B}$, utiliser une fenêtre plus grande diminue le nombre de spectres nécessaires. Contrairement aux résultats précédents, il n'existe pas d'écart significatif entre les distributions empiriques et les ajustements à des distributions du $\chi^2$ au niveau des queues de distributions.

Les résultats présentés dans cette section ont confirmé les propriétés exprimées dans la section 3.4. En pratique, ils nous permettent d'établir un lien entre une probabilité de fausse alarme cible maximale et les seuils des tests statistiques.

## 3.6 Stratégie de détection et performances

Nous avons pu évaluer individuellement les performances des tests proposés dans la partie précédente. Dans cette partie, nous présentons une stratégie globale de détection (section 3.6.1), puis l'évaluons sur des données simulées (section 3.6.2) et réelles (section 3.6.3).

### 3.6.1 Stratégie de détection

Nous présentons dans un premier temps la stratégie de détection employée. La démarche comporte deux étapes : la détection de sources brillantes avec le test $\mathcal{T}_{\text{n}}$ (3.25), suivie par la détection de sources ténue avec le test $\mathcal{T}_{\text{ncs}}$ (3.28). Nous faisons en sorte d'utiliser à chaque étape des données blanchies, de sorte à contraindre les estimations des matrices de covariance à être diagonales. Pour la première étape, la matrice de covariance est calculée sur tout le cube. Dans la seconde étape, elle est évaluée sur les spectres qui ne font pas partie de la détection initiale (ensemble $\mathcal{F}$). En conséquence, la matrice de covariance n'est pas contaminée par les spectres les plus brillants de l'ensemble $\mathcal{B}$. De plus, bien que la détection de tous les pixels lors de la première étape ne soit pas nécessaire, il est en revanche souhaitable d'avoir le moins de fausses alarmes possibles afin d'éviter une éventuelle propagation des erreurs. Pour renforcer la robustesse de la première étape en termes de $P_{\text{FA}}$, une étape supplémentaire est ajoutée. Elle consiste à retirer de $\mathcal{B}$ tous les spectres qui ne sont pas spatialement connexes au centre de l'image. Cela ne gêne pas la détection, car les spectres pertinents vis-à-vis du premier test le seront par rapport au second test. L'ensemble de la stratégie de détection est résumé dans l'algorithme 1.

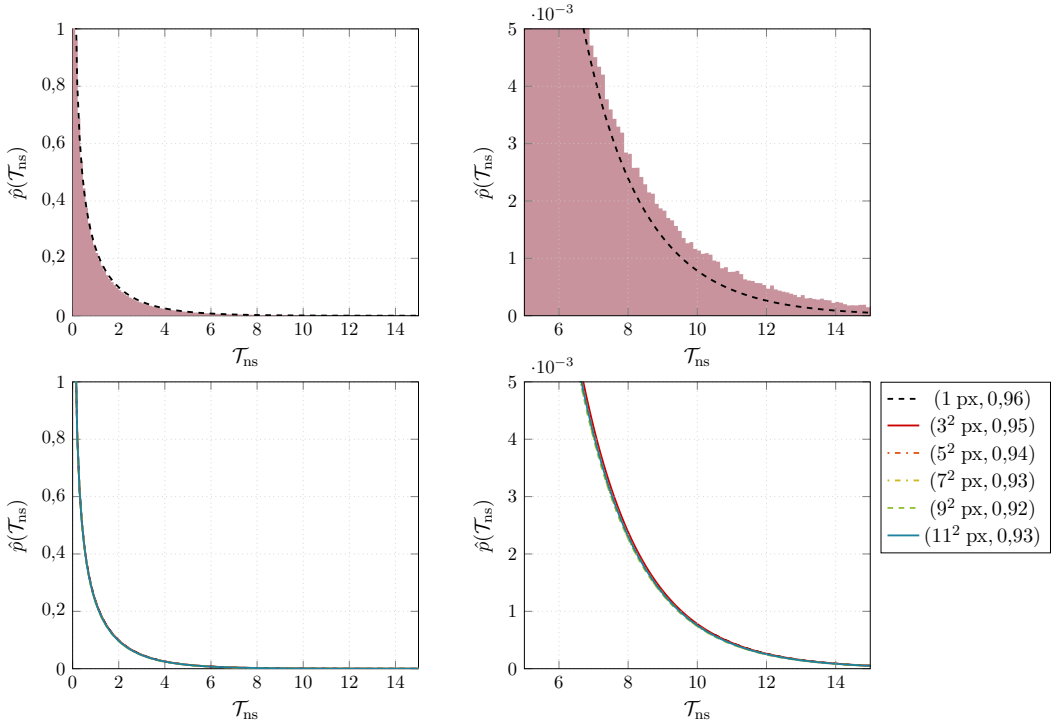
### 3.6.2 Performances

Les résultats de cette stratégie sont présentés en figure 3.11. Le comportement de cette méthode est proche de ceux décrits dans la section 3.5.2 :

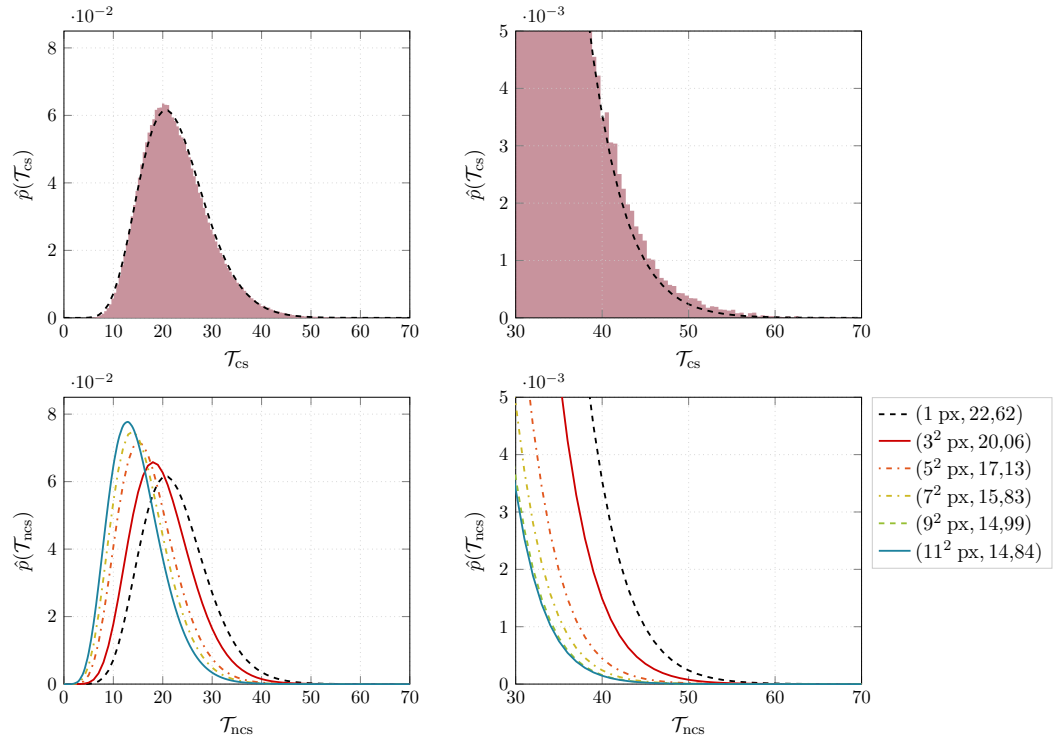
— prendre en compte les composantes spatiales permet un gain allant jusqu'à $71\%$ dans les probabilités de détection $P_{\text{D}}$ ;

— l'utilisation d'observations multiples ne met pas en évidence de gain significatif.

Remarquons de plus les résultats particulièrement intéressants à faible RSB : $\hat{P}_{\text{D}} = 70\%$ à $-10$ dB et $\hat{P}_{\text{D}} = 49\%$ à $-15$ dB lorsque $\hat{P}_{\text{FA}} = 5\%$.




(a) Distributions empiriques des statistique des tests $\mathcal{T}_s^{\hat{\mathbf{x}}_b}$ et $\mathcal{T}_{ns}^{\hat{\mathbf{x}}_b}$ sous $\mathcal{H}_0$. Ces résultats ont été obtenus avec $10^6$ réalisations de bruit blanc. Première ligne : histogramme en rouge et estimation au sens du maximum de vraisemblance de la distribution du $\chi^2$ (tirets noirs) obtenus pour $\mathcal{T}_s^{\hat{\mathbf{x}}_b}$. Seconde ligne : estimations de la distribution du $\chi^2$ pour $\mathcal{T}_{ns}^{\hat{\mathbf{x}}_b}$ avec différentes tailles de FSF. La légende associe à chaque cas le degré de liberté estimé pour la distribution du $\chi^2$. La colonne de droite concerne les queues de distributions.




(b) Distributions empiriques des statistiques de tests $\mathcal{T}_{cs}^{\mathcal{B}}$ et $\mathcal{T}_{ncs}^{\mathcal{B}}$ sous $\mathcal{H}_0$. Ces résultats ont été obtenus avec $10^5$ réalisations de spectres de bruit blanc. La légende est identique à celle de la figure 3.10a.

Figure 3.10 – Distributions empiriques des tests proposés.

**Remarque 3.6.1.** *Les résultats de la figure 3.11 semblent améliorés par rapport à ceux de la figure 3.9 évalués pour la détection ténue, dans lesquels la galaxie est pourtant connue. Cet effet est dû à l'utilisation des parties ténues seulement comme référence dans les premières expériences, et de l'adjonction de la partie brillante à la référence dans les deuxièmes.*

Dans un second temps, nous proposons de comparer la méthode proposée avec d'autres méthodes correspondant à la même problématique. La première méthode utilisée pour la comparaison est basée sur une modélisation par champs gaussiens aléatoires [Ahmad et al., 2015]. Dans ce cadre, l'arrière-plan est modélisé par un champ aléatoire gaussien spatio-spectral (SS-GRF), qui permet d'établir un lien théorique entre le nombre de composantes connexes résultant d'un seuillage de ce champ et une probabilité théorique de fausse alarme.

Pour l'application aux images hyperspectrales de télédétection, les détecteurs les plus courants sont les détecteurs AMF [Reed et al., 1974] et ACE [Kraut et al., 2001]. Tous deux requièrent la connaissance d'un spectre de référence. Nous construisons un détecteur basé sur le test ACE à des fins de comparaison. Pour ce faire, nous utilisons la stratégie de détection proposée et remplaçons la dernière étape par un détecteur ACE. Le spectre de référence est le spectre moyen évalué sur la région $\mathcal{B}_1$.

La figure 3.12 illustre la comparaison entre la méthode que nous proposons, sa variante basée sur le test ACE, et le modèle SS-GRF. Ce dernier semble peu efficace par rapport à notre modèle. Ce constat se justifie par le fait que notre méthode est plus spécifique : elle cible des spectres d'une forme particulière, prend en compte explicitement le phénomène d'étalement dû à la FSF. En ce sens, notre méthode de détection est davantage supervisée que la méthode SS-GRF.

La comparaison avec une variante basée sur le test ACE montre quant à elle des résultats équivalents à ceux de notre méthode lorsque $\mathrm{RSB} = 0$ dB. Ce n'est pas le cas à tous les RSB : lorsque $\mathrm{RSB} < -12$ dB la méthode proposée donne de meilleurs résultats que l'alternative basée sur le test ACE. Ce phénomène est dû à une diminution de la qualité de la référence utilisée dans ACE lorsque le RSB diminue.

Les résultats des expériences sur les images de synthèse peuvent être résumés comme suit :

1. l'utilisation des composantes spatiales améliore significativement les résultats de détection, amenant jusqu'à un gain de $71\%$ dans les taux de détection ;
2. l'utilisation des observations multiples n'est pas, dans ce contexte, avantageuse. Ceci est probablement dû au très faible RSB par observation.

**Algorithme 1** Stratégie de détection.

**Entrée :** Image hyperspectrale $(\mathbf{y}_s)_{s\in\mathcal{S}}$, $P_{\mathrm{FA}}$ maximum cible, catalogue $\mathbf{c}$, modèle de FSF $\mathbf{f}$.
**Sortie :** Ensemble des détections $\mathcal{D}$.

1: Blanchiment de $(\mathbf{y}_s)_{s\in\mathcal{S}}$ en estimant la covariance sur tous les spectres.
2: Détection de sources brillantes : le test $\mathcal{T}_{\mathrm{n}}(\mathbf{y}_s)$ (3.25) est appliqué à tous les spectres de $(\mathbf{y}_s)_{s\in\mathcal{S}}$, et isole l'ensemble des spectres $\mathcal{B}$ pour lequel $\mathcal{H}_1$ est décidée.
3: Étape intermédiaire : suppression dans $\mathcal{B}$ de tous les spectres qui ne sont pas connexes au pixel central, produisant l'ensemble $\mathcal{B}_1$.
4: Blanchiment de $(\mathbf{y}_f)_{f\in\mathcal{S}\setminus\mathcal{B}_1}$ en estimant la covariance sur l'ensemble des spectres de $\mathcal{S}\setminus\mathcal{B}_1$.
5: Détection de sources ténues : le test $\mathcal{T}_{\mathrm{ncs}}^{\mathcal{B}_1}(\mathbf{y}_f)$ (3.28) est appliqué à tous les spectres de $(\mathbf{y}_f)_{f\in\mathcal{S}\setminus\mathcal{B}_1}$, et isole l'ensemble $\mathcal{D}$ des spectres pour lesquels $\mathcal{H}_1$ est décidée.

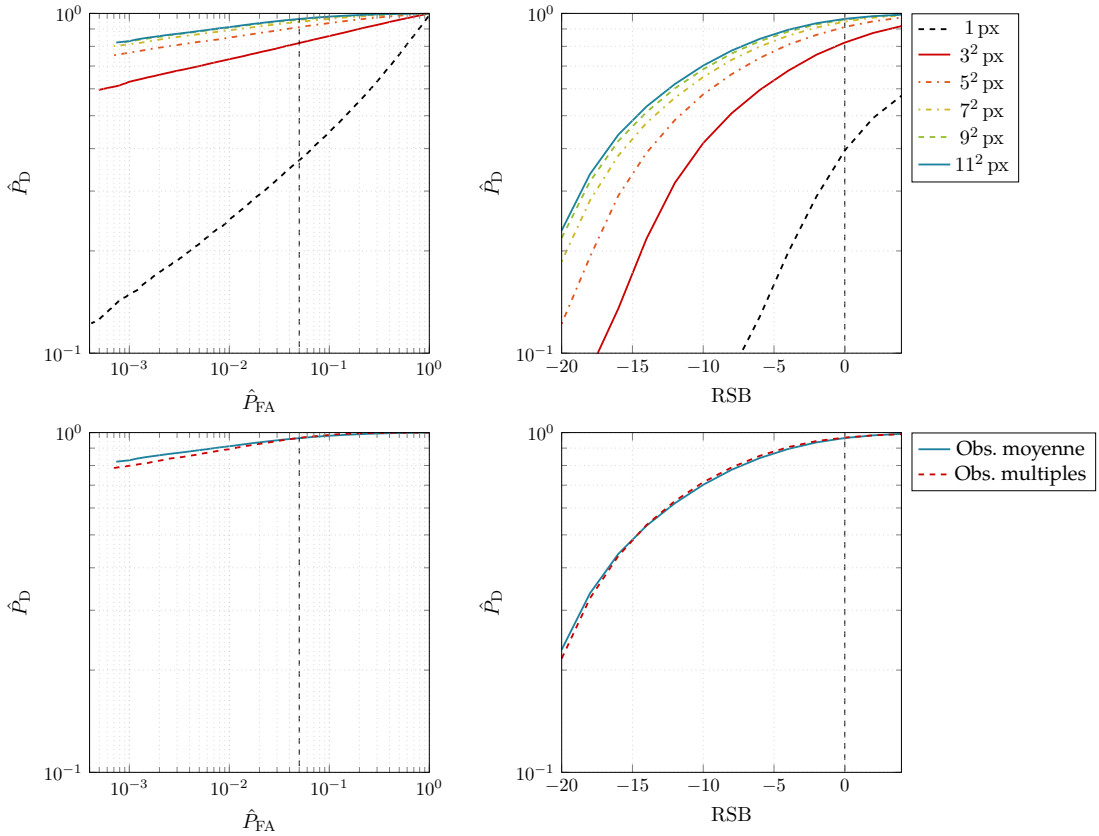



Figure 3.11 – Résultats numériques pour la stratégie de détection de l'algorithme 1. La légende est identique à celles des figures 3.8 et 3.9.

### 3.6.3 Résultats sur les données réelles MUSE

Nous présentons dans cette section des résultats de détection obtenus dans une observation MUSE, dont une description peut être trouvée dans le chapitre 9. Ces résultats seront détaillés et comparés aux résultats obtenus par d'autres méthodes dans le chapitre 10.

Les hypothèses que nous retenons pour l'instrument sont :

— le bruit est de distribution normale, multivariée spectralement ;
— la PSF est séparable : la FSF est modélisée par une distribution de Moffat 3.44, et la LSF couvrant $2{,}1 \pm 0{,}2$ bandes spectrales est négligée.

Nous utilisons ici les données d'observation du champ *Hubble deep field south* (HDFS). Ces données ont été acquises durant la période de mise en service de l'instrument (« *commissionning* ») en 2014. La description de ces données et de leur formation est disponible dans [Bacon et al., 2015].

Les 90 sources lumineuses susceptibles de comporter un halo ont été localisées manuellement par des experts. Ces données sont particulièrement bruitées, et l'émission lumineuse étendue n'est visible qu'en fusionnant des images obtenues en bande étroite. C'est la démarche adoptée, par exemple, dans [Wisotzki et al., 2016] pour permettre une visualisation de la cartographie des halos. Ce type d'approche est cependant trop peu robuste par rapport au bruit pour permettre une interprétation astrophysique des « formes » obtenues.

Pour chaque source, une image hyperspectrale centrée spatialement et spectralement autour de la raie d'émission est isolée. Pour éviter la contamination de la détection par les sources les plus brillantes présentes dans le voisinage projeté (galaxies, étoiles), un filtrage médian est effectué sur chaque spectre. Cela permet de centrer en $\mathbf{0}$ les spectres sous $\mathcal{H}_0$.

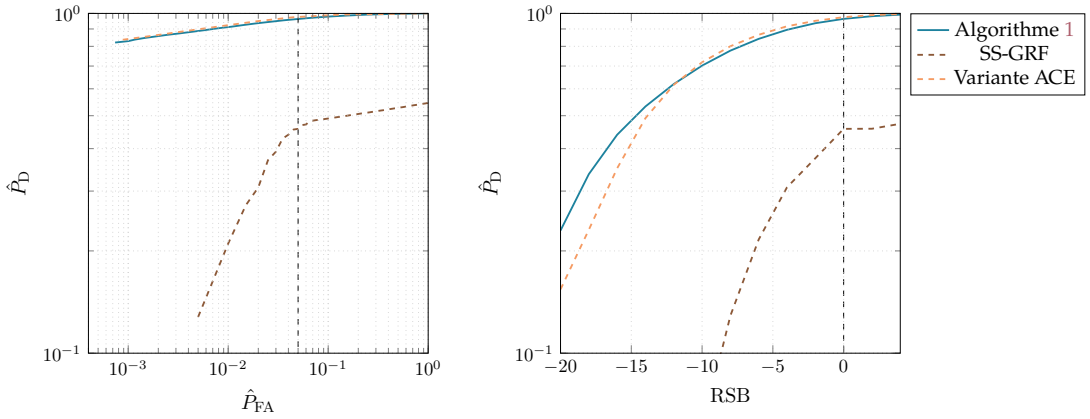



Figure 3.12 – Résultats numériques comparatifs entre la méthode proposée dans l'algorithme 1, sa variante basée sur un test ACE, et la détection base sur un modèle SS-GRF [Ahmad et al., 2015]. La légende est identique à celles des premières lignes des figures 3.8, 3.9 et 3.11.

L'algorithme 1 est ensuite appliqué à chaque image.

Nous présentons ici des résultats sur 6 objets, en sélectionnant une $P_\mathrm{FA}$ cible maximale à $10^{-4}$. Ces résultats illustrent la variété des cas se présentant avec des données réelles :

— la forme spatiale des résultats varie fortement. Les formes peuvent être limitées (ID 552) ou étendues (ID 43, 92, 139) ; isotropes (ID 43, 144) ou non (ID 92, 200) ; partagées en plusieurs parties (ID 139, 552) ou non (ID 43, 92, 144, 200) ;

— les spectres moyens (parties brillantes, parties ténues) montrent tous une raie d'émission asymétrique caractéristique des émissions de Lyman-alpha. Ces raies d'émissions diffèrent cependant d'un objet à l'autre, notamment en terme d'intensité et de nombre de bandes spectrales couvertes ;

— les variations d'intensité traduisent de fortes variations de RSB entre les objets.

Ces constats illustrent les défis rencontrés lors de l'étude de ces objets. Nous pouvons en particulier constater qu'il n'est pas souhaitable d'utiliser des *a priori* sur la forme des objets. De plus, le comportement spectral est variable et inconnu *a priori*, ce qui ne permet pas de fournir un modèle précis des formes spectrales. Dans le contexte de ces problématiques, la méthode proposée semble robuste : les contours de détection varient très peu lorsque la $P_\mathrm{FA}$ varie. Cela signifie que la stratégie de détection proposée fournit un fort contraste entre le signal modélisé par $\mathcal{H}_1$ et l'arrière-plan représenté par $\mathcal{H}_0$.

## 3.7 Conclusion

Dans ce chapitre, nous avons présenté une méthode de détection statistique pour la détection de sources ténues et étendues dans les images hyperspectrales. Cette méthode est basée sur l'introduction de deux tests GLR avec contrainte de similarité. Les propriétés théoriques de ces tests ont également été établies. Nous avons de plus exploré les extensions de ces tests à la présence de signaux cohérents spatialement et à l'exploitation de plusieurs observations d'une même scène.

La validation sur des données de synthèse est convaincante et met en évidence de bons résultats à faible RSB. Des résultats complémentaires seront présentés dans les chapitre 5 et 10, pour comparaison avec les autres méthodes introduites dans ce manuscrit. Nous y montrons notamment que les résultats de la méthode présentée dans ce chapitre ne sont pas, à très faible RSB, totalement satisfaisants. Ce phénomène a conduit au développement

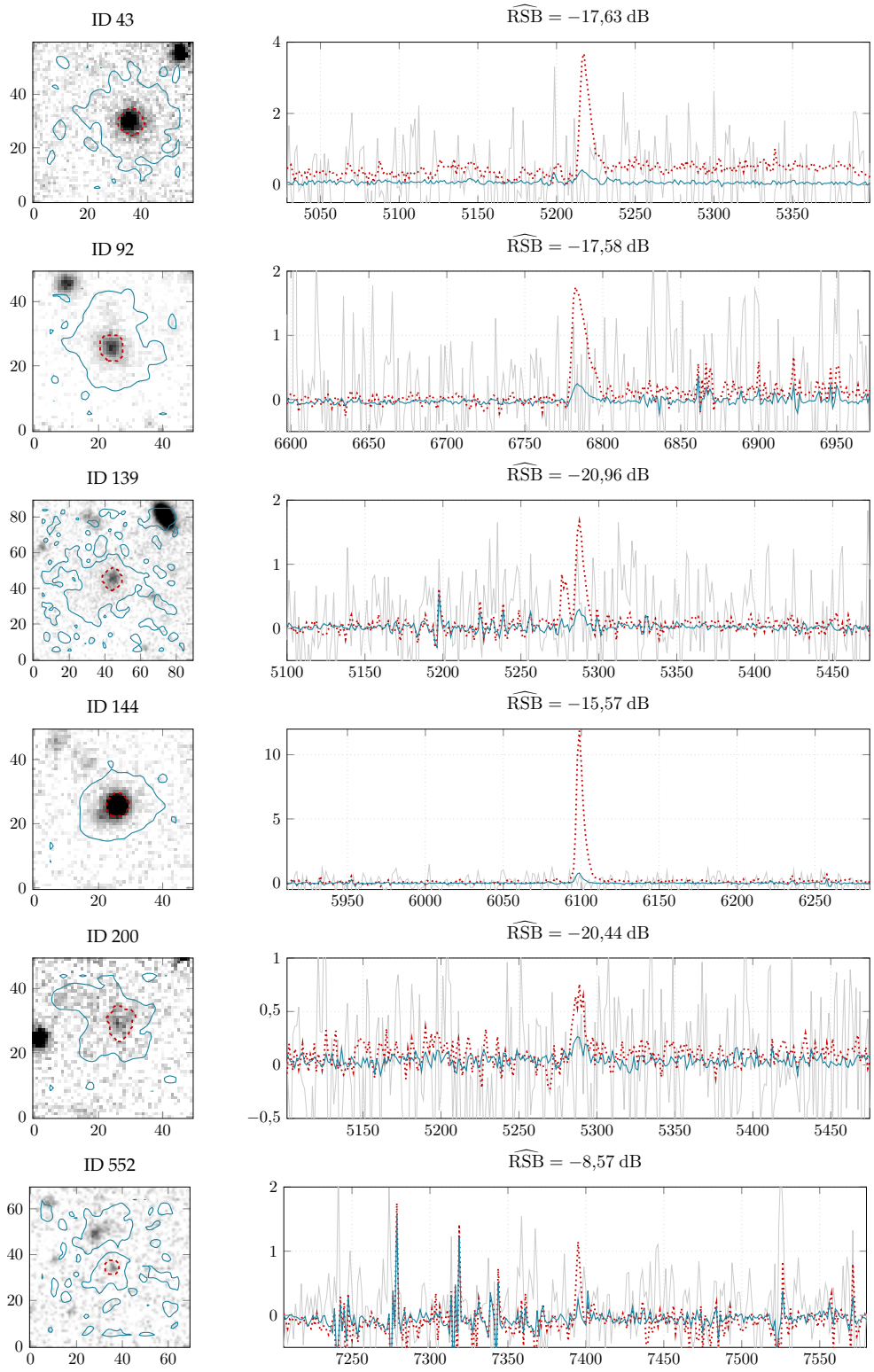



Figure 3.13 – Résultats obtenus sur 6 objets de l'observation HDFS, identifiés par leur numéro de catalogue « ID » [Bacon et al., 2015]. À gauche : images en bande étroite (moyenne sur les 10 bandes centrales de l'image hyperspectrale). Les contours rouges et bleus sont ceux de la première partie de la détection et de la méthode de l'algorithme 1 respectivement. Les graphiques de droite montrent les spectres moyens sur ces régions. Un spectre localisé en-dehors des régions détectées est illustré en gris à titre d'exemple. Les longueurs d'onde sont en Angström et les intensités lumineuses en $10^{-20}\mathrm{erg\ s^{-1}cm^{-2}\AA^{-1}}$.

d'une méthode alternative pour la détection dans les images hyperspectrales.

# II

# Modèles markoviens pour la segmentation d'images

**Résumé.** Cette partie présente les travaux liés aux modèles de Markov. Dans un premier temps, le chapitre 4 présente une synthèse sur les modèles markoviens qu'il est possible d'employer en traitement d'images. Dans le chapitre 5, nous introduisons un modèle de Markov couple permettant la formulation du problème de détection comme un cas particulier de segmentation d'image. Le contexte d'application est identique à celui de la partie I. Nous nous intéressons ensuite à la possibilité de modéliser des structures orientées à grande échelle dans les images. Ainsi, dans le chapitre 6, nous proposons la modélisation et la segmentation simultanée des classes et orientations dans les images par un modèle de champs de Markov triplet orienté. Dans le chapitre 7, nous introduisons un modèle d'arbre de Markov permettant la segmentation de structures hiérarchiques homogènes dans les images. Ces deux dernières études sont formulée dans le cadre général de la segmentation d'images. Leur application à la détection dans des images hyperspectrales astronomiques est présentée dans le chapitre 10.

# 4

# Synthèse sur les modèles markoviens pour la segmentation d'images



Les modèles de Markov portent le nom du mathématicien russe Andrei Andreiivitch Markov, qui les étudia au début du vingtième siècle [Markov, 1913]. Un aperçu de l'historique de ces travaux peut être trouvé dans [Hayes et al., 2013; Seneta, 1996]. Historiquement, les modèles de Markov sont d'abord considérés comme des séquences « temporelles ». La considération de processus markoviens avec deux dimensions spatiales a son origine dans les travaux du physicien allemand Ernst Ising, qui propose un modèle mathématique des interactions ferromagnétiques. Ce modèle, introduit en mécanique statistique, a été généralisé pour représenter des processus mathématiques sur une grille, notamment par les travaux de [Spitzer, 1971]. Ce type de processus a été, par extension, baptisé *champ de Markov*. Les variables contenues dans un processus de champ de Markov ne sont pas ordonnées avec un premier et un dernier élément, comme le sont les séquences. Ce point rend complexe les calculs associés à ces modèles. C'est dans ce cadre qu'ont été développés les processus en *arbre de Markov*, afin d'approcher les modèles de champ tout en facilitant les calculs associés. Un très bon historique de ces développements peut être trouvé dans [Kindermann et Snell, 1980].

L'application des modèles de Markov au traitement d'image est relativement récente. Les travaux fondateurs dans ce contexte sont ceux de [Besag, 1974], [Geman et Geman,

1984] et [Marroquin et al., 1987]. Cette application considère une image comme un processus aléatoire associé à un processus de Markov, et permet de déduire des informations sur cette image par inférence statistique.

La section 4.1 donne une introduction Les modèles markoviens qu'il est possible d'utiliser dans le cadre de la segmentation sont tout d'abord introduits en section 4.1, puis la section 4.2 présente leur application pour la segmentation non supervisée d'images. Les modèles de Markov couples et triplets, permettant d'enrichir les modélisations considérées, sont enfin présentés en section 4.3.

## 4.1 Introduction aux modèles markoviens

### 4.1.1 Chaînes de Markov

La markovianité d'un vecteur de variables aléatoires propose une certaine restriction des dépendances possibles entre variables.

**Définition 4.1.1** (Chaîne de Markov). *Soit* $\mathbf{X} = (X_1, \ldots, X_S)$ *un vecteur de* $S$ *variables aléatoires* [1]. $\mathbf{X}$ *est une chaîne de Markov si* $\forall s \in \{2, \ldots, S\}$ *:*

$$p(X_s|X_1, \ldots, X_{s-1}) = p(X_s|X_{s-1}) \tag{4.1}$$

*Cela signifie que chaque* $X_s$ *est indépendante de* $(X_1, \ldots, X_{s-2})$*, conditionnellement à* $X_{s-1}$*.*

En conséquence, si $\mathbf{X}$ forme une chaîne de Markov, sa distribution est définie par la loi du premier terme $X_1$ et par toutes les lois de transitions apparaissant dans le dernier terme de l'équation suivante :

$$\begin{aligned} p(\mathbf{X}) &\triangleq p(X_1, \ldots, X_S) \\ &\triangleq p(X_S|X_1, \ldots, X_{S-1})p(X_{S-1}|X_1, \ldots, X_{S-2}) \ldots p(X_2|X_1)p(X_1) \\ &= p(X_1)p(X_2|X_1) \ldots p(X_S|X_{S-1}). \end{aligned} \tag{4.2}$$

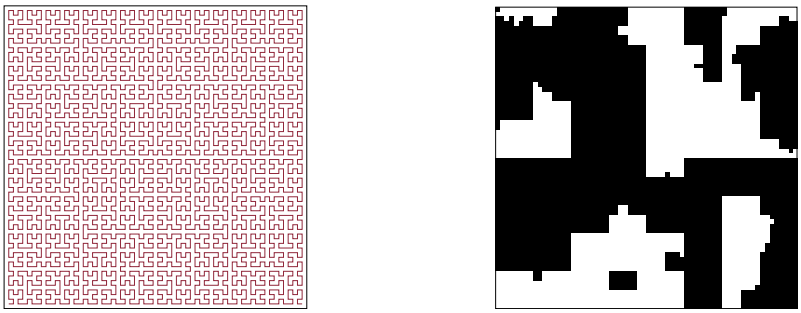
Les nombreux exemples de la littérature font intervenir, pour faciliter la modélisation, la propriété d'*homogénéité*.

**Propriété 4.1.1** (Homogénéité). *Soit* $\mathbf{X} = (X_1, \ldots, X_S)$ *une chaîne de Markov.* $\mathbf{X}$ *est homogène si* $\forall s \in \{2, \ldots, S\}$*, la distribution de transition* $p(X_s|X_{s-1})$ *ne dépend pas de* $s$*.*

Ainsi, la distribution d'une chaîne de Markov homogène $\mathbf{X}$ est définie uniquement par $p(X_1)$ et $p(X_s|X_{s-1})$ en tout $s \in \{2, \ldots, S\}$. Le nombre de paramètres caractérisant cette chaîne est alors beaucoup plus petit.

Pour permettre le traitement d'image par des modèles de chaînes de Markov, il est nécessaire d'associer à chaque site de l'image un élément de la chaîne. Cela peut se faire à l'aide d'un parcours de Hilbert-Peano [Sagan, 1994]. La figure 4.1 illustre une réalisation de chaîne de Markov dans ce cadre : la signature du parcours est visible, et forme des « blocs » marqués de réalisations identiques. Cette représentation des chaînes est assez courante : la littérature présente de nombreux travaux concernant la segmentation d'images avec un modèle par chaîne de Markov, par exemple d'images 2D ou de séquences d'images [Giordana et Pieczynski, 1997; Yahiaoui et al., 2016], d'images 3D [Bricq et al., 2006; Courbot et al., 2016c] ou d'images hyperspectrales [Mercier et al., 2003].

1. La propriété de Markov existe également sur des séquences indexées de manière continue (voir par exemple [Ethier et Kurtz, 2009]). Cette situation n'est pas abordée dans ce manuscrit.



(a) (b)

Figure 4.1 – (a) Parcours de Hilbert-Peano permettant de former une image à partir d'une chaîne de Markov de $64^2$ variables. (b) Réalisation d'une chaîne de Markov de $64^2$ variables, à deux classes, reformées en image selon un parcours de Hilbert-Peano. Ce type de procédure peut être généralisé au cas 3D [Bricq et al., 2006].

### 4.1.2 Markovianité sur graphe

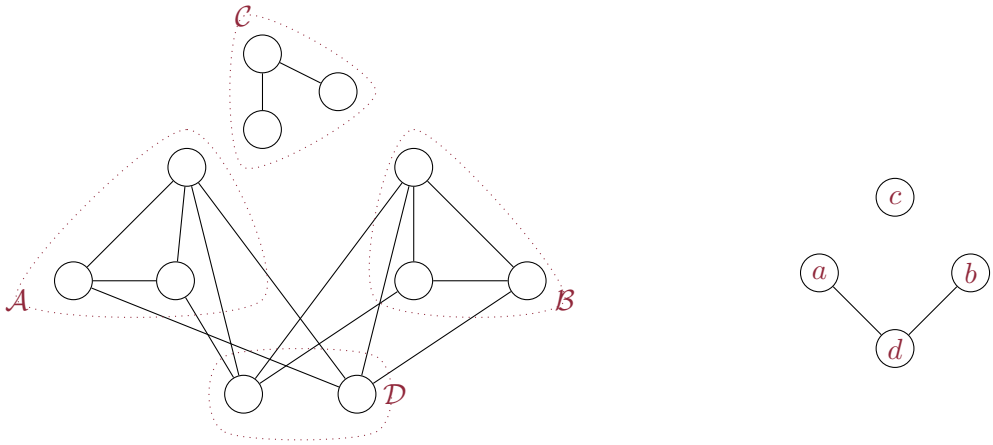
Pour comprendre la complexité des processus stochastiques composés de plusieurs variables aléatoires corrélées, comme une chaîne de Markov, il est possible de recourir à la représentation par graphes de dépendance.

**Définition 4.1.2** (Markovianité sur graphe)**.** *Soit $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ un vecteur de variables aléatoires indexé par l'ensemble fini $\mathcal{S}$, et soit $G = (\mathcal{S}, \Gamma)$ un graphe.*

*La loi de $\mathbf{X}$ est markovienne par rapport au graphe $G$ dans les conditions suivantes. Soient trois sous-ensembles $\mathcal{A}, \mathcal{B}, \mathcal{C} \subset \mathcal{S}$ :*

1. *$\forall a \in \mathcal{A}$ et $\forall b \in \mathcal{B}$, s'il n'existe pas de chemin entre $a$ et $b$, alors $(X_a)_{a\in\mathcal{A}}$ et $(X_b)_{b\in\mathcal{B}}$ sont indépendants ;*
2. *$\forall a \in \mathcal{A}$ et $\forall c \in \mathcal{C}$, si tout chemin reliant $a$ à $c$ passe par $\mathcal{B}$, alors $(X_a)_{a\in\mathcal{A}}$ et $(X_c)_{c\in\mathcal{C}}$ sont indépendants conditionnellement à $(X_b)_{b\in\mathcal{B}}$.*

*On dit alors que le graphe $G$ est un* graphe de dépendance *pour $\mathbf{X}$.*




(a) Cas quelconque : $\mathcal{A}$ et $\mathcal{C}$ sont indépendants, $\mathcal{A}$ et $\mathcal{B}$ sont indépendants conditionnellement à $\mathcal{D}$.

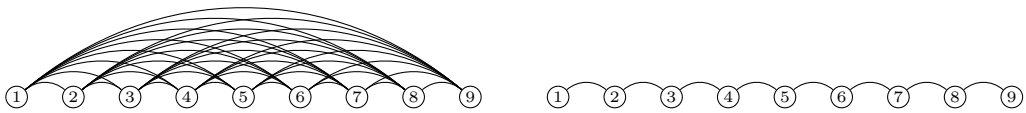
(b) Cas des singletons : $a$ et $c$ sont indépendants, $a$ et $b$ sont indépendants conditionnellement à $d$.

Figure 4.2 – Exemple de graphes de dépendances.

La figure 4.2 illustre ces situations, dans un cas quelconque et dans le cas où les sous-ensembles considérés sont des singletons. L'absence d'arête entre deux sous-ensembles de

$\mathcal{S}$ dans le graphe implique une indépendance conditionnelle, mais la présence d'arête est non informative : les variables impliquées peuvent être dépendantes ou indépendantes.

La figure 4.3 propose le graphe de dépendances illustrant les concessions nécessaires, en terme d'indépendance conditionnelle, pour faire d'un vecteur $\mathbf{X}$ quelconque une chaîne de Markov. Intuitivement, nous pouvons concevoir que ces concessions soient peu satisfaisantes en traitement d'image. Un parcours en chaîne ne considère en effet, conditionnellement à chaque pixel, que les dépendances à 2 pixels voisins, alors que chaque pixel a « au moins » 4 voisins. En prenant en compte pour chaque site un voisinage plus complexe, les modèles par champs de Markov permettent de faire moins de concessions que les chaînes de Markov.




(a) Vecteur aléatoire sans hypothèse d'indépendance conditionnelle.

(b) Chaîne de Markov.

Figure 4.3 – graphe de dépendances illustrant les concessions d'indépendances conditionnelles à faire au sein d'un vecteur $(X_1, \ldots, X_9)$.

### 4.1.3 Champs de Markov

Les champs de Markov, qui généralisent les chaînes de Markov, permettent une modélisation qui apparaît plus intuitive pour le traitement d'images.

**Définition 4.1.3** (Champ de Markov). *Soit $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ un vecteur aléatoire aléatoire. $\mathbf{X}$ est un champ de Markov si $\forall s \in \mathcal{S}$ :*

$$p(X_s|(X_{s'})_{s'\in\mathcal{S},s'\neq s}) = p(X_s|\mathbf{X}_{N_s}); \tag{4.3}$$

*où $N_s$ représente un voisinage (ensemble des sites voisins) de $s$ de « forme géométrique », $N$, fixée.*

Le voisinage $N$ peut représenter par exemple les 4 ou les 8 plus proches voisins de chaque variable de $\mathcal{S}$. Une illustration de graphes de dépendances dans ces deux cas est proposée en figure 4.4. Les graphes sous forme de vecteurs sont à comparer avec les graphes de la figure 4.3b. Les représentations graphiques montrent que les champs de Markov généralisent les chaînes de Markov. Cependant, la complexité d'un vecteur sans hypothèses d'indépendance n'est pas atteinte. La figure 4.5 illustre deux réalisations de champs de Markov en considérant un 4-voisinage ou un 8-voisinage.

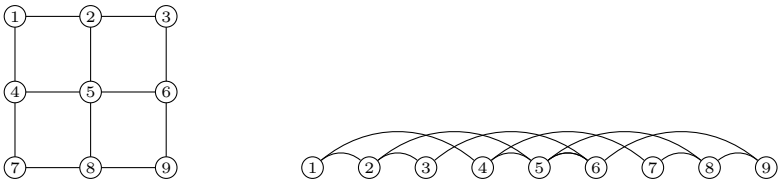
Pour obtenir des réalisations de champs de Markov, nous pouvons utiliser la forme analytique donnée par le théorème de Hammersley-Clifford. Cette forme repose sur la notion de *clique*.

**Définition 4.1.4** (Clique). *Soit $N$ un voisinage associé à $\mathcal{S}$. Une clique $c \subset \mathcal{S}$ est soit :*

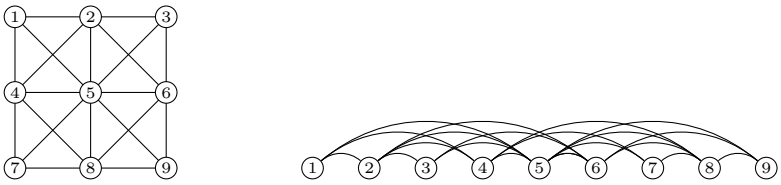
- *un singleton ;*
- *un ensemble d'éléments mutuellement voisins dans $N$.*

**Théorème 4.1.1** (Hammersley-Clifford). *Soit $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ un vecteur aléatoire dont les probabilités de chaque réalisation sont non nulles. $\mathbf{X}$ est un champ de Markov par rapport au système de voisinages $(N_s)_{s\in\mathcal{S}}$ si et seulement si la fonction énergie $U$ caractérisant la distribution $p(\mathbf{X}) = \gamma \exp[-U(\mathbf{X})]$ s'écrit comme une fonction de Gibbs :*

$$U(\mathbf{X}) = \sum_{c\in\mathcal{C}} \phi_c(\mathbf{X}_c); \tag{4.4}$$

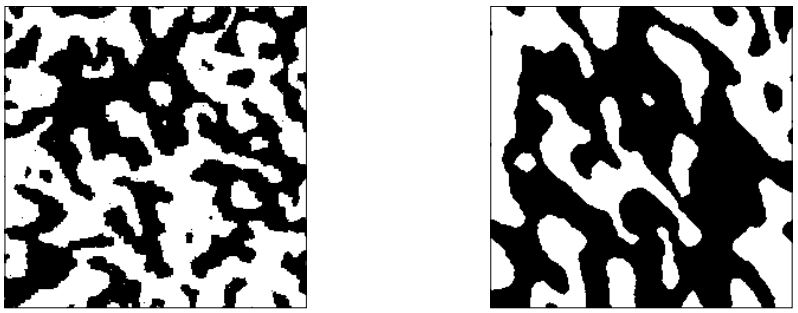



(a) 4-voisinage




(b) 8-voisinage

Figure 4.4 – Graphes de dépendances associés à $\mathbf{X} = (X_1, \ldots, X_9)$ selon un modèle de champ de Markov à 4 ou 8 voisins. Les graphes de droite sont sous forme vectorielle pour comparaison avec les graphes de la figure 4.3.



(a) (b)

Figure 4.5 – Réalisation de champs de Markov de $256 \times 256$ variables à deux classes, reposant sur un modèle d'Ising/Potts avec un système à 4 voisins (a) et 8 voisins (b).

*où $\mathcal{C}$ est l'ensemble des cliques associées au système de voisinages $(N_s)_{s \in \mathcal{S}}$.*

Ce théorème nous renseigne sur la forme de la loi de $\mathbf{X}$ lorsque le système de voisinages et les fonctions $\phi_c$ sont connus. La constante de normalisation $\gamma$ est en revanche ardue à calculer : en effet, elle requiert le calcul des distributions selon toutes les combinaisons possibles de $\mathbf{X}$. Le coût de calcul devient combinatoire, ce qui le rend dans la plupart des cas inenvisageable au-delà de quelques variables dans $\mathbf{X}$. En revanche, les distributions $p(X_s|X_{N_s})$ peuvent être calculées exactement. Ce point est essentiel pour la segmentation d'images, car il permet de simuler des réalisations $\mathbf{X} = \mathbf{x}$ autrement inaccessibles grâce à des procédures itératives, comme l'échantillonneur de Gibbs [Geman et Geman, 1984] ou de Metropolis [Metropolis et al., 1953]. La littérature portant sur la segmentation d'images par champs de Markov abonde [Benboudjema et Pieczynski, 2005, 2007; Besag, 1986; Dass, 2004; Derin et Elliott, 1987; Descombes et al., 1995; Fjortoft et al., 2003; Geman et Geman, 1984; Marroquin et al., 1987; Pieczynski et Tebbache, 2000; Salzenstein et Pieczynski, 1997; Wu et al., 2011; Zhang et al., 2012, 2001] et reflète le succès de ce type de méthode dans une très grande variété d'applications.

Les procédures itératives nécessaires à la simulation de réalisations de $\mathbf{X}$ induisent des temps de calculs plus importants que lors de l'utilisation de chaînes de Markov, et traduisent l'impossibilité de connaître exactement $p(\mathbf{X})$. La modélisation des corrélations

est cependant plus intuitive. Les modèles d'arbres de Markov offrent une alternative à ces deux modèles, tant en richesse de modélisation qu'en possibilités calculatoires.

### 4.1.4 Arbres de Markov

Nous présentons la définition d'un arbre de Markov à l'aide du graphe de dépendances associé, qui forme un *arbre*.

**Définition 4.1.5** (Arbre). *Un arbre est un graphe :*
- *connexe : pour tout couple de sites*[2] $(s,t)$ *il existe un chemin permettant d'aller de* $s$ *à* $t$ *;*
- *partitionné en résolutions (ou échelles) numérotées de* $0$ *pour la plus grossière à* $N$ *pour la plus fine ;*
- *ne contenant pas de cycle.*

*De plus, un arbre est homogène si chaque site, à l'exception de ceux de l'échelle* $N$*, est relié à un même nombre de sites.*

Nous appelons *racine* la résolution $0$ et *feuilles* les sommets de la résolution $N$. Chaque site de toutes les autres résolutions possède un site *parent* (échelle plus grossière) et plusieurs sites *enfants* (échelle plus fine). L'ensemble $\mathcal{S}$ des sites du graphe peut être subdivisé selon les échelles : $\mathcal{S} = \{\mathcal{S}^0, \ldots, \mathcal{S}^N\}$.

**Définition 4.1.6** (Arbre de Markov). *Soit* $G = (\mathcal{S}, \Gamma)$ *un graphe formant un arbre homogène, avec la subdivision en échelle* $\mathcal{S} = \{\mathcal{S}^0, \ldots, \mathcal{S}^N\}$. $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ *est un arbre de Markov si* $\forall s \in \mathcal{S} \setminus \mathcal{S}_0$ *:*

$$p(X_s|(X_{s'})_{s'\in\mathcal{S},s'\neq s}) = p(X_s|X_{s^-}) \tag{4.5}$$

*où* $X_{s^-}$ *est la variable parent de* $X_s$*.* $\mathbf{X}$ *est alors markovien par rapport au graphe* $G$*.*

En conséquence, la loi d'un arbre de Markov $\mathbf{X}$ est :

$$p(\mathbf{X}) = p(X_0) \prod_{s\in\mathcal{S}\setminus\mathcal{S}^0} p(X_s|X_{s^-}); \tag{4.6}$$

en assimilant à la variable $X_0$ le singleton de la racine $(X_s)_{s\in\mathcal{S}^0}$. Un graphe de dépendances d'arbre de Markov est donné en figure 4.6. La définition des liens de parentés permet de construire différents types d'arbres : si chaque site a deux descendants il s'agit d'un diarbre, si chaque site possède quatre descendants il s'agit d'un quadabre. Un exemple de réalisation d'arbre de Markov structuré en quadarbre est donné en figure 4.7, qui reflète bien la transmission hiérarchique de l'information (classes identiques) d'une échelle à l'autre.

Nous pouvons aussi constater que la structure en chaîne de Markov forme un cas particulier d'arbre de Markov dans lequel chaque nœud n'a qu'un seul descendant. Ce modèle en arbre peut permettre de représenter de manière naturelle les dimensions spatiales d'une image au sein de la résolution la plus fine, et d'éventuelles caractéristiques hiérarchiques selon les dépendances considérées dans l'arbre. La littérature présente là aussi plusieurs travaux sur la segmentation d'images avec des modèles d'arbres de Markov [Laferté et al., 2000; Monfrini et al., 1999; Pieczynski, 2003a, 2002; Provost et al., 2004].

Nous avons décrit les modèles de chaînes, arbres et champs de Markov. Ces modèles ne permettent pas d'effectuer directement une classification dans des images : il faut en effet pouvoir distinguer l'image à traiter de la classification recherchée. Cela conduit à définir deux processus : le processus d'observation $\mathbf{Y}$ et le processus de classes $\mathbf{X}$.

2. Par analogie avec les modèles de champs de Markov, nous appelons *site* d'un graphe ses sommets.

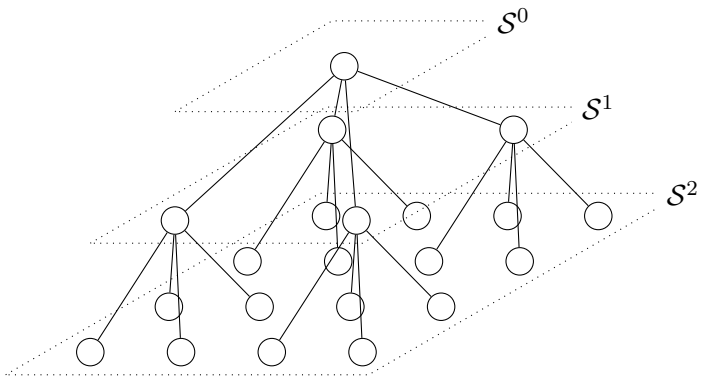



Figure 4.6 – Graphe de dépendances associé à un arbre de Markov, dans une structure en quadarbre.

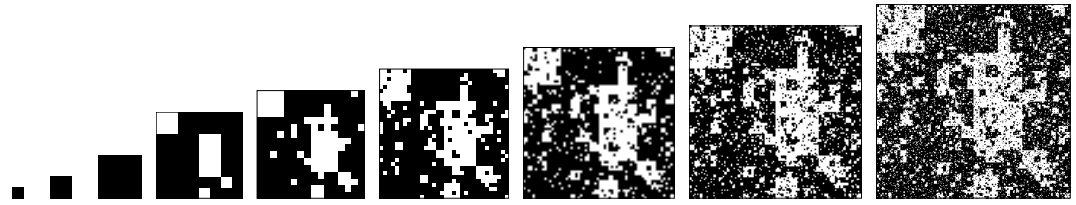



Figure 4.7 – Réalisation d'un arbre de Markov à deux classes et à 9 résolutions. De gauche à droite, les résolutions occupent $1, 2^2, 4^2, \ldots, 256^2$ pixels (échelle non respectée).

## 4.2 Segmentation non supervisée d'images

Dans cette partie, nous décrivons les méthodes de segmentation non supervisée d'images pour les modèles utilisés dans ce manuscrit. Nous nous plaçons ici dans le cadre de la segmentation par modèles de Markov cachés à bruit indépendant [3], qui sont présentés en section 4.2.1. La section 4.2.2 décrit les outils de la segmentation bayésienne reposant sur des modélisations de Markov, et la section 4.2.3 décrit les différentes stratégies d'estimation des paramètres qui peuvent être employées en contexte non supervisé.

### 4.2.1 Modèles de Markov cachés à bruit indépendant

Le problème de la segmentation bayésienne est celui de l'estimation de la réalisation d'un processus de classes $\mathbf{X} = \mathbf{x}$, inobservable, à partir d'une réalisation d'une observation

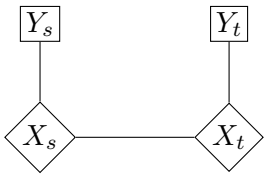



Figure 4.8 – Graphe de dépendances dans le cadre de modèle de Markov caché à bruit indépendant. Le graphe est limité à deux sommets voisins dans les graphes de dépendances présentés dans les figures 4.3b, 4.4 et 4.6.

3. Les modèles couples et triplets, formant une généralisation du modèle caché, sont décrits dans la section 4.3.

$\mathbf{Y} = \mathbf{y}$. Nous supposons que $\mathbf{x}$ est à valeurs dans l'ensemble discret $\Omega_x$ [4] et que $\mathbf{y}$ est, dans le cas général, à valeurs réelles $K-$variées. Nous avons $\mathbf{x} \in \Omega_x^{|\mathcal{S}|}$ et $\mathbf{y} \in \mathbb{R}^{|\mathcal{S}| \times K}$ où $\mathcal{S}$ est l'ensemble des sites du processus (chaîne, arbre ou champ).

La nature des liens statistiques entre $\mathbf{X}$ et $\mathbf{Y}$ est notamment définie par le choix du modèle de Markov *caché*. Soient deux variables localisées en deux sites connectés (« voisins ») sur l'un des graphes de dépendances présentés dans les figures 4.3b, 4.4 et 4.6. Ces variables seront indicées par $s$ et $t$, telles que :

— dans le cas d'une chaîne de Markov, $t$ correspond à $s-1$ ;
— dans un champ de Markov, $t$ est un élément du voisinage $N_s$ de $s$ ;
— pour un arbre de Markov, $t$ correspond à $s^-$.

**Définition 4.2.1** (Modèle de Markov Caché à bruit indépendant)**.** $(\mathbf{X}, \mathbf{Y})$ *correspond à un modèle de Markov caché si* $\mathbf{X}$ *est markovien d'après les définitions 4.1.1, 4.1.3 ou 4.1.5, et si pour tous* $s$ *et* $t$ *voisins :*

$$p(Y_s, Y_t | X_s, X_t) = p(Y_s|X_s)p(Y_t|X_t). \tag{4.7}$$

Le graphe de dépendances pour le modèle de Markov caché à bruit indépendant est présenté dans la figure 4.8. Les modèles de Markov cachés à bruit indépendant représentent la vaste majorité des modèles markoviens employés en segmentation d'images [Besag, 1986; Dass, 2004; Derin et Elliott, 1987; Descombes et al., 1995; Fjortoft et al., 2003; Geman et Geman, 1984; Giordana et Pieczynski, 1997; Laferté et al., 2000; Marroquin et al., 1987; Mercier et al., 2003; Monfrini et al., 1999; Provost et al., 2004; Salzenstein et Pieczynski, 1997; Zhang et al., 2001]

### 4.2.2 Segmentation bayésienne

En vertu de la loi de Bayes, nous pouvons écrire :

$$p(\mathbf{X}|\mathbf{Y}) \propto p(\mathbf{X})p(\mathbf{Y}|\mathbf{X}) \tag{4.8}$$

La distribution *a posteriori* de $\mathbf{X}$ est connue lorsque les deux termes du membre de droite sont connus.

Un point essentiel pour la segmentation des images est la propriété de *markovianité a posteriori*, qui stipule que le loi $p(\mathbf{X}|\mathbf{Y})$ est markovienne. En effet, cette loi peut souvent être exprimée (au moins à une constante de normalisation près), ce qui permet d'en générer des réalisations à des fins de segmentation. Cette propriété est vérifiée dans le cadre des modèles cachés présentés en section 4.2.1.

Plusieurs critères permettent la segmentation bayésienne des images, dans le cadre de modèles de Markov. L'estimateur du *maximum a posteriori* (MAP) [Geman et Geman, 1984] recherche la configuration globale de $\mathbf{X}$ maximisant la distribution *a posteriori* $p(\mathbf{X}|\mathbf{Y})$ :

$$\hat{\mathbf{x}}^{\text{MAP}} = \arg\max_{\boldsymbol{\omega} \in \Omega_x^{|\mathcal{S}|}} p(\mathbf{X} = \boldsymbol{\omega} | \mathbf{Y} = \mathbf{y}); \tag{4.9}$$

où $\boldsymbol{\omega}$ est une réalisation de $\mathbf{X}$ et $\mathbf{y}$ est la réalisation fixée de $\mathbf{Y}$.

L'estimateur du MAP minimise une fonction de coût binaire :

$$L^{\text{MAP}}(\hat{\mathbf{x}}, \mathbf{x}) = \left| \begin{array}{l} 0 \text{ si } \hat{\mathbf{x}} = \mathbf{x} \\ 1 \text{ si } \hat{\mathbf{x}} \neq \mathbf{x} \end{array} \right. ; \tag{4.10}$$

4. Le cas où les classes sont continues est également envisageable, mais fait appel à d'autres catégories de méthodes qui relèvent davantage de la *restauration* d'images que de la segmentation [Li, 2009].

où $\hat{\mathbf{x}}$ est une estimation de la « réalité » $\mathbf{x}$. Une autre fonction de coût peut être définie, de sorte à minimiser le nombre de classifications erronées :

$$L^{\mathrm{MPM}}(\hat{\mathbf{x}}, \mathbf{x}) = \sum_{s \in \mathcal{S}} \left[1 - \delta_{x_s}(\hat{x}_s)\right]; \tag{4.11}$$

où $\delta$ est le symbole de Kronecker de $x_s$. Optimiser ce critère conduit à formuler la segmentation au sens du *maximum posterior mode* (MPM)[5] [Marroquin et al., 1987] :

$$\forall s \in \mathcal{S} : \hat{x}_s^{\mathrm{MPM}} = \arg\max_{\omega \in \Omega_x} p(X_s = \omega | \mathbf{Y} = \mathbf{y}). \tag{4.12}$$

Sans autre information supplémentaire sur la distribution de $(\mathbf{X}, \mathbf{Y})$, il n'existe aucune raison de choisir un estimateur plutôt que l'autre, malgré la différence entre les fonctions de coût qu'ils optimisent. Le calcul des distributions *a posteriori* requises pour ces estimateurs dépend du choix parmi les modélisations décrites dans la section précédente. Ces calculs sont présentés dans le cadre des modèles développés dans les chapitres 5, 6 et 7.

### 4.2.3 Estimation des paramètres

La mise en place du modèle repose sur des choix de distributions, régies en général par plusieurs paramètres. Dans le contexte de la segmentation non supervisée, il est nécessaire d'estimer ces paramètres à partir de l'observation $\mathbf{Y} = \mathbf{y}$ uniquement en amont de la segmentation. Nous décrivons dans cette section plusieurs stratégies envisageables à cet effet.

L'algorithme *expectation-maximization* (EM) est une méthode générale permettant l'estimation des paramètres $\theta$ lorsque des données sont manquantes [Dempster et al., 1977]. Cet algorithme part du constat que la distribution $p(\mathbf{X}, \mathbf{Y})$ est inconnue, mais que la distribution $p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ peut être obtenue car elle ne dépend que de l'observation $\mathbf{Y} = \mathbf{y}$. L'algorithme EM utilise au lieu de $p(\mathbf{X}, \mathbf{Y})$ l'espérance $\mathbb{E}$ de la loi $p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$, et maximise cette espérance conditionnelle. L'algorithme est itératif et génère une séquence $(\theta_1, \ldots, \theta_Q)$ de paramètres, où à l'itération $q$ est formée des étapes :

1. *expectation* (E) : calcul de $\mathbb{E}_{\theta_{q-1}}\left[p_\theta(\mathbf{X}|\mathbf{Y} = \mathbf{y})\right]$ ;
2. *maximization* (M) : choix des paramètres par $\theta_q = \arg\max_\theta \mathbb{E}_{\theta_{q-1}}\left[p_\theta(\mathbf{X}|\mathbf{Y} = \mathbf{y})\right]$.

L'algorithme s'arrête lorsqu'un critère d'arrêt est vérifié. L'estimateur est le dernier terme de la suite $\theta_Q$. L'algorithme EM est déterministe et converge [Wu, 1983]. Cependant, il n'y a pas de garantie sur la valeur de convergence : il peut converger vers un minimum local de la vraisemblance.

L'objectif de l'algorithme *stochastic expectation-maximization* (SEM) est la convergence vers le minimum global en introduisant une étape de simulation [Broniatowski et al., 1983; Celeux et Diebolt, 1986]. L'algorithme SEM se formule à chaque itération $q$ comme :

1. *stochastic/expectation* (SE) : simulation de $\hat{\mathbf{x}}^q$ selon $p_{\theta_{q-1}}(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ ;
2. *maximization* (M) : estimation des paramètres grâce aux estimateurs du maximum de vraisemblance sur $(\hat{\mathbf{x}}^q, \mathbf{y})$.

Tout comme EM, SEM est stoppé lorsqu'un critère d'arrêt est vérifié. SEM permet à la suite de paramètres, dans certains cas, « d'échapper » à un maximum local. En contrepartie, SEM n'a pas de propriété de convergence bien que son application donne de très bons résultats. Le critère d'arrêt de la méthode repose donc sur une mesure de « stabilité » à définir.

5. Cet estimateur est parfois appelé estimateur du MAP marginal.

Mentionnons de plus l'existence d'algorithmes dérivés de SEM : SAEM (pour « stochastic approximation EM ») [Celeux et Diebolt, 1992] et MCEM (pour « Monte-Carlo EM ») [Wei et Tanner, 1990]. Dans [Celeux et al., 1995], les auteurs montrent que les performances offertes par SEM sont meilleures que celles de SAEM et MCEM.

Enfin, l'algorithme itératif *iterative conditional estimator* (ICE) repose sur une formulation plus générale, dans laquelle la vraisemblance n'intervient pas [Pieczynski, 1992, 1994]. Sa formulation, à l'étape $q$, est la suivante :

$$\theta_q = \mathbb{E}_{q-1}\left[\hat{\theta}|\mathbf{Y}=\mathbf{y}\right]; \tag{4.13}$$

où $\hat{\theta}$ est un estimateur de $\theta$. Cette espérance n'est pas toujours calculable directement, il faut alors procéder à une estimation par les fréquences sur un grand nombre de réalisations $\hat{\theta}$. L'algorithme ICE, tout comme SEM, ne bénéficie pas de propriété de convergence, mais son application donne des résultats satisfaisants.

Les algorithmes EM, SEM et MCMC sont comparés dans [Dias et Wedel, 2004]. Une autre comparaison de EM, SEM et ICE peut être trouvée dans [Monfrini et Pieczynski, 2005]. Ces deux études ont montré que l'algorithme SEM offre des performances légèrement supérieures aux alternatives envisagées. Nous utiliserons donc des algorithmes inspirés de la méthode SEM pour l'estimation des paramètres dans les chapitres 5, 6 et 7, en précisant dans chaque cas les estimateurs nécessaires à son fonctionnement.

## 4.3 Modèles de Markov couples et triplets

Les modèles couples et triplets permettent des généralisations du modèle caché pour une modélisation plus riche tout en autorisant la segmentation et l'estimation des paramètres de manière similaire à l'exposé des sections 4.2.2 et 4.2.3.

### 4.3.1 Modèles de Markov couples

**Définition 4.3.1** (Modèle de Markov couple)**.** $(\mathbf{X}, \mathbf{Y})$ *correspond à un modèle de Markov couple si le processus* $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ *est markovien d'après les définitions 4.1.1, 4.1.3 ou 4.1.5.*

Cela signifie que les hypothèses d'indépendances évoquées dans le cas du modèle caché sont relaxées, et notamment que $\mathbf{X}$ n'est pas nécessairement markovien. La figure 4.9 donne le graphe de dépendances restreint aux sites voisins $(s,t)$ pour un modèle couple. Nous pouvons en particulier constater que les modèles cachés et cachés à bruit indépendant sont des cas particuliers d'un modèle de Markov couple. En effet, nous pouvons constater notamment d'après (4.7) qu'un modèle de Markov couple est un modèle de Markov caché à bruit indépendant si $p(Y_s, Y_t|X_t, X_s) = p(Y_s|X_s)p(Y_t|X_t)$, et un modèle de Markov caché lorsque $p(Y_s|X_t, X_s) = p(Y_s|X_s)$.

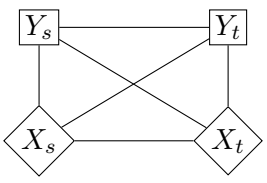



Figure 4.9 – Graphe de dépendances associé à un modèle de Markov couple, limité à deux sommets connectés dans les graphes de dépendances présentés dans les figures 4.3b, 4.6 ou 4.4.

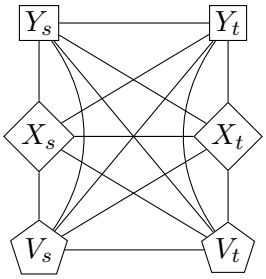



Figure 4.10 – Graphe de dépendances pour un modèle de Markov triplet, en deux sommets voisins des graphes de dépendances présentés dans les figures 4.3b, 4.6 et 4.4

La propriété de markovianité *a posteriori* est toujours valable dans le cas des modèles de Markov couple [Pieczynski et Tebbache, 2000]. Cela permet l'utilisation de formes plus générales des estimateurs du MAP et du MPM pour la segmentation, et des méthodes d'estimation des paramètres pour les cas non supervisés. En conclusion, les modèles de Markov couples permettent des modélisations plus riches que leurs homologues cachés, sans nécessiter le recours à des algorithmes dédiés pour la segmentation non supervisée d'images.

### 4.3.2 Modèles de Markov triplets

L'approche retenue dans le cadre des modèles de Markov triplets est l'enrichissement de la nature des informations inobservables par l'adjonction d'un processus $\mathbf{V}$ auxiliaire à $(\mathbf{X}, \mathbf{Y})$. $\mathbf{V}$ est inobservable comme $\mathbf{X}$, et prend valeurs dans $\Omega_v$ sur le même ensemble de sites : $\mathbf{V} = (V_s)_{s \in \mathcal{S}}$.

**Définition 4.3.2** (Modèle de Markov triplet)**.** $(\mathbf{X}, \mathbf{Y}, \mathbf{V})$ *correspond à un modèle de Markov triplet si le processus* $\mathbf{T} = (\mathbf{X}, \mathbf{Y}, \mathbf{V})$ *est markovien d'après les définitions 4.1.1, 4.1.3 ou 4.1.5.*

Cette définition ne spécifie pas le rôle de $\mathbf{V}$, ce qui le rend très polyvalent. À ce titre, $\mathbf{V}$ peut permettre de caractériser des stationnarités différentes dans $(\mathbf{X}, \mathbf{Y})$, ce qui n'est pas envisageable dans les cas couples et cachés. $\mathbf{V}$ peut aussi modéliser des changements dans les distributions $p(\mathbf{Y}|\mathbf{X})$ en fonction des régions de $s$.

Le graphe de dépendances d'un modèle couple relatif aux sites $(s, t)$ voisins est présenté en figure 4.10. Nous pouvons notamment constater que les modèles couples sont un cas particulier de modèle triplet, dans lequel le processus $\mathbf{V}$ est ignoré. De plus, l'hypothèse du modèle couple, selon laquelle $(\mathbf{Y}, \mathbf{X})$ est markovien, est relâchée. Enfin, la markovianité *a posteriori* reste valable : $p(\mathbf{X}, \mathbf{V}|\mathbf{Y})$ est de Markov selon (4.2), (**??**) ou (4.6). Cela permet là aussi d'envisager la segmentation conjointe des processus $\mathbf{X}$ et $\mathbf{V}$ à partir de l'image $\mathbf{Y} = \mathbf{y}$ uniquement.

Plus récents que leur homologue caché, les modèles couples et triplets sont par comparaison moins bien représentés dans la littérature [Benboudjema et Pieczynski, 2005, 2007; Bricq et al., 2006; Lanchantin et al., 2008; Pieczynski, 2003a; Pieczynski et al., 2003; Pieczynski, 2002, 2003b; Wu et al., 2011; Zhang et al., 2012].

## 4.4 Conclusion

Nous avons présenté les modèles de chaînes, champs et arbres de Markov qu'il est possible d'utiliser dans le cadre de la segmentation d'images. Le formalisme des modèles de Markov cachés nous a permis d'introduire les méthodes permettant la segmentation

bayésienne, ainsi que l'estimation non supervisée des paramètres. Nous avons ensuite présenté les modèles de Markov couples et triplets, permettant une plus grande richesse de modélisation que le modèle caché, tout en préservant la possibilité de la segmentation bayésienne non supervisée.

Dans le chapitre 5, nous présentons un modèle de champ de Markov couple avec pour cadre applicatif la segmentation de halos dans les images hyperspectrales astronomiques. Ensuite, un modèle de champ de Markov triplet permettant la description des orientations au sein d'une image est présenté dans la chapitre 6. Enfin, une approche par arbre de Markov triplet permettant de gérer l'homogénéité spatiale lors de la segmentation des images est présentée dans le chapitre 7.

# 5

# Champs de Markov couples convolutifs



## 5.1 Introduction

Nous avons vu dans la partie I comment le problème de détection peut être pris en charge par l'utilisation de tests d'hypothèses, ce qui a fait l'objet de plusieurs travaux récents au sein d'images hyperspectrales astronomiques [Paris et al., 2013a; Courbot et al., 2017a; Bacher et al., 2017a]. Alternativement, il est possible de modéliser explicitement la forme des objets recherchés. C'est ce que proposent les travaux basés sur des processus ponctuels marqués [Meillier et al., 2015]. Il est aussi possible de modéliser le problème de détection comme un cas particulier d'un problème de segmentation. Plus particulièrement, nous nous intéressons à une formulation bayésienne du problème, reposant sur des modèles markoviens.

La littérature présente de nombreux travaux sur la segmentation bayésienne des images hyperspectrales [Eches et al., 2013; Li et al., 2012, 2014; Rellier et al., 2004; Schweizer et Moura, 2000; Xia et al., 2015]. Tous sont basés sur des modèles de champs de Markov cachés. Concernant l'application astronomique, quelques travaux portent sur la segmentation d'images multispectrales [Salzenstein et Collet, 2006; Vollmer et al., 2013]. Bien que ces modèles puissent dans certain cas être généralisés à l'imagerie hyperspectrale, aucun ne propose la segmentation et la détection d'objets ténus dans des images hyperspectrales astronomiques. De plus, aucun de ces modèles ne propose de prendre en compte les phénomènes de convolution susceptibles d'intervenir dans la formation des images.

L'objectif de ce chapitre est d'introduire un modèle de Markov couple permettant la modélisation du problème de détection de sources convoluées comme un problème de segmentation ; Les résultats numériques sont présentés avec des données de synthèse, correspondant au problème du point de vue de la modélisation. Des résultats complémentaires sur données synthétiques et réelles sont quant à eux présentés dans le chapitre 10.

Les contributions de ce chapitre sont les suivantes :

— nous introduisons un modèle de champ de Markov couple *convolutif* (CMco) permettant la modélisation du phénomène de convolution ;
— nous présentons la segmentation bayésienne non supervisée dans le cadre de ce modèle ;
— une mesure de confiance associée à la segmentation est également proposée ;
— les performances des méthodes de segmentation sont établies sous plusieurs configurations expérimentales, et comparées avec l'approche développée dans le chapitre 3 [Courbot et al., 2017a] et la méthode présentée dans [Bacher et al., 2017a].

Le modèle de champ de Markov couple est introduit dans la section 5.2, de même que le modèle d'observation et la segmentation bayésienne non supervisée. Les résultats numériques et leurs modalités d'obtention sont quant à eux reportés en section 5.3.

## 5.2 Modèle

### 5.2.1 Champs de Markov couples convolutifs

Soient deux processus : $\mathbf{Y} = (\mathbf{Y}_s)_{s\in\mathcal{S}}$ est le processus d'observation et $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ est le processus de classes. $\mathcal{S}$ est l'ensemble des sites $s$ de l'image, qui est régie par un système de voisinage noté $(N_s)_{s\in\mathcal{S}}$. Les variables $X_s$ sont à valeurs dans $\Omega_x$ et les vecteurs $\mathbf{Y}_s$ sont à valeurs dans $\mathbb{R}^\Lambda$. Dans le cas de la modélisation d'une image hyperspectrale, les vecteurs $\mathbf{Y}_s$ représentent les spectres de l'image et $\Lambda$ est le nombre de bandes spectrales de ces spectres.

Dans le cadre de la modélisation par champs de Markov couples (CMC) [Pieczynski et Tebbache, 2000], le couple $(\mathbf{X}, \mathbf{Y})$ a une distribution de champ de Markov :

$$p(\mathbf{x}, \mathbf{y}) \propto \prod_{s\in\mathcal{S}} p(x_s, \mathbf{y}_s|\mathbf{x}_{N_s}, \mathbf{y}_{N_s}). \tag{5.1}$$

Nous supposons que $(\mathbf{X}, \mathbf{Y})$ est stationnaire, et retenons uniquement les cliques de deux éléments. Dans le cadre du modèle de Markov couple que nous proposons, nous faisons l'hypothèse que $\mathbf{Y}_{N_s}$ et $(X_s, \mathbf{Y}_s)$ sont indépendants conditionnellement à $\mathbf{X}_{N_s}$ :

$$p(x_s, \mathbf{y}_s|\mathbf{x}_{N_s}, \mathbf{y}_{N_s}) = p(x_s, \mathbf{y}_s|\mathbf{x}_{N_s}). \tag{5.2}$$

Cette densité peut s'écrire sous la forme suivante :

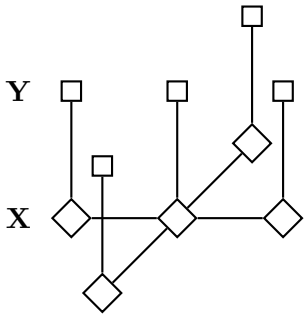
$$p(x_s, \mathbf{y}_s|\mathbf{x}_{N_s}) = p(\mathbf{y}_s|x_s, \mathbf{x}_{N_s})p(x_s|\mathbf{x}_{N_s}). \tag{5.3}$$

La figure 5.1 illustre le graphe de dépendances du modèle correspondant à (5.2).

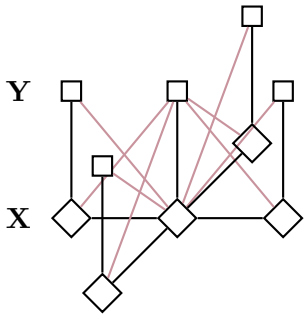
Nous détaillons maintenant les deux termes formant la densité (5.3). Nous supposons que $\mathbf{X}$ a une distribution de champ de Markov, reposant sur une fonction de Ising/Potts :

$$p(x_s|\mathbf{x}_{N_s}) \propto \exp\left(-\sum_{s'\in N_s} \alpha\left(1 - 2\delta_{x_s}^{x_{s'}}\right)\right); \tag{5.4}$$

où $\alpha$ est un paramètre du modèle et $\delta_{x_s}$ est la fonction de Kronecker de $x_s$.




(a) Champ de Markov caché à bruit indépendant.




(b) Champ de Markov couple convolutif.

Figure 5.1 – Graphes de dépendances dans le cadre d'un 4-voisinage. Les arêtes en couleur sont les arêtes additionnelles par rapport au modèle de champ de Markov caché à bruit indépendant.

La distribution $p(\mathbf{y}_s|x_s, \mathbf{x}_{N_s})$ régit la manière dont les classifications $\mathbf{X}$ influencent l'observation $\mathbf{Y}$. De manière générique, nous l'écrivons :

$$p(\mathbf{y}_s|x_s, \mathbf{x}_{N_s}) = g_{x_s,\mathbf{x}_{N_s}}(\mathbf{y}_s). \tag{5.5}$$

Le choix de la fonction $g$ dépend notamment du modèle de bruit souhaité, et donc de l'application recherchée. Dans la suite de ce chapitre, $g$ permettra la prise en compte d'un phénomène de convolution dans la modélisation.

### 5.2.2 Segmentation et mesure de confiance

#### Segmentation bayésienne

La segmentation bayésienne des images peut être obtenue entre autres par les estimateurs du MAP [Geman et Geman, 1984] et du MPM [Marroquin et al., 1987]. Rappelons que l'expression du MAP est :

$$\hat{\mathbf{x}}^{\text{MAP}} = \underset{\boldsymbol{\omega}\in\Omega_x^{|\mathcal{S}|}}{\arg\max}\, p(\mathbf{X} = \boldsymbol{\omega}|\mathbf{Y} = \mathbf{y}); \tag{5.6}$$

et que le MPM s'exprime $\forall s \in \mathcal{S}$ comme :

$$\hat{x}_s^{\text{MPM}} = \underset{\omega\in\Omega_x}{\arg\max}\, p(X_s = \omega|\mathbf{Y} = \mathbf{y}). \tag{5.7}$$

Le calcul de ces deux estimateurs repose sur la markovianité *a posteriori* du couple $(\mathbf{X}, \mathbf{Y})$ qui rend possible la simulation de $p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$. Ces deux estimateurs sont comparés dans la section 5.3.

#### Mesure de confiance pour le MPM

Nous nous intéressons à la détermination d'une mesure de confiance associée à la segmentation en chaque site $s \in \mathcal{S}$. Une telle mesure peut aisément être déduite dans le cadre de la segmentation au sens du MPM. En effet, les estimateurs de $X_s$ sont par construction les classes les plus probables par rapport à la distribution *a posteriori* $p(x_s|\mathbf{y})$. Le choix du MPM est une décision « dure » dans $\Omega_x$ qui ne reflète pas la confiance avec laquelle la décision a été prise. En effet, nous pouvons rencontrer des situations variables, comme :

— le cas le plus favorable : la classe retenue est beaucoup plus probable que les autres ;
— les cas ambigus, plus complexes : les probabilités *a posteriori* sont très proches entre les classes.

Nous introduisons ici une mesure de confiance complémentaire du résultat de segmentation. Elle est notée $\mathbf{u}^x$ et s'appuie sur la différence de probabilités estimée entre les classes retenues par le MPM et les autres classes. Nous avons ainsi $\forall s \in \mathcal{S}$ :

$$u_s^x = \min_{\omega \neq \hat{x}_s^{\mathrm{MPM}}} \left[ p(X_s = \hat{x}_s^{\mathrm{MPM}} | \mathbf{y}) - p(X_s = \omega | \mathbf{y}) \right] \tag{5.8}$$

La carte de confiance résultante $\mathbf{u}^x = (u_s^x)_{s \in \mathcal{S}}$ prend ses valeurs entre $0$ (segmentation très incertaine) et $1$ (segmentation très sûre).

### 5.2.3 Modèle d'observation

#### Distribution du bruit

Le choix d'un modèle d'observation permet notamment de spécifier la forme de $g_{x_s, \mathbf{x}_{N_s}}(\mathbf{y}_s)$ (5.5). Nous supposons que le bruit est de distribution normale multivariée :

$$g_{x_s, \mathbf{x}_{N_s}}(\mathbf{y}_s) \sim \mathcal{N}(\boldsymbol{\mu}_{x_s, \mathbf{x}_{N_s}}, \boldsymbol{\Sigma}); \tag{5.9}$$

où $\boldsymbol{\mu}_{x_s, \mathbf{x}_{N_s}} \in \mathbb{R}^\Lambda$ est un paramètre de moyenne et $\boldsymbol{\Sigma} \in \mathbb{R}^{\Lambda \times \Lambda}$ est la matrice de covariance du bruit.

Celle-ci est supposée identique pour toutes les classes. Nous supposons qu'elle est pentadiagonale et que les paramètres ne varient pas d'une composante de $\mathbf{y}_s$ à l'autre [1] :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho_1 & \rho_2 & 0 & \cdots & 0 \\ \rho_1 & \ddots & \ddots & \ddots & \ddots & \\ \rho_2 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \rho_2 \\ & \ddots & \ddots & \ddots & \ddots & \rho_1 \\ 0 & & 0 & \rho_2 & \rho_1 & \sigma^2 \end{bmatrix} \tag{5.10}$$

$\sigma$ correspond à l'écart-type des intensités, et $\rho_1$, $\rho_2$ sont les paramètres de corrélation entre les éléments consécutifs des observations $\mathbf{y}_s$.

#### Modèle de convolution

Nous nous intéressons aux phénomènes de convolution, pour lesquels même en l'absence de bruit l'image présente une dégradation par rapport à la classification recherchée. Cette dernière s'exprime comme un mélange des contributions associées à $\mathbf{x}_s$ et à $\mathbf{x}_{N_s}$ :

$$\boldsymbol{\mu}_{x_s, \mathbf{x}_{N_s}} = f_s \boldsymbol{\mu}_{x_s} + \sum_{r \in N_s} f_r \boldsymbol{\mu}_{x_r}; \tag{5.11}$$

où $\boldsymbol{\mu}_{x_s} \in \mathbb{R}^\Lambda$ est la moyenne associée à la classe de $x_s$ dans $\Omega_x$. Ce mélange est pondéré par une fonction formée des coefficients $\mathbf{f} = (f_r)_{r \in \{s\} \cup N_s}$, qui représente la fonction d'étalement spatiale du chapitre 3. Notons que $\mathbf{f}$ ne dépend pas de $s$ car le champ $(\mathbf{X}, \mathbf{Y})$ est stationnaire. Le modèle de champ de Markov couple que nous proposons peut ainsi permettre la prise en compte de phénomènes de convolution.

1. Ces hypothèses sont issues de l'application aux images de l'instrument MUSE : le bruit est stationnaire spectralement, et présente des corrélations limitées spectralement (cf. chapitre 9).

**Remarque 5.2.1.** *Dans (5.11), si tous les termes $(f_r)_{r\in N_s}$ sont nuls, le modèle devient un modèle de champ de Markov caché à bruit indépendant.*

#### Modèle de détection

Nous nous intéressons à l'application de ce modèle dans un contexte de détection ; qui ne prend en général que deux possibilités en compte : la présence et l'absence d'un signal d'intérêt. Cela se traduit par une segmentation à $K = 2$ classes :

— l'absence de signal, que nous associons à la classe $\omega_0$, se manifeste par la présence de bruit seul : $\boldsymbol{\mu}_{\omega_0} = \mathbf{0}_\Lambda$ ;
— la présence d'un signal associé à la classe $\omega_1$ correspond à une moyenne $\boldsymbol{\mu}$.

**Remarque 5.2.2.** *Ce cadre peut être enrichi dans le cadre de la segmentation, en prenant en compte différents niveaux d'atténuation dans le signal recherché. Cela peut permettre de modéliser $K-1$ classes intermédiaires par $\boldsymbol{\mu}_{\omega_k} = \frac{k}{K}\boldsymbol{\mu}_{\omega_K}$.*

### 5.2.4 Formulation alternative du modèle

La présentation du modèle dans le contexte des champs de Markov couple s'est appuyée sur la markovianité du couple $(\mathbf{x}, \mathbf{y})$ pour détailler immédiatement les distributions $p(x_s, \mathbf{y}_s|\mathbf{x}_{N_s}, \mathbf{y}_{N_s})$. Dans cette partie, nous présentons une formulation différente de ce modèle.

Par commodité, nous notons ici $z_s$ la valeur prise par $i$ lorsque $x_s = \omega_i$ [2], et le vecteur $\mathbf{z} = (z_s)_{s\in\mathcal{S}}$. En vectorisant $\mathbf{z}$ et $\mathbf{y}$, le modèle des observations est :

$$\mathbf{y} = \mathbf{h}(\mathbf{z} \otimes \boldsymbol{\mu}) + \mathbf{b} \tag{5.12}$$

où $\otimes$ est le produit de Kronecker, $\mathbf{b} \in \mathbb{R}^{\Lambda|\mathcal{S}|}$ représente le bruit et $\mathbf{h} \in \mathbb{R}^{\Lambda|\mathcal{S}|\times\Lambda|\mathcal{S}|}$ est l'opérateur de convolution :

$$\mathbf{h} = \mathbf{I}_\Lambda \otimes \mathbf{f}; \tag{5.13}$$

où $\mathbf{f} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ représente la fonction d'étalement spatial.

Pour restreindre la convolution à une région de $3 \times 3$ sites (c'est-à-dire $\{s\} \cup N_s$), il faut rendre chaque colonne de $\mathbf{f}$ 9-parcimonieuse [3]. Dans le cadre du modèle présenté, $\mathbf{f}$ ne contient que ces 9 coefficients.

Nous supposons que $\mathbf{b}$ est un bruit normal multivarié, de matrice de covariance $\boldsymbol{\Xi} \in \mathbb{R}^{\Lambda|\mathcal{S}|\times\Lambda|\mathcal{S}|}$. Nous supposons que le bruit n'est pas corrélé entre sites. La matrice $\boldsymbol{\Xi}$ est donc bloc-diagonale :

$$\boldsymbol{\Xi} = \mathbf{I}_{|S|} \otimes \boldsymbol{\Sigma}; \tag{5.14}$$

où $\boldsymbol{\Sigma}$ est la matrice de covariance introduite en (5.9).

Nous disposons de l'expression de la vraisemblance du modèle :

$$p(\mathbf{y}|\mathbf{z}) \sim \mathcal{N}(\mathbf{h}(\mathbf{z} \otimes \boldsymbol{\mu}), \boldsymbol{\Xi}) \tag{5.15}$$

Nous supposons de plus que $p(\mathbf{Z})$ est une distribution de champs de Markov régulée par la distribution (5.4). L'expression de cet *a priori* et de la vraisemblance permet enfin de calculer la distribution *a posteriori* $p(\mathbf{z}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z})$. Ainsi, le problème de segmentation par CMco présenté dans ce chapitre peut être assimilé à un problème de déconvolution bayésienne d'images hyperspectrales, dans le cas où l'inconnue $\mathbf{z}$ est à valeurs dans $\{0, 1\}$ et a une distribution de champs de Markov.

2. C'est-à-dire $z_s = \underset{i\in\{0,1\}}{\arg}\ (x_s = \omega_i)$.
3. Cela revient le plus souvent à ne conserver que les 9 termes de $\mathbf{f}$ les plus élevés.

**Remarque 5.2.3.** *Cette formulation est directement généralisable à la présence de $K > 2$ classes, traduisant par exemple des niveaux d'atténuation (cf. remarque 5.2.2).*

### 5.2.5 Estimation des paramètres

L'ensemble des paramètres du modèle présenté dans les sections 5.2.1 et 5.2.3 est :

$$\theta = \{\mathbf{f}, \boldsymbol{\mu}, \sigma, \rho_1, \rho_2, \alpha\}\,. \tag{5.16}$$

Nous considérons $\mathbf{f}$ connu en amont de la segmentation : dans le cas de la détection de sources convoluées, il s'agit d'un paramètre de l'instrument qui peut être connu *a priori*.

Nous présentons tout d'abord les estimateurs basés sur les données complètes $(\mathbf{x}, \mathbf{y})$, nécessaires ensuite pour l'estimation non supervisée. $\alpha$ est estimé au sens des moindres carrés d'après l'estimateur de [Derin et Elliott, 1987]. Soit le processus $\mathbf{x}^f$, résultant de la convolution de $\mathbf{x}$ par $\mathbf{f}$ :

$$\mathbf{x}^f = \mathbf{x} * \mathbf{f}. \tag{5.17}$$

Ce processus prends valeurs en chaque site $s$ : nous avons $\mathbf{x}^f = (x_s^f)_{s\in\mathcal{S}}$ avec $x_s^f = (\mathbf{x} * \mathbf{f})_s$.

L'estimateur de $\boldsymbol{\mu}$ au sens du maximum de vraisemblance s'écrit comme :

$$\hat{\boldsymbol{\mu}} = \frac{\sum\limits_{s\in\mathcal{S}} x_s^f \mathbf{y}_s}{\sum\limits_{s\in\mathcal{S}} (x_s^f)^2}. \tag{5.18}$$

*Démonstration.* La vraisemblance s'écrit :

$$\begin{aligned}\ell(\mathbf{x}, \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{s\in\mathcal{S}} g_{x_s, \mathbf{x}_{N_s}}(\mathbf{y}_s) \\ &= \frac{1}{((2\pi)^\Lambda \det(\boldsymbol{\Sigma}))^{|\mathcal{S}|/2}} \prod_{s\in\mathcal{S}} \exp\left(-\frac{1}{2}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})\right).\end{aligned} \tag{5.19}$$

Par commodité, nous employons la log-vraisemblance pour les calculs :

$$\log \ell(\mathbf{x}, \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{|\mathcal{S}|}{2} \log\left((2\pi)^\Lambda \det(\boldsymbol{\Sigma})\right) - \frac{1}{2}\sum_{s\in\mathcal{S}} (\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_s - x_s^f\boldsymbol{\mu}). \tag{5.20}$$

En notant $\hat{\boldsymbol{\mu}}$ l'estimation de $\boldsymbol{\mu}$ au sens du maximum de vraisemblance, nous avons :

$$\left.\frac{\partial \log \ell(\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}} = 0. \tag{5.21}$$

Détaillons l'expression de cette dérivée partielle :

$$\begin{aligned}\frac{\partial \log \ell(\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\mu}}\left(\sum_{s\in\mathcal{S}} (\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})\right) \\ &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\mu}}\left(\sum_{s\in\mathcal{S}} \mathbf{y}_s^\top \boldsymbol{\Sigma}^{-1}\mathbf{y}_s - 2x_s^f\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\mathbf{y}_s + (x_s^f)^2\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) \\ &= -\frac{1}{2}\sum_{s\in\mathcal{S}} -2x_s^f\boldsymbol{\Sigma}^{-1}\mathbf{y}_s + 2(x_s^f)^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.\end{aligned} \tag{5.22}$$

En conséquence, nous pouvons réécrire (5.21) comme :

$$\begin{aligned}\sum_{s\in\mathcal{S}} 2x_s^f \boldsymbol{\Sigma}^{-1}\mathbf{y}_s &= \sum_{s\in\mathcal{S}} 2(x_s^f)^2\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}\\ \sum_{s\in\mathcal{S}} x_s^f \mathbf{y}_s &= \sum_{s\in\mathcal{S}} (x_s^f)^2\hat{\boldsymbol{\mu}}\\ \hat{\boldsymbol{\mu}} &= \frac{\sum\limits_{s\in\mathcal{S}} x_s^f\mathbf{y}_s}{\sum\limits_{s\in\mathcal{S}}(x_s^f)^2};\end{aligned} \tag{5.23}$$

où le passage de la première à la deuxième ligne se fait en multipliant à gauche par $\boldsymbol{\Sigma}$. □

Nous pouvons en déduire une estimation de la matrice de covariance $\boldsymbol{\Sigma}$ au sens du maximum de vraisemblance :

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{|\mathcal{S}|-1}\sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\hat{\boldsymbol{\mu}})(\mathbf{y}_s - x_s^f\hat{\boldsymbol{\mu}})^\top \tag{5.24}$$

*Démonstration.* Soit $\hat{\boldsymbol{\Sigma}}$ l'estimation de $\boldsymbol{\Sigma}$ au sens du maximum de vraisemblance, vérifiant :

$$\left.\frac{\partial \log \ell(\mathbf{x},\mathbf{y},\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial\boldsymbol{\Sigma}}\right|_{\boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}} = 0. \tag{5.25}$$

D'après (5.20), en notant $c = -\Lambda|\mathcal{S}|\log(2\pi)$, cette dérivée s'écrit comme :

$$\begin{aligned}\frac{\partial \log \ell(\mathbf{x},\mathbf{y},\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial\boldsymbol{\Sigma}} &= \frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}}\left(c - |\mathcal{S}|\log\left[\det(\boldsymbol{\Sigma})\right] - \sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})\right)\\ &= \frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}}\left(c + |\mathcal{S}|\log\left[\det(\boldsymbol{\Sigma}^{-1})\right] - \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top(\mathbf{y}_s - x_s^f\boldsymbol{\mu})\right]^\top\right)\\ &= \frac{|\mathcal{S}|}{2}\boldsymbol{\Sigma} - \frac{1}{2}\left(\sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top(\mathbf{y}_s - x_s^f\boldsymbol{\mu})\right)^\top.\end{aligned} \tag{5.26}$$

Nous pouvons donc, à partir de (5.25), écrire que

$$\begin{aligned}\frac{|\mathcal{S}|}{2}\hat{\boldsymbol{\Sigma}} &= \frac{1}{2}\sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})(\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top\\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\boldsymbol{\mu})(\mathbf{y}_s - x_s^f\boldsymbol{\mu})^\top.\end{aligned} \tag{5.27}$$

Comme nous employons à la place de $\boldsymbol{\mu}$ son estimée $\hat{\boldsymbol{\mu}}$, l'estimateur du maximum de vraisemblance non biaisé pour la covariance est :

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{|\mathcal{S}|-1}\sum_{s\in\mathcal{S}}(\mathbf{y}_s - x_s^f\hat{\boldsymbol{\mu}})(\mathbf{y}_s - x_s^f\hat{\boldsymbol{\mu}})^\top. \tag{5.28}$$

□

Comme nous pouvons écrire $\Sigma = f(\sigma^2,\rho_1,\rho_2)$, avec $f$ une fonction bijective, inverser cette relation permet d'obtenir l'estimateur de $\sigma^2,\rho_1,\rho_2$ au sens du maximum de vraisemblance [4]. L'estimation des paramètres $\sigma^2$, $\rho_1$ et $\rho_2$ est alors déduite de $\hat{\boldsymbol{\Sigma}}$ de la manière

4. L'estimateur du maximum de vraisemblance bénéficie en effet de la propriété d'invariance fonctionnelle : si $\hat{\theta}$ est l'estimation de $\theta$ au sens du maximum de vraisemblance, alors $f(\hat{\theta})$ est l'estimateur de $f(\theta)$ au sens du maximum de vraisemblance.

suivante :

$$\hat{\sigma}^2 = \frac{1}{\Lambda}\sum_{i=1}^{\Lambda}\hat{\mathbf{\Sigma}}_{i,i}; \tag{5.29}$$

$$\hat{\rho}_1 = \frac{1}{\Lambda-1}\sum_{i=1}^{\Lambda-1}\hat{\mathbf{\Sigma}}_{i,i+1}; \tag{5.30}$$

$$\hat{\rho}_2 = \frac{1}{\Lambda-2}\sum_{i=1}^{\Lambda-2}\hat{\mathbf{\Sigma}}_{i,i+2}. \tag{5.31}$$

L'ensemble des paramètres peut ainsi être estimé à partir des données complètes $(\mathbf{x}, \mathbf{y})$. En contexte non supervisé, nous employons un algorithme inspiré de l'algorithme SEM [Celeux et Diebolt, 1986] (cf. section 4.2.3). L'algorithme est itératif, et se formule à chaque itération $q$ comme :

1. la simulation de $\hat{\mathbf{x}}^q$ selon $p_{\theta_{q-1}}(\mathbf{X}|\mathbf{Y}=\mathbf{y})$ ;
2. l'estimation des paramètres grâce aux estimateurs sur les données complètes simulées $(\hat{\mathbf{x}}^q, \mathbf{y})$.

**Remarque 5.2.4.** *Cet algorithme n'est pas un algorithme SEM. En effet, celui-ci repose sur des estimateurs au sens du maximum de vraisemblance uniquement dans l'étape 2. À notre connaissance, il n'existe pas d'estimateurs au sens du maximum de vraisemblance pour $\alpha$. C'est pourquoi nous avons employé l'estimateur de [Derin et Elliott, 1987].*

**Remarque 5.2.5.** *Cet algorithme est en revanche un cas particulier d'un algorithme ICE (cf. section 4.2.3), dans lequel l'espérance (4.13) est calculée avec une unique réalisation de l'estimateur des paramètres.*

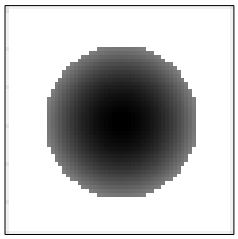
## 5.3 Résultats numériques

### 5.3.1 Configuration expérimentale

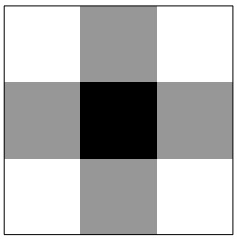
Nous nous plaçons dans un cadre expérimental semblable à celui présenté dans le chapitre 3. Nous nous donnons une image hyperspectrale comportant $60 \times 60$ spectres et $\Lambda = 10$ bandes spectrales. Les intensités lumineuses sont construites à partir d'une gaussienne bi-dimensionnelle tronquée à mi-hauteur, de sorte à avoir des intensités nulles en périphérie de l'objet (cf. figure 5.2a). Ces dernières sont associées à $\omega_0$, et les intensités non nulles sont associées à $\omega_1$. Cette carte d'intensité subit une convolution par la FSF, qui prend la forme d'une fonction de Moffat (cf. chapitre 3) sur $3 \times 3$ pixels et est illustrée en figure 5.2b. Nous choisissons $\boldsymbol{\mu}$ de sorte avoir 3 coefficients non nuls parmi 10 (cf. figure 5.2e).

Nous évaluons dans ce contexte trois méthodes différentes :

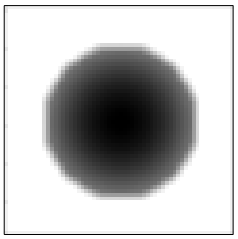
1. la segmentation par modèle de champs de Markov couples convolutifs (CMco) présentée dans ce chapitre, avec le critère du MAP (5.6) et du MPM (5.7) et dans le cas binaire, c'est-à-dire avec $K = 1$ ;
2. la détection par tests introduite dans le chapitre 3, décrite dans l'algorithme 1. Nous nous plaçons dans la même configuration que dans le chapitre 3 pour le choix du catalogue $\mathbf{c}$ et de la $P_{\text{FA}}$ cible (fixée à $0{,}05$), et utilisons une FSF de Moffat de $3^2$ pixels employée dans les simulations ;
3. la détection par tests d'hypothèses proposée dans [Bacher et al., 2017b] et étendue dans [Bacher et al., 2017a]. Dans ces travaux, les auteurs utilisent un prétraitement par filtrage adapté pour prendre en compte la structure spatiale des images, et



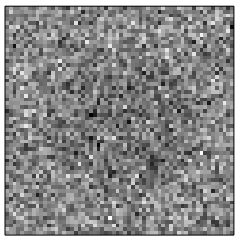
(a) Carte des intensités en gaussienne 2D tronquée.



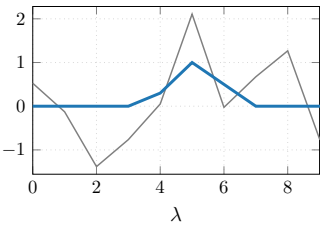
(b) FSF employée dans les expériences.



(c) Carte des intensités convoluée par la FSF



(d) Moyenne spectrale d'une réalisation de $\mathbf{y}$.




(e) Spectre d'origine (bleu) et réalisation de spectre $\mathbf{y}_s$ (gris).

Figure 5.2 – Illustration de la formation des images de synthèse. Les images sont représentées en inverse vidéo : le noir représente des intensités lumineuses élevées.

les erreurs sont gérées globalement par un contrôle de FDR prenant en compte la connexité spatiale des détections. Nous employons là aussi pour FSF une fonction de Moffat de $3^2$ pixels et choisissons un paramètre de FDR valant $0{,}1$.

Ces deux dernières méthodes traitent les pixels d'un point de vue local : l'information concernant la dimension spatiale n'est prise en compte qu'avec des informations contextuelles. Une étude préliminaire a montré que la segmentation par un modèle « standard » de champ de Markov caché à bruit indépendant n'est pas compétitif par rapport aux trois méthodes évaluées.

### 5.3.2 Résultats comparatifs

Les performances sont qualifiées selon trois critères :
— le taux d'erreur, c'est-à-dire la proportion de classifications erronées ;
— le taux de fausse alarme (faux positifs ou erreurs de type I) : la proportion de décisions à $\omega_1$ alors que la vérité terrain vaut $\omega_0$ ;
— le taux de faux négatif (ou erreur de type II) : la proportion de résultats à $\omega_0$ alors que la vérité terrain vaut $\omega_1$.

Les performances sont évaluées avec un RSB variable, défini par :

$$\mathrm{RSB} = 10\log_{10}\left(\frac{\|\overline{\boldsymbol{\mu}}\|_2^2}{\mathrm{Tr}(\boldsymbol{\Sigma})}\right) = 10\log_{10}\left(\frac{\|\overline{\boldsymbol{\mu}}\|_2^2}{\Lambda\sigma^2}\right); \tag{5.32}$$

où $\overline{\boldsymbol{\mu}}$ est la moyenne des spectres associés à $\omega_1$ en l'absence de bruit.

Un exemple de résultats obtenus est présenté en figure 5.3. Ces résultats illustrent la robustesse des trois modèles considérés à faible RSB. De plus, la carte d'incertitude $\mathbf{u}^x$ associée à la segmentation au sens du MPM donne une illustration claire des régions de changement de stationnarité dans les estimations de $\mathbf{x}$. Les expériences sont répétées avec

100 réalisations différentes de $\mathbf{Y} = \mathbf{y}$, afin d'évaluer les erreurs susceptibles de se produire en moyenne, et leur variabilité.

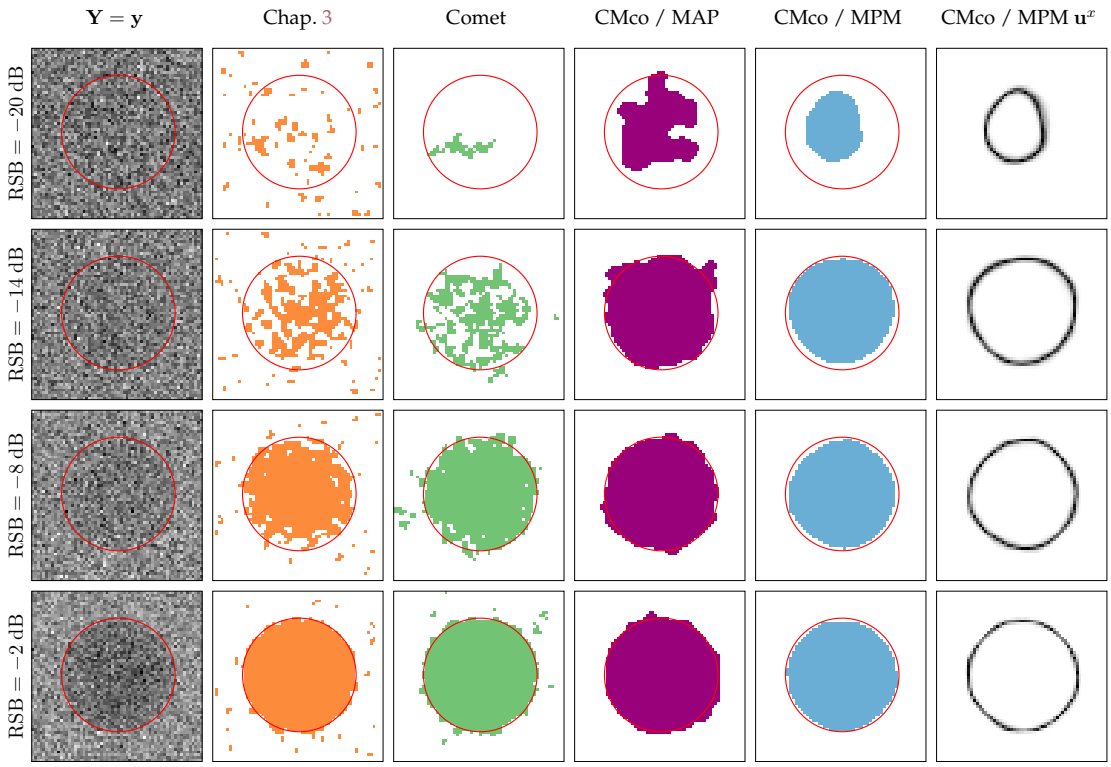
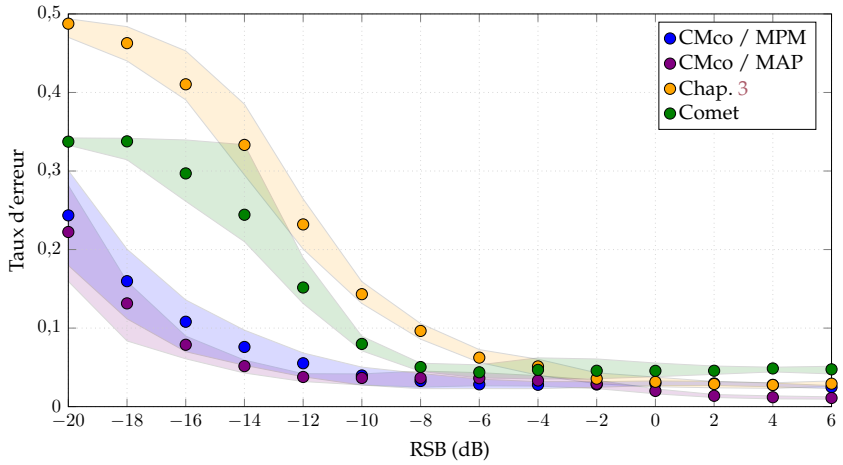



Figure 5.3 – Exemples de résultats avec des RSB valant $-20$, $-14$, $-8$ et $-2$ dB avec la méthode présentée dans le chapitre 3 [Courbot et al., 2017a], avec la méthode de [Bacher et al., 2017a] intitulée « Comet » et avec la segmentation par champs de Markov couples convolutifs. Les cercles rouges représentent le contour de la vérité terrain (cf. figure 5.2a). Remarquons les bons résultats obtenus dans le cadre du modèle CMco, en particulier à faible RSB.

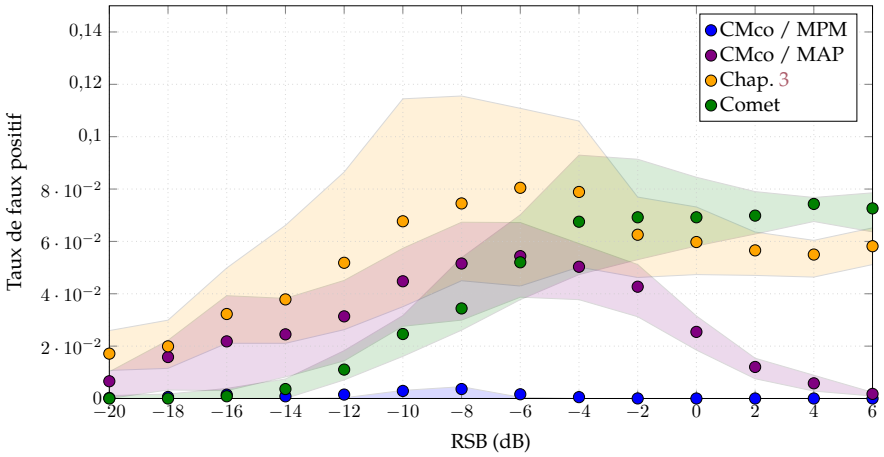
Les résultats sont reportés en figure 5.4. Plusieurs éléments peuvent être déduits de l'évolution des taux d'erreurs, de faux positif et de faux négatifs :

— la segmentation au sein du modèle CMco apporte, à tout RSB, le meilleur taux d'erreur par rapport aux deux alternatives. De plus, les segmentations au sens du MAP et du MPM donnent, en moyenne, des taux d'erreurs très semblables ;
— en termes de taux de faux positif et de faux négatifs, ces deux segmentations dans le modèle CMco ont des comportements bien distincts. En effet, la segmentation au sens du MAP permet de manquer peu de classifications dans $\omega_1$ mais conserve un taux de faux positif non négligeable. À l'opposé, la segmentation au sens du MPM permet de n'avoir presque aucune fausse alarme, au prix d'un faible taux de faux négatif.
— la méthode proposée par [Bacher et al., 2017a] montre en revanche des taux d'erreurs qui restent relativement élevés lorsque $\mathrm{RSB} < -8$ dB. Ceux-ci sont dus à des faux négatifs : la méthode décide en excès des non-détections à faible RSB, ce qui permet un contrôle du taux de fausse découverte [5].
— une tendance de croissance du taux de fausse alarme en fonction du RSB peut être visible dans les résultats, à l'exception de la méthode basée sur le MPM dans le cadre

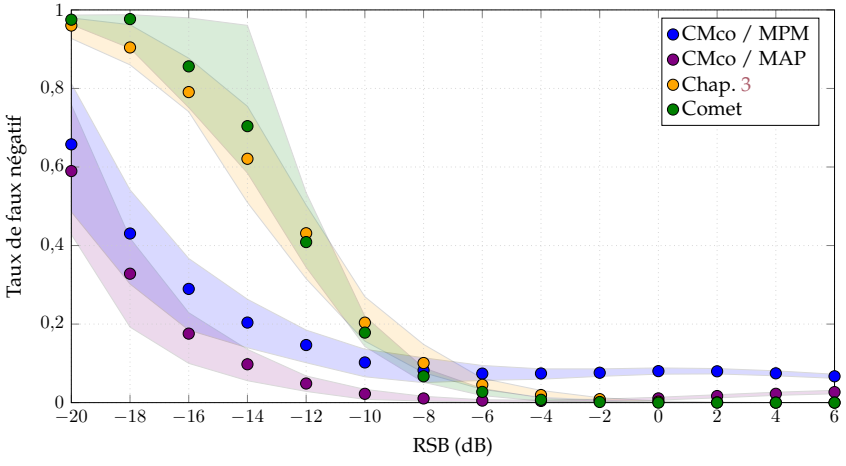
5. Les expériences montrent en effet un taux de fausse découverte inférieur ou égal à 0,1, ce qui correspond au paramètre fixé en entrée.




(a) Taux d'erreur.




(b) Taux de faux positif.




(c) Taux de faux négatif.

Figure 5.4 – Performances moyennes mesurés sur 100 réalisations de $\mathbf{Y} = \mathbf{y}$ pour la méthode présentée dans le chapitre 3 [Courbot et al., 2017a], de la méthode « Comet » de [Bacher et al., 2017a] et avec la segmentation par champs de Markov convolutifs au sens du MAP (5.6) et du MPM (5.7). Chaque série de point représente les résultats moyens, et les régions de couleurs associées sont délimitées par les premiers et troisièmes quartiles des valeurs.

du modèle proposé. Cela peut être interprété comme une influence trop grande des classifications dans $\omega_1$ sur leur voisinage. Ce phénomène peut être observé en figure 5.3, dans laquelle un excès de classification dans $\omega_1$ peut être observé dans la périphérie immédiate de la région à détecter.

Ces résultats donnent des éléments de compréhension sur le comportement des méthodes évaluées, que nous pouvons résumer comme suit :

1. la méthode introduite dans le chapitre 3 présente quelques limites. En effet, le taux d'erreur diminue très peu lorsque $\mathrm{RSB} > -12\ \mathrm{dB}$, et ces erreurs consistent surtout en des faux positifs ;
2. la méthode de [Bacher et al., 2017a] présente un très faible taux de fausse alarme à $\mathrm{RSB} < -8\ \mathrm{dB}$, permettant un contrôle du FDR au prix d'une disproportion de faux négatifs par rapport aux faux positifs. À $\mathrm{RSB} \geq -8\ \mathrm{dB}$ RSB, elle est en revanche plus performante que la méthode proposée dans le chapitre 3 ;
3. les segmentations basées sur le modèle CMco montrent des performances particulièrement intéressantes. En effet, dans le cadre de la segmentation au sens du MAP, le taux d'erreur reste inférieur à $0{,}1$ jusqu'à $\mathrm{RSB} = -14\ \mathrm{dB}$ environ, tout en permettant un faible taux de faux négatif. Cette méthode semble fournir le plus faible taux d'erreur par rapport aux autres méthodes évaluées. La segmentation au sens du MPM a quant à elle un taux d'erreur légèrement plus important que son homologue. En revanche, la plupart de ces erreurs sont des faux négatifs, car la segmentation offre un taux de fausse alarme inférieur à $0{,}03$ quel que soit le RSB.

## 5.4 Conclusion

Dans ce chapitre, nous nous sommes intéressés au problème de détection sous l'angle de la segmentation bayésienne des images. La modélisation par champs de Markov a permis de proposer une méthode de segmentation prenant en compte l'homogénéité spatiale au sein de l'image entière. C'est sur ce point que le modèle proposé diffère de ses alternatives, qui ne prennent l'information de voisinage en compte que localement, conduisant à des méthodes de détection contextuelles. La modélisation par CMco a de plus permis de prendre en compte le phénomène de convolution pouvant intervenir au sein des images.

Les résultats numériques ont montré l'intérêt de cette modélisation, en mettant en évidence l'efficacité par rapport aux alternatives des segmentations basées sur le modèle CMco. Les résultats obtenus dans ce cadre sont de plus complémentaires : le MAP permet de faibles taux d'erreurs, et le MPM offre des taux de fausse alarme très faibles. Des résultats complémentaires sur données synthétiques et sur des images issues de l'instrument MUSE sont présentés dans le chapitre 10.

# 6

# Champs de Markov triplets orientés



## 6.1 Introduction

Dans ce chapitre, nous nous intéressons à la segmentation de textures orientées au sein d'images, avec pour problématique applicative celle de la segmentation de filaments formant des structures homogènes à grande échelle dans les images MUSE. Ce chapitre se place dans le contexte plus général de la segmentation d'images, et nous nous intéressons à une modélisation conjointe des classes et des orientations, afin de pouvoir les retrouver simultanément à partir d'une image observée.

### 6.1.1 Contexte de travail

Hormis les travaux de modélisation de champs de Markov prenant en compte les effets de bords [August et Zucker, 2003; Descombes et al., 1995; Smits et Dellepiane, 1997], peu de travaux de la littérature proposent une modélisation markovienne des structures orientées. Les travaux les plus adaptés à ce problème sont présentés dans [Dass, 2004]. Dans cet article, les auteurs proposent une modélisation par « champs directionnels » pour retrouver les orientations seules au sein d'images d'empreintes digitales.

La littérature présente plusieurs travaux portant sur la recherche d'orientation dans les images, basés sur des méthodes relevant du domaine de la vision par ordinateur. Mentionnons par exemple des méthodes de filtrage [Rigamonti et Lepetit, 2012], de chemin minimal

sur graphe [Benmansour et Cohen, 2011], ou de coupe de graphe (*graph cut*) [Bauer et al., 2010].

Nous introduisons dans ce chapitre une nouvelle modélisation markovienne qui permet de prendre en compte les caractéristiques d'orientation dans les images. Plus précisément, nous présentons dans ce chapitre :

— la distribution du modèle ;
— un algorithme de type SEM introduit pour estimer les paramètres du triplet $(\mathbf{X}, \mathbf{Y}, \mathbf{V})$ à partir de la seule réalisation $\mathbf{Y} = \mathbf{y}$ ;
— deux méthodes pour restaurer $\mathbf{X}$ et $\mathbf{V}$ à partir de $\mathbf{Y} = \mathbf{y}$ uniquement ;
— une carte de confiance complémentaire de la segmentation ;
— les résultats numériques sur des images de synthèse, permettant de mesurer les performances de segmentation et de les comparer avec la segmentation par champs de Markov cachés ;
— l'application à des images réelles.

La section 6.1.2 présente la segmentation dans le cadre des modèles de champs de Markov triplets. Nous présentons ensuite en section 6.2 le modèle de champ de Markov triplet *orienté* (CMTO), conçu pour tenir compte des orientation locales. Cette partie permet, de plus, la description de méthodes de segmentation, du critère de confiance associable aux segmentations, et d'estimation des paramètres. Le détail des procédures implémentées est reporté en annexe A.

Les résultats expérimentaux sont enfin décrits dans la section 6.3 à l'aide d'images de synthèse et d'images réelles. Des résultats complémentaires sont également présentés dans le chapitre 10 et en annexe B.1.

Rappelons quelques notations employées dans ce chapitre :

— l'ensemble des sites $s$ de l'image est noté $\mathcal{S}$, et nous avons le processus d'observation $\mathbf{Y} = (Y_s)_{s\in\mathcal{S}}$, le processus de classes $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$, et le processus auxiliaire $\mathbf{V} = (V_s)_{s\in\mathcal{S}}$. Le processus triplet est $\mathbf{T} = (\mathbf{T}_s)_{s\in\mathcal{S}}$, avec $T_s = (Y_s, X_s, V_s)$ ;
— nous nous donnons un système de voisinage $(N_s)_{s\in\mathcal{S}}$. Nous considérons ici que $N_s$ est l'ensemble des 8 voisins de $s$.
— $\mathcal{C}$ désigne l'ensemble des cliques $c$ de l'image, une clique étant soit un ensemble de sites mutuellement voisins, soit un singleton. Nous noterons $\mathbf{x}_c$ le sous-ensemble de $\mathbf{x}$ indexé par la clique $c$, et de même pour $\mathbf{v}$, $\mathbf{y}$ et $\mathbf{t}$.

Nous considérons de plus que $X_s$ et $V_s$ sont à valeurs discrètes dans $\Omega_x$ et $\Omega_v$ respectivement, et que $Y_s \in \mathbb{R}$.

### 6.1.2 Segmentation bayésienne par champs de Markov triplets

Les modèles de champs de Markov triplets permettent, comme leurs homologues en chaîne ou en arbre, l'introduction d'un processus auxiliaire $\mathbf{V}$. Ce processus est inobservable et doit, dans le cadre de la segmentation bayésienne, être déduit d'une observation $\mathbf{Y} = \mathbf{y}$ seule, au même titre que $\mathbf{X}$.

Dans un modèle de champ de Markov triplet, $\mathbf{T} = (\mathbf{Y}, \mathbf{X}, \mathbf{V})$ est un champ de Markov. Grâce au théorème de Hammersley-Clifford [Besag, 1974], sa distribution peut être écrite comme une distribution de Gibbs :

$$p(\mathbf{t}) \triangleq p(\mathbf{y}, \mathbf{x}, \mathbf{v}) = \gamma \exp\left[-\sum_{c\in\mathcal{C}} \phi(\mathbf{y}_c, \mathbf{x}_c, \mathbf{v}_c)\right] ; \tag{6.1}$$

où $\phi$ est une fonction « potentiel » et $\gamma$ est la constante de normalisation qui est inconnue dans le cas général (cf. section 4.2.3). La propriété de markovianité *a posteriori* permet, dans

le cadre de la segmentation, de contourner cette limitation, et cette constante est souvent omise pour écrire :

$$p(\mathbf{t}) \propto \exp\left[-\sum_{c\in\mathcal{C}} \phi(\mathbf{y}_c, \mathbf{x}_c, \mathbf{v}_c)\right]. \tag{6.2}$$

Nous pouvons ici constater la généralisation par rapport aux modèles de champs de Markov cachés et de champs de Markov couples. Par exemple, en notant $\phi'$ et $\phi''$ des fonctions de potentiel :

— $\mathbf{T}$ est un champ de Markov couple lorsque $\forall c \in \mathcal{C}$ :

$$\phi(\mathbf{y}_c, \mathbf{x}_c, \mathbf{v}_c) = \phi'(\mathbf{y}_c, \mathbf{x}_c); \tag{6.3}$$

c'est-à-dire lorsque le processus auxiliaire $\mathbf{V}$ n'est pas exploité ;

— $\mathbf{T}$ est un champ de Markov caché à bruit indépendant lorsque :

$$p(\mathbf{t}) \propto \exp\left[-\sum_{c\in\mathcal{C}} \phi''(\mathbf{x}_c) - \sum_{s\in\mathcal{S}} \log(f(x_s, y_s))\right]. \tag{6.4}$$

où $f$ est une fonction paramétrique. $\mathbf{V}$ est là aussi ignoré, et des hypothèses d'indépendance conditionnelle plus restrictives que dans le modèle couple interviennent (cf. section 4.3.1).

Le modèle de champ de Markov triplet bénéficie de la propriété de markovianité *a posteriori* : à $\mathbf{Y} = \mathbf{y}$ fixé, la distribution $p(\mathbf{x}, \mathbf{v}|\mathbf{y})$ est de Markov. Cela permet la simulation de réalisations de $(\mathbf{X}, \mathbf{V})$ à partir de l'observation $\mathbf{Y} = \mathbf{y}$ seule, grâce à la connaissance des fonctions $\phi$ et de leurs paramètres, par exemple à l'aide d'un échantillonneur de Gibbs.

**Remarque 6.1.1.** *En vertu du théorème de Hammersley-Clifford, la possibilité de simuler un champ de Markov triplet repose sur la connaissance de la fonction $\phi$ ou de manière équivalente sur la connaissance des distributions $p(x_s, v_s, y_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, \mathbf{y}_{N_s})$.*

Comme des simulations selon $p(\mathbf{x}, \mathbf{v}|\mathbf{y})$ peuvent être réalisées, la segmentation est possible. Au sens du MAP, elle s'écrit [Geman et Geman, 1984] :

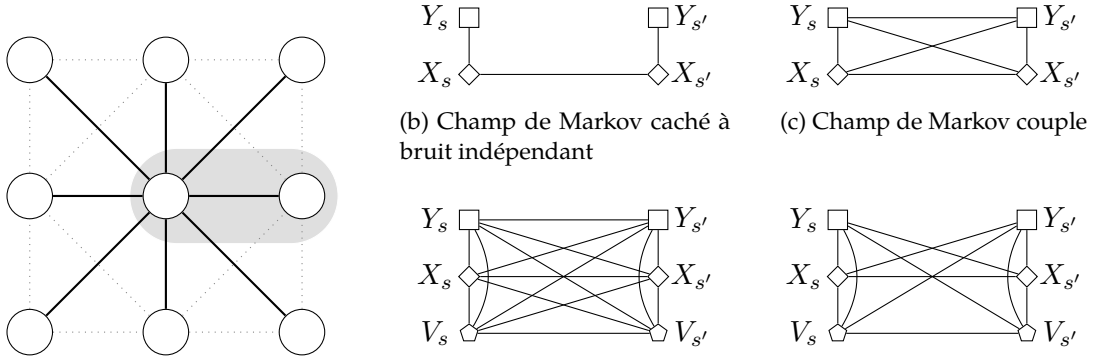
$$(\hat{\mathbf{x}}, \hat{\mathbf{v}})^{\mathrm{MAP}} = \underset{(\boldsymbol{\omega},\boldsymbol{\nu})\in(\Omega_x\times\Omega_v)^{|\mathcal{S}|}}{\arg\max}\ p(\mathbf{X} = \boldsymbol{\omega}, \mathbf{V} = \boldsymbol{\nu}|\mathbf{Y} = \mathbf{y}); \tag{6.5}$$

et la segmentation au sens du MPM [Marroquin et al., 1987] est $s \in \mathcal{S}$ :

$$(\hat{x}_s, \hat{v}_s)^{\mathrm{MPM}} = \underset{(\omega,\nu)\in\Omega_x\times\Omega_v}{\arg\max}\ p(X_s = \omega, V_s = \nu|\mathbf{Y} = \mathbf{y}). \tag{6.6}$$

Les distributions $p(\mathbf{x}, \mathbf{v}|\mathbf{y})$ et $p(x_s, v_s|\mathbf{y})$ peuvent être simulées par des méthodes de Monte-Carlo par chaînes de Markov (MCMC) . De plus, la segmentation au sens de l'estimateur du MAP peut être approchée par l'algorithme *iterated conditional modes* (ICM) [Besag, 1986], et la segmentation par le MPM peut être obtenue en adaptant l'algorithme de Marroquin [Marroquin et al., 1987]. Nous utiliserons dans ce chapitre ces deux derniers algorithmes, dont l'implémentation est détaillée en annexe A.2.

En contexte non supervisé, les paramètres régissant la distribution $p(\mathbf{t})$ sont inconnus. Il faut donc recourir à des méthodes dédiées pour les estimer en amont de la segmentation. L'adjonction du processus auxiliaire rend le problème plus difficile que dans le cas d'un modèle couple ou caché, car les paramètres à estimer sont plus nombreux. Les paramètres sont estimés grâce à un algorithme inspiré de la méthode SEM, décrite en section 6.2.3 et dont l'implémentation algorithmique est présentée en annexe A.3.1.




(a) (b) Champ de Markov caché à bruit indépendant (c) Champ de Markov couple (d) Champ de Markov triplet (e) Champs de Markov triplet orienté

Figure 6.1 – Graphe de dépendances pour un champ de Markov. (a) Dépendances d'un champ de Markov au sein des 8 voisins d'un site $s$. Les connections en pointillés représentent les liens ne concernant pas $s$ directement. (b-e) Sous-graphes pour les liens entre le site central et un site voisin $s'$ (région grisée de (a)) dans les modèles de champ de Markov caché à bruit indépendant, couple, triplet et triplet orienté.

## 6.2 Champs de Markov triplets orientés

Dans cette partie, nous présentons un modèle de champ de Markov triplet *orienté* (CMTO) conçu pour modéliser les orientations dans les images. Un graphe de dépendances illustre les liens retenus dans la figure 6.1, et la comparaison avec les modèles cachés, couples et triplets.

### 6.2.1 Modèle

Soit $\mathbf{T} = (\mathbf{Y}, \mathbf{X}, \mathbf{V})$ un champ de Markov triplet stationnaire (6.2), où $\mathbf{V}$ modélise les orientations privilégiées au sein de $\mathbf{X}$. Nous avons vu que comme $\mathbf{T}$ est de Markov, sa distribution est caractérisée en chaque site $s$ par $p(\mathbf{t}_s|\mathbf{t}_{N_s})$. Ainsi, pour décrire ce modèle, nous détaillons dans cette section la distribution $p(\mathbf{t}_s|\mathbf{t}_{N_s})$.

Nous supposons que $\mathbf{T}_s$ et $\mathbf{Y}_{N_s}$ sont indépendants conditionnellement à $(\mathbf{X}_{N_s}, \mathbf{V}_{N_s})$. Nous pouvons donc écrire :

$$\begin{aligned} p(\mathbf{t}_s|\mathbf{t}_{N_s}) &= p(\mathbf{t}_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}) \\ &= p(y_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, x_s, v_s)p(x_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, v_s)p(v_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}). \end{aligned} \tag{6.7}$$

Nous supposons de plus que $V_s$ et $\mathbf{X}_{N_s}$ sont indépendants conditionnellement à $\mathbf{V}_{N_s}$ : l'orientation $V_s$ en un site $s$ ne dépend pas directement du voisinage de $X_s$. Le troisième terme de (6.7) s'écrit alors :

$$p(v_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}) = p(v_s|\mathbf{v}_{N_s}). \tag{6.8}$$

Nous supposons de plus que $X_s$ ne dépend directement que de son voisinage $X_{N_s}$ et de $V_s$. Cela revient à dire que $X_s$ et $\mathbf{V}_{N_s}$ sont indépendants conditionnellement à $(\mathbf{X}_{N_s}, V_s)$. En conséquence, le deuxième terme de (6.7) s'écrit :

$$p(x_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, v_s) = p(x_s|\mathbf{x}_{N_s}, v_s). \tag{6.9}$$

Ces hypothèses d'indépendance conditionnelle sont représentées en figure 6.1e.

Nous détaillons maintenant les distributions (6.8), (6.9) ainsi que le premier terme de (6.7) :

— $\mathbf{V}$ a une distribution de champ de Markov. La distribution (6.8) s'écrit à l'aide d'un potentiel de Potts :

$$p(v_s|\mathbf{v}_{N_s}) \propto \exp\left(-\alpha \sum_{s' \in N_s} [1 - 2\delta_{v_s}^{v_{s'}}]\right); \tag{6.10}$$

où $\alpha$ est un paramètre du modèle et $\delta_{v_s}$ est la mesure de Dirac pour $v_s$.

— $p(\mathbf{x}|\mathbf{v})$ est une distribution de champ de Markov. La distribution (6.9) s'écrit :

$$p(x_s|\mathbf{x}_{N_s}, v_s) \propto \exp\left(-\beta \sum_{s' \in N_s} \varphi^{s'}(v_s)[1 - 2\delta_{x_s}^{x_{s'}}]\right). \tag{6.11}$$

où $\beta$ est un paramètre du modèle, et $\varphi^{s'}$ est une *fonction d'orientation* qui permet de gérer les orientations au sein d'un voisinage local. Dans le cas d'un $8-$voisinage, $N_s = \{0, \ldots, 7\}$ (cf. figure 6.2). En considérant que les cliques d'ordre 2 ont 4 orientations différentes, celles-ci peuvent être prises en compte par :

$$\varphi^k(v) = \left|\cos\left(v - \frac{k\pi}{4}\right)\right|; \tag{6.12}$$

où $k \in N_s$.

**Remarque 6.2.1.** *En pratique, l'ensemble des $\Omega_v$ dans lequel les $v_s$ prennent valeurs est une subdivision de $[0, \pi]$ de la forme $\{\frac{\pi}{12}, \frac{3\pi}{12}, \ldots, \frac{11\pi}{12}\}$ (cf. section 6.3).*

— pour fournir une modélisation plus intuitive de $p(y_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, x_s, v_s)$, nous écrivons ce terme comme :

$$p(y_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, x_s, v_s) \propto f(x_s, y_s) g(x_s, v_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s}). \tag{6.13}$$

$f$ requiert une réflexion particulière lors de sa définition afin d'être adéquat à l'application considérée : $f$ doit en particulier être déterminée par rapport au modèle de bruit. Dans la suite, nous supposons que $f$ est une gaussienne de moyenne $\mu_{x_s}$ et de variance $\sigma_{x_s}^2$, ce qui est une hypothèse très courante en segmentation des images [Besag, 1986; Geman et Geman, 1984; Marroquin et al., 1987] :

$$f(x_s, y_s) = \frac{1}{\sigma_{x_s}\sqrt{2\pi}} \exp\left(-\frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2}\right). \tag{6.14}$$

De plus, $g(x_s, v_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$ est une distribution liée aux configurations (nombre de voisins identiques à la valeur centrale) des variables de voisinage dans $\mathbf{X}$ et $\mathbf{V}$. Nous l'écrivons sous la forme :

$$g(x_s, v_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s}) \propto \sum_{(s',r') \in N_s^2} \delta_{x_s, v_s}(x_{s'}, v_{r'}). \tag{6.15}$$

$g$ permet de pondérer les contributions de $f(x_s, y_s)$ par rapport aux configurations du voisinage, tout en gardant l'expression du terme $p(y_s|\mathbf{x}_{N_s}, \mathbf{v}_{N_s}, x_s, v_s)$ modérément complexe.

En conclusion, la distribution $p(\mathbf{t}_s|\mathbf{t}_{N_s})$ régissant $\mathbf{T}$ s'écrit maintenant :

$$p(\mathbf{t}_s|\mathbf{t}_{N_s}) \propto \frac{1}{\sigma_{x_s}\sqrt{2\pi}} \exp\left(-\frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2}\right) \sum_{(s',r') \in N_s^2} \delta_{x_s, v_s}(x_{s'}, v_{r'})$$
$$\times \exp\left(-\alpha \sum_{s' \in N_s} [1 - 2\delta_{v_s}^{v_{s'}}] - \beta \sum_{s' \in N_s} \varphi^{s'}(v_s) \left[1 - 2\delta_{x_s}^{x_{s'}}\right]\right). \tag{6.16}$$

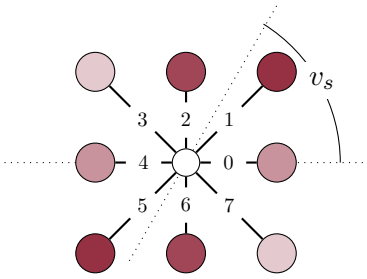



Figure 6.2 – Repérage des cliques binaires pour la mesure d'orientation au sein d'un voisinage local. L'opacité des sites est proportionnelle à la valeur prise par la fonction $\varphi^{s'}$ (6.12) lorsque $v_s = \pi/3$.

Cette distribution étant définie, il est maintenant possible de simuler des réalisations de $\mathbf{T}$ par échantillonnage de Gibbs [Geman et Geman, 1984; Robert et Casella, 2013]. Nous utilisons en pratique l'échantillonneur de Gibbs *chromatique* [Gonzalez et al., 2011], qui permet d'effectuer des tirages simultanés sur des ensembles de sites (« couleurs ») non mutuellement voisins, tout en conservant la propriété d'ergodicité de l'échantillonneur de Gibbs classique. Cet algorithme est détaillé en annexe A.2.

Il est également possible de simuler des réalisations selon les distributions marginales de $p(\mathbf{t})$. Cela est notamment pertinent pour :

1. échantillonner la distribution *a posteriori* $p(\mathbf{x}, \mathbf{v}|\mathbf{y})$ à des fins de segmentation (cf. section 6.1.2). Cet algorithme est détaillé en annexe A.2 ;
2. échantillonner $p(\mathbf{x}, \mathbf{y}|\mathbf{v})$ afin de simuler des images de synthèse (cf. section 6.3.1).

**Remarque 6.2.2.** *Dans le cas le plus général, la distribution du triplet* $\mathbf{T}$ *peut être écrite :*

$$p(\mathbf{t}) = p(\mathbf{y}, \mathbf{x}, \mathbf{v}) = p(\mathbf{y}|\mathbf{x}, \mathbf{v})p(\mathbf{x}|\mathbf{v})p(\mathbf{v}). \tag{6.17}$$

*Dans le modèle que nous proposons,* $p(\mathbf{y}|\mathbf{x}, \mathbf{v})$*,* $p(\mathbf{x}|\mathbf{v})$*, et* $p(\mathbf{v})$ *sont toutes trois des distributions de champs de Markov.*

### 6.2.2 Mesures de confiance

En se basant sur le même principe que dans la section 5.2.2, nous souhaitons établir une carte de confiance associée aux segmentations de $\mathbf{X}$ et de $\mathbf{V}$. Cette mesure est ici notée $\mathbf{u} = (\mathbf{u}^x, \mathbf{u}^v)$, et s'appuie ici aussi sur l'estimateur du MPM. Nous avons ainsi $\forall s \in \mathcal{S}$ :

$$\left| \begin{array}{lcl} u_s^x & = & \min\limits_{\omega \neq \hat{x}_s^{\mathrm{MPM}}} \left[ p(X_s = \hat{x}_s^{\mathrm{MPM}}|\mathbf{y}) - p(X_s = \omega|\mathbf{y}) \right] ; \\ u_s^v & = & \min\limits_{\nu \neq \hat{v}_s^{\mathrm{MPM}}} \left[ p(V_s = \hat{v}_s^{\mathrm{MPM}}|\mathbf{y}) - p(V_s = \nu|\mathbf{y}) \right] . \end{array} \right. \tag{6.18}$$

Cela permet de former les deux *cartes de confiance* $\mathbf{u}^x = (u_s^x)_{s\in\mathcal{S}}$ et $\mathbf{u}^v = (\mathbf{u}_s^v)_{s\in\mathcal{S}}$. Toutes deux prennent leurs valeurs dans $[0, 1]$ : 0 correspond au cas le plus incertain (classes équiprobables), et 1 est le cas le plus sûr (toutes les classes non retenues ont une probabilité nulle).

Remarquons enfin que cette mesure de confiance est particulièrement adaptée à la segmentation d'orientations dans des images : en effet, les cartes de confiance décrites en (6.18) peuvent permettre de repérer l'absence de direction privilégiée.

### 6.2.3 Estimation des paramètres

Le modèle décrit en section 6.2.1 comporte plusieurs paramètres qu'il est nécessaire de pouvoir estimer de manière supervisée et non supervisée afin de procéder à la segmentation d'images. Ces paramètres interviennent tous dans (6.16) :

— $\alpha$ et $\beta$ régulent les potentiels de type Potts dans (6.10) et (6.11) respectivement ;
— $g$, décrite en (6.15) ;
— les paramètres de $f$ (6.14) : $K = |\Omega_x|$ variances $\sigma^2_{x_s}$ et $K$ moyennes $\mu_{x_s}$.

L'ensemble de ces paramètres est noté $\theta$. Remarquons par ailleurs que ces paramètres ne dépendent pas du site $s$ considéré, car $\mathbf{T}$ est stationnaire.

#### Estimation supervisée

Nous supposons dans un premier temps qu'un ensemble de données complètes $(\mathbf{y}, \mathbf{x}, \mathbf{v})$ est disponible. Les $K$ moyennes $\mu_{x_s}$ et variances $\sigma^2_{x_s}$ sont estimées par les estimateurs du maximum de vraisemblance [Besag, 1975]. Les paramètres $\alpha$ et $\beta$ sont quant à eux estimés à partir de $\mathbf{x}$ et $\mathbf{v}$ à l'aide d'estimateurs au sens des moindres carrés inspiré de [Derin et Elliott, 1987]. Enfin, $g(x_s, v_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$ est estimé grâce aux fréquences empiriques :

$$\hat{g}(x_s, v_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s}) = \frac{1}{|\mathcal{S}'|} \sum_{r \in \mathcal{S}'} \mathbb{1}_{h_{r,N_r} = h_{s,N_s}} ; \tag{6.19}$$

avec

$$h_{s,N_s} = \sum_{(s',r') \in N_s^2} \delta_{x_s,v_s}(x_{s'}, v_{r'}) \tag{6.20}$$

et où $\mathcal{S}'$ est l'ensemble $\mathcal{S}$ dont sont retirés les bords de l'image, qui ne comportent pas $8$ voisins.

**Remarque 6.2.3.** *En pratique, la fonction indicatrice peut prendre des valeurs nulles dans l'estimateur (6.19). Cela peut gêner les simulations en empêchant les configurations rares de se réaliser. C'est pourquoi l'estimation est dans ce cas remplacée par une valeur non nulle $\epsilon \ll 1$.*

#### Estimation non supervisée

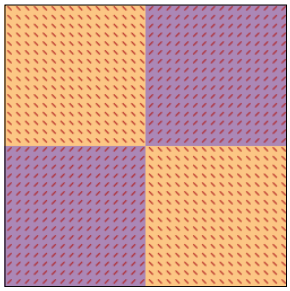
Dans le cadre de la segmentation non supervisée d'image, seule l'observation $\mathbf{Y} = \mathbf{y}$ est disponible. Nous utilisons un algorithme inspiré de l'algorithme SEM [Celeux et Diebolt, 1992] qui permet, à chaque itération, une ré-estimation des paramètres à partir de données complètes conçues par simulation grâce aux paramètres estimés à l'étape précédente.

L'algorithme proposé pour les champs de Markov triplets orientés est le suivant. À chaque itération $k$, et en notant $\hat{\theta}^k$ l'ensemble des paramètres estimés à cette étape, il faut :

1. simuler $(\mathbf{x}^k, \mathbf{v}^k)$ selon $p_{\hat{\theta}^{k-1}}(\mathbf{x}, \mathbf{v}|\mathbf{y})$ ;
2. estimer $\hat{\theta}^k$ sur les données complètes simulées formées par $(\mathbf{y}, \mathbf{x}^k, \mathbf{v}^k)$.

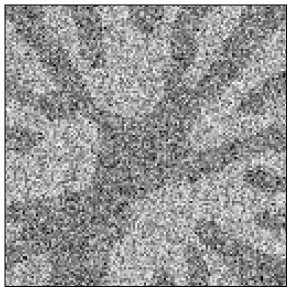
**Remarque 6.2.4.** *Tout comme l'algorithme décrit en section 5.2.5, cet algorithme n'est pas un algorithme SEM car nous n'utilisons pas les estimateurs au sens du maximum de vraisemblance pour $\alpha$, $\beta$ et $g(x_s, v_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$, qui ne sont pas définis à notre connaissance.*

En conclusion, l'estimation non supervisée des paramètres est possible au sein du modèle CMTO. Cela permet la segmentation non supervisée des images, afin de restaurer conjointement les classes de l'image et les orientations associées. La section suivante présente les résultats expérimentaux de segmentation sur des données de synthèse et sur des images réelles.



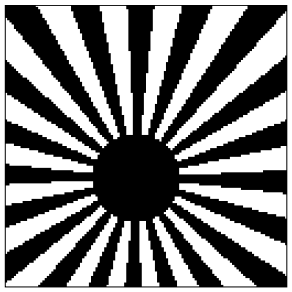
(a) $\mathbf{v}$ spécifié.

(b) Réalisation de $\mathbf{x}$ selon $p(\mathbf{x}|\mathbf{v})$.

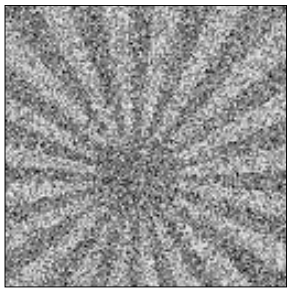


(c) Réalisation de $p(\mathbf{y}|\mathbf{x},\mathbf{v})$ (RSB $= -6$ dB).

Figure 6.3 – Expérience A.



(a) Réalisation $\mathbf{x}$ spécifiée.



(b) Réalisation de $p(\mathbf{y}|\mathbf{x})$ (RSB $= -6$ dB).

Figure 6.4 – Expérience B.

## 6.3 Résultats numériques

Dans cette partie, nous présentons les résultats numériques de segmentation d'images basée sur le modèle CMTO. Lorsque le processus $\mathbf{V}$ est retiré du modèle CMTO, nous aboutissons à un modèle de champ de Markov caché à bruit indépendant (CMC). En conséquence, celui-ci est utilisé comme base de comparaison pour tous les résultats.

### 6.3.1 Segmentation d'images de synthèse

Nous définissons dans cette section $\Omega_x = \{\omega_0, \omega_1\}$ et $\Omega_v = \{\nu_0, \nu_1\} = \{\pi/4, 3\pi/4\}$. Nous nous donnons deux cas expérimentaux, de difficulté croissante par rapport aux hypothèses du modèle :

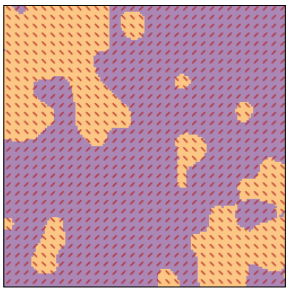
A. $\mathbf{v}$ est connu et $\mathbf{x}, \mathbf{y}$ sont simulés conditionnellement à $\mathbf{v}$ (figure 6.3).

B. $\mathbf{v}$ est inconnu, et $\mathbf{x}$ est fixé de sorte à présenter des structures orientées (figure 6.4).

La distribution du bruit est une gaussienne (6.14) de moyennes $\mu_0 = 0$ ou $\mu_1 = 1$ et d'écart-type $\sigma_0 = \sigma_1 = \sigma$. Nous mettons en place plusieurs configurations expérimentales :

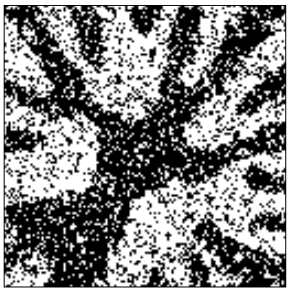
— segmentation supervisée ou non supervisée ;
— critère du MAP (6.5) et du MPM (6.6) ;
— niveaux de bruit variables, mesurés par le RSB :

$$\mathrm{RSB} = 20 \log_{10} \left( \frac{\overline{\mu}}{\sigma} \right) ; \tag{6.21}$$

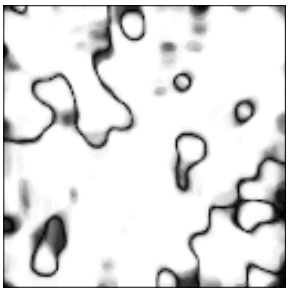


(a) CMTO : $\hat{\mathbf{v}}$, $\mathcal{E}_v = 26{,}04\%$.

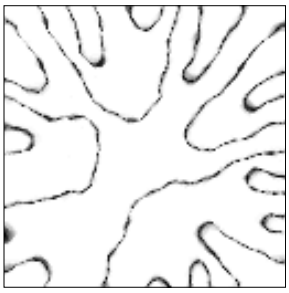
(b) CMTO : $\hat{\mathbf{x}}$, $\mathcal{E}_x = 4{,}84\%$.



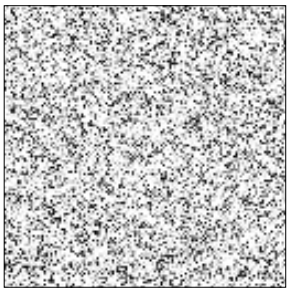
(c) CMC : $\hat{\mathbf{x}}$, $\mathcal{E}_x = 21{,}53\%$.



(d) CMTO : $\mathbf{u}^v$ .



(e) CMTO : $\mathbf{u}^x$.



(f) CMC : $\mathbf{u}^x$.

Figure 6.5 – Expérience A avec $\mathrm{RSB} = -6$ dB : résultats de segmentation non supervisée basée sur le critère du MPM de l'image présentée en figure 6.3c. Dans les cartes de confiance $\mathbf{u}^v$ et $\mathbf{u}^x$, les pixels blancs et noirs valent 1 et 0 respectivement. Ces cartes illustrent bien les contours des régions segmentées dans $\mathbf{x}$ et $\mathbf{v}$, qui correspondent aux changements de stationnarité dans les processus sous-jacents.

où $\overline{\mu}$ est défini comme :

$$\overline{\mu} = \frac{\sum_{s \in \mathcal{S}} \delta_{x_s}^{\omega_0} \mu_0 + \delta_{x_s}^{\omega_1} \mu_1}{|\mathcal{S}|}. \tag{6.22}$$

Pour comparaison, la segmentation par champ de Markov caché est également effectuée dans les mêmes conditions. L'évaluation des résultats est basée sur les taux d'erreurs de segmentation sur $\mathbf{x}$ et $\mathbf{v}$, notés respectivement $\mathcal{E}_x$ et $\mathcal{E}_v$. Les taux d'erreurs sont évalués sur 100 réalisations de $\mathbf{Y} = y$. Deux exemples de segmentation sont illustrés en figures 6.5 et 6.7, et l'ensemble des résultats est reporté en figures 6.6 et 6.8.

De manière générale, la segmentation par modèle CMTO fournit dans toutes les situations des résultats au moins aussi bons que la segmentation basée sur CMC. Les points de comparaison sont les suivants.

**Segmentation supervisée et non-supervisée.** Les résultats dans le contexte supervisé montrent dans tous les cas une très nette différence entre l'utilisation des modèles CMTO et CMC. En effet, les segmentations de $\mathbf{x}$ basées sur modèle CMTO semblent très robustes à la dégradation du RSB, permettant des gains de taux d'erreurs allant d'un facteur 3 (Expérience B, MAP) à un facteur 10 (Expérience A, MPM) par rapport aux segmentations par CMC.

**MAP et MPM.** À fort RSB, les performances des segmentations basées sur le MPM et le MAP ne présentent pas de différence significative. Lorsque le RSB diminue, nous observons un gain du second sur le premier, en particulier dans les segmentations supervisées : dans l'expérience A par exemple, le MPM permet $\mathcal{E}_x = 2{,}3\%$ contre $\mathcal{E}_x = 10{,}1\%$ pour le MAP à $\mathrm{RSB} = -20$ dB. Nous nous trouvons donc dans le cadre d'un modèle pour lequel la segmentation de $\mathbf{x}$ au sens du MAP semble surpassée par la segmentation au sens du MPM.

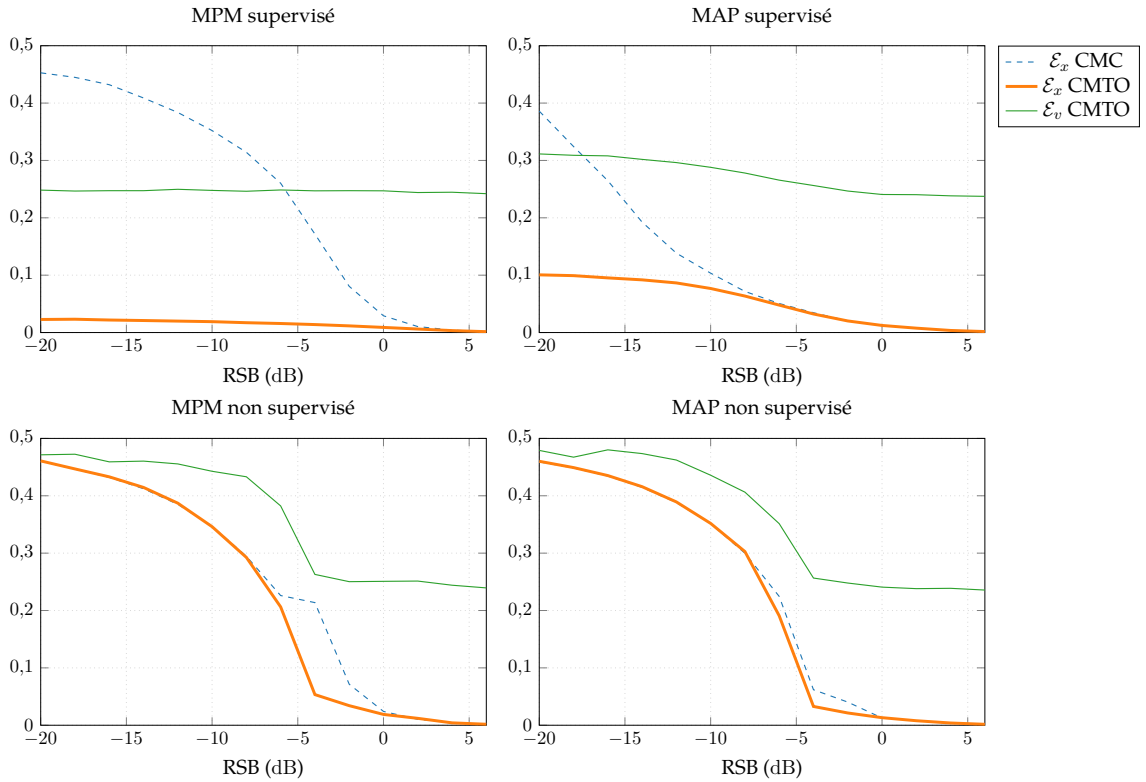



Figure 6.6 – Résultats numériques obtenus sur l'expérience A. Chaque point est la moyenne des taux d'erreur de segmentation obtenus avec 100 réalisations différentes de $\mathbf{y}$.

Le constat est différent pour le modèle de CMC : les segmentations supervisées de $\mathbf{x}$ au sens du MAP sont plus efficaces que leurs homologues basées sur le MPM.

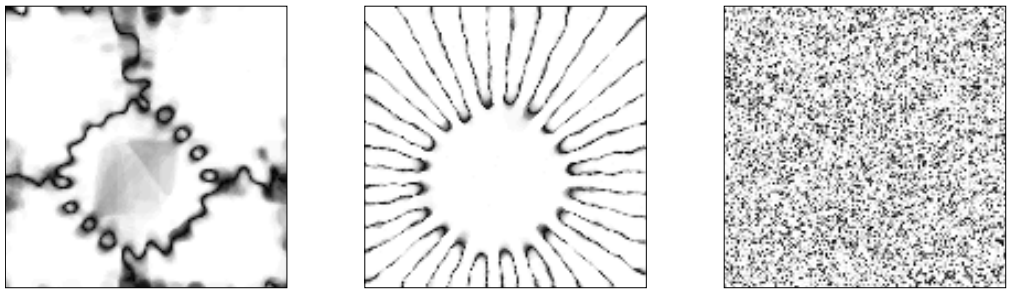
**Segmentation des orientations.** Seule l'expérience A permet une mesure des performances de segmentation des orientations, car $\mathbf{v}$ est inconnu dans l'expérience B. Nous pouvons constater que le taux d'erreur $\mathcal{E}_v$ est constant dans le cas de la segmentation supervisée basée sur le MPM. Cette valeur n'est cependant pas négligeable, car $\mathcal{E}_v$ vaut environ $24{,}6\%$. L'apparition d'une telle « erreur minimale » dans la segmentation de $\mathbf{v}$ a comme origine probable le caractère peu texturé de $\mathbf{x}$ : les motifs formés laissent de nombreuses régions qui ne semblent pas favoriser de direction particulière. La segmentation supervisée basée sur le MAP est de moindre performance que son homologue basée sur le MPM : une évolution en fonction du RSB est visible, et donne $\mathcal{E}_v = 31{,}2\%$ à $\mathrm{RSB} = -20$ dB, avant d'atteindre $24{,}0\%$ à $\mathrm{RSB} = 6$ dB.

**Sensibilité au RSB.** Nous avons constaté que, en contexte supervisé, l'évolution des taux d'erreurs des segmentations basées sur le modèle CMTO, $\mathcal{E}_x$ et $\mathcal{E}_v$, est modérée. Les résultats obtenus en contexte non supervisé montrent quant à eux une variation nette, située à $-5$ dB environ. En effet, $\mathcal{E}_x$ et $\mathcal{E}_v$ atteignent au-delà de ce seuil des valeurs proches de celles observées en segmentation supervisée.

Deux conclusions principales peuvent être tirées de ces résultats :

— les très bons résultats de segmentation dans le cadre du modèle CMTO en contexte supervisé montrent que l'adjonction du processus auxiliaire $\mathbf{v}$ offre une modélisation particulièrement robuste pour la segmentation d'images.
— la nette différence entre les résultats de segmentations supervisées et non supervisées illustre les défis rencontrés lors de la segmentation non supervisée d'image.

(a) CMTO : $\hat{\mathbf{v}}$. (b) CMTO : $\hat{\mathbf{x}}$, $\mathcal{E}_x = 9{,}31\%$. (c) CMC : $\hat{\mathbf{x}}$, $\mathcal{E}_x = 20{,}01\%$.



(d) CMTO : $\mathbf{u}^v$ . (e) CMTO : $\mathbf{u}^x$. (f) CMC : $\mathbf{u}^x$.

Figure 6.7 – Expérience B avec RSB $= -6$ dB : résultats de segmentation non supervisée basée sur le critère du MPM de l'image présentée en figure 6.4b. (même légende qu'en figure 6.5).

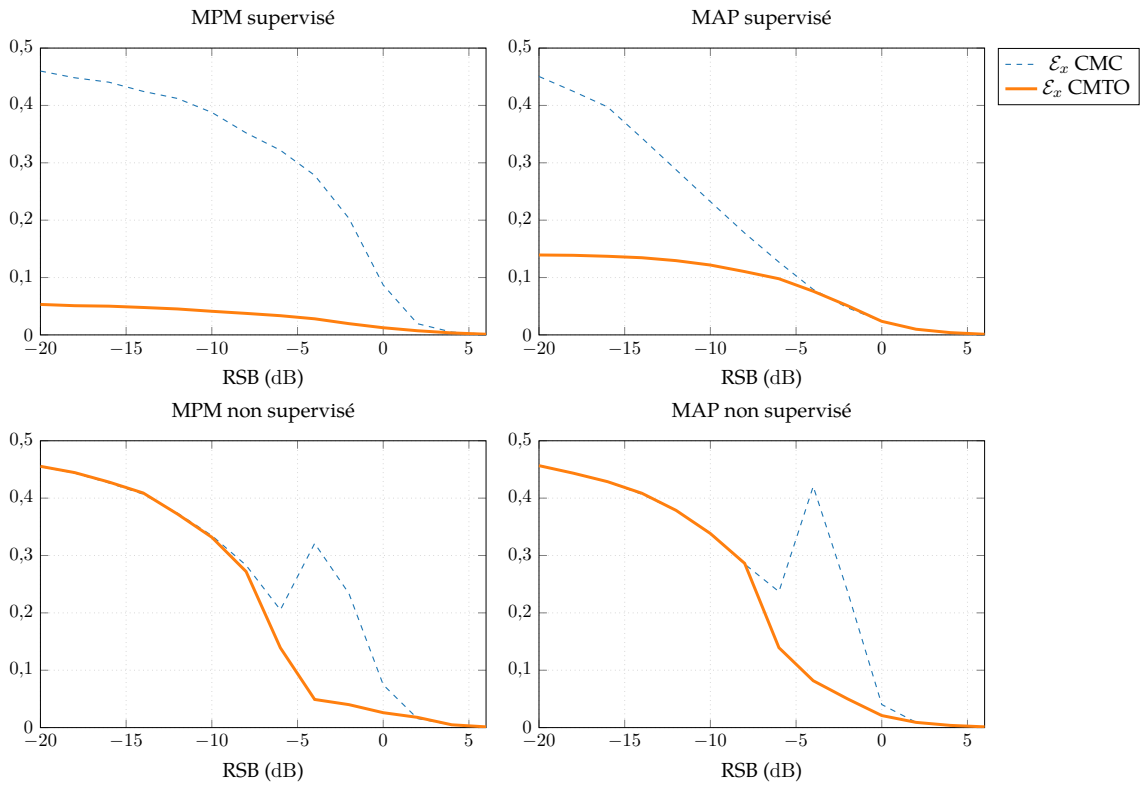



Figure 6.8 – Résultats numériques obtenus sur l'expérience B. Chaque point est la moyenne des taux d'erreur de segmentation obtenus avec 10 réalisations différentes de $\mathbf{y}$.

### 6.3.2 Segmentation d'images réelles

Le problème de la segmentation conjointe de $\mathbf{x}$ et $\mathbf{v}$ peut être posé dans de nombreux cas applicatifs. Il est en effet possible de rencontrer des images présentant de fortes structures orientées en télédétection, en imagerie médicale, etc. Nous présentons dans cette section deux résultats complémentaires obtenus sur des images de télédétection. Ces résultats sont obtenus par segmentation au sens du MAP et du MPM, avec les modèles de champs de Markov cachés et CMTO. Des résultats additionnels obtenus sur des textures de Brodatz [Brodatz, 1966; nor, 2017] sont présentés en annexe B.1. Dans tous les cas, nous choisissons $|\Omega_v| = 6$ avec $\Omega_v = \{\pi/12, 3\pi/12, \ldots, 11\pi/12\}$.

#### Image satellitaire de vignes

La figure 6.9a représente une image panchromatique d'une culture de vignes acquise par le système Pléiades [ple, 2017] en Alsace en 2012. Cette image présente des textures orientées qui semblent homogènes sur plusieurs régions mais nettement distinctes entre régions. Pour tenir compte des différentes intensités, nous prenons $|\Omega_x| = 3$.

Les résultats obtenus avec le modèle CMTO sont présentés dans les figures 6.9d–6.9i, ainsi que les cartes de confiances associées aux segmentations MPM. Les résultats obtenus par champs de Markov cachés sont quant à eux présentés dans les figures 6.9b et 6.9c pour comparaison. Plusieurs commentaires peuvent être faits sur ces résultats :

— les segmentations basées sur le modèle CMTO (figures 6.9e et 6.9f) semblent modéliser correctement les caractéristiques de l'image $\mathbf{y}$. La différence avec la segmentation par champs de Markov cachés est très marquée : en effet, ce dernier modèle semble mal modéliser ou ne pas prendre en compte un certain nombre d'éléments de petite taille. Ce constat est identique sur les segmentations au sens du MAP et du MPM.

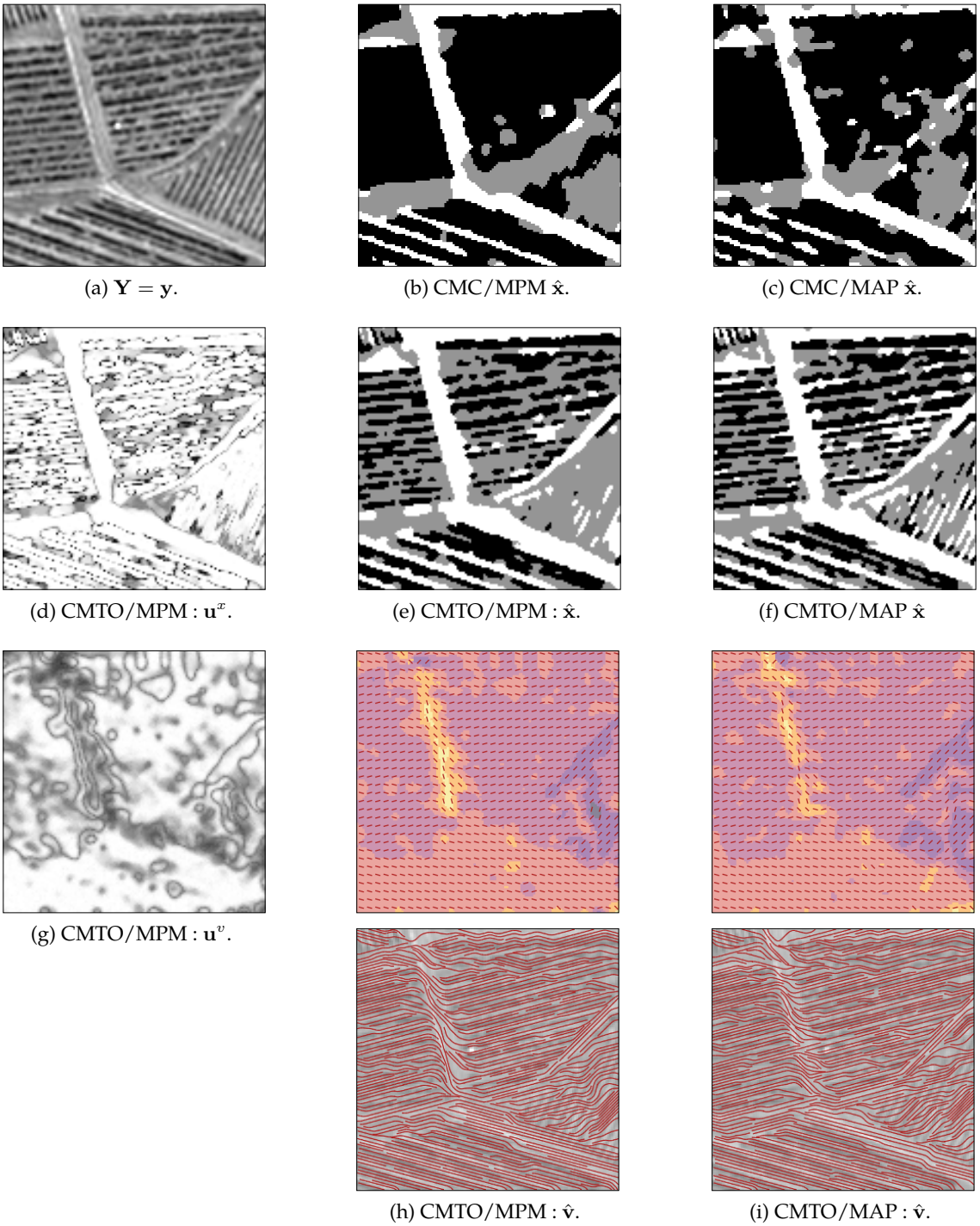
— la carte de confiance $\mathbf{u}^x$ (figure 6.9d) illustre bien les difficultés rencontrées par la modélisation markovienne pour gérer les changements de stationnarité. Cela est en particulier le cas pour toutes les régions « limite » entre deux classes dans la segmentation associée de la figure 6.9e. Les régions à l'intérieur de ces frontières sont au contraire segmentées avec une plus grande certitude.

— les segmentations de $\mathbf{v}$ basées sur le modèle CMTO (figures 6.9h et 6.9i) présentent plusieurs régions homogènes. Elles semblent toutes deux cohérentes avec les stationnarités observées dans $\hat{\mathbf{x}}$ et les caractéristiques d'orientation apparentes de $\mathbf{y}$.

— enfin, la carte de confiance $\mathbf{u}^v$ (figure 6.9g) associée à la segmentation au sens du MPM de $\hat{\mathbf{v}}$ illustre là aussi les changements de stationnarités dans l'estimation de $\mathbf{v}$. Les régions larges et homogènes semblent plus ardues à segmenter du point de vue des caractéristiques directionnelles. Cette carte de confiance permet en effet une discrimination des textures : elle renforce l'hypothèse d'une direction privilégiée dans certaines régions (les cultures de vigne) et l'atténue dans d'autres (les chemins).

#### Dépression atmosphérique

La figure 6.10a présente une image satellitaire d'un phénomène météorologique de dépression atmosphérique, acquise notamment au-dessus de l'Islande par le satellite Aqua/MODIS de la NASA (domaine public). Deux classes principales sont visibles : les nuages et le reste de l'image. Nous choisissons donc ici $|\Omega_x| = 2$. Le bruit affectant la classe « nuage » correspond à des intensités variables, tandis que l'autre classe est affectée par des éléments de petite taille et par l'Islande apparaissant en bas à droite de l'image.



(a) $\mathbf{Y} = \mathbf{y}$. (b) CMC/MPM $\hat{\mathbf{x}}$. (c) CMC/MAP $\hat{\mathbf{x}}$.

(d) CMTO/MPM : $\mathbf{u}^x$. (e) CMTO/MPM : $\hat{\mathbf{x}}$. (f) CMTO/MAP $\hat{\mathbf{x}}$

(g) CMTO/MPM : $\mathbf{u}^v$.

(h) CMTO/MPM : $\hat{\mathbf{v}}$. (i) CMTO/MAP : $\hat{\mathbf{v}}$.

Figure 6.9 – Image panchromatique Pléiades ($0{,}5$ m) [ple, 2017], couvrant des cultures de vignes d'orientations différentes. ©CNES 2012, distribution Astrium Services, France, tous droits réservés. Dans les cartes de confiance $\mathbf{u}^x$ (d) et $\mathbf{u}^v$ (g), les valeurs faibles sont en noir (incertitude élevée) et les valeurs fortes sont en blanc (confiance élevée). Les orientations sont représentées d'une part par valeur (images du haut en (h) et (i)) et d'autre part par les courbes tangentes superposées à l'observation $\mathbf{y}$ (images du bas).

Les figures 6.10b–6.10i représentent les résultats de segmentation par champs de Markov cachés, par CMTO et les cartes de confiances associées lorsqu'elles sont disponibles. Quelques commentaires complémentaires peuvent être faits sur ces résultats :

— les segmentations de $\mathbf{x}$ par champs de Markov cachés et par CMTO (figures 6.10b, 6.10c, 6.10e et 6.10f) semblent détecter les caractéristiques principales de l'image. En effet, les différences entre les deux segmentations ne sont pas aussi nettes que dans la figure 6.9. Remarquons cependant que plusieurs petits éléments présents dans la segmentation CMTO semblent ignorés dans la segmentation par CMC.

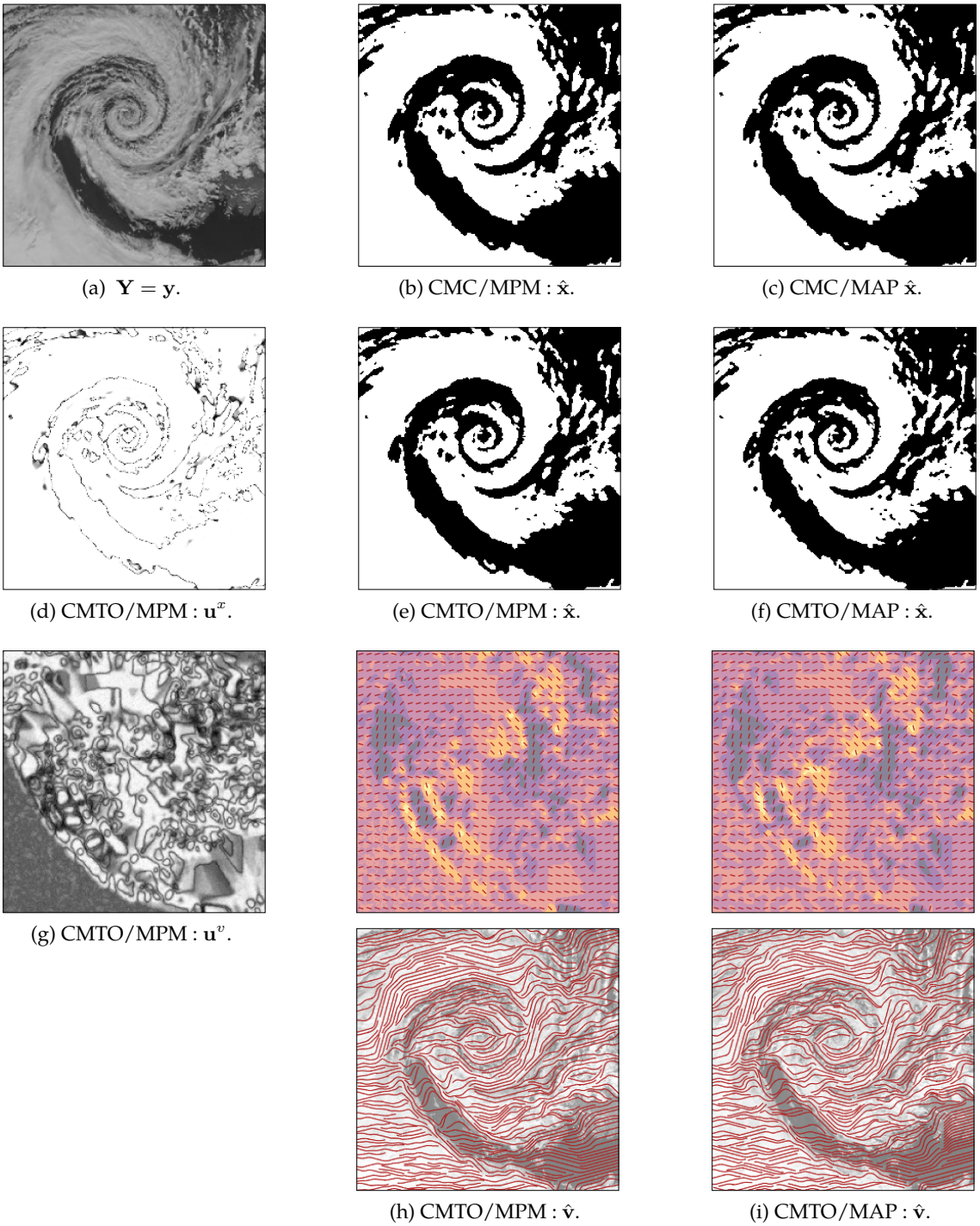
— en revanche, les résultats obtenus pour la segmentation de $\mathbf{v}$ (figure 6.10h et 6.10i) sont particulièrement intéressants. En effet, la plupart des orientations des segmentations correspondent à des caractéristiques visibles dans l'image $\mathbf{y}$, et semblent cohérentes par rapport à la dynamique « en spirale » de l'image.

— enfin, les régions homogènes, ne présentant pas d'orientations marquées (comme la partie en bas à gauche de l'image) sont bien retransmises dans la carte d'incertitude de la figure 6.10g. Cela permet, de manière plus marquée que pour l'image de vigne, d'identifier rapidement les régions pertinentes dans la recherche de structures orientées.

## 6.4 Conclusion

Dans ce chapitre, nous avons présenté un modèle de champ de Markov triplet original prenant conjointement en compte les classes et les orientations au sein d'une image. Nous avons également présenté comment simuler, estimer les paramètres, et segmenter conjointement classes et orientations dans le cadre du modèle CMTO. Nous avons de plus proposé une évaluation de l'incertitude associée aux segmentations dans le cadre de la segmentation au sens du MPM. Les résultats expérimentaux ont montré que le modèle est adapté au problème de segmentation de structures orientées, et est plus efficace que le modèle de champ de Markov caché.

Plusieurs perspectives sont envisageables à partir de ces travaux. En effet, il serait possible de faire du processus d'orientation $\mathbf{V}$ un processus à valeurs continues dans $[0, \pi]$. Cela conduirait à une segmentation « mixte » continu (pour $\mathbf{V}$) – discret (pour $\mathbf{X}$) et ferait appel à d'autres critères de segmentation, tout comme la segmentation d'un $\mathbf{X}$ à valeurs réelles. Il est également envisageable d'étendre le modèle à la segmentation d'images tridimensionnelles, comme les images médicales où peuvent intervenir des structures orientées (fibres, vaisseaux sanguins, etc.).



(a) $\mathbf{Y} = \mathbf{y}$. (b) CMC/MPM : $\hat{\mathbf{x}}$. (c) CMC/MAP $\hat{\mathbf{x}}$.

(d) CMTO/MPM : $\mathbf{u}^x$. (e) CMTO/MPM : $\hat{\mathbf{x}}$. (f) CMTO/MAP : $\hat{\mathbf{x}}$.

(g) CMTO/MPM : $\mathbf{u}^v$.

(h) CMTO/MPM : $\hat{\mathbf{v}}$. (i) CMTO/MAP : $\hat{\mathbf{v}}$.

Figure 6.10 – Région de dépression atmosphérique dans une image satellitaire, couvrant des nuages de direction variable. La légende est la même que dans la figure 6.9.

# 7

# Arbres de Markov triplets spatiaux



## 7.1 Introduction

Nous avons présenté, dans les chapitres 5 et 6, des modèles de champs de Markov couples et triplets permettant la segmentation bayésienne des images. Ces modèles permettent une grande richesse de modélisation, mais dépendent d'algorithmes stochastiques requis pour estimer la distribution *a posteriori* des classifications. Nous avons vu, dans le chapitre 4, que la modélisation par chaîne de Markov offre quant à elle la possibilité de calculer exactement les distributions *a posteriori*, au prix d'un modèle semblant moins « intuitif » pour la modélisation d'images.

Un compromis entre la richesse de modélisation et la possibilité d'effectuer les calculs exacts, évitant le recours à des méthodes stochastiques, semble exister avec les modèles par arbres de Markov cachés à bruit indépendant (AMC) [Laferté et al., 2000]. Cette modélisation permet d'exploiter, dans un contexte de segmentation, une hiérarchie traduisant l'homogénéité spatiale des classifications recherchées. Cet aspect a pour principal défaut d'introduire, dans les cas difficiles, des résultats de segmentation présentant des « mouchetages » relativement marqués. Plusieurs types de modèles on été proposés pour compenser cet effet. Citons par exemple les modèles d'arbres de Markov *évolutifs* [Monfrini et Pieczynski, 2005], dans lesquels les probabilités de transition parent-enfant changent selon

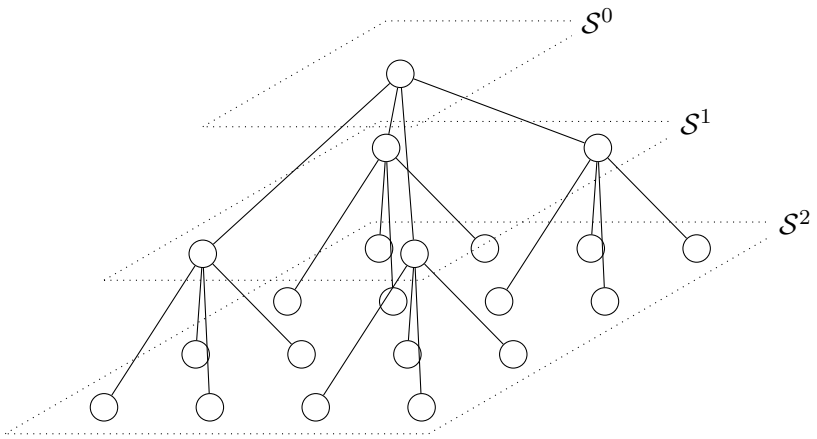



Figure 7.1 – Graphe de dépendances associé à un quadabre à 3 niveaux.

l'échelle. Il existe également des modélisations par *champs hiérarchiques* [Mignotte et al., 2000] où chaque *a priori* est markovien spatialement, sur chaque échelle de l'arbre.

Les modèles basés sur des arbres de Markov, tout comme les modèles par chaînes ou par champs, on été enrichis par l'introduction de modèles triplets [Pieczynski, 2003a] (cf. chapitre 4). Dans ce cadre, en plus du processus observé $\mathbf{Y}$ et du processus de classe $\mathbf{X}$, un troisième processus $\mathbf{V}$ est adjoint, dans le but de modéliser des phénomènes plus complexes, par exemple comme nous l'avons proposé en modélisant des directions privilégiées (cf. chapitre 6).

Dans ce chapitre, nous introduisons un modèle d'arbre de Markov triplet (AMT) dans lequel le processus auxiliaire permet de moduler les probabilités de transitions parent/enfant, en tenant compte des classifications des voisins du parent. Cette modulation a pour objectif de permettre la modélisation de structures homogènes à grande échelle, et d'éviter lors de la segmentation les effets de « mouchetage » précédemment évoqués.

Ce modèle sera appelé « AMT spatial » (AMTS) et sera introduit en section 7.2, avec la segmentation au sens du MPM. Nous appliquons ensuite ce modèle dans le cas de la segmentation d'une observation mono-résolution en section 7.3. Enfin, les résultats présentés en section 7.4 permettent la comparaison avec les méthodes basées sur les modèles d'arbres et de champs de Markov cachés classiques.

## 7.2 Arbre de Markov triplet spatial

### 7.2.1 Modèle général

Soit $\mathbf{T} = (\mathbf{T}_s)_{s\in\mathcal{S}}$ un processus stochastique. Nous nous intéressons au cas d'un arbre de Markov dans lequel chaque parent génère quatre enfants. $\mathcal{S}$ est donc l'ensemble des résolutions d'un quadarbre : $\mathcal{S} = \{\mathcal{S}^0, \ldots, \mathcal{S}^N\}$. Comme illustré dans le graphe de dépendances de la figure 7.1, chaque $\mathcal{S}^n$ contient $4^n$ sites : $n = 0$ représente la racine de l'arbre, et $n = N$ la plus fine résolution. De plus, $\mathbf{T}$ comporte trois processus distincts : $\mathbf{T} = (\mathbf{X}, \mathbf{Y}, \mathbf{V})$. De manière analogue au modèle présenté dans le chapitre 6, $\mathbf{Y}$ est un processus d'observation, $\mathbf{X}$ est un processus de classe et $\mathbf{V}$ est un processus auxiliaire discret. Chaque $\mathbf{T}_s$ peut aussi s'écrire comme $\mathbf{T}_s = (Y_s, X_s, \mathbf{V}_s)$, où $Y_s \in \mathbb{R}$, $X_s \in \Omega_x$ et $\mathbf{V}_s \in \Omega_\mathbf{v}$ avec $\Omega_x$ et $\Omega_\mathbf{v}$ des ensembles finis.

Formellement, $\mathbf{T}$ est un arbre de Markov [Laferté et al., 2000], et vérifie :

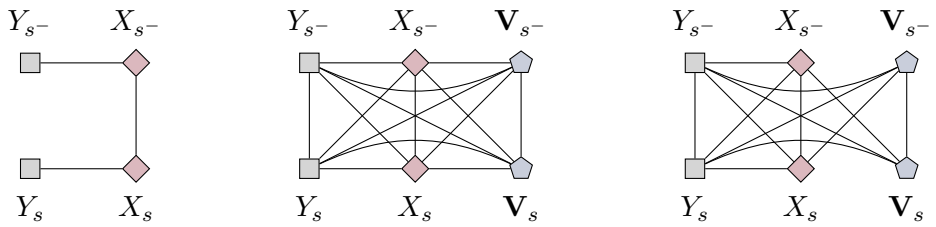
$$p(\mathbf{T}) = p(\mathbf{T}_0) \prod_{s \in \mathcal{S} \setminus \mathcal{S}^0} p(\mathbf{T}_s | \mathbf{T}_{s^-}); \tag{7.1}$$

où $s^-$ est le site parent de $s$, et $\mathbf{T}_0$ est la valeur de $\mathbf{T}_s$ en la racine ($s \in \mathcal{S}^0$).

$\mathbf{V}$ est un processus auxiliaire, qui permet de faire varier la loi de $\mathbf{X}$ à chaque transition parent-enfant. Dans ce qui suit, nous supposons que $X_s$ et $\mathbf{V}_s$ sont indépendants conditionnellement à $\mathbf{T}_{s^-}$. Cela nous permet d'écrire $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^N\}$ :

$$p(\mathbf{T}_s | \mathbf{T}_{s^-}) = p(Y_s | X_s, \mathbf{V}_s, \mathbf{T}_{s^-}) p(X_s | \mathbf{T}_{s^-}) p(\mathbf{V}_s | \mathbf{T}_{s^-}). \tag{7.2}$$

Les trois densités de cette équation seront précisées dans la section 7.3. La figure 7.2 représente le graphe de dépendances intervenant au sein d'une transition parent/enfant au sein des modèles AMC, AMT et AMTS.




(a) AMC. (b) AMT. (c) AMTS.

Figure 7.2 – Graphes de dépendances pour une transition parent-enfant pour les différents modèles d'arbres de Markov.

### 7.2.2 Segmentation au sens du MPM

Nous considérons ici que $\mathbf{Y}$ est défini pour toutes les résolutions : $\mathbf{Y} = (Y_s)_{s \in \mathcal{S}^0, \ldots, \mathcal{S}^N}$. Nous nous plaçons dans le contexte de la segmentation, et choisissons l'estimateur du MPM. Celui-ci requiert le calcul, en tout site $s \in \mathcal{S}$, de :

— $p(X_s = \omega_i, \mathbf{V}_s = \boldsymbol{\nu}_j | X_{s^-} = \omega_k, \mathbf{V}_{s^-} = \boldsymbol{\nu}_l, \mathbf{y}) = p(\omega_i, \boldsymbol{\nu}_j | \omega_k, \boldsymbol{\nu}_l, \mathbf{y})$ ;

— $p(X_s = \omega_i, \mathbf{V}_s = \boldsymbol{\nu}_j, y_0) = p(\omega_i, \boldsymbol{\nu}_j, y_0)$ ;

pour toutes les valeurs $(\omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l) \in (\Omega_x \times \Omega_v)^2$. Nous présentons un algorithme inspiré de [Monfrini et al., 1999] et adapté au cas des AMT, permettant le calcul du MPM. Il comporte les cinq étapes décrites ci-dessous, et repose sur l'utilisation de fonctions auxiliaires notées $A_n, 0 \leq n \leq N$, qui sont calculées dans la *passe montante* suivante :

1. À la plus fine résolution, $\forall s \in \mathcal{S}^N$ :

$$A_N(s; \omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l) = p(y_s, \omega_i, \boldsymbol{\nu}_j | y_{s^-}, \omega_k, \boldsymbol{\nu}_l) \tag{7.3}$$

où $p(y_s, \omega_i, \boldsymbol{\nu}_j | y_{s^-}, \omega_k, \boldsymbol{\nu}_l)$ vaut $p(Y_s = y_s, X_s = \omega_i, \mathbf{V}_s = \boldsymbol{\nu}_j | Y_{s^-} = y_{s^-}, X_{s^-} = \omega_k, \mathbf{V}_{s^-} = \boldsymbol{\nu}_l)$.

2. Sur les résolutions intermédiaires, $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^{N-1}\}$ :

$$A_n(s; \omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l) = p(y_s, \omega_i, \boldsymbol{\nu}_j | y_{s^-}, \omega_k, \boldsymbol{\nu}_l) \times \prod_{s^+ \in \mathcal{E}_s} \left( \sum_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_{n+1}(s^+; \omega, \boldsymbol{\nu}, \omega_i, \boldsymbol{\nu}_j) \right) \tag{7.4}$$

où $\mathcal{E}_s$ est l'ensemble des enfants de $s$.

3. En la racine, $r \in \mathcal{S}^0$ :

$$A_0(\omega_i, \boldsymbol{\nu}_j) = p(y_0, \omega_i, \boldsymbol{\nu}_j) \times \prod_{s^+ \in \mathcal{E}_0} \left( \sum_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_1(s^+; \omega, \boldsymbol{\nu}, \omega_i, \boldsymbol{\nu}_j) \right) ; \tag{7.5}$$

où $\mathcal{E}_0 = \mathcal{S}^1$.

Un AMT étant markovien *a posteriori* [Pieczynski, 2003a], il est possible de calculer $p(X_s = \omega_i, \mathbf{V}_s = \boldsymbol{\nu}_j | \mathbf{y}) = p(\omega_i, \boldsymbol{\nu}_j | \mathbf{y})$ pour tout $s \in \mathcal{S}^N$ et pour tout $(\omega_i, \boldsymbol{\nu}_j) \in \Omega_x \times \Omega_u$. Ce calcul s'effectue dans la *passe descendante* suivante :

4. Sélection de la classe la plus probable en la racine, $s \in \mathcal{S}^0$ :

$$\hat{x}_s^{\text{MPM}}, \hat{\mathbf{v}}_s^{\text{MPM}} = \underset{\omega_i \in \Omega_x, \boldsymbol{\nu}_j \in \Omega_v}{\arg\max}\ p(X_s = \omega_i, \mathbf{V}_s = \boldsymbol{\nu}_j | \mathbf{y}) \tag{7.6}$$

avec la distribution *a posteriori* obtenue par :

$$p(\omega_i, \boldsymbol{\nu}_j | \mathbf{y}) = \frac{A_0(\omega_i, \boldsymbol{\nu}_j)}{\sum\limits_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_0(\omega, \boldsymbol{\nu})}. \tag{7.7}$$

5. les estimations de $X_{s^-}$ et $\mathbf{V}_{s^-}$ étant fixés par la segmentation à la résolution précédente, sélection en chaque site de la classe la plus probable, $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^N\}$ :

$$\hat{x}_s^{\text{MPM}}, \hat{\mathbf{v}}_s^{\text{MPM}} = \underset{\omega_i \in \Omega_x, \boldsymbol{\nu}_j \in \Omega_v}{\arg\max}\ p(X_s = \omega_i, \mathbf{V}_s = \boldsymbol{\nu}_j | X_{s^-} = \hat{x}_{s^-}^{\text{MPM}}, \mathbf{V}_{s^-} = \hat{\mathbf{v}}_{s^-}^{\text{MPM}}, \mathbf{y}) \tag{7.8}$$

avec les distributions *a posteriori* suivantes :

$$p(\omega_i, \boldsymbol{\nu}_j | \omega_k, \boldsymbol{\nu}_l, \mathbf{y}) = \frac{A_n(s; \omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l)}{\sum\limits_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_n(s; \omega, \boldsymbol{\nu}, \omega_k, \boldsymbol{\nu}_l)}. \tag{7.9}$$

Une fois cette cascade de sélections effectuée, la segmentation au sens du MPM est l'ensemble des $\hat{x}_s^{\text{MPM}}, \hat{\mathbf{v}}_s^{\text{MPM}}$ retenus dans l'ensemble des résolutions $\mathcal{S}$.

## 7.3 Application à la segmentation d'images

### 7.3.1 Spécificités du modèle

Le modèle présenté dans la section 7.2 correspond, en ce qui concerne le processus d'observation $\mathbf{Y}$, au cas le plus général. Nous détaillons ici l'application de ce modèle à la segmentation d'une image observée, qui correspond à $(Y_s)_{s \in \mathcal{S}^N}$. Par ailleurs, nous supposons que :

$$p(Y_s | X_s, \mathbf{V}_s, \mathbf{T}_{s^-}) = p(Y_s | X_s) \tag{7.10}$$

Cette distribution pourra être notée $f(x_s, y_s)$, où $f$ est une fonction paramétrique. De plus, $\forall s \in \{\mathcal{S}^0, \ldots, \mathcal{S}^{N-1}\}$, nous considérons formellement que $p(Y_s | X_s) \propto 1$.

Nous choisissons de travailler avec des composantes 8-variées pour le processus auxiliaire $\mathbf{V}$, et notons $\mathbf{V}_s = (V_{s,i})_{i \in \{1,\ldots,8\}}$. Nous supposerons de plus que les $(\mathbf{V}_{s,i})_{i \in \{1,\ldots,8\}}$ sont indépendants conditionnellement à $\mathbf{T}_{s^-}$. En conséquence, nous pouvons écrire :

$$p(\mathbf{V}_s | \mathbf{T}_{s^-}) = \prod_{i \in \{1,\ldots,8\}} p(V_{s,i} | \mathbf{T}_{s^-}). \tag{7.11}$$

Finalement, nous avons $\forall s \in \mathcal{S}^N$ :

$$p\left(\mathbf{T}_s|\mathbf{T}_{s^-}\right) = p\left(Y_s|X_s\right) p\left(X_s|X_{s^-}, \mathbf{V}_{s^-}\right) \prod_{i\in\{1,\dots,8\}} p\left(V_{s,i}|X_{s^-}, \mathbf{V}_{s^-}\right), \tag{7.12}$$

et $\forall s \in \{\mathcal{S}^1, ..., \mathcal{S}^{N-1}\}$ :

$$p\left(\mathbf{T}_s|\mathbf{T}_{s^-}\right) \propto p\left(X_s|X_{s^-}, \mathbf{V}_{s^-}\right) \prod_{i\in\{1,\dots,8\}} p\left(V_{s,i}|X_{s^-}, \mathbf{V}_{s^-}\right). \tag{7.13}$$

Nous présentons maintenant le lien entre $(X_{s^-}, \mathbf{V}_{s^-})$ et $(X_s, \mathbf{V}_s)$ pour $s \in \mathcal{E}_{s^-}$. Cela revient à déterminer $p\left(X_s|X_{s^-}, \mathbf{V}_{s^-}\right)$ d'une part et $p\left(V_{s,i}|X_{s^-}, \mathbf{V}_{s^-}\right)$ pour $i \in \{1, \dots, 8\}$ d'autre part. L'objectif est de tenir compte de l'homogénéité entre les classes prises par $\mathbf{X}$ au sein d'une résolution. Pour ce faire, nous faisons intervenir les classes des voisins spatiaux des parents dans la paramétrisation des probabilités de transition parent/enfants [1].

$p(X_s|X_{s^-}, \mathbf{V}_{s^-})$ représente la distribution d'un des quatre enfants de $s-$. Nous supposons que $X_s$ est issu de $X_{s^-}$ et de trois composantes de $\mathbf{V}_{s^-}$ par :

$$p(X_s|X_{s^-}, \mathbf{V}_{s^-}) \propto \exp\left(\alpha\delta_{X_{s^-}}^{X_s} + \beta \sum_{V^-\in\mathbf{V}'_{s^-}} \delta_{V^-}^{X_s}\right); \tag{7.14}$$

où $\delta_a^b$ est la fonction de Kronecker et vaut 1 si $a = b$ et 0 sinon, et $\alpha$ et $\beta$ sont des paramètres réels du modèle. $\mathbf{V}'_{s^-}$ est l'ensemble de trois composantes de $\mathbf{V}_{s^-}$ dans la direction de $X_s$. Le choix de ces trois $V_{s_i^-}$ est représenté en figure 7.3.

Nous détaillons maintenant la distribution $p(V_{s,i}|X_{s^-}, \mathbf{V}_{s^-})$, pour laquelle deux cas existent. Dans le premier cas, les $V_{s,i}$ sont liés aux $X_{s^-}$ selon le même principe que précédemment, comme l'illustre la figure 7.4.

Dans le second cas, il est nécessaire de remonter à la résolution supérieure pour comprendre comment sont générés les $X_t$, non enfants de $X_{s^-}$ et voisins spatiaux de $X_s$. $X_t$ est donc un enfant de $X_{t^-}$ où $X_{t^-}$ est un voisin spatial de $X_{s^-}$ . Comme l'illustre la figure 7.5, les composantes de $(X_{s^-}, \mathbf{V}_{s^-})$ permettent de caractériser les $V_{s,i}$ correspondant à ces $X_t$ à partir de quadruplets ayant la même loi.

Cette méthode de génération est complexe mais n'est finalement que très formelle du point de vue de l'utilisateur. En effet il faut en retenir que lorsque les processus $(X_s)_{s\in\mathcal{S}^n}$ et $(\mathbf{V}_s)_{s\in\mathcal{S}^n}$ sont connus à la résolution $n$, la loi de $(X_s)_{s\in\mathcal{S}^{n+1}}$ peut être définie et celle de $(\mathbf{V}_s)_{s\in\mathcal{S}^{n+1}}$ peut être déduite, ce qui permet de passer à la résolution suivante.

**Remarque 7.3.1.** *Lorsque le processus* $\mathbf{V}$ *est ignoré, la distribution de* $\mathbf{T}$ *est celle d'un arbre de Markov caché. En effet, nous pouvons dans ce cas écrire :*

$$p(X_s|\mathbf{T}_{s^-}) = p(X_s|X_{s^-}) \propto \exp\left(\alpha\delta_{X_{s^-}}^{X_s}\right); \tag{7.15}$$

*ou de manière équivalente pour tous* $\omega_i, \omega_j \in \Omega_x$ *tels que* $\omega_i \neq \omega_j$ *:*

$$\left|\begin{array}{l} p(X_s = \omega_i|X_{s^-} = \omega_i) = \dfrac{e^\alpha}{e^\alpha + |\Omega_x| - 1} \\ p(X_s = \omega_i|X_{s^-} = \omega_j) = \dfrac{1}{e^\alpha + |\Omega_x| - 1} \end{array}\right. \tag{7.16}$$

1. Dans les AMC, ceci est impossible sans créer de cycle dans le quadarbre et perdre la markovianité du processus. Le processus $\mathbf{V}$ permet de relever ce défi tout en récupérant la markovianité du processus a posteriori permettant le calcul du MPM.

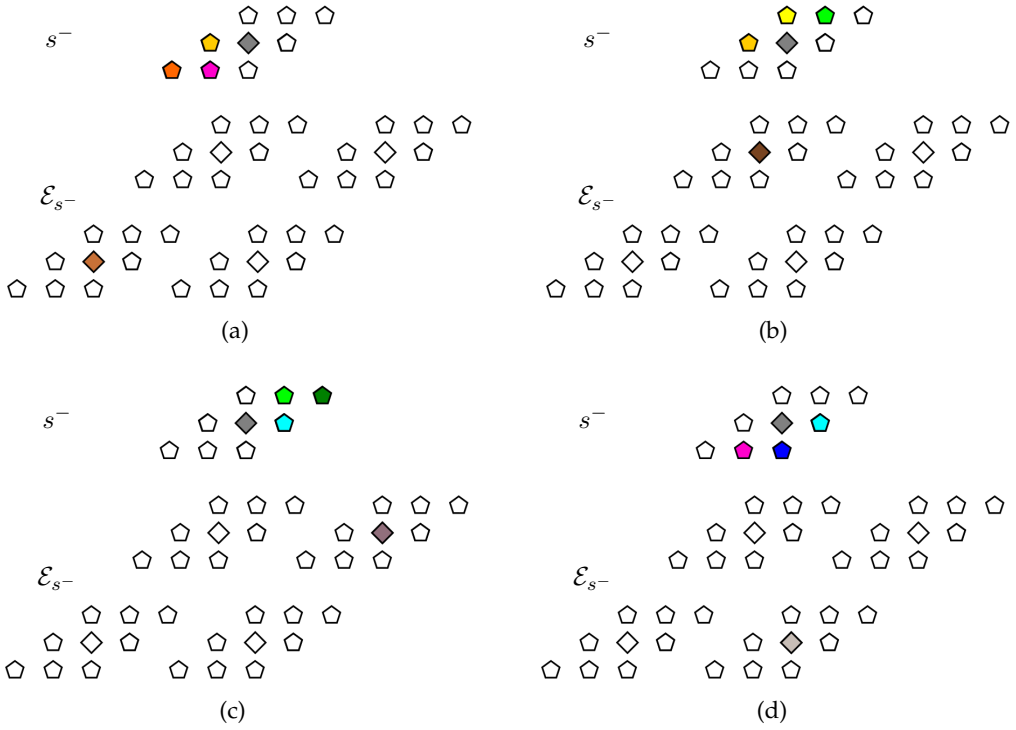



Figure 7.3 – Lien entre les variables $X_s$ et $(X_{s^-}, \mathbf{V}_{s^-})$. Sur ces graphes, les carrés représentent les variables $X$ et les pentagones les variables $V$. Dans chaque figure, seules les variables nécessaires à la génération des $X_s \in \mathcal{E}_{s^-}$ sont colorées au site $s^-$.

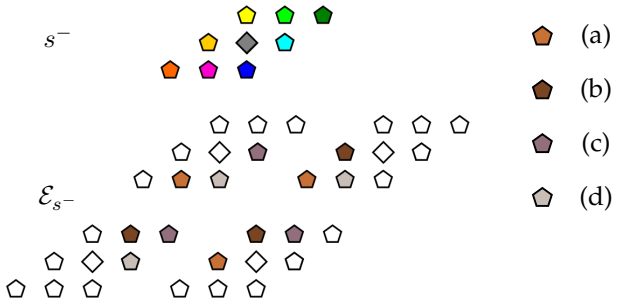



Figure 7.4 – Liens entre les $V_{s,i}$ (pentagones colorés) et $(X_{s^-}, \mathbf{V}_{s^-})$ lorsque ces variables correspondent aux $X_s$ de la figure 7.3. La légende à droite indique quels $V_{s,i}$ sont de même distribution que les $X_s$ de la figure 7.3.

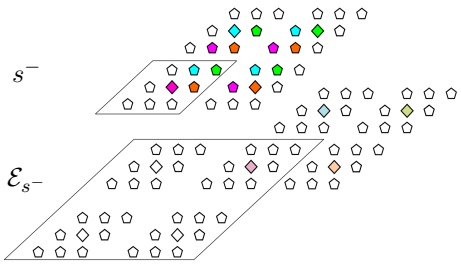



Figure 7.5 – Liens entre les $V_{s,i}$ (pentagones colorés) et $(X_{s^-}, \mathbf{V}_{s^-})$ dans le cas complémentaire de celui de la figure 7.4.

### 7.3.2 Segmentation

Nous reprenons ici l'algorithme de segmentation au sens du MPM de la section 7.2.2, dans le cas où $\mathbf{Y}$ n'est défini qu'à la résolution la plus fine : $\mathbf{Y} = (Y_s)_{s \in \mathcal{S}^N}$. Les fonctions auxiliaire $A_n, 0 \leq n \leq N$ sont calculées dans la passe montante :

1. À la plus fine résolution, $\forall s \in \mathcal{S}^N$ :

$$A_N(s; \omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l) = p(y_s, \omega_i, \boldsymbol{\nu}_j | \omega_k, \boldsymbol{\nu}_l) \tag{7.17}$$

2. Sur les résolutions intermédiaires, $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^{N-1}\}$ :

$$A_n(s; \omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l) = p(\omega_i, \boldsymbol{\nu}_j | \omega_k, \boldsymbol{\nu}_l) \times \prod_{s^+ \in \mathcal{E}_s} \left( \sum_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_{n+1}(s^+; \omega, \boldsymbol{\nu}, \omega_i, \boldsymbol{\nu}_j) \right) \tag{7.18}$$

3. En la racine, $r \in \mathcal{S}^0$ :

$$A_0(\omega_i, \boldsymbol{\nu}_j) = p(\omega_i, \boldsymbol{\nu}_j) \times \prod_{s^+ \in \mathcal{E}_0} \left( \sum_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_1(s^+; \omega, \boldsymbol{\nu}, \omega_i, \boldsymbol{\nu}_j) \right); \tag{7.19}$$

où $\mathcal{E}_0 = \mathcal{S}^1$.

L'estimation du MPM repose sur les mêmes calculs que dans la section 7.2.2 :

4. Sélection de la classe la plus probable en la racine, $s \in \mathcal{S}^0$, d'après :

$$p(\omega_i, \boldsymbol{\nu}_j | \mathbf{y}) = \frac{A_0(\omega_i, \boldsymbol{\nu}_j)}{\sum\limits_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_0(\omega, \boldsymbol{\nu})}. \tag{7.20}$$

5. Sélection de la classe la plus probable en chaque site des résolutions intermédiaires, $\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^N\}$, d'après :

$$p(\omega_i, \boldsymbol{\nu}_j | \omega_k, \boldsymbol{\nu}_l, \mathbf{y}) = \frac{A_n(s; \omega_i, \boldsymbol{\nu}_j, \omega_k, \boldsymbol{\nu}_l)}{\sum\limits_{\substack{\omega \in \Omega_x \\ \boldsymbol{\nu} \in \Omega_v}} A_n(s; \omega, \boldsymbol{\nu}, \omega_k, \boldsymbol{\nu}_l)}. \tag{7.21}$$

### 7.3.3 Estimation des paramètres

Dans la suite, nous retenons un modèle de bruit gaussien de paramètres de moyenne $\mu_k$ et d'écart-type $\sigma_k$ et nous considérons une segmentation en $|\Omega_x| = 2$ classes. Les paramètres de ce modèle sont :

$$\theta = \{\mu_0, \mu_1, \sigma_0, \sigma_1, \boldsymbol{\pi}, \beta, \alpha\}. \tag{7.22}$$

En contexte non supervisé, les paramètres sont estimés à partir de $\mathbf{Y} = \mathbf{y}$ seulement. Nous présentons dans un premier temps les estimateurs sur les données complètes $(\mathbf{x}, \mathbf{y}, \mathbf{v})$. L'estimation de $\mu_k$ et $\sigma_k$ se fait avec les estimateurs du maximum de vraisemblance sur le couple $(\mathbf{x}, \mathbf{y})$. L'estimation des composantes $\pi_{i,j}$ de $\boldsymbol{\pi}$ se fait directement à partir du calcul des probabilités *a posteriori* (7.7) :

$$\hat{\pi}_{i,j} = p(X_r = \omega_i, \mathbf{V}_r = \boldsymbol{\nu}_j | \mathbf{Y} = \mathbf{y}). \tag{7.23}$$

Nous détaillons maintenant les estimations de $\alpha$ et $\beta$, inspirées de l'estimation au sens des moindres carrés [Derin et Elliott, 1987] pour des distributions de champs de Markov. Pour $s \in \{\mathcal{S}^1, \dots, \mathcal{S}^N\}$ et pour tous $\omega_i, \omega_j \in \Omega_x$ tels que $\omega_i \neq \omega_j$ :

$$\frac{p(X_s = \omega_i | T_{s^-})}{p(X_s = \omega_j | T_{s^-})} = \exp\left[\alpha\left(\delta_{X_{s^-}}^{\omega_i} - \delta_{X_{s^-}}^{\omega_j}\right) + \beta \sum_{V^- \in \mathbf{V}'_{s^-}} (\delta_{V^-}^{\omega_i} - \delta_{V^-}^{\omega_j})\right]. \tag{7.24}$$

Le terme de gauche est estimé grâce aux distributions *a posteriori* (7.9) [2].

Cela nous amène à considérer l'estimation « partielle » de $\alpha$ pour tout couple $(x_s, \mathbf{t}_{s^-})$ :

$$\hat{\alpha}_{s,s^-} = \frac{1}{\delta_{x_{s^-}}^{\omega_i} - \delta_{x_{s^-}}^{\omega_j}} \left(\log\left[\frac{p(x_s = \omega_i | \mathbf{t}_{s^-}, \mathbf{Y} = \mathbf{y})}{p(x_s = \omega_j | \mathbf{t}_{s^-}, \mathbf{Y} = \mathbf{y})}\right] - \beta \sum_{v^- \in \mathbf{v}'_{s^-}} (\delta_{v^-}^{\omega_i} - \delta_{v^-}^{\omega_j})\right). \tag{7.25}$$

**Remarque 7.3.2.** *Lorsque le modèle comporte deux classes, le dénominateur* $(\delta_{x_{s^-}}^{\omega_i} - \delta_{x_{s^-}}^{\omega_j})$ *est toujours différent de zéro.*

L'estimation de $\alpha$ au sens des moindres carrés est donc :

$$\hat{\alpha} = \frac{1}{|\mathcal{S} \setminus \mathcal{S}^0|} \sum_{s \in \mathcal{S} \setminus \mathcal{S}^0} \hat{\alpha}_{s,s^-}. \tag{7.26}$$

En suivant le même raisonnement, nous pouvons avec (7.9) poser pour tout $(x_s, \mathbf{t}_{s^-})$ :

$$\hat{\beta}_{s,s^-} = \frac{1}{\sum_{v^- \in \mathbf{v}'_{s^-}} (\delta_{v^-}^{\omega_i} - \delta_{v^-}^{\omega_j})} \left(\log\left[\frac{p(x_s = \omega_i | \mathbf{t}_{s^-}, \mathbf{Y} = \mathbf{y})}{p(x_s = \omega_j | \mathbf{t}_{s^-}, \mathbf{Y} = \mathbf{y})}\right] - \alpha\left(\delta_{x_{s^-}}^{\omega_i} - \delta_{x_{s^-}}^{\omega_j}\right)\right). \tag{7.27}$$

Lorsque le dénominateur $\sum_{v^- \in \mathbf{v}'_{s^-}} (\delta_{v^-}^{\omega_i} - \delta_{v^-}^{\omega_j}) = 0$, cette estimation « partielle » n'est pas définie. En notant $C(s^-, i, j)$ cette grandeur, l'estimation de $\beta$ au sens des moindres carrés est donc :

$$\hat{\beta} = \frac{\sum_{s \in \mathcal{S} \setminus \mathcal{S}^0} \mathbb{1}_{\{C(s^-,i,j) \neq 0\}} \hat{\beta}_{s,s^-}}{\sum_{s \in \mathcal{S} \setminus \mathcal{S}^0} \mathbb{1}_{\{C(s^-,i,j) \neq 0\}}}. \tag{7.28}$$

L'ensemble des paramètres $\theta$ peut donc être estimé à partir des données complètes $(\mathbf{x}, \mathbf{y}, \mathbf{v})$. De la même manière que dans les chapitres 5 et 6, en contexte non supervisé les paramètres sont estimés par une méthode inspirée de l'algorithme SEM. Celui-ci est itératif, et produit la suite d'estimations $\{\theta_0, \dots, \theta_Q\}$ en effectuant à chaque itération $q$ :

1. la simulation de $(\hat{\mathbf{x}}^q, \hat{\mathbf{v}}^q)$ selon $p_{\theta_{q-1}}(\mathbf{X}, \mathbf{V} | \mathbf{Y} = \mathbf{y})$ (cf. section 7.2.2),
2. l'estimation des paramètres grâce aux estimateurs sur les données complètes simulées $(\hat{\mathbf{x}}^q, \mathbf{y}, \hat{\mathbf{v}}^q)$.

Cet algorithme n'est pas un algorithme SEM car nous n'utilisons pas les estimateurs au sens du maximum de vraisemblance pour les paramètres $\alpha$ et $\beta$.

2. Cet estimateur s'avère, en pratique, plus robuste que l'estimation par les fréquences.

## 7.4 Résultats numériques

### 7.4.1 Simulations selon le modèle AMTS

Nous nous intéressons dans un premier temps aux réalisations du modèle AMTS. Celui-ci permet la prise en compte simultanée de dépendances hiérarchiques et d'homogénéité au sein des résolutions, traduites par les paramètres $\alpha$ et $\beta$. La figure 7.6 illustre plusieurs réalisations de $\mathbf{X}$ au sein du modèle AMTS avec les paramètres $\alpha$ et $\beta$ variables :

— lorsque la valeur des deux paramètres diminue, nous observons une réalisation plus « mouchetée ». Cela indique que les probabilités de conserver les classes dans les transitions parents/enfants sont plus faibles.
— lorsque $\alpha$ augmente, nous observons une structure hiérarchique plus prononcée, similaire à celle observable au sein d'un modèle d'AMC.
— alternativement, lorsque $\beta$ augmente, nous observons l'apparition d'une certaine homogénéité spatiale semblable à celle observée dans un modèle de champ de Markov.

L'inspection visuelle des réalisations simulées ne permet pas, à elle seule, de juger la qualité des segmentations possibles dans ce modèle. Ces réalisations sont en revanche une bonne illustration de la richesse de modélisation engendrée par le modèle AMTS.

### 7.4.2 Performances

Pour mesurer les performances du modèle, nous utilisons deux images de synthèse de taille $128 \times 128$ pixels :

— cas A : $\mathbf{x}_A$ est générée à l'aide d'une simulation d'AMT avec $\alpha = 3{,}0$ et $\beta = 0{,}3$ (figure 7.7a) ;
— cas B : $\mathbf{x}_B$ présente de grandes régions homogènes avec des contours plus nets et elliptiques (figure 7.7b).

Nous prenons par ailleurs $\mu_0 = 0$, $\mu_1 = 1$, et $\sigma_0 = \sigma_1 = \sigma$ définie en fonction du RSB, de sorte à avoir :

$$\mathrm{RSB} = 20 \log_{10} \left( \frac{\overline{\mu}}{\sigma} \right) ; \tag{7.29}$$

où $\overline{\mu}$ est défini comme :

$$\overline{\mu} = \frac{\sum_{s \in \mathcal{S}^N} \delta_{x_s}^{\omega_0} \mu_0 + \delta_{x_s}^{\omega_1} \mu_1}{|\mathcal{S}^N|}. \tag{7.30}$$

Nous comparons les résultats de segmentations obtenus d'après les modèles :

— de champs de Markov cachés à bruit indépendant (CMC) [Marroquin et al., 1987] ;
— d'arbres de Markov caché à bruit indépendant (AMC) [Laferté et al., 2000] ;
— AMTS présenté dans ce chapitre.

Toutes les segmentations sont obtenues au sens du MPM, et les paramètres sont estimés avec des algorithmes SEM ou inspirés de l'algorithme SEM [3].

Des expériences préliminaires ont montré que le modèle CMC offrait des performances de segmentation de mauvaise qualité. Nous proposons donc un modèle « mixte » CMC/AMC pour concurrencer le modèle AMTS qui cumule les avantages des arbres et des champs de Markov. Pour cela, nous choisissons d'initialiser les valeurs de $\mathbf{x}$, dans les échantillonneurs de Gibbs de l'algorithme de Marroquin (cf. Annexe A.2.1), par le résultat de la segmentation AMC.

3. Il ne s'agit en effet pas d'algorithmes SEM dans le cadre de champs de Markov cachés, car le paramètre de « granularité » ne dispose pas, à notre connaissance, d'estimateurs au sens du maximum de vraisemblance (cf. chapitres 5 et 6).

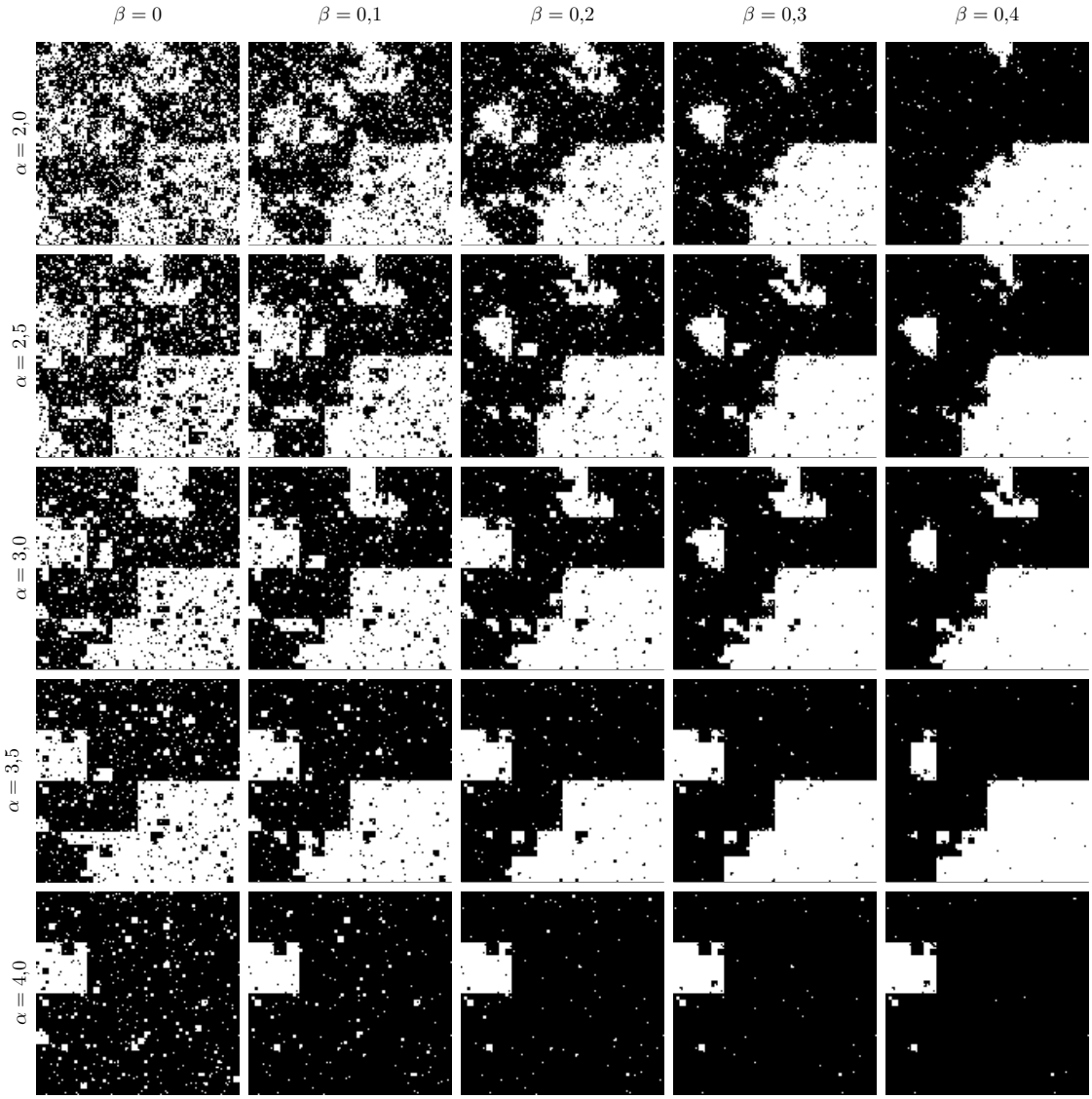
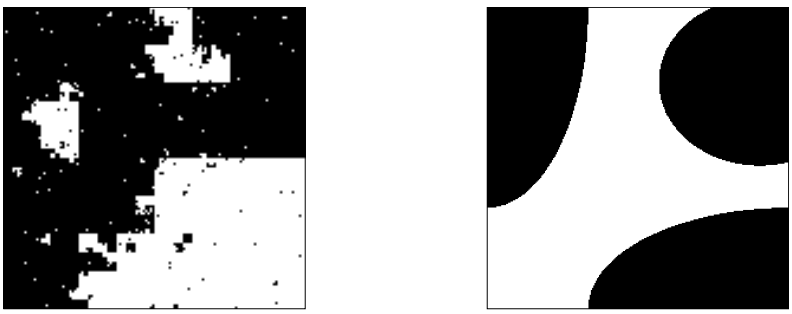


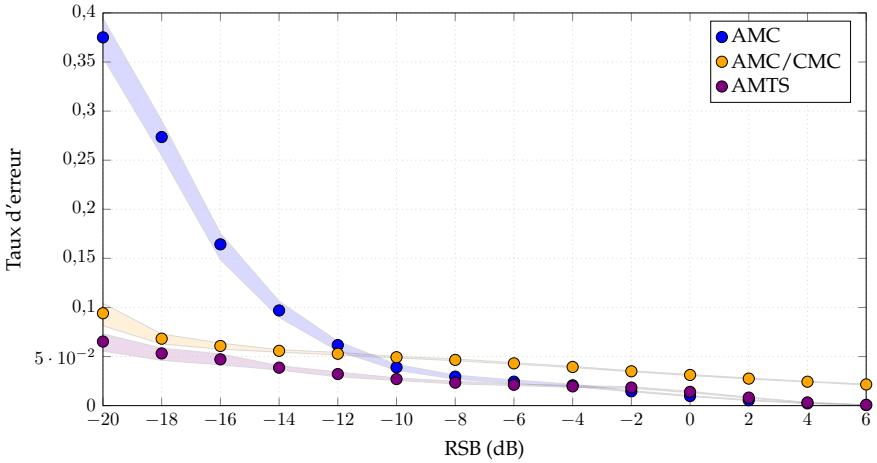

Figure 7.6 – Réalisations de $\mathbf{X}^N$ selon le modèle AMTS avec les paramètres $\alpha$ et $\beta$ variables, avec 7 résolutions ($N = 6$). Le générateur de nombre pseudo-aléatoire a la même valeur initiale pour toutes les réalisations.



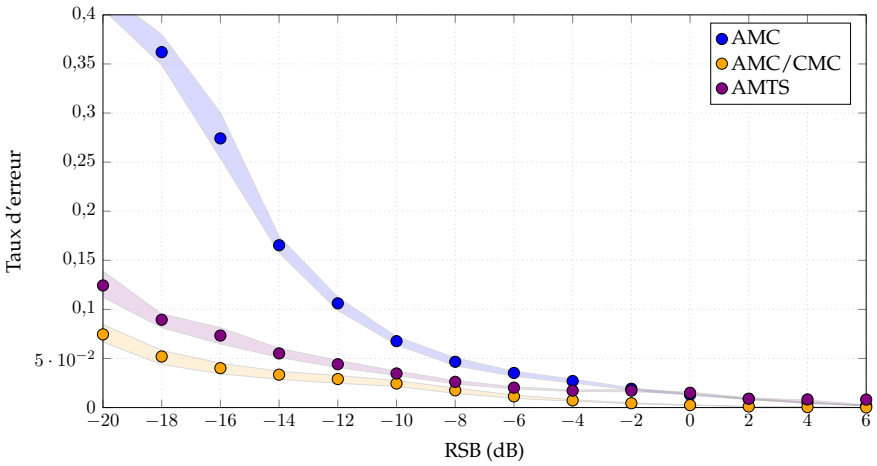
(a) Réalisation $\mathbf{x}_A$ simulée à partir du modèle AMTS.

(b) Réalisation $\mathbf{x}_B$.

Figure 7.7 – Réalisations de $\mathbf{x}$ employées pour l'obtention des résultats numériques.




(a) Taux d'erreur moyens selon le RSB (dB) avec $\mathbf{x}_A$ (figure 7.7a).




(b) Taux d'erreur moyens selon le RSB (dB) avec $\mathbf{x}_B$ (figure 7.7b).

Figure 7.8 – Taux d'erreur moyens obtenus par les modèles AMC, mixte AMC/CMC et AMTS en fonction du RSB. Les résultats sont obtenus sur 100 réalisations $\mathbf{Y} = \mathbf{y}$, et les enveloppes colorées représentent les intervalles entre le premier et le troisième quartile des résultats.

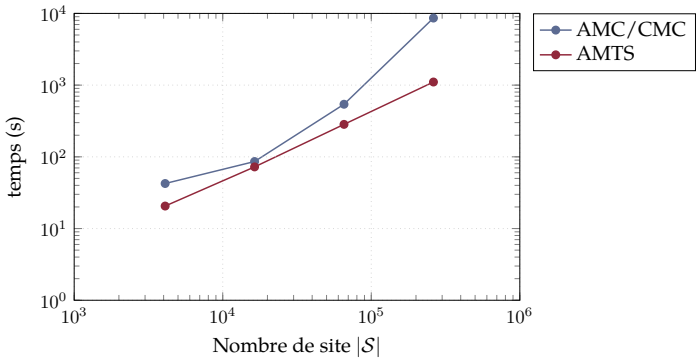



Figure 7.9 – Temps de calculs moyens requis pour la segmentation selon les modèles mixte AMC/CMC et AMTS (10 réalisations).

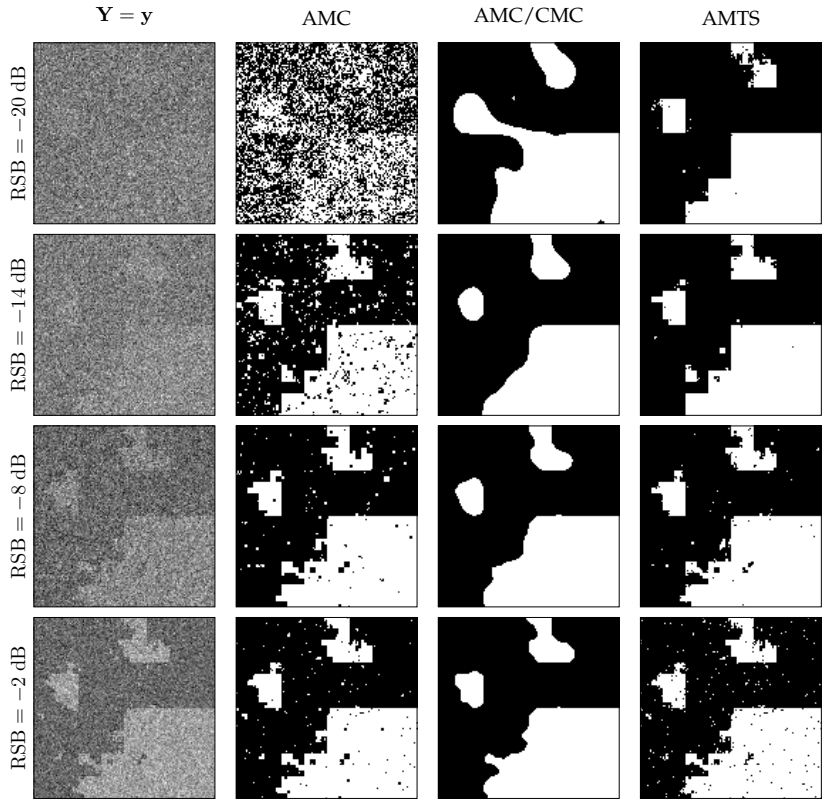
Les expériences ont été répétées avec 100 réalisations différentes $\mathbf{Y} = \mathbf{y}$. L'ensemble des taux d'erreur moyens mesurés est reporté en figure 7.8, et la figure 7.10 illustre quelques résultats de segmentation obtenus selon ces deux modèles. L'étude de ces résultats met en évidence plusieurs phénomènes :

— pour les deux réalisations de $\mathbf{x}$ étudiées, le modèle AMTS offre des performances systématiquement meilleures que le modèle AMC. Ce dernier rencontre en effet des difficultés lors de la segmentation d'images très bruitées ($\mathrm{RSB} < -14$ dB), pour lesquelles le modèle semble privilégier les informations provenant de $\mathbf{Y} = \mathbf{y}$ par rapport à celles du processus $\mathbf{X}$.
— *a contrario*, la méthode basée sur le modèle mixte AMC/CMC permet une évolution relativement « douce » des taux d'erreur en fonction du RSB. Notons en particulier qu'à fort RSB, le taux d'erreur moyen est presque nul (inférieur à $1\%$) pour le cas B : en effet, cette image semble bien se prêter à un *a priori* de champ de Markov sur $\mathbf{x}$. La situation est opposée concernant les réalisations du cas A, pour lesquelles le taux d'erreur ne devient pas aussi faible.
— le modèle AMTS offre des performances de segmentation comparables à celles fournies par le modèle mixte AMC/CMC. En effet, ce modèle offre les meilleurs taux d'erreur à tout RSB pour la segmentation dans le cas A, qui correspond précisément à une réalisation selon le modèle AMTS. Dans le cadre de la segmentation dans le cas B, le modèle est légèrement moins performant que le modèle mixte AMC/CMC. Rappelons cependant que ce dernier a été « renforcé » par l'utilisation des résultats AMC. Les résultats obtenus bénéficient, pour partie, de cette dernière modélisation. En conséquence, la segmentation basée sur le modèle AMTS égale et peut surpasser la segmentation par champs de Markov. De plus, le type d'erreur commise diffère : à faible RSB le modèle AMTS semble préserver les structures à grande échelle, au contraire de la segmentation par AMC/CMC.

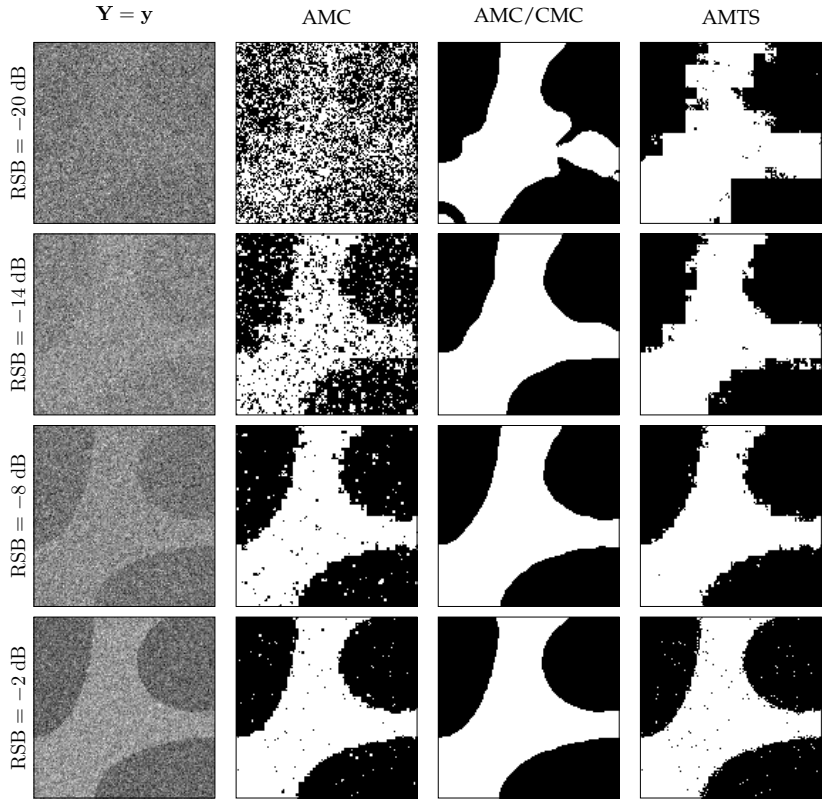
Rappelons enfin qu'au contraire du modèle mixte basé sur un modèle par champs de Markov, le modèle AMTS permet le calcul analytique des distributions nécessaires à la segmentation, ce qui agit de manière favorable sur les temps de calculs nécessaires à la segmentation. Ce phénomène est illustré en figure 7.9. Notons également que le modèle mixte repose sur une version optimisée de l'échantillonneur de Gibbs, présentée en annexe A.1.2.

## 7.5 Conclusion

Dans ce chapitre, nous avons introduit un modèle d'arbres de Markov triplets permettant la prise en compte conjointe de l'aspect hiérarchique entre résolutions et de l'homogénéité spatiale au sein d'une résolution. Nous avons de plus présenté les estimateurs des paramètres, permettant la mise en place d'un algorithme d'estimation non supervisée inspiré de l'algorithme SEM. La confrontation des résultats de segmentation avec l'alternative basée sur le modèle AMC a mis en évidence l'apport de cet enrichissement de la modélisation. De plus, nous avons montré que la méthode de segmentation basée sur le modèle proposé offre des résultats comparables à ceux fournis par la segmentation mixte AMC/CMC, tout en permettant le calcul analytique des distributions requises pour la segmentation. Ce point a pour avantage immédiat une nette réduction des temps de calculs nécessaires à la segmentation. Les résultats illustrent également la robustesse du modèle, avec notamment des résultats de segmentation de bonne qualité même au sein d'images très bruitées. Des résultats complémentaires, concernant l'application à la détection dans des images hyperspectrales astronomiques, sont reportés dans le chapitre 10.




(a) Résultats de segmentation le cas A (figure 7.7a).




(b) Résultats de segmentation pour le cas B (figure 7.7b).

Figure 7.10 – Exemples de segmentations obtenues selon les modèles AMC, CMC et AMTS sous RSB variable.

# III

# Application aux données réelles MUSE

**Résumé.** Dans cette partie, nous présentons l'application des méthodes développées dans les chapitres 3, 5, 6 et 7 aux images hyperspectrales astronomiques issues du spectro-imageur MUSE. La problématique astronomique, ainsi que l'instrument MUSE, sont tout d'abord présentés dans le chapitre 8. Nous proposons ensuite, dans le chapitre 9, une analyse du bruit affectant les données réelles, en mettant en évidences les phénomènes de corrélations et la distribution des intensités lumineuses. Le chapitre 10 propose enfin une évaluation comparative des méthodes proposées sur des images de synthèse et sur les images MUSE.

# 8

# Problématique astronomique



Ce chapitre présente le contexte astronomique des travaux présentés dans ce manuscrit. Ce cadre applicatif répond, d'une part, à une problématique astronomique que nous présentons dans la section 8.1. D'autre part, il correspond à des observations de l'instrument MUSE que nous présentons dans la section 8.2.

## 8.1 Contexte cosmologique

La problématique astronomique consiste à détecter des filaments inter-galactiques et des halos circum-galactiques. Nous présentons dans un premier temps la place de ces structures dans l'histoire de l'Univers, telles que prédites par les modèles cosmologiques actuels.

### 8.1.1 Une brève chronologie de l'Univers jeune

La chronologie de l'Univers selon le modèle standard [Liddle, 2015] [1] commence avec le Big Bang, point auquel l'Univers est extrêmement dense. Cet événement date d'environ 13,82 milliards d'années, cette date étant évaluée d'après les contraintes cosmologiques établies avec les observations et le modèle physique standard [2].

Le Big Bang n'est pas observable directement. En effet, l'Univers jeune est trop dense pour permettre à la lumière de circuler librement par rapport à la matière : il est opaque. En se dilatant, la densité de l'Univers diminue, ce qui conduit à l'étape dite de *recombinaison*

1. Une ressource tutorielle sur la cosmologie peut être également être trouvée sur `http://www.astro.ucla.edu/~wright/cosmolog.htm`.

2. Il s'agit du modèle ΛCDM. Il postule notamment que l'Univers est homogène à grande échelle, que la relativité générale s'applique et que l'Univers est en expansion accélérée. Le lecteur pourra se reporter à [Liddle, 2015] pour une introduction aux principes cosmologiques correspondants.

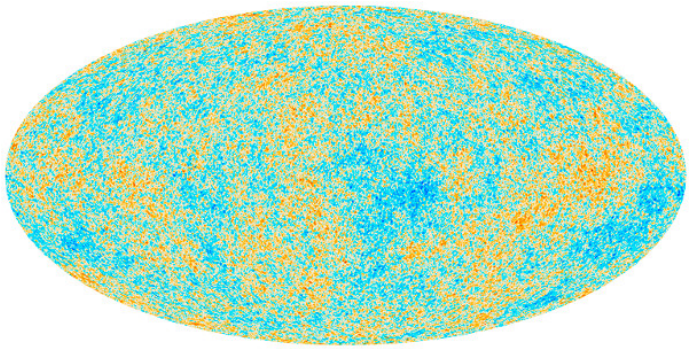


Figure 8.1 – Carte du fond diffus cosmologique issue des observations du satellite Planck en 2012. Les inhomogénéités sont représentées en $\mu$K par rapport à la température de fond de 2,728 $K$. Source : http://www.cosmos.esa.int/web/planck.

à laquelle la lumière est découplée de la matière et circule librement : l'Univers devient transparent. En conséquence, le plus ancien rayonnement observable est celui émis lors de la recombinaison, et constitue le *fond diffus cosmologique* daté d'environ 380 000 ans après le Big Bang. Ce fond est homogène à grande échelle, mais présente de légères inhomogénéités illustrées dans la figure 8.1. Ces irrégularités sont importantes à deux titres : elles traduisent les phénomènes antérieurs à la recombinaison, et forment les premières variations de densités observables dans l'Univers.

Après l'émission du fond diffus cosmologique, l'Univers ne possède pas de sources lumineuses. Cette période est parfois appelée *âges sombres*, et se termine avec la formation des premières étoiles et galaxies environ 200 millions d'années après le Big Bang. Le principal moteur de ce changement est la gravité, qui permet l'accrétion de matériau brut (hydrogène principalement) puis l'ignition des réactions nucléaires alimentant les étoiles.

La gravité reste le moteur de l'évolution à grande échelle de l'Univers, permettant la formation des galaxies et d'amas de galaxies. Le phénomène conséquent est celui d'une accrétion progressive de la matière, pour former des régions très denses d'une part, et très peu denses d'autre part. Cette accrétion est illustrée en figure 8.2. Les régions d'accrétion prennent successivement la forme de *parois*, de *filaments* puis d'*amas* dans le temps [3].

La compréhension de la formation de ces structures est importante car elles caractérisent la distribution de la matière à très grande échelle dans l'Univers. De plus, le modèle standard postule l'existence de la matière noire. Bien qu'invisible, elle représente 90% de la masse

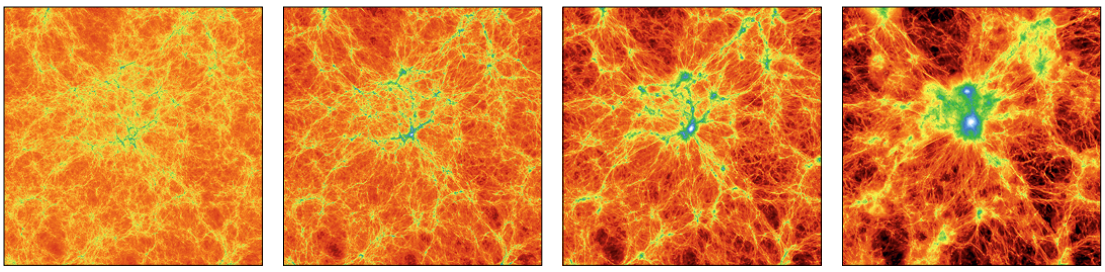


Figure 8.2 – Effet de l'accrétion gravitationnelle sur des masses de matière (gaz) selon le temps. Figures adaptées depuis http://www.illustris-project.org/ [Vogelsberger et al., 2014].

3. Cette succession correspond au formalisme développé par Zel'dovich. Le lecteur pourra se reporter à [Hidding et al., 2013] pour une description analytique des géométries d'accrétion.

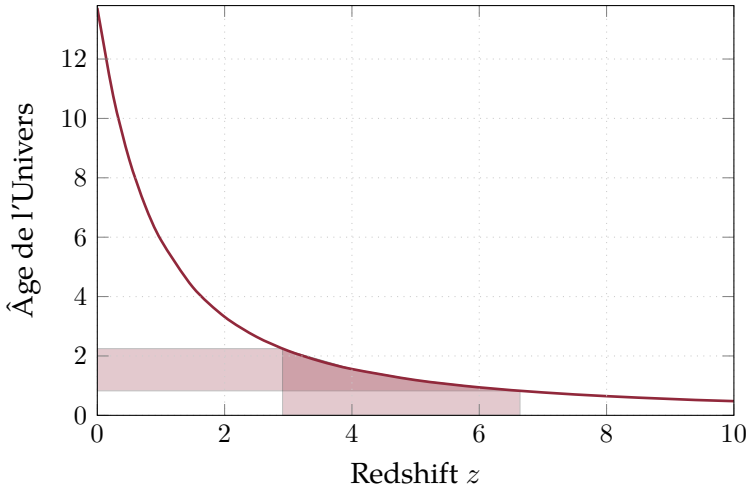



Figure 8.3 – Âge de l'Univers (en milliards d'années) lors de l'émission de la lumière en fonction du redshift observé. La région opaque correspond aux objets étudiés dans cette thèse (cf. section suivante et chapitre 10). Valeurs issues de `http://www.astro.ucla.edu/~wright/ACC.html` [Wright, 2006].

de matière dans l'Univers. Elle a donc un rôle majeur dans l'accrétion gravitationnelle, et observer des structures à grande échelle permettrait d'inférer la distribution de matière noire dans l'espace et selon l'âge de l'Univers.

L'ignition des premières étoiles a déclenché, par le biais de la photo-ionisation, le passage d'un hydrogène neutre à un hydrogène ionisé. Ce phénomène, nommé *ré-ionisation*, est le dernier grand événement cosmologique. Cet événement est situé à un âge de l'Univers proche d'un milliard d'années. Bien que le principe et la période soient connus, les sources et la durée de la ré-ionisation restent encore mal comprises.

Pour mesurer la distance des objets très lointains de l'Univers jeune, le seul outil disponible est la mesure du *redshift*. Celui-ci permet de repérer la distance d'un objet par rapport à l'observateur *via* le décalage de son spectre vers le rouge, dû à l'effet Doppler-Fizeau [Liddle, 2015]. Pour un phénomène caractérisé en laboratoire à une longueur d'onde $\lambda_0$, et observé sur un corps céleste à $\lambda_{\text{obs}}$, le redshift $z$ est défini comme :

$$z = \frac{\lambda_{\text{obs}} - \lambda_0}{\lambda_0} \tag{8.1}$$

Un redshift nul correspond à un objet très proche, et un redshift élevé correspond à un objet lointain. Plusieurs phénomènes contribuent au redshift :

— l'expansion de l'Univers ;
— le mouvement particulier de l'objet ;
— la cinématique de l'objet.

Pour les objets lointains, les deux derniers phénomènes sont négligeables par rapport au premier. Ainsi, la connaissance du redshift, d'un modèle cosmologique régissant l'évolution de l'Univers et de ses paramètres permettent d'inférer l'âge de l'Univers lorsque la lumière a été émise. Cette relation est illustrée en figure 8.3.

### 8.1.2 Halos circum-galactiques et filaments inter-galactiques

Le phénomène d'accrétion gravitationnelle mentionné dans la partie précédente a permis la formation des étoiles et galaxies, mais toute la matière n'a pas été immédiatement consommée dans ce processus. Il existe donc dans l'Univers des structures à grande échelle,

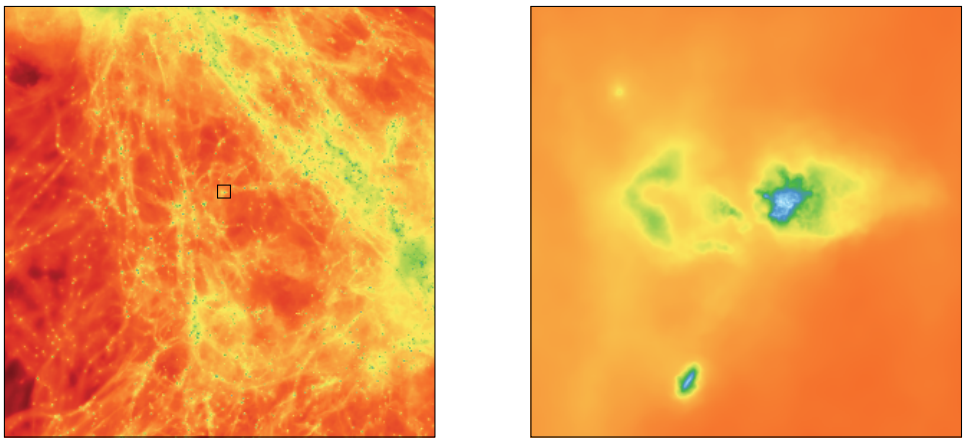


Figure 8.4 – IGM (à gauche) et CGM (à droite), représentés par la densité de gaz. L'image du CGM est issue de la région encadrée de l'image de gauche. Cette dernière illustre en particulier les différentes échelles dans la structure filamentaire. Source : http://www.illustris-project.org/ [Vogelsberger et al., 2014].

résiduelles et composées de gaz uniquement. La structure à grande échelle de la matière forme la *toile cosmique* (*cosmic web*). Elle peut être décrite comme la combinaison de régions :

— très peu denses, formant de très grands « vides » ;
— très denses : les galaxies et les amas de galaxies ;
— intermédiaires : il s'agit de *filaments* de matière très peu visibles, que l'on appelle l'*inter-galactic medium* (IGM).

À plus petite échelle, les jeunes galaxies elles-mêmes sont hébergées par un amas de gaz. Cet amas a lui aussi pour origine l'hydrogène qui n'a pas encore été consommé par la galaxie. Nous parlons dans ce cas de *circum-galactic medium* (CGM) ou encore de *halos*. Le CGM est en interaction plus forte avec la galaxie hébergée, et peut alors comporter une fraction non négligeable de matériau plus complexe que l'hydrogène. La figure 8.4 donne un exemple de structures de l'IGM et du CGM, issues d'une simulation de l'évolution de la matière dans l'univers [Vogelsberger et al., 2014].

IGM et CGM sont tous deux composés de gaz neutre, et n'émettent pas de lumière spontanément. C'est pourquoi leur observation ne peut être qu'indirecte. Les premières mises en évidences ont été faites grâce à un phénomène d'absorption de la lumière : si de l'hydrogène neutre se situe entre une source lumineuse (le plus souvent, un quasar) et l'observateur, le spectre observé de cette source sera atténué aux longueurs d'ondes de l'hydrogène. La longueur d'onde de l'absorption renseigne sur la composition et la vitesse du milieu traversé. Les absorptions peuvent être multiples : elles traduisent alors la traversée de plusieurs régions d'hydrogène neutre par la lumière. Bien que la caractérisation spectrale puisse être très fine, cette technique dépend du bon alignement de l'observateur, d'un quasar et d'un élément de l'IGM ou du CGM. Elle est donc limitée spatialement : la caractérisation ne peut être faite que dans la ligne de visée.

Bien que CGM et IGM n'émettent pas de rayonnement par eux-même, ils peuvent diffuser le rayonnement de galaxies de leur voisinage. C'est en particulier le cas pour le rayonnement de Lyman-alpha, pour lequel on parle de rayonnement *résonnant* : la lumière provenant de jeunes étoiles est successivement absorbée et retransmise par l'hydrogène neutre jusqu'à être assez décalée spectralement pour échapper à l'absorption. Le spectre observé dans ces régions est caractéristique des phénomènes conjoints d'émission, d'absorption et de diffusion. En fonction de l'importance de chacun de ces phénomènes, la

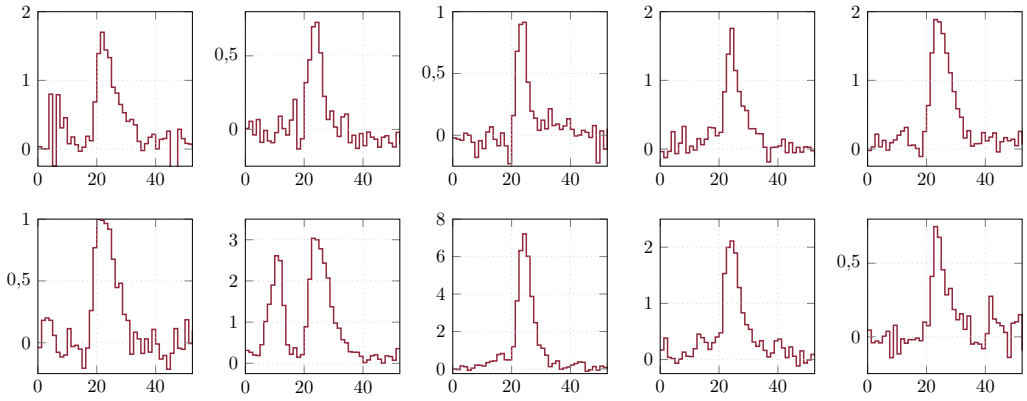


FIGURE 8.5 – Échantillon de raies d'émission de Lyman-alpha observées par l'instrument MUSE. Les longueurs d'ondes en abscisses sont graduées en Angströms ($0{,}1$ nm) et les ordonnées représentent le flux lumineux, en $10^{-20}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}}$.

forme du spectre peut significativement varier, comme l'illustre la figure 8.5.

Le rayonnement de l'hydrogène ionisé dans la raie de Lyman-alpha est donc un traceur important de l'IGM et du CGM. Son observation a d'abord été faite à l'aide de spectrographes, fournissant un échantillon du spectre lumineux d'une région ponctuelle du ciel. Cette approche permet d'affirmer la présence d'une raie de Lyman-alpha, mais pas de cartographier sa répartition spatiale. C'est pourquoi l'utilisation d'un spectrographe intégral de champ, capable de fournir une image hyperspectrale du ciel, est pertinente dans ce contexte. Notons enfin que le rayonnement de Lyman-alpha transmis par l'IGM et le CGM est extrêmement ténu, ce qui est une des principales difficultés rencontrées lors de leur observation. En conséquence, une grande sensibilité instrumentale est nécessaire pour pouvoir accéder à ces flux lumineux. Il est aussi nécessaire d'accéder à un très bon contraste pour distinguer la raie de Lyman-alpha du reste du spectre de la galaxie. L'instrument MUSE, présenté dans la partie suivante, propose une solution à cette problématique.

## 8.2 L'instrument MUSE

Plusieurs processus consécutifs interviennent dans la formation des images MUSE. Cette partie décrit la formation des données MUSE en suivant d'abord le trajet de la lumière jusqu'au capteur (section 8.2.1), puis en suivant les données du capteur à l'image hyperspectrale (section 8.2.2).

### 8.2.1 La lumière entre l'espace et le capteur

Dans la configuration actuelle [4], l'instrument MUSE permet l'observation en une pose d'une région du ciel d'une superficie de $1 \times 1$ arc-minute, c'est-à-dire un carré dont le côté vaut un trentième du diamètre apparent de la Lune.

Avant d'arriver dans le plan focal du télescope, la lumière traverse l'atmosphère terrestre. Cela a deux effets importants :

— au rayonnement lumineux de l'objet s'ajoute la *lumière du ciel nocturne*, dont l'origine se situe dans la haute atmosphère. C'est un rayonnement parasite non négligeable, en particulier pour l'observation d'objets ténus. Comme sa localisation spectrale

4. Il s'agit du « Wide field mode » (cf. [Bacon et al., 2010]).

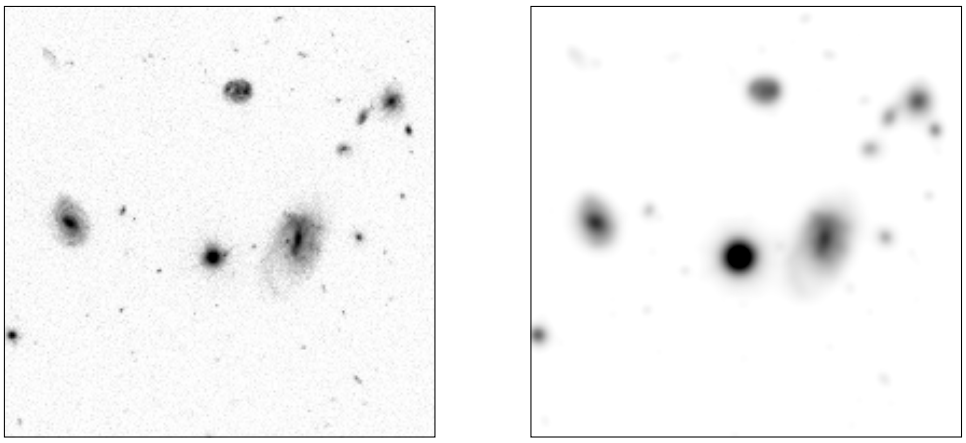


Figure 8.6 – Exemple d'étalement spatial induit par les turbulences atmosphériques. Ici, l'étalement spatial de MUSE est appliqué à une image issue du télescope spatial Hubble [Beckwith et al., 2006].

est connue, il est possible de le compenser lors du traitement des données (cf. section 8.2.2) ;

— les turbulences atmosphériques provoquent un étalement spatial de la lumière, qui est lui aussi non négligeable, comme l'illustre la figure 8.6. Cet étalement peut être évalué relativement finement, mais il peut difficilement être compensé. Ce phénomène doit donc être pris en compte lors du traitement et de l'analyse des images. D'un point de vue instrumental, cet étalement sera fortement réduit suite à l'installation d'un système d'*optique adaptative*, dont la mise en fonctionnement est prévue pour l'automne 2017.

La modélisation de la PSF de MUSE a fait l'objet de plusieurs travaux préliminaires [Serre et al., 2010; Villeneuve et al., 2011]. Le modèle retenu actuellement est une PSF séparable en :

— une composante spatiale prenant la forme d'une fonction de Moffat, de paramètres variant en fonction de la longueur d'onde [Bacon et al., 2015, 2017].

— une LSF de largeur faible par rapport à la résolution spectrale. Cela la rend ardue à modéliser : le modèle actuel est celui d'une LSF gaussienne, de largeur à mi-hauteur variant entre 2.4 et 3.2 Å [Bacon et al., 2017].

L'instrument MUSE a pour objectif de produire une image hyperspectrale du ciel à partir de la lumière acquise par le télescope. Former directement une image hyperspectrale est techniquement délicat : en effet, les capteurs sont bidimensionnels, alors que les données à acquérir sont tridimensionnelles. C'est pourquoi il est nécessaire de recourir à un réarrangement géométrique de la lumière en amont des capteurs. L'instrument sépare la lumière en 24 sous-champs, qui correspondent aux 24 *integral field units* (IFU) composant l'instrument. Au sein de chaque IFU, les sous-champs sont chacun décomposés en $12 \times 4$ blocs, de sorte à atteindre la finesse de la résolution selon une dimension spatiale [5]. La lumière des sous-blocs d'un IFU est ensuite transférée à un spectrographe, qui permet l'acquisition sur un capteur CCD d'une image spatio-spectrale du sous-champ. La figure 8.7 présente une vue des séparations géométriques de la lumière en amont des capteurs CCD. Pour une description détaillée de l'instrument, le lecteur intéressé pourra également se reporter à [Bacon et al., 2010].

5. C'est-à-dire 1 arcminute répartie en $24 \times 12$ blocs, représentés verticalement sur la figure 8.7. Cela donne une résolution de 0,208 arcseconde répartie sur 288 lignes horizontales.

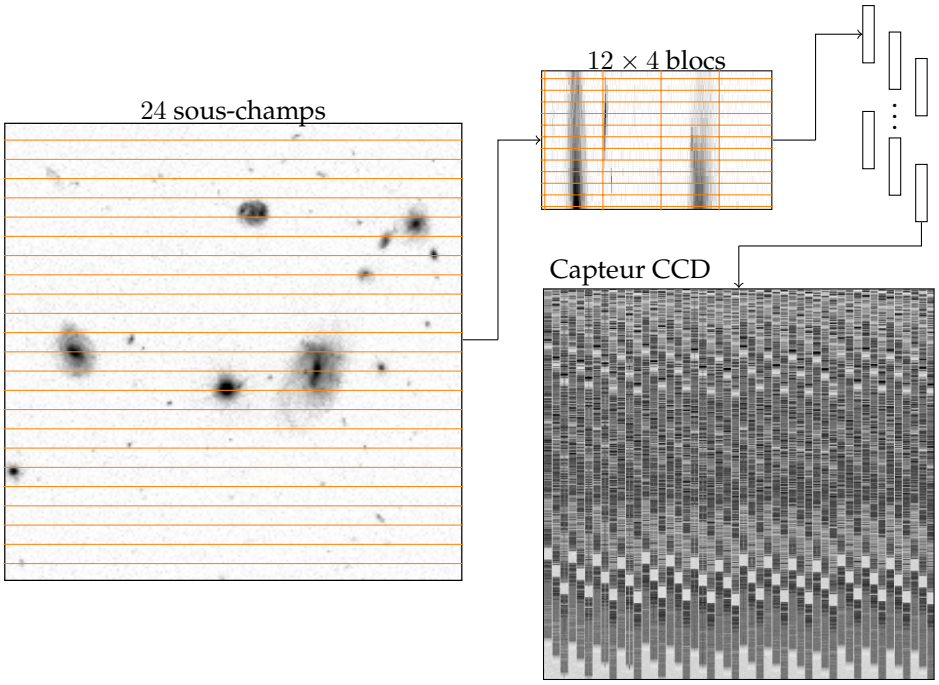



Figure 8.7 – Illustration des séparations géométriques du champ MUSE jusqu'au capteur CCD. Adapté depuis [Weilbacher et al., 2012].

### 8.2.2 Des capteurs à l'image hyperspectrale

Nous avons vu que la lumière subit un certain nombre de transformations pour pouvoir être collectée par les capteurs CCD. Il est nécessaire, pour pouvoir former une image hyperspectrale, d'inverser les processus de séparation, de sorte à regrouper les données spatialement et spectralement. De plus, plusieurs perturbations biaisent les acquisitions et nécessitent d'être corrigées pour une exploitation scientifique des données. L'ensemble des processus qui ont pour but d'effacer la signature de l'instrument constitue la *réduction des données*, que nous décrivons dans cette section.

La première partie de la réduction des données consiste en des traitements liés aux systématiques induites par l'instrument. En effet, chaque capteur CCD possède ses propres caractéristiques, qu'il faut compenser avant de pouvoir regrouper les données. Ces caractéristiques incluent le courant d'obscurité (*dark*), qui est lié à la température du capteur et peut être facilement calibré lors des observations. Une autre caractéristique importante est le *bias*, qui décrit la réponse du capteur en l'absence de lumière et forme donc un « niveau 0 » de chaque pixel du capteur. Il est plus délicat de l'estimer, et peut aussi varier dans le temps. Il est également nécessaire que les différentes sorties des 24 IFU soient homogènes : c'est l'objectif du processus de « champ plat » (*flat-fielding*). La correction de l'estimation du *flat-fielding* a permis une amélioration significative de la qualité des données sur les observations de champs profonds [Bacon et al., 2017, section 3.1.2] : cette amélioration est illustrée en figure 8.8. Les autres traitements liés à l'instrument consistent notamment en un repérage des pixels défectueux et en une calibration en longueur d'onde.

Les traitements de cette première partie de la réduction des données permettent de former un ensemble de tables de pixels, nommées *pixtables*, correctement calibrées en intensité et en longueur d'onde. La seconde partie de la réduction des données a pour objectif la formation d'une image hyperspectrale à partir de l'ensemble des *pixtables*. Cela nécessite en premier lieu une calibration spatiale des spectres observés, qui ne sont pas positionnés régulièrement en raison des agencements physiques des capteurs. Connaissant

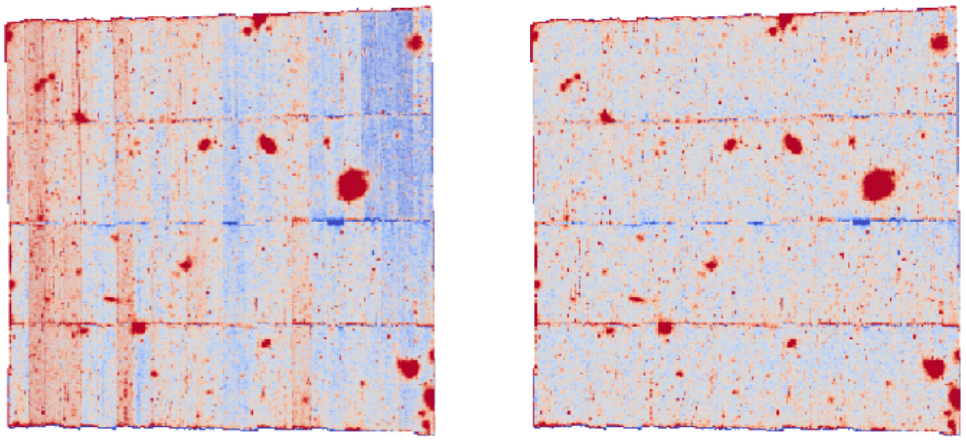


Figure 8.8 – Résultat du processus d'auto-calibration pour la correction des inhomogénéités d'une observation MUSE. La figure montre la moyenne spectrale du cube avant (à gauche) et après (à droite) auto-calibration, le rouge et le bleu correspondant aux fortes et faibles intensités respectivement. Images adaptées de [Bacon et al., 2017].

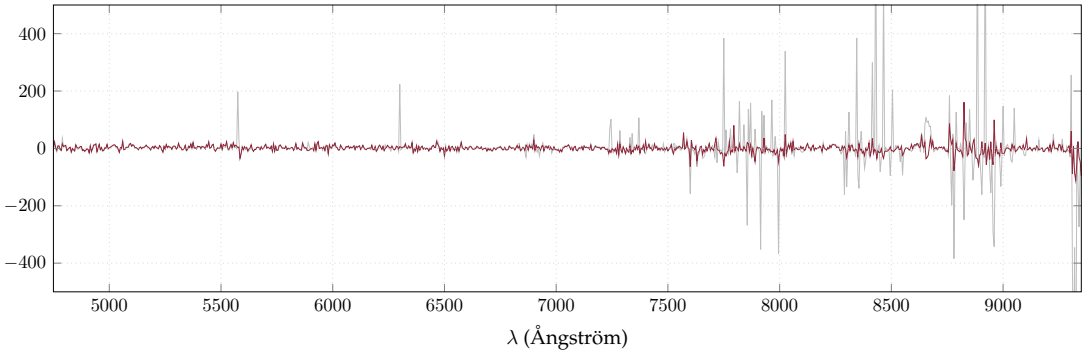



Figure 8.9 – Spectre d'une région vide dans une observation MUSE, sans (gris) et avec (rouge) l'affinement de la soustraction des raies du ciel. Les résidus de la première étape sont relativement importants et prennent aussi bien des valeurs positives que négatives. Le résultat de la seconde étape apporte une amélioration substantielle du spectre : les résidus de raies du ciel sont d'amplitude bien moindre.

la position spatiale des spectres, une étape d'interpolation permet de les ré-échantilloner sur un maillage régulier. L'interpolation est menée avec une méthode de *drizzling*. Il s'agit d'une adaptation de celle utilisée dans le cadre des observations de champs profonds Hubble [Fruchter et Hook, 1997], étendue notamment aux données hyperspectrales. Cette procédure induit des corrélations spatiales et spectrales non négligeables dans les images hyperspectrales. Dans le chapitre 9, nous proposons une évaluation de ces corrélations.

Nous avons vu que la traversée de l'atmosphère induit la présence d'un rayonnement parasite, bien localisé spectralement. Le processus de *soustraction des raies du ciel* a pour but de compenser la plus grande partie de ce rayonnement. La principale difficulté réside dans la méconnaissance de l'intensité de ces raies d'émission. Une méthode combinatoire reposant sur les modèles de raies est employée dans un premier temps. Une méthode basée sur une analyse en composantes principales [Soto et al., 2016] permet dans un second temps d'affiner le traitement. La figure 8.9 illustre cette étape de la soustraction des raies du ciel.

Dans le cadre d'observations multiples, une dernière étape de fusion des données est nécessaire pour produire une unique image hyperspectrale. Cette fusion peut concerner plusieurs observations d'une même région du ciel, ou une mosaïque d'observations. La

méthode de fusion consiste en une moyenne des pixels de la grille [6]. Cette moyenne se fait dans le cadre d'un rejet de données aberrantes [Bacon et al., 2017, section 3.2], celles-ci étant détectées par une méthode de *sigma-clipping*, consistant à rejeter les intensités trop éloignées de leur moyenne.

Une description récente du processus de réduction des données peut également être trouvée dans [Bacon et al., 2017] ainsi que dans la documentation du logiciel MPDAF [Bacon et al., 2016]. L'ensemble du processus de réduction des données aboutit à la formation d'une image hyperspectrale, comportant environ $300 \times 300$ spectres, chacun étant formé d'environ $3680$ composantes spectrales. Un champ MUSE couvre $1\ \mathrm{arcmin}^2$ avec une résolution spatiale de $0{,}2\ \mathrm{arcsec}^2$ environ. Le domaine de longueur d'onde de MUSE va de $4750$ à $9350$ Å avec une résolution spectrale de $1.25$ Å.

## Conclusion

Nous avons présenté la problématique de détection de structures lointaines, ténues et caractérisées par l'émission de Lyman-alpha de l'hydrogène ionisé. Nous avons ensuite décrit l'instrument MUSE, qui réunit les conditions nécessaires pour pouvoir cartographier spatialement ces structures grâce à leur émission caractéristique. Le processus de formation des images est complexe : le bruit de l'image est affecté par les conditions d'observation, la structure de l'instrument, et la réduction des données. C'est pourquoi nous proposons dans le chapitre suivant une étude des propriétés statistiques du bruit MUSE.

6. Dans le cas d'observations multiples, la grille est commune pour toutes les interpolations effectuées en amont.

# 9

# Caractérisation du bruit dans les données MUSE



Les données MUSE peuvent présenter, en raison des traitements nécessaires à la formation des images hyperspectrales, des structures que nous appelons *systématiques*. Il s'agit essentiellement de variations des paramètres de la distribution du bruit selon la position spatiale et spectrale au sein de l'image hyperspectrale. Nous étudions d'abord en section 9.1 les corrélations spatiales et spectrales qui apparaissent dans ces observations, afin de caractériser leur structure. La section 9.2 propose ensuite une description de la distribution statistique du bruit. Rappelons que les données hyperspectrales de champs profonds MUSE sont issues de plusieurs dizaines d'observations d'une même région du ciel, qui sont ensuite fusionnées. Nous étudierons donc à la fois les observations seules et l'observation fusionnée.

## 9.1 Évaluation des corrélations

Nous nous intéressons dans un premier temps à évaluer les corrélations présentes au sein des données MUSE. Comme présenté en figure 9.1, les corrélations sont étudiées entre :

— les spectres : position spatiale fixée et longueur d'onde variable ;
— les lignes et les colonnes du cube : la position spectrale et une coordonnée spatiale sont fixées.

Nous étudions tout d'abord les corrélations de manière qualitative dans les sections 9.1.1 et 9.1.2, ce qui met en évidence plusieurs systématiques. Ensuite, nous les étudions de manière quantitative grâce à une formulation par test d'hypothèses en section 9.1.3, afin de décrire plus exactement ces structures.

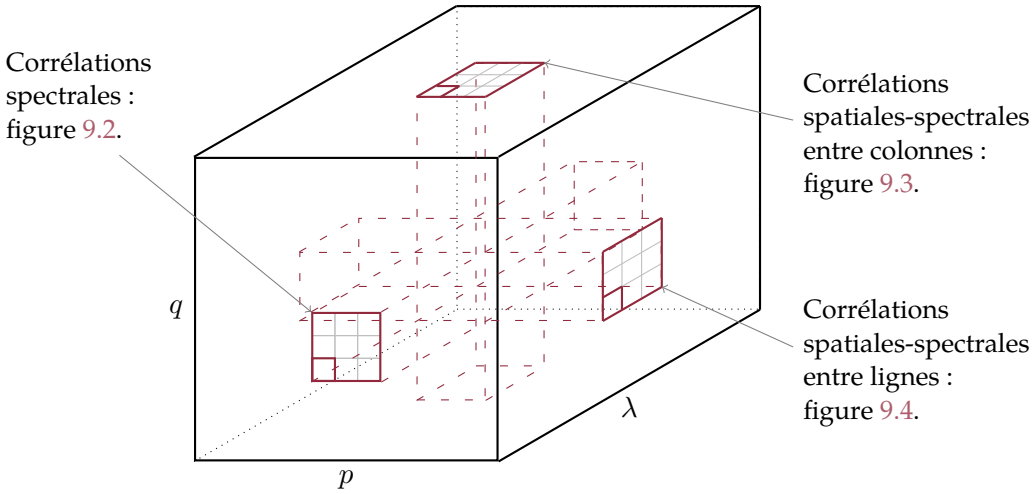



Figure 9.1 – Représentation des mesures de corrélation. Chaque grille de $3 \times 3$ éléments illustre les mesures de corrélation avec l'élément isolé en bas à gauche de chaque grille.

### 9.1.1 Mesure des corrélations spectrales

Soient deux spectres, notés $\mathbf{y}_s$ et $\mathbf{y}_t$. L'estimation de la corrélation réelle $\rho$ est la mesure de corrélation empirique suivante :

$$r_{\mathbf{y}_s,\mathbf{y}_t} = \frac{\langle \overline{\mathbf{y}}_s, \overline{\mathbf{y}}_t \rangle}{\|\overline{\mathbf{y}}_s\| \|\overline{\mathbf{y}}_t\|}; \tag{9.1}$$

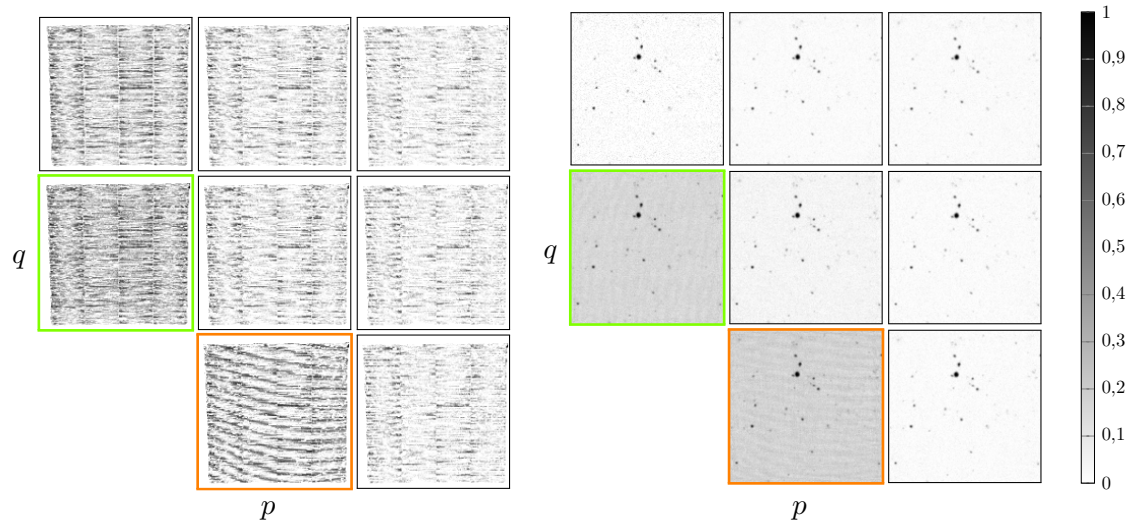
où $\langle ., . \rangle$ représente le produit scalaire et $\overline{\mathbf{y}}$ est le vecteur $\mathbf{y}$ duquel la moyenne est soustraite. $r_{\mathbf{y}_s,\mathbf{y}_t}$ peut être calculé pour toutes les combinaisons $(s, t)$ existantes. Par exemple, si nous choisissons pour $t$ le voisin horizontal immédiat de $s$ pour tout $s \in \mathcal{S}$, nous obtenons un jeu de valeurs qui peut se représenter comme une image : nous parlerons alors de *carte de corrélations* mesurées. Cette opération peut être réitérée en choisissant un autre type de spectres voisins : vertical, diagonal, séparé horizontalement d'un site, etc. La figure 9.2 illustre les cartes de corrélation mesurées sur une observation individuelle et fusionnée pour un voisinage variant entre $-2$ et 2 sites, verticalement et horizontalement. Les mesures de corrélations ne sont présentées que sur un seul quadrant, car la mesure de corrélation (9.1) est symétrique.

Plusieurs phénomènes se manifestent sur les cartes concernant l'observation individuelle (figure 9.2a). Les mesures traduisent une corrélation à la fois importante et structurée, en motifs. Ces motifs diffèrent selon la direction considérée :

— verticalement (cadre vert en figure 9.2a), nous observons en particulier les formes des quatre séparations verticales du champ décrites en section 8.2, ainsi qu'une structure horizontale plus fine qui correspond aux subdivisions du champ ;
— horizontalement (cadre orange en figure 9.2a) apparaît un motif structuré sur tout le champ. Il a pour origine l'étape d'interpolation permettant la formation de l'image hyperspectrale à partir de l'ensemble des spectres observés.

Cette différence s'explique par le fait que, par construction de l'instrument (cf. section 8.2), les deux dimensions spatiales ne sont pas traitées de la même manière.

Ces deux phénomènes sont atténués sur l'observation fusionnée, dont les cartes de corrélations sont présentées en figure 9.2b. En effet, les observations individuelles dont elle résulte sont décalées spatialement, et observées avec des orientations différentes de l'instrument. Cela permet d'atténuer les structures dues aux séparations du champ, et de produire des systématiques similaires verticalement et horizontalement. Des corrélations




(a) Cube individuel.

(b) Cube fusionné.

Figure 9.2 – Corrélations spectrales observées entre spectres voisins sur les 300 premières bandes spectrales du premier cube individuel *udf10* et du cube *udf10* fusionné. Les anti-corrélations ne sont pas affichées car elles sont très peu présentes, comme l'illustrent les histogrammes de la figure 9.5. La carte pour $(0, 0)$ n'est pas représentée car elle vaut $1$ en tout site.

non négligeables subsistent néanmoins entre voisins immédiats, au sein desquelles le motif dû à l'interpolation reste visible. Celles-ci dépendent en effet de l'orientation de l'instrument, mais pas du décalage spatial des observations. Les corrélations entre voisins non immédiats sont quant à elles négligeables. Remarquons enfin que sur les cartes caractérisant l'observation fusionnée (figure 9.2b), nous pouvons observer de petites régions fortement corrélées : elles correspondent aux galaxies observées, présentant des spectres similaires entre pixels voisins. Ces corrélations sont invisibles pour l'observation individuelle (figure 9.2a) car le RSB est trop faible pour visualiser ces objets.

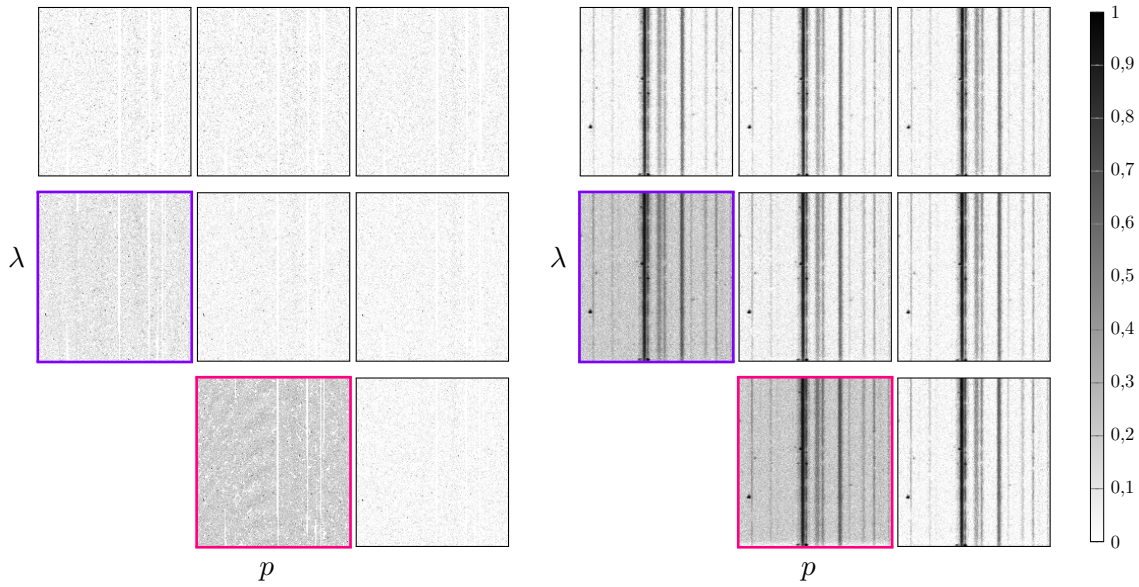
Les cartes de corrélation spatiale de la figure 9.2 sont mesurées sur les 300 premières bandes spectrales. En effet, les motifs évoluent en fonction des bandes spectrales considérées, ce qui conduit à sous-évaluer les corrélations si nous nous intéressons, par exemple, à toutes les bandes spectrales [1]. En conséquence, mesurer la corrélation sur un plus grand nombre de bandes peut mener à conclure en une absence de corrélation. Il est donc pertinent de mesurer les corrélations spectrales au sein des cubes MUSE, et de mesurer leur évolution selon les bandes spectrales considérées.

### 9.1.2 Mesure des corrélations spatio-spectrales

Les données MUSE étant tri-dimensionnelles, les cartes de la figure 9.2 ne permettent pas de visualiser la variation spectrale de ces phénomènes. C'est pourquoi nous établissons maintenant des cartes de corrélation *spatio-spectrales* en mesurant les corrélations (9.1) entre les vecteurs représentant les colonnes ou les lignes de l'image hyperspectrale.

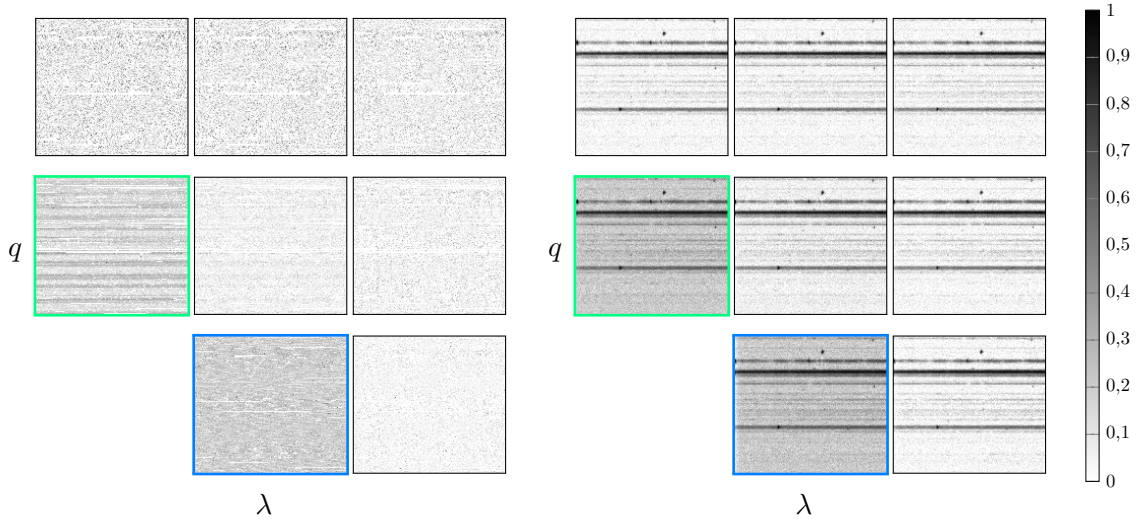
Les corrélations correspondantes sont présentées en figure 9.3 et 9.4. Ces figures montrent que les deux types de corrélations spatio-spectrales sont très semblables. De plus, les cartes

1. Remarquons également que les résidus de raies du ciel, énergétiques et homogènes spatialement, ont dans ce cas une influence importante sur les cartes de corrélations.



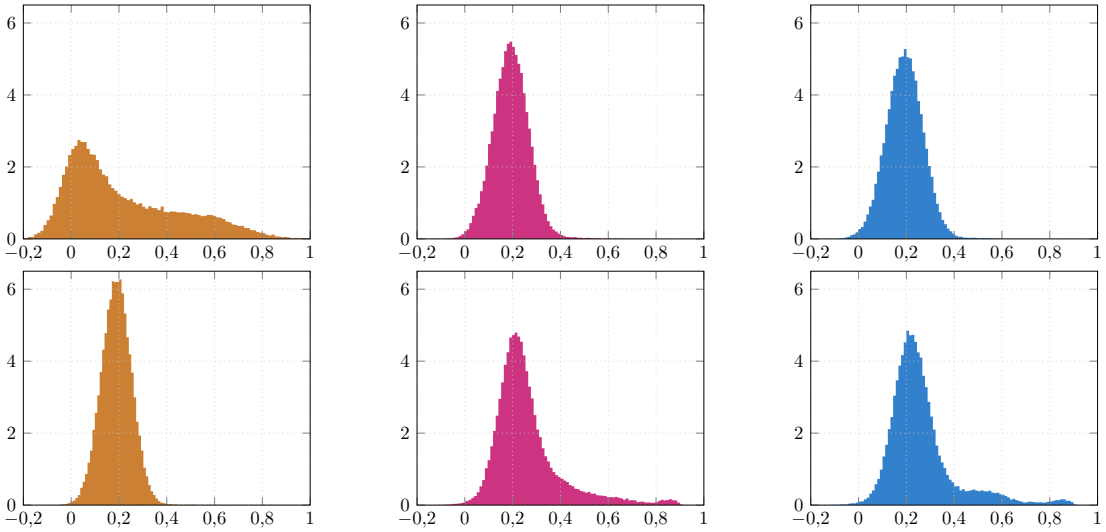

(a) Cube individuel. (b) Cube fusionné.

Figure 9.3 – Cartes des corrélations entre colonnes (même légende que la figure 9.2).




(a) Cube individuel. (b) Cube fusionné.

Figure 9.4 – Cartes des corrélations spatiales-spectrales (même légende que la figure 9.2).




(a) Corrélations de spectres avec leurs voisins horizontaux (cadres oranges en figure 9.2).

(b) Corrélations de colonnes avec leurs voisines horizontales (cadres magentas en figure 9.3).

(c) Corrélations des colonnes avec leurs voisines spectrales (cadres bleus en figure 9.4).

Figure 9.5 – Histogrammes des corrélations entre voisins immédiats sur une observation individuelle (première ligne) et fusionnée (seconde ligne).

de corrélations spatio-spectrales concernant les observations individuelles sont moins marquées que les cartes de corrélations spatio-spectrales illustrées en figure 9.2. En effet, seules les corrélations entre voisins immédiats (cadres violets et magentas en figure 9.3, cadres verts et bleus en figure 9.4) présentent un niveau de corrélation élevé : les autres corrélations sont similaires, en intensité, à celles mesurées sur le cube fusionné (figure 9.2b). Remarquons en particulier les structures de corrélation avec le voisinage horizontal immédiat (cadres magenta en figure 9.3a et bleu en figure 9.4a), présentant des motifs couvrant toute l'image, ayant eux aussi pour origine l'interpolation dans les images.

De plus, les cartes de corrélations mesurées sur l'image fusionnée (figures 9.3b et 9.4b) font apparaître plusieurs bandes verticales de corrélation. Ces bandes correspondent principalement à des galaxies, qui présentent une émission lumineuse forte sur toutes les bandes spectrales concernées. Notons aussi l'observation de « taches » de corrélations élevées, traduisant la présence d'une raie d'émission brillante d'un objet invisible par ailleurs : il est localisé spatialement et spectralement.

En conclusion, bien que les mesures fassent apparaître des corrélations marquées et structurées sur les observations individuelles, ces corrélations semblent perdre une grande partie de leur structure sur les observations fusionnées. Nous pouvons donc retenir que sur les images fusionnées, les corrélations sont non négligeables en-dehors du voisinage immédiat spatial et spectral.

La figure 9.5 présente les histogrammes des corrélations entre vecteurs immédiatement voisins. Plusieurs conclusions peuvent être tirées de l'examen de ces histogrammes :

— le principal mode de corrélation se situe à une valeur approximative de $0{,}2$ pour les données fusionnées.
— les corrélations importantes observées en figure 9.2a (cadres oranges) sont fortement atténuées lors du passage à l'observation fusionnée. Cette atténuation n'est cependant pas totale, car le mode des corrélations n'est pas centré en $0$. En conséquence, les corrélations entre spectres immédiatement voisins ne sont pas négligeables ;

— les corrélations spatio-spectrales sont déjà centrées en $0{,}2$ sur une observation individuelle (cadres magentas et bleus dans les figures 9.3a et 9.4a respectivement). Le passage au cube fusionné fait apparaître des corrélations plus importantes : ce phénomène n'est pas dû au bruit, mais à la présence d'astres formant des objets fortement corrélés au sein du cube.

D'un point de vue applicatif, nous pouvons de plus remarquer que ces structures de corrélations peuvent permettre un premier repérage spatial des astres, et un repérage spectral de leur raies d'émission.

### 9.1.3 Structures de corrélations entre spectres

Les corrélations que nous avons mesurées dans les sections 9.1.1 et 9.1.2 sont à valeurs réelles entre $-1$ et $1$. De plus, elles présentent des motifs complexes : certains éléments (spectres, lignes ou colonnes illustrés en figure 9.1) sont décorrélés, d'autres sont corrélés avec un ou plusieurs éléments voisins, etc. L'objet de cette partie est la caractérisation des structures de corrélation.

Nous nous intéressons dans cette partie aux structures concernant les corrélations entre spectres, qui forment la description « naturelle » des images hyperspectrales. Pour caractériser les structures de corrélations, il est d'abord nécessaire d'établir si chaque corrélation mesurée est significative ou non. Nous utilisons à cet effet un test de corrélation qui repose sur deux hypothèses complémentaires portant sur la corrélation théorique $\rho$ :

$$\left| \begin{array}{l} \mathcal{H}_0 : \rho = 0 \\ \mathcal{H}_1 : \rho \neq 0 \end{array} \right. \tag{9.2}$$

Le test résultera en une validation, ou une impossibilité de valider $\mathcal{H}_0$. Le coefficient de corrélation théorique $\rho$ est inconnu, nous le remplaçons par l'estimation $r$ (9.1). La distribution de cette estimation dépend du nombre d'éléments dans les vecteurs considérés. Sous $\mathcal{H}_0$, la distribution de $r$ est approximativement normale (bornée entre $-1$ et $1$) pour de grands échantillons et suit une loi de Student pour de petits échantillons [Lehmann et Romano, 2006]. Lorsque les corrélations sont mesurées sur des spectres, le nombre d'échantillons est le nombre de bandes spectrales.

Pour une corrélation calculée sur $\Lambda$ bandes spectrales, la statistique de Student correspondant à la corrélation mesurée $r$ est [Casella et Berger, 2002] :

$$t = (r - \rho)\sqrt{\frac{\Lambda - 2}{1 - r^2}} \overset{\mathcal{H}_0}{=} r\sqrt{\frac{\Lambda - 2}{1 - r^2}} \tag{9.3}$$

Soit $A$ une variable aléatoire suivant la loi de Student de $\Lambda - 2$ degrés de libertés, d'écart-type $1$. La p-valeur associée à la statistique de test est donnée par $p_t$, par symétrie de la distribution par rapport à sa moyenne nulle :

$$\begin{aligned} p_t =& p(|A| > t) \\ =& 2p(A > t) \\ =& 2(1 - \Phi_A(t)); \end{aligned} \tag{9.4}$$

où $\Phi$ est la fonction de répartition de la loi de Student. Nous nous donnons un seuil de signification $\alpha$ mesurant la probabilité d'accepter $\mathcal{H}_0$ de manière erronée. La décision sur la validité de $\mathcal{H}_0$ est la suivante :

$$\left| \begin{array}{l} p_t > \alpha : \mathcal{H}_0 \text{ valide}, \rho = 0; \\ p_t < \alpha : \mathcal{H}_0 \text{ non valide}, \rho \neq 0. \end{array} \right. \tag{9.5}$$

Cette démarche permet donc d'affirmer ou non la dé-corrélation entre deux spectres. Nous avons vu dans la partie précédente que ces corrélations opèrent principalement, dans les données fusionnées, au sein d'un voisinage local. Plusieurs combinaisons de corrélation sont envisageables entre un spectre et son voisinage : aucune corrélation, corrélation uniquement avec le voisin de droite, supérieur ou diagonal, ou bien une combinaison de ces possibilités. Nous appelons ces combinaisons, au nombre de 8, des *configurations de corrélation*. Comme les structures de corrélation varient spatialement et spectralement, il est pertinent d'étudier l'évolution des configurations spatiales en fonction des bandes spectrales.

La figure 9.6a présente des cartes de configuration de corrélations établies en différentes bandes spectrales, pour une observation individuelle :

— les structures de bandes horizontales sont très bien visibles et reprennent la forme des mesures de corrélations de la figure 9.2 ;
— ces structures varient en fonction des bandes spectrales : les configurations restent globalement identiques, mais se « déplacent » légèrement en changeant de bande spectrale ;
— la présence de résidus de soustraction de raie du ciel induit des structures de corrélation très marquées sur les deux dernières images de la figure. En effet, les autres phénomènes (bruit, astres) deviennent négligeables par rapport à ces perturbations, qui affectent toute l'image.

La figure 9.6b présente les cartes de configuration établies sur une observation fusionnée :

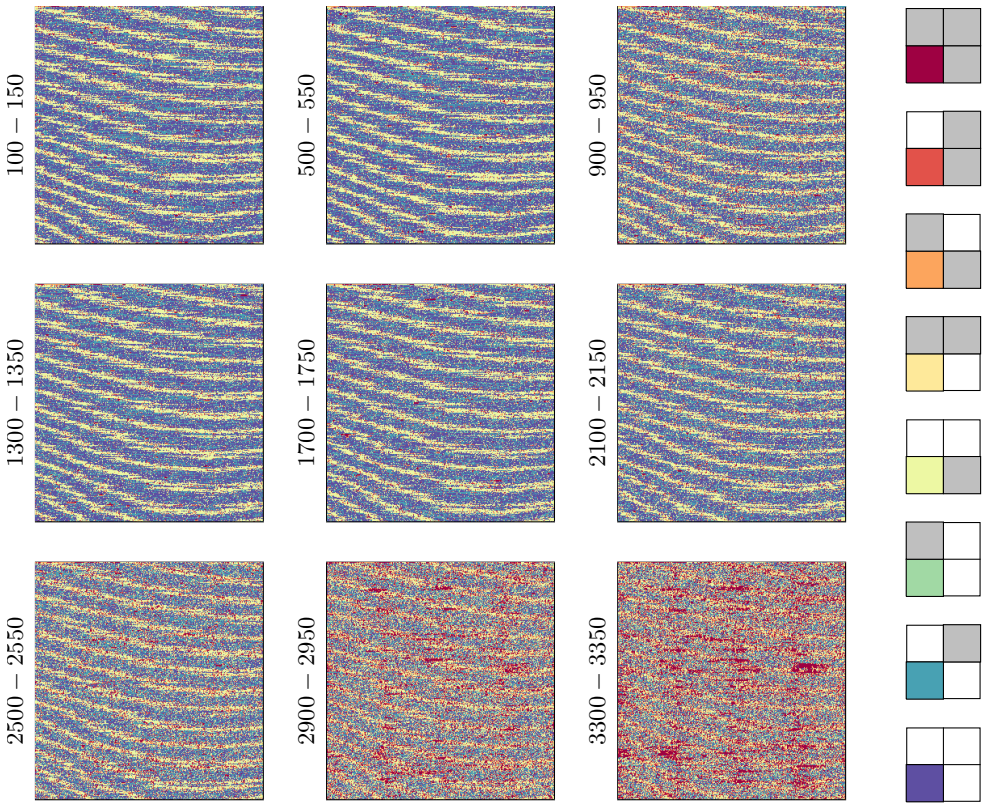
— l'effet de la soustraction des raies du ciel peut être retrouvé sous la forme d'un nombre élevé de configuration corrélées avec tous leurs voisins sur les deux dernières images ;
— les structures de corrélations ne semblent pas apparentes par ailleurs : le processus de fusion atténue la plupart des structures spatiales rencontrées dans les observations individuelles ;
— pour toutes les bandes spectrales, les corrélations sont rarement négligeables : cela étend le constat fait dans la figure 9.2 ;
— de même, nous constatons la présence de régions fortement corrélées, circulaires ou elliptiques, traduisant la présence d'objets structurés : ici, des galaxies.

Ces cartes de configuration permettent une meilleure visualisation des structures de corrélation intervenant dans les données. De plus, elles permettent une caractérisation exacte de la nature des corrélations affectant chaque spectre de l'image. Remarquons enfin qu'elles pourraient être utilisées dans un contexte de simulation des données MUSE, afin de générer des réalisations d'image de synthèse bruitée. Nous nous proposons, dans la seconde partie de ce chapitre, d'établir la distribution régissant le bruit au sein des images MUSE.

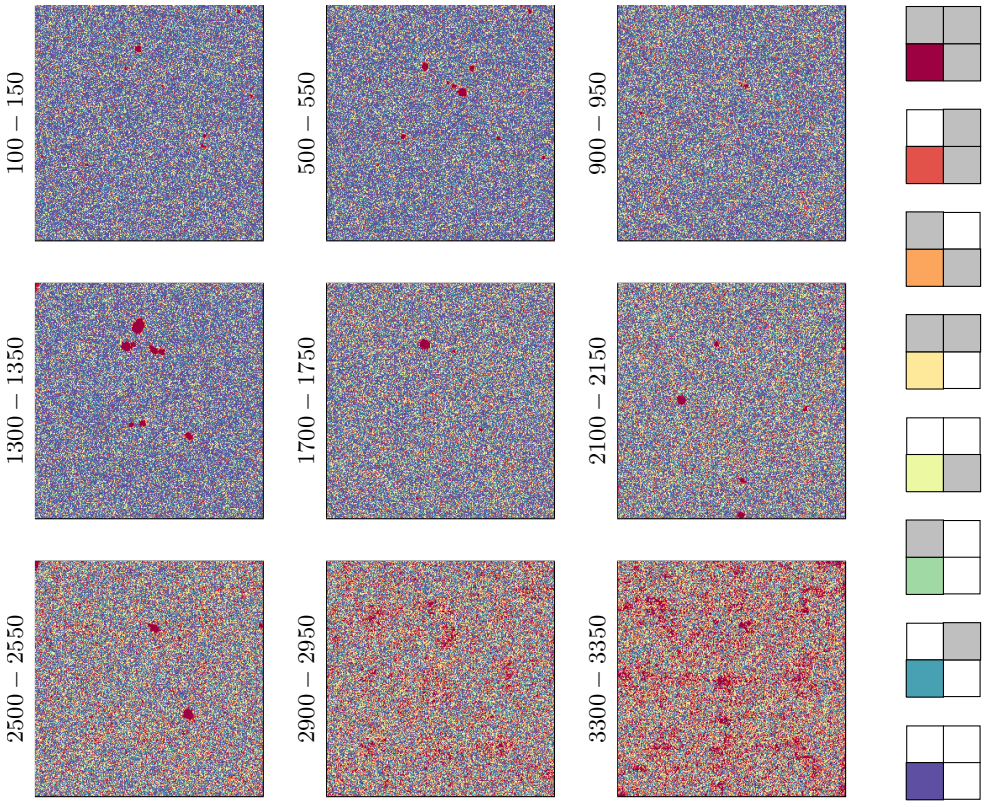
## 9.2 Lois du bruit

Nous avons vu que les intensités lumineuses présentent des corrélations spatiales et spectrales marquées qui doivent être prises en compte pour avoir une modélisation du bruit la plus précise. Cependant, la loi des données est de très grande dimension et n'est pas séparable. Par exemple, pour une image de taille $50 \times 50$ spectres et $20$ bandes spectrales, cela conduirait à la modélisation d'une loi :

— multivariée à $5{,}0 \cdot 10^4$ composantes ;
— comportant plus de $1{,}3 \cdot 10^5$ corrélations, en ne considérant que les termes liant les plus proches voisins.

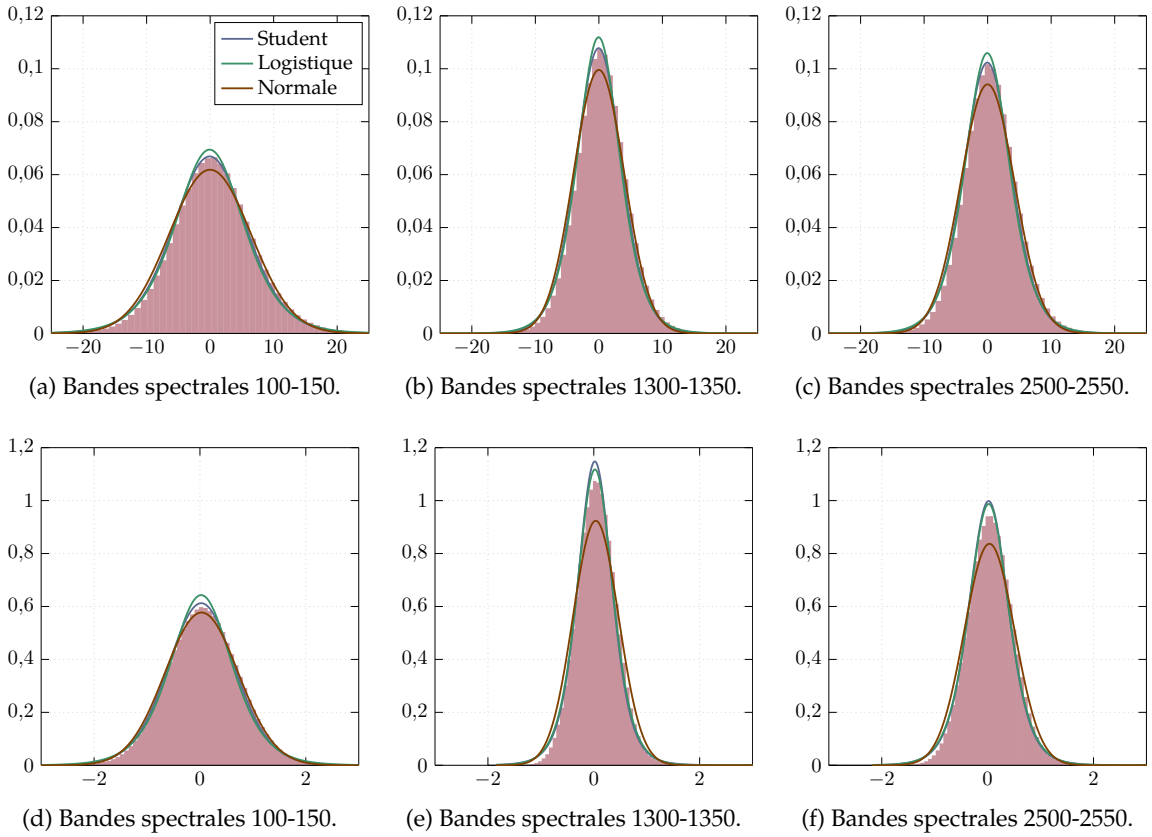



(a) Cube individuel.




(b) Cube fusionné.

Figure 9.6 – Configuration de corrélation par régions spectrales (indiquées à gauche). Les couleurs correspondent au type de corrélation, comme illustré dans les légendes à droite : sur quatre sites, le site de couleur est corrélé avec les voisins qui sont grisés.




(a) Bandes spectrales 100-150. (b) Bandes spectrales 1300-1350. (c) Bandes spectrales 2500-2550.

(d) Bandes spectrales 100-150. (e) Bandes spectrales 1300-1350. (f) Bandes spectrales 2500-2550.

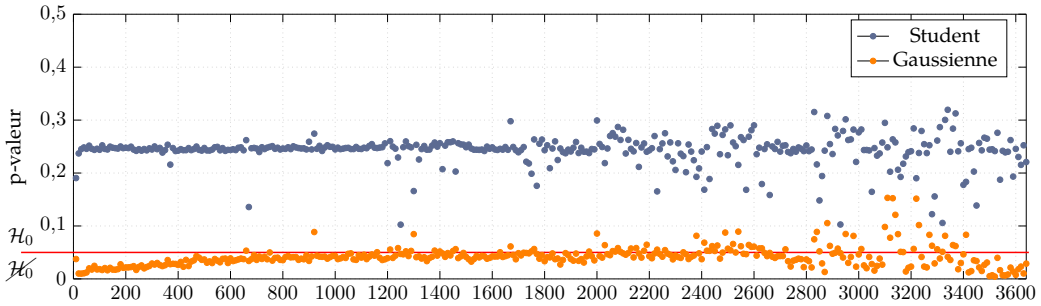
Figure 9.7 – Histogrammes des intensités des régions décorrélées pour quelques régions spectrales et meilleur ajustement au sens du maximum de vraisemblance par les lois normale, de Student et logistique. Première ligne : cube individuel, seconde ligne : cube fusionné.

Ce type de loi est trop complexe pour être évalué directement. C'est pourquoi nous l'évaluons marginalement, en étudiant la distribution des intensités dans l'image hyperspectrale. Pour éviter un biais dans l'estimation des distributions, nous nous intéresserons uniquement aux pixels dont la dé-corrélation avec leurs voisins est avérée, via le test de corrélation (9.5).

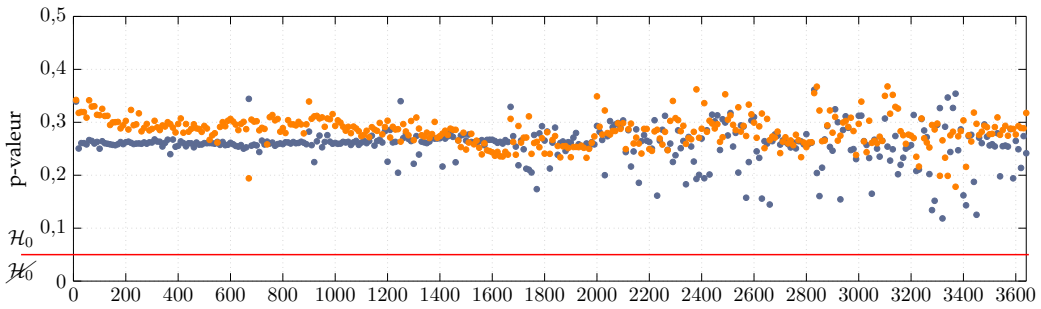
### 9.2.1 Distributions empiriques

Nous nous intéressons aux fréquences d'apparition des intensités lumineuses, que nous représentons sous la forme d'un histogramme normalisé. Ces fréquences sont mises en correspondance avec les lois gaussienne, logistique et de Student. Cet ajustement est fait via les estimateurs du maximum de vraisemblance sur les paramètres de moyenne et d'écart-type pour les distributions gaussienne et logistique. L'ajustement de la loi de Student est moins direct. En effet, il n'existe pas de formule explicite pour le calcul du degré de liberté $\nu$ au sens du maximum de vraisemblance. C'est pourquoi l'ensemble des paramètres ($\nu$, moyenne $\mu$ et écart-type $\sigma$) est évalué via l'algorithme EM [Liu et Rubin, 1995]. Nous présentons en figure 9.7 les histogrammes et ajustements, dont l'inspection visuelle permet trois constats :

— les propriétés de ces données varient spectralement : bien que les modes soient proches de 0 dans tous les cas, leur dispersion varie clairement selon la bande spectrale considérée ;
— dans le cas d'une observation individuelle, le meilleur modèle pour les données est




(a) Image individuelle.




(b) Image fusionnée.

Figure 9.8 – p-valeurs moyennes par bande spectrale (abscisses), selon les tests d'adéquation à une loi normale (Shapiro-Wilk [Shapiro et Wilk, 1965]) et à une loi de Student. La droite rouge représente le seuil de rejet de $\mathcal{H}_0$ avec un risque $\alpha = 0{,}05$.

visuellement la distribution de Student ;

— les intensités de l'observation fusionnée ne permettent pas de tirer une conclusion aussi tranchée. En effet, la légère asymétrie des intensités autour de la moyenne semble être mal modélisée par les trois distributions proposées [2]. De plus, le mode principal semble sous-estimé par un modèle de distribution gaussienne, et surestimé dans les autres cas.

Ce dernier point nous incite à adopter une démarche de qualification de ces distributions, par le biais de tests statistiques.

### 9.2.2 Mesure de l'adéquation par tests d'hypothèses

Nous avons constaté visuellement l'adéquation empirique des intensités avec les distributions gaussienne et de Student [3]. Ce constat est qualitatif : l'objet de cette section est d'obtenir des résultats quantitatifs sur ces adéquations. Nous recourons à un raisonnement par test d'hypothèse afin de confirmer ou infirmer ces adéquations. Ainsi, pour la distribution normale, la forme des hypothèses est la suivante :

$$\begin{aligned} &\mathcal{H}_0 : \text{l'échantillon a une distribution gaussienne de paramètres inconnus,} \\ &\mathcal{H}_1 : \text{l'échantillon n'a pas une distribution normale.} \end{aligned} \tag{9.6}$$

Nous établissons le choix entre ces hypothèses à l'aide du test de Shapiro-Wilk [Shapiro et Wilk, 1965]. Les propriétés de l'image étant variables spectralement, nous constituons 368 subdivisions de 10 bandes spectrales dans lesquelles effectuer le test. Pour éviter les biais

2. Cette asymétrie est due à la présence de rayonnements lumineux provenant des objets les plus ténus.
3. Nous retenons ces deux distributions car la distribution logistique n'apporte pas une modélisation significativement différente de celles-ci.

statistiques liés à la taille de l'échantillon, nous effectuons le test sur $1000$ valeurs consécutives, calculons la p-valeur, et effectuons une moyenne de ces p-valeurs par subdivision. La valeur obtenue est donc la moyenne des probabilités d'obtenir les intensités testées dans le cas où $\mathcal{H}_0$ est valide. Cela signifie que plus la p-valeur est faible, plus il est souhaitable de rejeter $\mathcal{H}_0$.

Le même raisonnement peut être développé pour l'adéquation à une distribution de Student, en se donnant :

$$\left|\begin{array}{l}\mathcal{H}_0 : \text{l'échantillon a une distribution de Student de paramètres inconnus,}\\ \mathcal{H}_1 : \text{l'échantillon n'a pas une distribution de Student.}\end{array}\right. \tag{9.7}$$

L'estimation des paramètres de la loi de Student repose comme précédemment sur l'algorithme EM [Liu et Rubin, 1995]. En se donnant une variable aléatoire $A$ suivant une loi de Student reposant sur les paramètres estimés, la p-valeur $p_t$ se calcule par [4] :

$$p_t = p(A > y) = 1 - \Phi_A(y); \tag{9.8}$$

où $y$ représente ici une intensité lumineuse et $\Phi$ la fonction de répartition de la loi de Student.

La figure 9.8 présente les p-valeurs résultantes de ces deux tests, sur une image individuelle et fusionnée, et selon les bandes spectrales considérées :

— les p-valeurs calculées sur l'observation individuelle illustrent clairement la mauvaise adéquation des intensités lumineuses à une loi gaussienne, et la meilleure modélisation par une loi de Student.
— les résultats sur l'observation fusionnée sont moins tranchés : les deux distributions semblent adéquates, avec un léger avantage pour la distribution gaussienne.

Dans tous les cas, la présence de résidus de soustraction de raie du ciel provoque une plus forte dispersion des mesures sur les bandes spectrales d'indice élevé.

En conclusion, nous retenons l'hypothèse d'une distribution de Student pour les observations individuelles. Précisons de plus que cette distribution ne résulte pas d'un processus physique particulier : la loi de Student permet principalement un meilleur ajustement par l'utilisation du paramètre supplémentaire de degré de liberté. Concernant les observations fusionnées, l'adéquation similaire des distributions nous incite à retenir un modèle gaussien, car ce modèle est plus simple et semble légèrement plus adéquat.

### 9.2.3 Estimation des paramètres

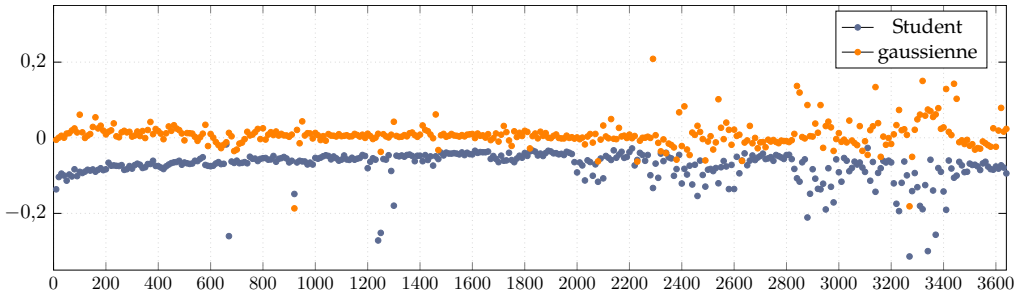
Dans cette section, nous présentons une évaluation des paramètres de ces distributions. Les paramètres de la distribution gaussienne sont estimés au sens du maximum de vraisemblance, et les paramètres de la distribution de Student sont estimées grâce à l'algorithme EM [Liu et Rubin, 1995].

La figure 9.9 présente les valeurs estimées des paramètres régissant les deux distributions retenues. Plusieurs constats peuvent être tirés de ces estimations :

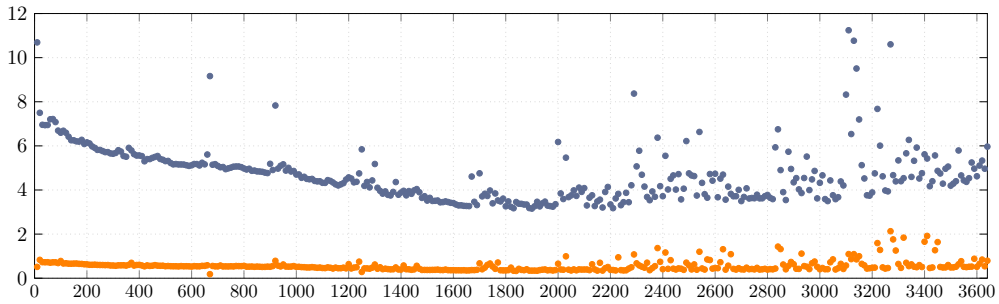
— l'influence de la soustraction des raies du ciel est visible dans tous les cas, sous la forme d'une plus grande variabilité des paramètres ;
— hormis cet effet, les paramètres présentent tous une évolution spectrale douce : à l'échelle de quelques bandes spectrales, ils évoluent très peu.

Nous pouvons de plus remarquer l'apport des observations fusionnées sur les observations individuelles :

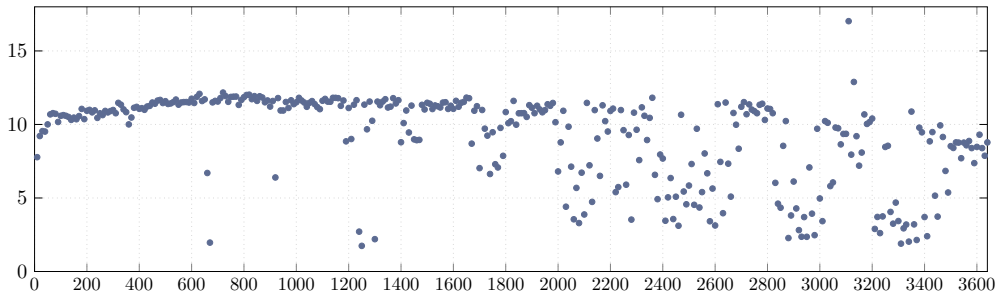
4. Le calcul diffère de (9.4) car la loi n'est pas nécessairement centrée en $0$.




(a) Moyennes $\mu$




(b) Écart-types $\sigma$




(c) Degrés de liberté $\nu$

Figure 9.9 – Estimation par bande spectrale (abscisses) des paramètres selon une loi de Student sur une observation individuelle (bleu), et selon une loi normale pour une observation fusionnée (orange).

— les moyennes $\mu$, qui ne sont pas centrées dans le premier cas, le sont dans le second ;
— les écarts-types $\sigma$ sont réduits d'un facteur $10$ environ, ce qui traduit une augmentation nette du RSB dans les images.

Les mesures présentées sur une observation individuelle proviennent toutes de la même image hyperspectrale. Cependant, chaque observation individuelle a ses caractéristiques propres en raison des conditions d'observations. Cela conduit à une variabilité inter-observations supplémentaire des propriétés des images. Cette variabilité n'a pas été illustrée dans ce chapitre, car les traitements sont effectués sur l'image fusionnée.

## 9.3 Conclusion

Les données MUSE présentent une structure complexe. Nous avons pu mettre en évidence des phénomènes de corrélations non négligeables dans les données. Le caractère massif de celles-ci nous empêche de considérer une distribution d'ensemble, c'est pourquoi nous avons étudié dans un second temps les intensités lumineuses en chaque point de l'image.

Les observations individuelles présentent une signature très marquée : les corrélations ne sont pas négligeables et sont structurées dans le cube. En conséquence, estimer la variance des données par la seule variance des intensités lumineuses induit nécessairement des biais. En ignorant les spectres corrélés, nous avons pu établir que la loi de Student apporte une bonne modélisation des intensités lumineuses. Les paramètres de cette loi varient selon les bandes spectrales considérées.

Les effets observés sur une pose individuelle sont nettement atténués lorsque nous nous intéressons à une pose fusionnée résultant d'environ $60$ observations individuelles. Néanmoins, les corrélations ne disparaissent pas dans le processus : elles sont limitées au voisinage immédiat de chaque point et perdent de leur caractère structuré. De même, les intensités des spectres non corrélés semblent correctement modélisées par une loi gaussienne, dont les paramètres varient là aussi selon les bandes spectrales considérées.

Enfin, remarquons que cette étude ne concerne que les cubes résultant de nombreuses observations, ici $60$. Cela a pour effet une forte atténuation des spécificités de chaque observation : corrélations en motifs, distribution de Student, etc. Il est donc vraisemblable qu'un cube résultant par exemple de $5$ observations individuelles ait des propriétés différentes de celles présentées ici. Les observations de champs profonds que nous utilisons dans la suite de ce manuscrit ne sont pas concernées par cette remarque, car elles reposent sur de très longs temps de pose, et donc sur plusieurs dizaines d'observations individuelles.

# 10

# Applications aux données réelles MUSE



## 10.1 Introduction

Dans ce chapitre, nous présentons l'application des modèles introduits dans ce manuscrit à la détection dans les images hyperspectrales de l'instrument MUSE. Nous spécifions d'abord les particularités des méthodes pour la détection dans des images hyperspectrales astronomiques.

Les méthodes évaluées dans ce chapitre sont les suivantes :

1. la stratégie de détection par tests d'hypothèses présentée dans le chapitre 3 (Algorithme 1). Nous choisissons pour cette méthode le même catalogue que dans le chapitre 3 avec une $P_{\text{FA}}$ cible fixée à $0{,}05$. La FSF employée est une fonction de Moffat de $3 \times 3$ pixels.
2. la détection reposant sur le modèle de champs de Markov couples convolutifs (CMco) du chapitre 5. Nous employons le même paramètre de FSF que pour l'Algorithme 1 ;
3. la segmentation dans le cadre du modèle de champs de Markov triplets orientés (CMTO) introduit dans le chapitre 6. Le modèle est présenté en section 6.2 ;
4. la segmentation selon le modèle d'arbres de Markov triplets spatiaux (AMTS) introduit dans le chapitre 7. Ce modèle est décrit dans les sections 7.2.1 et 7.3.1.

Pour les méthodes CMco, CMTO, et AMTS, nous choisissons d'employer la segmentation au sens du MPM. En effet, les résultats obtenus dans le chapitre 5 suggèrent que cet estimateur permet de réaliser très peu de fausses détections.

L'application des méthodes de l'Algorithme 1 et du modèle CMco aux images hyperspectrales MUSE est directe.

Les deux autres méthodes ont été présentées dans le cadre de la segmentation d'image, et doivent être adaptées au contexte de la détection en imagerie hyperspectrale. Nous considérons ici des observations vectorielles en chaque site. L'expression à adapter est celle de la fonction $f(x_s, y_s)$, décrite en (6.14) pour le modèle CMTO et en (7.10) pour le modèle AMTS. Pour l'application à des images hyperspectrales, elle s'exprime comme une distribution normale multivariée spectralement :

$$f(x_s, \mathbf{y}_s) = \frac{1}{(2\pi)^\Lambda \det(\mathbf{\Sigma})} \exp\left(-\frac{1}{2}(\mathbf{y}_s - \boldsymbol{\mu}_{x_s})^\top \mathbf{\Sigma}^{-1}(\mathbf{y}_s - \boldsymbol{\mu}_{x_s})\right); \tag{10.1}$$

où $\boldsymbol{\mu}_{x_s}$ est le paramètre de moyenne, dépendant de la classe prise par $x_s$. L'analyse du bruit MUSE, conduite dans le chapitre 9, permet également de supposer que :

— lorsque $\Lambda$ est de l'ordre de 20 bandes spectrales (largeur spectrale typique d'une émission de Lyman-alpha), les paramètres du bruit sont identiques pour toutes les composantes spectrales [1] ;
— les corrélations spectrales sont négligeables au-delà de 2 bandes spectrales.

En conséquence, $\mathbf{\Sigma}$ est pentadiagonale :

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \rho_1 & \rho_2 & 0 & \cdots & 0 \\ \rho_1 & & & & & \\ \rho_2 & & & & & 0 \\ 0 & & & & & \rho_2 \\ & & & & & \rho_1 \\ 0 & \cdots & 0 & \rho_2 & \rho_1 & \sigma^2 \end{bmatrix} \tag{10.2}$$

où $\sigma$ représente l'écart-type, et $\rho_1$ et $\rho_2$ sont des paramètres de corrélation. Dans le cadre de l'application à la détection, nous supposons enfin que :

— l'ensemble des classes $\Omega_x$ vaut $\{\omega_0, \omega_1\}$ avec $\omega_0$ et $\omega_1$ correspondant à « l'absence de signal » et la « présence de signal » respectivement. Dans le cadre du modèle CMTO, nous choisissons $\Omega_v = \{\pi/4, 3\pi/4\}$ ;
— le paramètre de moyenne spectrale associé à $\omega_0$ est le vecteur nul : $\boldsymbol{\mu}_{\omega_0} = \mathbf{0}_\Lambda$.

Les estimateurs des paramètres sont décrits en section 5.2.5. Les détections sont ensuite obtenues de manière identique aux segmentations non supervisées présentées dans les chapitres 6 et 7.

## 10.2 Évaluation sur des données synthétiques

Avant d'appliquer les méthodes proposées sur les images MUSE, nous procédons à une étape d'évaluation permettant de mesurer les performances des quatre méthodes dans le cadre de la détection dans des images hyperspectrales astronomiques.

### 10.2.1 Configuration expérimentale

Nous étudions les performances des méthodes dans le cadre de la détection d'émissions lumineuses étendues à proximité d'un objet (« halos ») et englobant plusieurs objets (« filaments »). Cela se traduit par la mise en place de deux expériences :

1. Remarquons que dans le cas de la présence de soustraction de raies du ciel, ce constat n'est plus valide.

A. la présence d'une source seule au sein d'une image $64 \times 64$ pixels, similaire à ce qui a été présenté dans le chapitre 5 (cf. figure 10.1a). Les halos de Lyman occupent en effet un espace similaire dans les observations MUSE. Dans cette image, les intensités non nulles varient en l'absence de bruit au plus d'un facteur $2$.
B. la présence d'une paire de sources et d'une structure les reliant au sein d'une image de $128 \times 128$ spectres (cf. figure 10.1b). Par rapport aux modèles prédits par la théorie astrophysique, nous nous plaçons donc dans un cas relativement simplifié concernant les éventuels échanges se produisant entre deux galaxies proches l'une de l'autre. Dans cette image, les intensités non nulles varient au plus d'un facteur $10$, et sont donc en moyenne plus faibles que celles de l'expérience A.

Ces intensités subissent une convolution par une FSF de taille $3 \times 3$ pixels, de la forme d'une fonction de Moffat. Spectralement, les sources d'émission partagent (à l'intensité près) le même spectre, illustré en figure 10.1c.

Les méthodes que nous évaluons dans ce chapitre ont été précédemment évaluées dans le contexte d'un bruit gaussien (CMTO et AMTS) ou normal multivarié (Chapitre 3 et CMco). Nous avons montré, dans le chapitre 9, que le bruit présent dans les données MUSE présente des structures de corrélation spatiales et spectrales. D'autres phénomènes sont susceptibles d'intervenir, comme la présence d'objets brillants, de données aberrantes ou de non stationnarités [2]. Les intensités lumineuses caractérisant les objets les plus brillants peuvent être retirées de l'image par filtrage médian, en revanche les autres phénomènes sont inévitables.

Afin d'évaluer nos méthodes dans le cadre le plus réaliste possible, nous choisissons de produire des réalisations du bruit à partir d'observations MUSE. Pour ce faire, une image hyperspectrale de taille adaptée ($64^2$ ou $128^2$ spectres) est extraite de l'image *udf10*. Cette image subit un filtrage médian spectral permettant de retirer la plupart des émissions lumineuses continues des sources brillantes (étoiles, galaxies). Une permutation aléatoire par blocs est ensuite effectuée afin d'obtenir une réalisation de bruit pseudo-aléatoire. L'image obtenue en combinant les contributions du bruit, des intensités lumineuses convoluées par la FSF et du spectre moyen en fonction du RSB forme une réalisation $\mathbf{Y} = \mathbf{y}$.

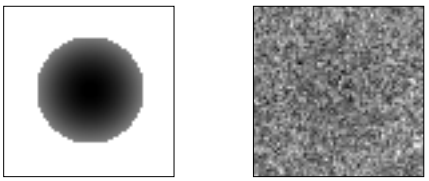
### 10.2.2 Performances

Nous nous intéressons aux performances des méthodes en termes de taux moyens d'erreur, de faux positif et de faux négatif. Des exemples de résultats de détection sont illustrés en figures 10.2a (cas A) et 10.2b (cas B). Les performances moyennes, obtenues sur $100$ réalisations de $\mathbf{y}$, sont quant à elles reportées en figure 10.3.

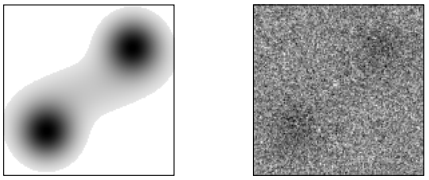
Les commentaires suivants peuvent être faits sur les résultats de l'expérience A (colonne gauche de la figure 10.3) :

— les meilleurs taux d'erreur sont fournis par les modèles CMco et CMTO : les valeurs sont inférieures à $10\%$ d'erreur pour $\mathrm{RSB} < -14$ dB. Les taux de faux positifs sont très faibles (inférieurs à $0{,}02$ en moyenne), ce qui implique que le taux d'erreur est dominé par le taux de faux négatif, qui est inférieur à $0{,}1$ pour $\mathrm{RSB} < -10$ dB ;
— la méthode de détection par tests du chapitre 3 fournit globalement peu de faux positifs, mais les performances à faible RSB sont peu satisfaisantes. Cela s'explique, d'une part, par l'absence d'un *a priori* sur l'ensemble des choix $\mathcal{H}_0/\mathcal{H}_1$ comme il en existe dans les modèles markoviens. D'autre part, cela s'explique par l'emploi d'une FSF de taille restreinte, qui correspond cependant à celle ayant permis la génération

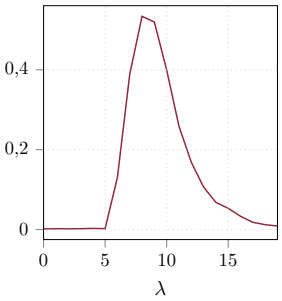
2. Ces deux derniers phénomènes apparaissent notamment dans les régions délimitant les $9$ parties de l'image *mosaic*.



(a) Expérience A ($64^2$ spectres).



(b) Expérience B ($128^2$ spectres).




(c) Moyenne spectrale.

Figure 10.1 – (a), (b) : formation des images de synthèse, avec les intensités lumineuses à gauche et la moyenne spectrale d'une image hyperspectrale $\mathbf{Y} = \mathbf{y}$ à $\mathrm{RSB} = -15$ dB à droite. En l'absence de bruit, les intensités non nulles varient entre $0{,}5$ et $1{,}0$ dans l'expérience A, et entre $0{,}1$ et $1{,}0$ dans l'expérience B. (c) : moyenne spectrale employée dans les expériences A et B.

des images ;

— la segmentation par AMTS offre, à RSB élevé, des taux d'erreurs plus élevés que les segmentations basées sur le modèle CMco ou CMTO. Ce constat est à nuancer : en effet, cette segmentation semble robuste à faible RSB, ce qui peut permettre de surpasser les méthodes par champs de Markov qui peuvent limiter la détection aux régions les plus brillantes.

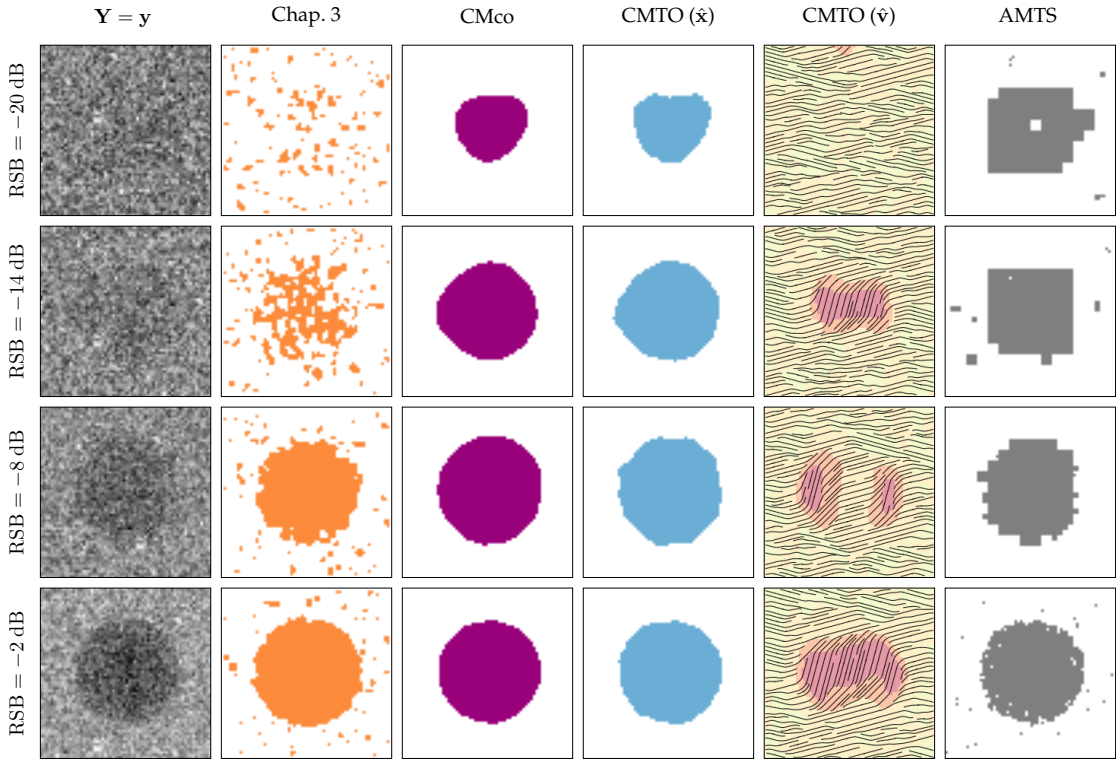
Dans l'expérience B, les intensités lumineuses des régions à détecter sont en moyenne beaucoup plus ténues que celles de l'expérience A. Cela se traduit dans les résultats de segmentation (colonne droite de la figure 10.3) :

— le nombre de faux positifs est très faible (inférieur à $0{,}1$) pour toutes les méthodes en raison des faibles intensités lumineuses de la région à détecter. En conséquence, les erreurs sont dominées par les faux négatifs.

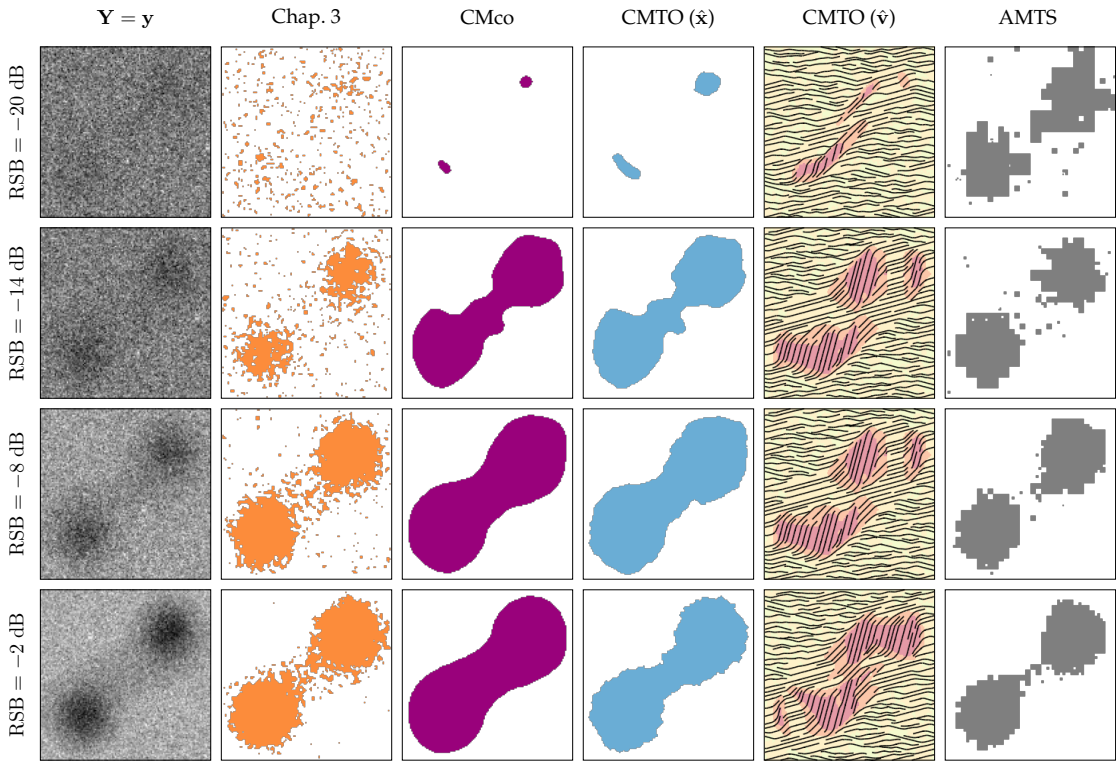
— les méthodes basées sur les modèles CMco et CMTO fournissent des résultats similaires à faible RSB. A fort RSB, la méthode CMTO fournit des résultats moins satisfaisants : la méthode de segmentation a tendance à privilégier les éléments les plus brillants, ce qui diminue la taille de la région détectée. Ce phénomène est visible en figure 10.2b.

— tout comme dans l'expérience A, la méthode basée sur le modèle AMTS fournit globalement un taux d'erreur plus élevé que les méthodes CMco ou CMTO. Cependant, les résultats à faible RSB sont là aussi remarquables : dans certaines situations, cette méthode permet la détection de régions ignorées par ces deux modèles (cf. figure 10.2b) ;

— la segmentation basée sur le modèle CMTO apporte un résultat additionnel concernant les orientations dans l'image hyperspectrale. Les performances de cette segmentation ne peuvent pas être quantifiées. Cependant, il apparaît nettement sur la figure 10.2b que la présence d'une structure orientée induit une segmentation des orientations caractéristiques : des régions homogènes dans les régions correspondant à une détection, et « mouchetées » en-dehors. En ce sens, la segmentation CMTO permet une caractérisation qualitative de la présence, ou non, de structures orientées dans les images.

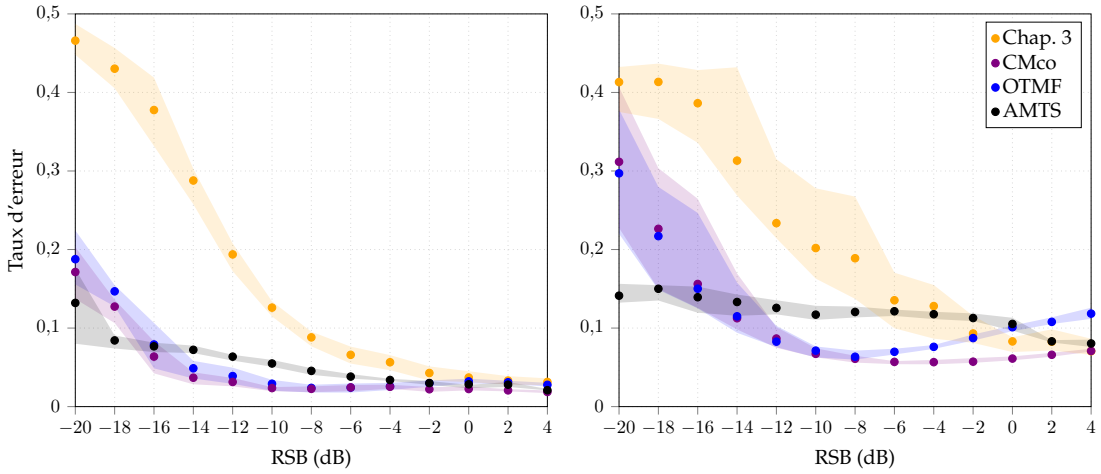



(a) Résultats pour l'expérience A.

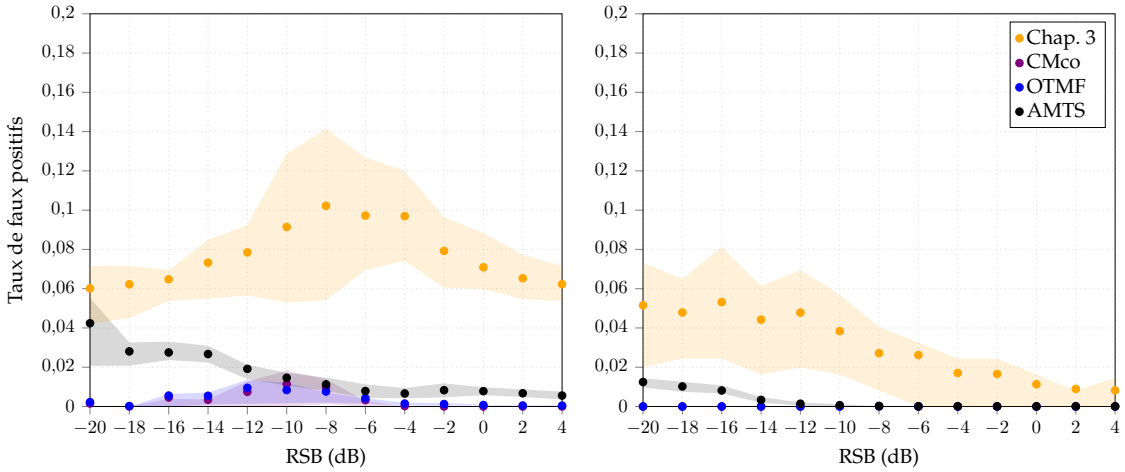



(b) Résultats pour l'expérience B.

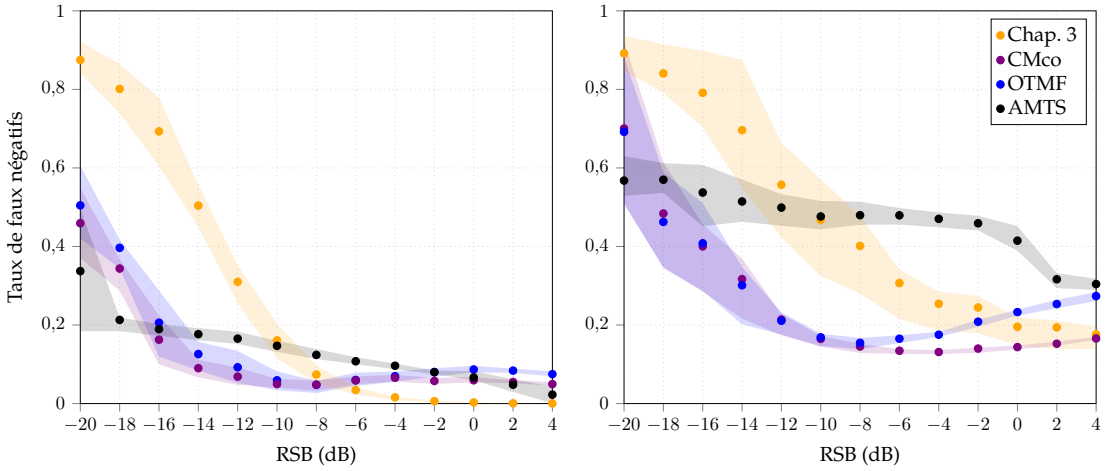
Figure 10.2 – Exemples de résultats de détection. Les images hyperspectrales $\mathbf{Y} = \mathbf{y}$ sont représentées par leurs moyennes spectrales (inverse vidéo), et les orientations $\hat{\mathbf{v}}$ sont représentées par leurs valeurs et par les courbes tangentes en chaque point (en rouge).




(a) Taux d'erreur.




(b) Taux de faux positif.




(c) Taux de faux négatif

Figure 10.3 – Performances moyennes des 4 méthodes évaluées dans ce chapitre pour l'expérience A (colonne de gauche) et l'expérience B (colonne de droite).

### 10.2.3 Bilan

Les résultats sur les images de synthèse ont permis de qualifier chaque méthode dans le contexte de la détection dans les images hyperspectrales astronomiques. La détection d'émissions de Lyman-alpha localisée autour d'un objet unique (« halos ») sera ainsi effectuée au mieux par :

— la méthode basée sur le modèle CMco, qui fournit un très faible taux de faux positif dans le cas général ;
— la segmentation du modèle AMTS pour les objets les plus ténus, car cette méthode est plus performante pour les très faibles RSB.

Comme le RSB des objets d'intérêt est inconnu *a priori*, nous emploierons ces deux méthodes pour les images MUSE.

La détection d'émissions plus étendues et plus ténues, englobant plusieurs objets (« filaments ») sera quant à elle réalisée au mieux par :

— la méthode basée sur le modèle CMTO, qui permet de bonnes performances à faible RSB tout en fournissant un critère qualitatif sur la présence de structures orientées à grande échelles ;
— la segmentation du modèle AMTS pour la détection à très faible RSB.

**Remarque 10.2.1.** *Dans les deux situations, la méthode basée sur le modèle AMTS peut permettre la détection à très faible RSB. Nous pouvons constater, d'après les figures 10.2a et 10.2b, que cela induit des formes « en blocs » dues à la structure de l'arbre de Markov. La détection ne sera donc pas aussi précise qu'à plus fort RSB, mais permettra néanmoins l'affirmation de la présence ou non des émissions lumineuses les plus ténues.*

## 10.3 Résultats sur les images réelles

### 10.3.1 Les images MUSE *Hubble ultra-deep field*

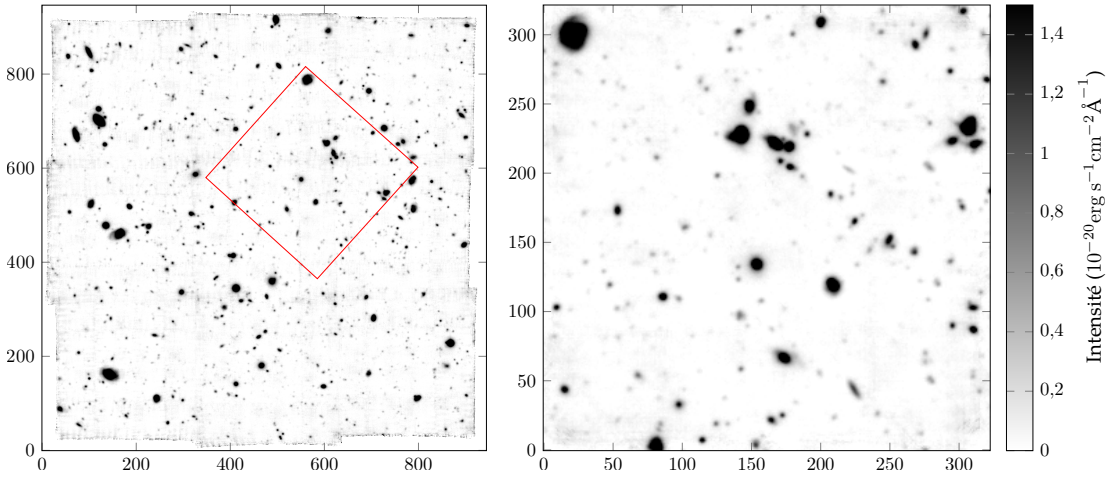
Les images MUSE employées correspondent à des observations de champs profonds : une région du ciel contenant très peu d'objets « proches » (galaxies ou étoiles dans la Voie Lactée), permettant notamment d'observer des objets très lointains de l'Univers jeune. Les observations couvrent la région *Hubble Ultra Deep Field* (HUDF), observée la première fois en 2003 par le télescope spatial Hubble [Beckwith et al., 2006]. Les observations de MUSE permettent pour la première fois une spectroscopie « en aveugle » des objets présents dans cette région [3], et sont particulièrement adaptées pour la détection d'émetteurs de Lyman-alpha.

Les observations HUDF produites par MUSE sont décrites dans [Bacon et al., 2017]. Dans leur version finale (après réduction des données), elles consistent en deux images hyperspectrales :

— une « mosaïque » couvrant l'étendue spatiale de $3 \times 3$ observations MUSE, avec un temps moyen de pose de $10$ h environ. Cette image, illustrée en figure 10.4a, comporte environ $900 \times 900$ spectres et sera appelée *mosaic* dans la suite.
— une observation « approfondie » d'une partie de ce champ avec un temps moyen de pose de $31$ h environ, qui est appelée *udf10* dans la suite. Elle est illustrée en figure 10.4b, et comporte environ $320 \times 320$ spectres.

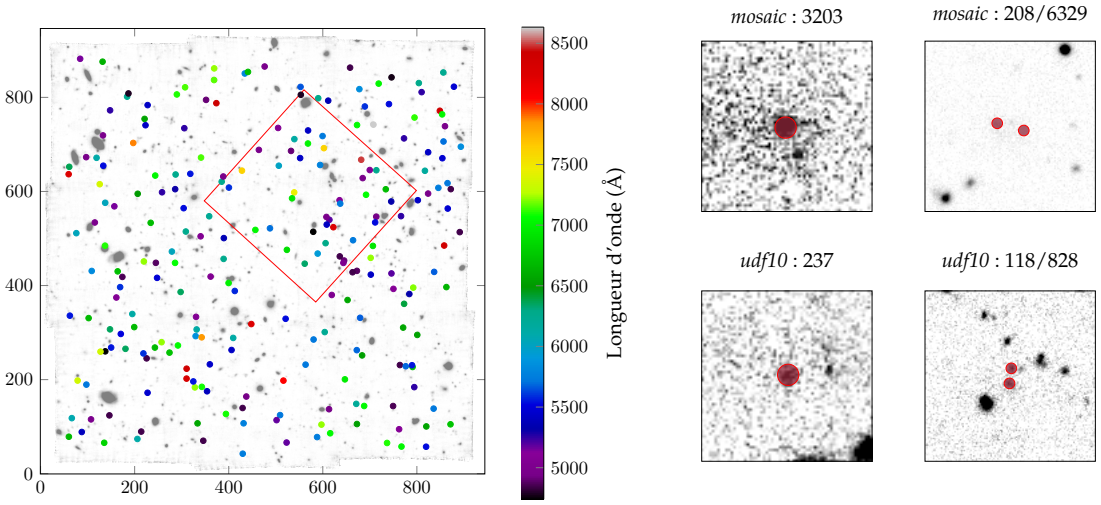
Lors de l'application des méthodes de détection aux données réelles, nous nous intéressons à des objets isolés et à des paires d'objets proches spatialement et en redshift.

3. C'est-à-dire indépendante d'une localisation ciblée *a priori*.




(a) Image *mosaic*. (b) Image *udf-10*.

Figure 10.4 – Images hyperspectrales MUSE employées dans ce chapitre, visualisées par leur moyenne spectrale. Le carré rouge représenté dans l'image *mosaic* représente la couverture spatiale de l'observation approfondie *udf10*.




(a) Localisation spatiale et spectrale des émetteurs de Lyman-alpha dans l'image *mosaic* (moyenne spectrale en inverse vidéo).

(b) Images hyperspectrales (moyennes spectrales en inverse vidéo) centrées sur des émetteurs de Lyman-alpha localisés par les cercles rouges.

Figure 10.5 – Localisations des émetteurs de Lyman-alpha et exemple d'images hyperspectrales associées.

La localisation spatiale et spectrale des émetteurs de Lyman-alpha dans les images a été obtenue indépendamment par des experts (cf. figure 10.5a). Nous isolons pour chaque objet une image hyperspectrale centrée sur le ou les émetteurs d'intérêt. Ces images ont une taille de $64 \times 64$ pour les objets isolés, car les émission lumineuses attendues sont plus petites (cf. [Leclercq et al., 2017; Wisotzki et al., 2016]. Pour les objets doubles, nous utilisons une image de taille $128 \times 128$ spectres afin d'étudier d'éventuelles émissions lumineuses à grande échelle. La largeur spectrale de ces images est définie par des experts (cf. [Leclercq et al., 2017] notamment), et varie entre 12 et 24 bandes spectrales de MUSE. La figure 10.5b donne une illustration de ces images hyperspectrales.

Dans cette partie, 242 émetteurs de Lyman-alpha ont été isolés dans les images *udf10* et *mosaic*. De manière similaire à [Leclercq et al., 2017], ces objets ne sont pas affectés par la présence d'un objet voisin, les bordures des images ni la présence d'une émission lumineuse de même localisation spectrale. Nous disposons de plus de 31 paires de sources formées par deux émetteurs de Lyman-alpha proches spatialement et spectralement. Tous ces objets sont repérés par leur numéro de catalogue (« ID ») [Inami et al., 2017].

Afin de traiter des images stationnaires [4], nous divisons chaque image par l'écart-type associé provenant de la réduction des données (cf. [Bacon et al., 2017]). Une soustraction de médiane est également effectuée afin de supprimer les objets présentant un spectre d'émission continu (galaxies proches, étoiles). Les résultats de détection sont obtenus avec les méthodes des modèles AMTS et CMco pour les objets individuels, et avec les méthodes AMTS et CMTO pour les paires d'objets.

### 10.3.2 Objets isolés

#### Résultats par objets

Pour qualifier les résultats, nous nous appuyons sur la moyenne des spectres des régions détectées : si aucune émission lumineuse localisée spectralement, caractéristique du rayonnement de Lyman-alpha n'apparaît, alors la détection a vraisemblablement échoué. De plus, la comparaison des surfaces détectées avec celle correspondant à la largeur à mi-hauteur de la FSF [5] fournit un critère d'appréciation sur le caractère étendu ou non de la détection.

Un échantillon de résultats de détection sur 8 objets est reporté en figure 10.6, et les résultats concernant 32 autres objets sont reportés en annexe B.2. Plusieurs commentaires peuvent être tirés de l'inspection visuelle de ces résultats :

— dans le cas général, la méthode basée sur le modèle CMco semble produire des résultats satisfaisants. Ce modèle ne semble cependant pas permettre la détection à très faible RSB, comme l'illustrent les objets 148 en figure 10.6 et les objets 208, 4598, et 6692 en section B.2.1 ;

— dans ces deux situations, la détection par AMTS semble fournir une alternative pertinente à celle proposée par le modèle CMco. En effet, à fort RSB la détection concerne des émissions plus ténues (ID 7085 en B.2.1). À faible RSB, comme dans le cas des images de synthèses, des détections sont effectuées par le modèle AMTS sans contrepartie du modèle CMco (ID 148 en figure 10.6 et 208, 4598, 6692 en B.2.1) ;

— la carte d'incertitude associée à la segmentation CMco forme enfin un outil pertinent pour évaluer qualitativement les résultats associés. En effet, elles peuvent permettre de repérer les régions ambiguës dans les détections des objets les plus ténus (ID 2069 en figure 10.6 et 1226, 1775, 1950, 3058 5613 en B.2.1).

— dans un certain nombre de cas, la surface détectée n'excède celle d'un disque de diamètre la largeur à mi-hauteur de la FSF : il est impossible de conclure à la présence d'une source lumineuse étendue spatialement (ID 2733 en figure 10.6 et 2307, 7085 en B.2.1).

En conséquence, nous choisissons pour chaque objet de retenir par défaut la détection proposée dans le cadre du modèle CMco. Lorsque celle-ci ne contient aucun pixel « détecté », nous optons pour la segmentation AMTS.

4. Il s'agit ici d'assurer au mieux la stationnarité selon les dimensions spatiales, qui peut notamment varier dans les régions délimitant les observations de l'image *mosaic*.

5. Se reporter à [Bacon et al., 2017, section 5.1] pour une étude de cette grandeur en fonction de la longueur d'onde dans les images *udf10* et *mosaic*.

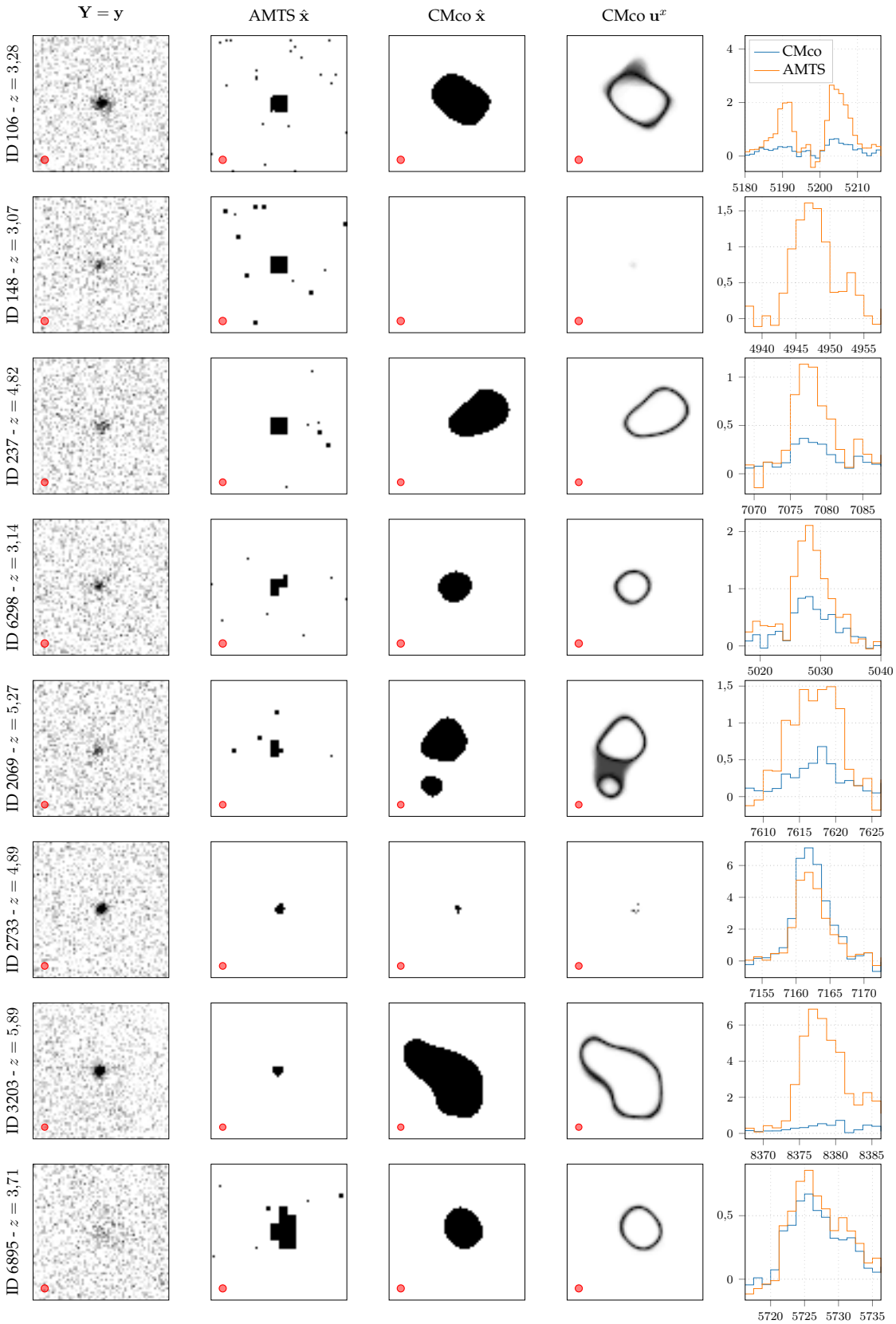



FIGURE 10.6 – Détections pour 8 objets *udf10* (ID 106, 148, 237, 6298) et *mosaic* (ID 2069, 2733, 3203, 6895). Colonne 1 : moyenne spectrale de l'image $\mathbf{Y} = \mathbf{y}$ (inverse vidéo), identifiant (« ID ») et redshift $z$. Colonnes 2-3 : résultats de détection. Colonne 4 : incertitudes associées à la méthode CMco (blanc : très certain, noir : très incertain). Colonne 5 : spectres moyens par détection, en $10^{-20}\mathrm{erg\ s^{-1}cm^{-2}\AA^{-1}}$ en fonction de la longueur d'onde (Å). Les disques rouges représentent la largeur à mi-hauteur de la FSF dans la raie d'émission de Lyman-alpha de l'objet.

**Ensemble des résultats**

Une synthèse des résultats obtenus avec ces choix est présentée en figure 10.7. Le tableau 10.1 dénombre les résultats de détection obtenus avec le modèle CMco, les cas pour lesquels la méthode du modèle AMTS est plus adaptée, ceux pour lesquels la détection ne permet pas d'affirmer la présence de sources étendues, et les absences de détection simultanées pour les deux méthodes.

Nous pouvons constater que la méthode basée sur le modèle CMco permet une détection de source étendue pour une majorité ($59{,}5\%$) des objets considérés. De plus, l'apport de la méthode AMTS permet de fournir des détections sur $16{,}5\%$ des objets restants, ce qui produit en tout un taux de détection de sources étendues de $76\%$.

L'inspection visuelle des résultats de non-détections ($12\%$) montre qu'elles se produisent lorsque :

— la largeur spectrale des cubes inclut un artefact de la réduction des données due à la soustraction de la lumière du ciel nocturne (cf. section 8.2.2) ;
— les intensités des objets brillants présents dans l'image hyperspectrale extraite n'ont pas été totalement supprimées par la soustraction de médiane ;
— un objet proche présente une raie d'émission lumineuse proche de la raie de Lyman-alpha recherchée.

Rappelons enfin que les méthodes employées ne font aucun *a priori* sur la forme spectrale des objets recherchés. Cela a pour avantage de pouvoir prendre en compte toutes les formes de raie d'émission de Lyman-alpha, notamment bimodales (ID 106, 148, ou 1301 par exemple). *A contrario*, cela signifie que la présence d'un signal homogène dans l'image, comme un artefact, sera considéré comme un signal à détecter.

Table 10.1 – Dénombrement des résultats de détection. La colonne « AMTS » indique les détections pour lesquelles cette méthode est plus adaptée que la méthode CMco, et la colonne « Non-étendu » celles ne permettant pas d'affirmer la présence de sources étendues.

| | CMco | | AMTS | | Non-étendu | | Non-détection | | Total |
|---|---|---|---|---|---|---|---|---|---|
| *mosaic* | 126 | 62,4% | 32 | 15,8% | 24 | 11,9% | 20 | 9,9% | 202 |
| *udf10* | 18 | 45,0% | 8 | 20,0% | 5 | 12,5% | 9 | 22,5% | 40 |
| Total | 144 | 59,5% | 40 | 16,5% | 29 | 12,0% | 29 | 12,0% | 242 |

### 10.3.3 Objets doubles

Les résultats concernant deux paires d'objets des images *mosaic* et *udf10* sont présentés en figure 10.8. Des résultats complémentaires, concernant 4 autres paires d'objets de l'image *mosaic*, sont reportés en annexe B.2.2.

Les résultats de détection ne mettent pas en évidence de détection étendue à l'échelle des images hyperspectrales traitées ($128 \times 128$ spectres). Nous pouvons cependant constater l'apparition des situations suivantes :

— la détection de régions séparées, comme pour la paire 412/6698 en section B.2.2 ;
— la mise en évidence, dans les autres cas illustrés, d'une région détectée de grande échelle par rapport aux objets individuels. Cette région englobe alors les deux objets, et est de taille comparable aux régions détectées dans le cas individuel ;
— l'apparition de régions éloignées des deux sources, comme dans les paires 265/633 en figure 10.8 et 1019/6488, 2071/6412 en section B.2.2. Dans la carte des incertitudes

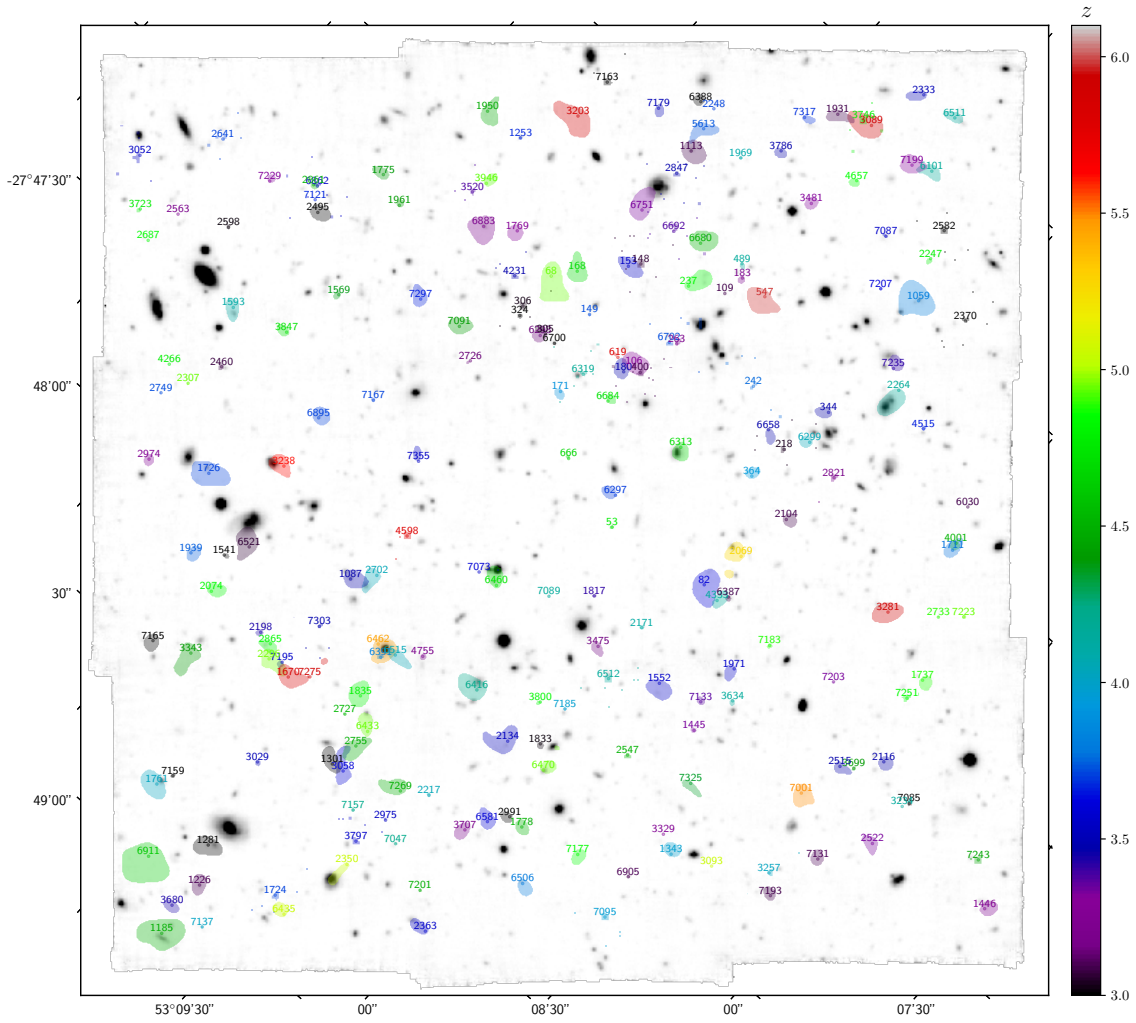



FIGURE 10.7 – Ensemble des 213 résultats de détection de sources individuelles superposés à la moyenne spectrale de l'image *mosaic* (inverse vidéo). Les abscisses et ordonnées sont dans le système de coordonnées équatoriales (J2000). Les objets détectés sont reportés par leur numéro de catalogue et le contour des détections associées, dont les couleurs correspondent au redshift $z$. Nous pouvons constater la grande variété des formes de détection produites : étendues ou non, circulaires ou non, etc. De plus, les objets ayant de grandes surfaces détectées semblent également être les objets de forme non circulaire.

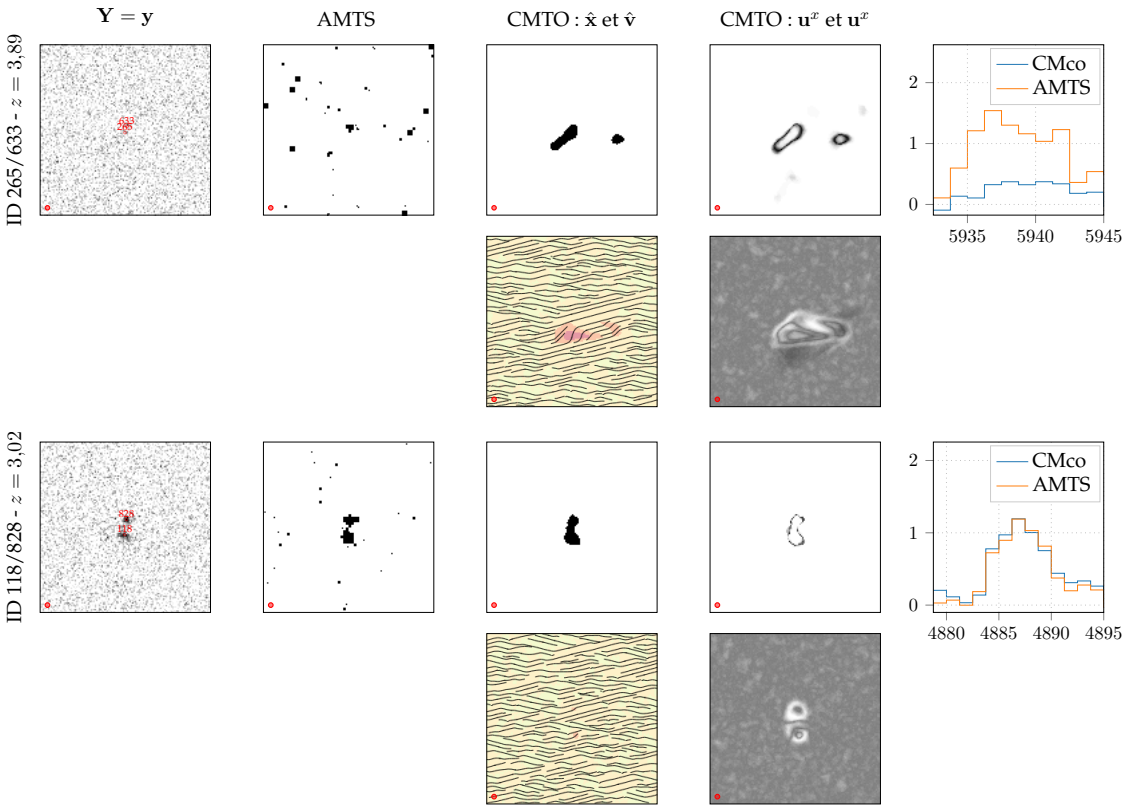



Figure 10.8 – Détections obtenues pour deux paires d'objets. Colonne 1 : moyenne spectrale de l'image $\mathbf{Y} = \mathbf{y}$ traitée, localisation (en rouge) des objets individuels, numéros de catalogue (« ID ») et redshift $z$. Colonne 2-3 : résultats de détection. Les orientations sont représentées par leur valeur et par les courbes tangentes en chaque point. Colonne 4 : incertitudes associées à la méthode CMTO (blanc : très certain, noir : très incertain). Colonne 5 : moyenne des spectres de $\mathbf{y}$ des régions détectées par les deux méthodes, exprimées en $10^{-20}\mathrm{erg\ s^{-1}cm^{-2}\AA^{-1}}$ en fonction de la longueur d'onde (Å). La largeur à mi-hauteur de la FSF dans la raie d'émission de Lyman-alpha des objets est représentée en rouge sur chaque image.

associées à la segmentation $\hat{\mathbf{x}}$ du modèle CMTO, ces régions apparaissent plus incertaines ;

— les orientations associées au modèle CMTO peuvent mettre en évidence des régions homogènes autour des objets détectés (paires 265/633, 1019/6488 et 2071/6412 en section B.2.2).

## 10.4 Conclusion

Ce chapitre a permis la comparaison puis l'application des modèles présentés dans ce manuscrit pour la détection d'émission de Lyman-alpha étendues spatialement au sein des images MUSE. L'analyse des méthodes sur des données synthétiques réalistes pour la formation des images a permis de mettre en évidence les méthodes les plus adaptées à la détection de structures localisées (« halos ») et de structures orientées à plus grande échelle (« filaments »).

L'application aux images hyperspectrales MUSE a permis de fournir une cartographie de l'émission Lyman-alpha étendue pour 76% des objets individuels traités. L'analyse détaillée

de ces résultats pourra permettre de mettre en évidence, dans les régions détectées, les contributions provenant de la source centrale des objets et celles provenant des émissions les plus étendues. Ensuite, il est envisageable de mettre en évidence les liens entre la forme des cartographies et les propriétés physiques des émetteurs de Lyman-alpha.

Les émetteurs de Lyman-alpha n'étant pas systématiquement isolés dans les images MUSE, une cartographie de paires d'émetteurs de Lyman a également été effectuée. Ces objets sont plus complexes, et requièrent là aussi une étude détaillée pour mettre en évidence les éventuelles interactions des émetteurs et leurs propriétés.

# Conclusions, perspectives

Les travaux présentés dans ce manuscrit ont permis d'explorer les thématiques de détection et de segmentation dans des images, avec pour application la détection d'objets dans les images hyperspectrales astronomiques. Quatre modèles ont ainsi été introduits, validés, puis comparés dans ce contexte. Pour chacun de ces modèles ainsi que pour l'application aux images réelles, cette partie présente un bref résumé des contributions puis les perspectives envisageables à l'issue des travaux de ce manuscrit.

## Détection par tests d'hypothèses

Dans le chapitre 3, nous avons proposé une méthode originale de détection par tests d'hypothèses, pour application à la détection de sources ténues et étendues dans des images hyperspectrales. Cette méthode repose sur des tests GLR contraints permettant la formulation explicite des caractéristiques recherchées : localisation spectrale, structure spatiale dépendant de la FSF, apparition au sein de multiples observations. Nous avons de plus introduit des tests GLR reposant sur une contrainte de similarité entre spectres et avec un ensemble de spectres. De plus, il a été possible de mettre en évidence les propriétés statistiques de ces tests, et de les vérifier expérimentalement. Après validation sur données de synthèse, l'application aux images hyperspectrales issues de l'instrument MUSE a permis de proposer de premières cartographies de détection des halos circum-galactiques.

Plusieurs perspectives sont envisageables à partir de ces travaux :

— comme cela a été montré dans le chapitre 9, les spectres des images MUSE présentent un certain degré de corrélation au sein d'un voisinage local. Il serait donc envisageable d'étudier les statistiques des tests proposés en tenant compte de cette corrélation spatiale ;

— la formulation analytique est présentée dans le cadre d'un bruit de distribution normale multivariée. Il serait possible d'étudier leur formulation, ainsi que l'estimation des paramètres, de distribution non centrée (cf. [Meillier et al., 2017]) puisque cela a déjà été abordé avec les images MUSE [Bacher et al., 2017a] ;

— les propriétés statistiques ont été étudiées en établissant la probabilité théorique de fausse alarme. Il serait envisageable de recourir à la théorie des tests d'hypothèses multiples et de proposer un contrôle du FDR, comme cela est proposé dans [Bacher et al., 2017a,b].

## Détection et champs de Markov couples convolutifs

Dans le chapitre 5, nous proposons un modèle de champs de Markov couples permettant la prise en compte de la convolution affectant les images. Ces travaux permettent la formulation du problème de détection du chapitre 3, comme un cas particulier d'un problème de

segmentation en contexte non supervisé. En ce sens, la méthode de détection du chapitre 3 s'apparente à une méthode de segmentation contextuelle.

*A contrario*, le modèle de champ de Markov couple intègre un *a priori* sur la structure du processus de classes $\mathbf{X}$ recherché. Nous avons dans ce cadre présenté la méthode de segmentation non supervisée, ainsi qu'une estimation des incertitudes liées à la segmentation du MPM. Les résultats sur données de synthèse ont illustré l'intérêt de ce modèle par rapport à la détection par test [Courbot et al., 2017a] et à la méthode développée dans [Bacher et al., 2017a,b]. Ce modèle offre de très bons résultats de segmentation, et en particulier un taux de faux positif très proche de 0 à tout RSB pour la segmentation par le MPM.

Ces travaux ont mis en évidence plusieurs pistes envisageables pour des travaux futurs. La comparaison du modèle CMco avec la méthode de détection par test d'hypothèse pourrait en effet permettre :

— la formulation explicite des tests proposés en tant que méthodes de segmentation contextuelles ;
— la description des méthodes de segmentation sous la forme de tests d'hypothèses avec *a priori* markovien. Cela permettrait également d'étudier le lien entre le modèle markovien et les probabilités de fausse alarmes attendues ;
— la comparaison des méthodes employées pour l'estimation des paramètres.

Nous avons également présenté, dans le chapitre 5, une formulation alternative du modèle de champ de Markov couple convolutif par le biais d'un modèle de mélange avec *a priori* markovien. Il serait donc intéressant d'étudier dans quel cadre les modèles de la littérature de cette forme peuvent être formulés comme des modèles de Markov couples particuliers, et de comparer les méthodes employées sur le plan de l'estimation des paramètres, de la décision et des résultats.

## Champs de Markov triplets orientés

Nous avons introduit dans le chapitre 6 un modèle de champ de Markov modélisant les orientations et des classes dans les images. Par rapport aux travaux des chapitres 3 et 5, cette étude s'inscrit dans le cadre plus général de la segmentation des images. Nous présentons de plus la segmentation bayésienne conjointe, en contexte non supervisé, des orientations et des classes dans les images. Les résultats sur des images synthétiques et réelles illustrent l'apport de ce modèle pour la segmentation, en offrant des performances meilleures que le modèle « classique » de champ de Markov caché à bruit indépendant. De plus, l'information fournie par la segmentation des orientations est une information complémentaire pertinente sur l'image traitée. Les résultats du chapitre 10 illustrent ce dernier point. En effet, les segmentations obtenues avec des images hyperspectrales synthétiques ont montré que la restauration des orientations permet de repérer la présence de structures ténues et orientées à grande échelle.

Ce modèle offre de nombreuses possibilités concernant la modélisation des orientations :

— l'extension du modèle à des images volumétriques (médicales, par exemple) permettrait la modélisation d'orientations dans un volume, et pourrait par exemple être applicable à la segmentation de vaisseaux sanguins ou de tissus fibreux ;
— les orientations, comme les classes, prennent leurs valeurs dans des ensembles discrets $\Omega_v$ et $\Omega_x$. Comme évoqué en fin du chapitre 3, il serait également envisageable de modéliser les orientations de manière continue, ce qui ferait appel à des classes d'algorithmes différentes, comme des méthodes de lissage ou de restauration d'images ;

— alternativement, il serait enfin possible d'étudier une modélisation floue des orientations, comme cela existe déjà pour les classes (cf. [Salzenstein et Collet, 2006]).

## Arbres de Markov triplets spatiaux

Dans le chapitre 7, nous avons proposé un modèle d'arbre de Markov triplet spatial permettant la prise en compte conjointe d'un *a priori* de hiérarchie et d'homogénéité spatiale dans les images. Ces travaux se placent également dans le contexte de la segmentation des images. Les résultats sur des données de synthèse ont montré que ce modèle était compétitif avec un modèle de champ de Markov caché, tout en permettant des calculs exacts pour réaliser la segmentation. De plus, les résultats obtenus dans le chapitre 10 pour la détection dans des images hyperspectrales ont montré que le modèle AMTS offre de meilleures capacités de détection que les modèles par champs dans des données extrêmement bruitées.

Deux possibilités d'enrichissement du modèle se dégagent :

— l'étude de paramètres variables selon la résolution permettrait d'élaborer un modèle évolutif (cf. [Monfrini et Pieczynski, 2005]), qui permettrait de renforcer la contribution de la hiérarchie à faible résolution et de renforcer la prise en compte de l'homogénéité spatiale à résolution élevée ;

— la distribution régulant l'homogénéité spatiale est isotrope dans le modèle proposé. Ce modèle peut être étendu à la modélisation des anisotropies, similaires à celle introduite dans le chapitre 6. Un second processus auxiliaire, représentant les orientations privilégiées, pourrait ainsi être adjoint aux processus $\mathbf{X}$ et $\mathbf{V}$ et permettrait la segmentation de structures hiérarchiques orientées.

## Application aux images hyperspectrales MUSE

La problématique applicative de ces travaux de thèse portait sur la détection de structures ténues (halos et filaments) au sein des images MUSE. Dans le chapitre 10, nous avons tout d'abord conduit une évaluation des quatre modèles présentés dans ce manuscrit sur des données synthétiques. Cette comparaison a permis de conclure que :

— la segmentation par champs de Markov couples convolutifs offre, dans le cas général, les meilleures performances concernant la détection de structures localisées autour des objets individuels (« halos »). Ce constat est à nuancer à très faibles RSB, pour lesquels la méthode AMTS offre de meilleures performances.

— la segmentation par le modèle CMTO est la plus adaptée à la détection de structures plus étendues, englobant plusieurs objets (« filaments »). En effet, la segmentation des orientations apporte une information complémentaire permettant de mettre en évidence ces structures orientées.

Les résultats de segmentation sur des images réelles issues de MUSE se sont également avérés satisfaisants : sur 242 objets isolés, 184 détections de sources étendues ont été réalisées (soit 76% de détection).

Les perspectives de ces travaux sont multiples :

— l'exploitation des méthodes de détection sur de très nombreux objets astronomiques permettra d'exploiter les caractéristiques de forme et d'orientation sur un large échantillon (plus de 800 dans l'observation *mosaic*) ;

— l'étude des objets multiples pourra également permettre une étude des interactions potentielles entre objets ;

— il est naturellement possible d'employer les modèles proposés dans d'autres données, et en particulier des observations de champs profonds obtenues en utilisant l'optique

adaptative ;
— les méthodes proposées ne disposent, à l'exception de la détection par test, d'aucun *a priori* sur la forme spectrale recherchée : cela signifie qu'elles peuvent être employées sur d'autres objets que les émetteurs de Lyman-alpha.

# Annexes

# A

# Algorithmes pour les modèles de champs de Markov



Cette annexe présente le détail des algorithmes employés pour la segmentation par champs de Markov couples convolutifs (chapitre 5) et par champs de Markov triplets orientés (chapitre 6). Les algorithmes de segmentation et d'estimation sur lesquels s'appuient les résultats présentés dans ces chapitres appartiennent à la même classe de méthode. Nous les présentons ici dans le cadre du modèle de champs de Markov triplets, leur adaptation dans le cadre du modèle couple étant immédiate.

Rappelons que dans le cadre d'un modèle par champ de Markov triplet, nous nous donnons trois processus $\mathbf{Y} = (Y_s)_{s\in\mathcal{S}}$, $\mathbf{X} = (X_s)_{s\in\mathcal{S}}$ et $\mathbf{V} = (V_s)_{s\in\mathcal{S}}$, où $\mathcal{S}$ est l'ensemble des sites de l'image.

## A.1 Échantillonneur de Gibbs

### A.1.1 Échantillonneur de Gibbs de la distribution *a posteriori*

Dans le cadre de la segmentation, il est nécessaire de connaître la distribution $p(\mathbf{x}, \mathbf{v}|\mathbf{y})$. Celle-ci est estimée par un échantillonneur de Gibbs, donc le principe est reporté dans l'algorithme 2.

Les valeurs initiales $(\mathbf{x}^0, \mathbf{v}^0)$ peuvent être choisies « proches » d'une solution souhaitée peut permettre une convergence plus rapide. En l'absence de telles informations, $(\mathbf{x}^0, \mathbf{v}^0)$ sont choisies au hasard dans $(\Omega_x \times \Omega_v)^{|\mathcal{S}|}$.

L'algorithme effectue autant de parcours que nécessaire afin que l'écart entre les paires $(\mathbf{x}^p, \mathbf{v}^p)$ consécutives soit faible (par exemple, moins de $5\%$) durant plusieurs itérations ($10$ par exemple), nous supposons que la limite est atteinte et stoppons l'algorithme.

Dans l'échantillonneur de Gibbs, le point critique en terme de coût calculatoire est le calcul de $p(X_s = \omega, V_s = \nu | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$ au sein de chaque pixel d'un parcours. Une première adaptation de l'algorithme consiste à calculer cette distribution en amont des parcours (avant l'étape 1) pour toutes les combinaisons possibles de $(\omega, \nu, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$. Cela transforme l'étape de calcul (étape 6) en une étape d'accès à un jeu de données, ce qui a pour effet d'accélérer significativement les calculs.

**Algorithme 2** Échantillonneur de Gibbs de la distribution *a posteriori* $p(\mathbf{x}, \mathbf{v}|\mathbf{y})$

**Requiert :** Observation $\mathbf{Y} = \mathbf{y}$, connaissance de la distribution $p(x_s, v_s | y_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$ et de ses paramètres.
**Produit :** Séquence $\{\mathbf{x}^0, \mathbf{v}^0, \mathbf{x}^1, \mathbf{v}^1, \ldots, \mathbf{x}^P, \mathbf{v}^P\}$
1: Choix de valeurs initiales $\mathbf{x}^0, \mathbf{v}^0$.
2: Initialiser $p$ à $0$.
3: **Tant que** critère de convergence non atteint :
4: Initialiser $\mathbf{x}^p, \mathbf{v}^p$ par $\mathbf{x}^{p-1}, \mathbf{v}^{p-1}$.
5: **Pour** chaque site $s \in \mathcal{S}$ :
6: Calcul de $p(X_s = \omega, V_s = \nu | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$ pour tout $\omega \in \Omega_x, \nu \in \Omega_v$.
7: Tirage de $x_s, v_s$ selon $p(X_s, V_s | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$.
8: Mise à jour de $x^p_s, v^p_s$ avec les valeurs simulées.
9: **fin Pour**
10: $p \leftarrow p + 1$
11: **fin Tant que**

### A.1.2 Échantillonneur de Gibbs chromatique

Le point critique subsistant dans l'échantillonneur de Gibbs est le parcours un à un des pixels de l'image. L'échantillonneur de Gibbs chromatique [Gonzalez et al., 2011] propose une solution de contournement de ce problème. En effet, il permet un tirage simultané de réalisations de $p(\mathbf{t}_s | \mathbf{t}_{N_s})$ en plusieurs sites non mutuellement voisins. Ces sites forment des subdivisions de $\mathcal{S}$ nommées « couleurs » dans [Gonzalez et al., 2011]. De telles subdivisions sont illustrées en figure A.1, et la procédure est reportée dans l'algorithme 3.

L'algorithme bénéficie des mêmes propriétés de convergence que l'algorithme de Gibbs « classique », et celle-ci est évaluée de la même manière que précédemment. La boucle comportant les calculs dans les ensembles $\mathcal{S}_i$ (étapes 7-9) peut être vectorisée. En terme de temps de calcul, le gain par rapport à l'échantillonnage de Gibbs standard est très important, comme illustré en figure A.2 : le gain de temps est d'un facteur $1700$ pour une image de taille $128 \times 128$ pixels.

## A.2 Algorithmes de segmentation

### A.2.1 Algorithme de Marroquin pour le calcul du MPM

Rappelons le critère du MPM [Marroquin et al., 1987] :

$$\forall s \in \mathcal{S}(\hat{x}_s, \hat{v}_s)^{\mathrm{MPM}} = \underset{(\omega,\nu)\in\Omega_x\times\Omega_v}{\arg\max}\ p(X_s = \omega, V_s = \nu | \mathbf{Y} = \mathbf{y}). \tag{A.1}$$

**Algorithme 3** Échantillonneur de Gibbs chromatique de la distribution *a posteriori*

**Requiert :** Observation $\mathbf{Y} = \mathbf{y}$, connaissance de la distribution $p(x_s, v_s | y_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$ et de ses paramètres, et subdivisions de $\mathcal{S}$ en $\mathcal{S}_0, \mathcal{S}_1, ..., \mathcal{S}_n$ formant des ensembles de sites non mutuellement voisins.
**Produit :** Séquence $\{\mathbf{x}^0, \mathbf{v}^0, \mathbf{x}^1, \mathbf{v}^1, \ldots, \mathbf{x}^P, \mathbf{v}^P\}$
1: Choix de valeurs initiales $\mathbf{x}^0, \mathbf{v}^0$.
2: Initialiser $p$ à 0.
3: **Tant que** critère de convergence non atteint :
4: Initialiser $\mathbf{x}^p, \mathbf{v}^p$ par $\mathbf{x}^{p-1}, \mathbf{v}^{p-1}$.
5: **Pour** chaque ensemble $\mathcal{S}_i$ :
6: **Pour** chaque site $s \in \mathcal{S}_i$ :
7: Calcul de $p(X_s = \omega, V_s = \nu | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$ pour tout $\omega \in \Omega_x, \nu \in \Omega_v$.
8: Simulation de $x_s, v_s$ selon $p(X_s, V_s | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$.
9: Mise à jour de $x^p_s, v^p_s$ avec les valeurs simulées.
10: **fin Pour**
11: **fin Pour**
12: $p \leftarrow p + 1$
13: **fin Tant que**

L'algorithme de Marroquin permet d'estimer la segmentation au sens du MPM. Il repose sur un grand nombre de tirages de la distribution *a posteriori* $p(\mathbf{x}, \mathbf{v} | \mathbf{y})$ réalisés par l'échantillonneur de Gibbs (algorithme 2 ou 3). La procédure est reportée dans l'algorithme 4.

**Algorithme 4** Algorithme de Marroquin pour le calcul du MPM

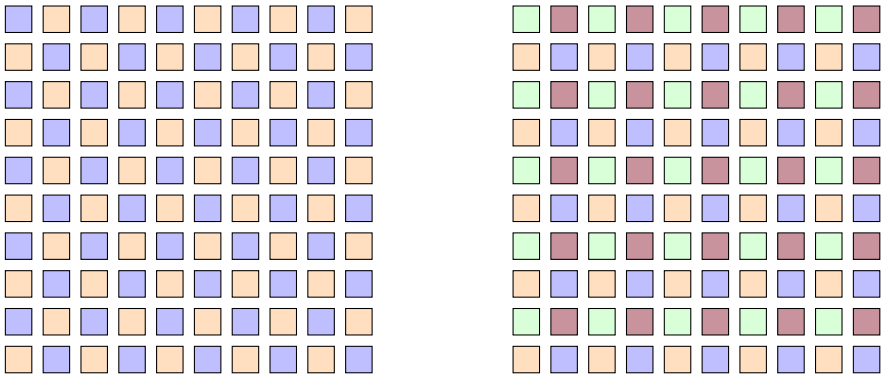
**Requiert :** Observation $\mathbf{Y} = \mathbf{y}$, connaissance de la distribution $p(x_s, v_s | y_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$ et de ses paramètres, nombre d'échantillons requis $N_{\text{Gibbs}}$.
**Produit :** Estimation $(\hat{\mathbf{x}}, \hat{\mathbf{v}})^{\text{MPM}} = \left((\hat{x}_s, \hat{v}_s)^{\text{MPM}}\right)_{s \in \mathcal{S}}$.
1: **Pour** chaque échantillon $e$ dans $1, \ldots, N_{\text{Gibbs}}$ :
2: Échantillonnage de la distribution *a posteriori* $p(\mathbf{x}, \mathbf{v} | \mathbf{y})$ : algorithme 2 ou 3.
3: Stockage de la dernière étape $(\mathbf{x}^P, \mathbf{v}^P)$ de la séquence en $(\mathbf{x}^e, \mathbf{v}^e)$.
4: **fin Pour**
5: Calcul de l'estimation $\hat{p}(X_s = \omega, V_s = \nu | \mathbf{Y} = \mathbf{y})$ par les fréquences sur l'ensemble des réalisations $(\mathbf{x}^e, \mathbf{v}^e)$.
6: Estimation du MPM $\forall s \in \mathcal{S}$ par $(\hat{x}_s, \hat{v}_s)^{\text{MPM}} = \underset{(\omega,\nu) \in \Omega_x \times \Omega_v}{\arg\max}\ \hat{p}(X_s = \omega, V_s = \nu | \mathbf{Y} = \mathbf{y})$.

### A.2.2 Algorithme ICM pour l'estimation du MAP

Rappelons que le critère du MAP s'écrit [Geman et Geman, 1984] :

$$(\hat{\mathbf{x}}, \hat{\mathbf{v}})^{\text{MAP}} = \underset{(\boldsymbol{\omega},\boldsymbol{\nu}) \in (\Omega_x \times \Omega_v)^{|\mathcal{S}|}}{\arg\max}\ p(\mathbf{X} = \boldsymbol{\omega}, \mathbf{V} = \boldsymbol{\nu} | \mathbf{Y} = \mathbf{y}). \tag{A.2}$$

L'algorithme ICM permet l'estimation du critère de segmentation du MAP. Il forme une version « déterministe » de l'échantillonneur de Gibbs. En effet, l'étape de simulation (étape 6 dans l'algorithme 2) selon $p(X_s, V_s | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$ devient un choix déterministe de la classe la plus probable selon cette distribution. Le choix des valeurs initiales est identique à celui décrit pour l'échantillonneur de Gibbs. La procédure est décrite dans l'algorithme 5.



(a) 4-voisinage (b) 8-voisinage

FIGURE A.1 – Illustration sur des ensembles $\mathcal{S}$ contenant $10 \times 10$ sites de l'échantillonneur de Gibbs chromatique (algorithme 3). Chaque carré représente un site et chaque couleur une subdivision de $\mathcal{S}$.

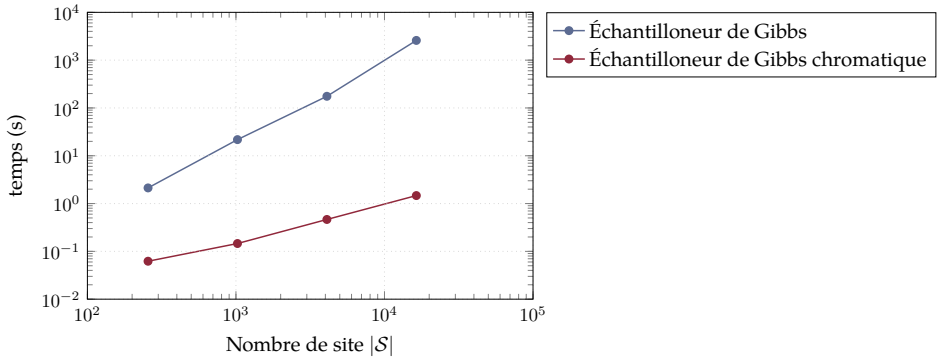



FIGURE A.2 – Temps de calculs mesurés pour un échantillonneur de Gibbs et un échantillonneur de Gibbs chromatique en fonction du nombre de sites considérés.

## A.3 Estimation des paramètres

### A.3.1 Algorithme

L'algorithme SEM est présenté en section 4.2.3 dans le cas général. Nous détaillons ici un algorithme inspiré de SEM pour l'estimation des paramètres, décrit en section 6.2.3 dans le cas du modèle CMTO. Rappelons que l'ensemble $\theta$ des paramètres à estimer est, d'après (6.16) :

— $\alpha$ et $\beta$ régulant les potentiels de type Gibbs dans (6.10) et (6.11) respectivement ;
— $g$, décrit en (6.15) ;
— $K = |\Omega_x|$ variances $\sigma^2_{x_s}$ et $K$ moyennes $\mu_{x_s}$.

La méthode implémentée est reportée dans l'algorithme 6. Ce dernier est itératif, et repose sur un jeu de paramètres initiaux $\theta^0$. Pour l'établir, nous effectuons une classification de l'image au sens des K-moyennes [MacQueen et al., 1967]. Cette étape permet de former une première segmentation $\mathbf{x}^0$. Pour former la segmentation $\mathbf{v}^0$ associée, nous estimons les orientations des contours de $\mathbf{x}^0$ et choisissons, en chaque site $s$, l'orientation la plus fréquente parmi les 5 plus proches de $s$ [1]. Ensuite, l'ensemble des paramètres $\theta^0$ est estimé

1. Cette démarche est un cas particulier d'algorithme des $K$ plus proches voisins utilisé pour compléter les données d'orientations.

**Algorithme 5** Modes conditionnels itératifs (ICM) pour l'estimation du MAP

**Requiert :** Observation $\mathbf{Y} = \mathbf{y}$, connaissance de la distribution $p(x_s, v_s | y_s, \mathbf{x}_{N_s}, \mathbf{v}_{N_s})$ et de ses paramètres.
**Produit :** Séquence $\{\mathbf{x}^0, \mathbf{v}^0, \mathbf{x}^1, \mathbf{v}^1, \ldots, \mathbf{x}^P, \mathbf{v}^P\}$
1: Choix de valeurs initiales $\mathbf{x}^0, \mathbf{v}^0$.
2: Initialiser $p$ à 0.
3: **Tant que** critère de convergence non atteint :
4: Initialiser $\mathbf{x}^p, \mathbf{v}^p$ par $\mathbf{x}^{p-1}, \mathbf{v}^{p-1}$.
5: **Pour** chaque site $s \in \mathcal{S}$ :
6: Calcul de $p(X_s = \omega, V_s = \nu | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$ pour tout $\omega \in \Omega_x, \nu \in \Omega_v$.
7: Choix déterministe de $x^p_s, v^p_s = \underset{(\omega,\nu)\in(\Omega_x \times \Omega_v)}{\arg\max} p(X_s = \omega, V_s = \nu | y_s, \mathbf{x}^p_{N_s}, \mathbf{v}^p_{N_s})$.
8: **fin Pour**
9: $p \leftarrow p + 1$
10: **fin Tant que**

sur les données complètes formées par $(\mathbf{y}, \mathbf{x}^0, \mathbf{v}^0)$ à l'aide des estimateurs présentés en section 6.2.3.

L'algorithme n'a pas de propriété de convergence, il faut donc décider de celle-ci empiriquement. L'algorithme étant stochastique, la convergence ne peut avoir lieu qu'en moyenne. En conséquence, nous mesurons l'écart entre les paramètres à une itération $q$ donnée et la moyenne des paramètres sur les $I$ itérations précédentes. Cela signifie par ailleurs que l'algorithme effectue au minimum $I$ itérations avant d'être stoppé. Nous posons en pratique $I = 10$ itérations. De plus, les paramètres ne sont pas commensurables, il faut donc mesurer un écart relatif. Nous posons par exemple pour le paramètre $\alpha$ :

$$\kappa_\alpha(q) = \frac{\left| \alpha^q - \frac{1}{I} \sum_{q'=1}^{I} \alpha^{q-q'} \right|}{\frac{1}{I} \sum_{q'=1}^{I} \alpha^{q-q'}}. \tag{A.3}$$

Nous associons de même une grandeur $\kappa$ à chaque paramètre. Notons que dans le cas de paramètres vectoriels ou matriciels, cette définition doit être adaptée. Il est par exemple possible de mesurer la moyenne des écarts terme à terme. Lorsque toutes ces grandeurs ont une valeurs inférieures à une valeur limite (0,05 dans notre implémentation), l'algorithme est arrêté.

**Algorithme 6** Estimation des paramètres.

**Requiert :** Observation $\mathbf{Y} = \mathbf{y}$.
**Produit :** Séquence de paramètres $\{\theta^1, \theta^2, \ldots, \theta^Q\}$.
1: Initialisation par un jeu de paramètres $\theta^0$.
2: **Tant que** Critère de convergence non atteint :
3: Simulation de $(\mathbf{x}^q, \mathbf{v}^q)$ selon $p_{\hat{\theta}^{q-1}}(\mathbf{x}, \mathbf{v} | \mathbf{y})$.
4: Estimation de $\hat{\theta}^q$ sur les « pseudo-données complètes » formées par $(\mathbf{y}, \mathbf{x}^q, \mathbf{v}^q)$ avec les estimateurs de la section 6.2.3.
5: **fin Tant que**

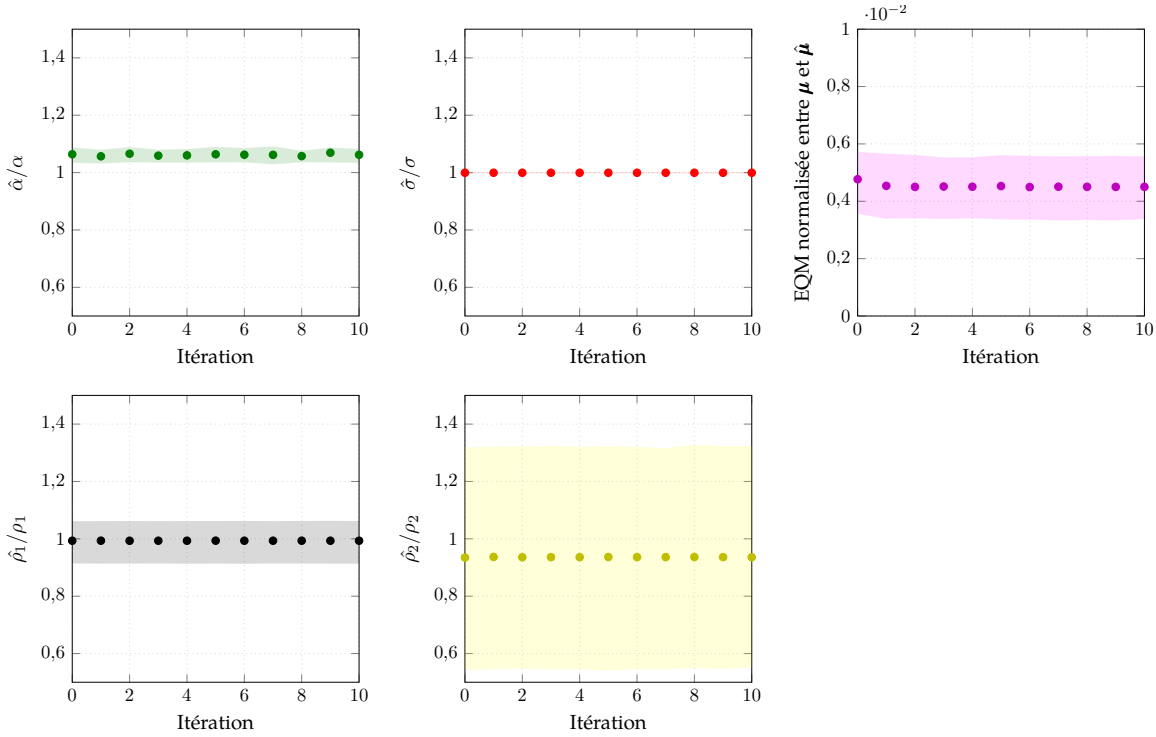



Figure A.3 – Exemple d'évolution de l'estimation des paramètres en fonction des itérations de l'algorithme 6. Chaque point représente la moyenne obtenue sur 100 réalisations différentes de $\mathbf{Y} = \mathbf{y}$ sur laquelle l'algorithme travaille. Les enveloppes des courbes correspondent aux premier et troisièmes quartiles des résultats.

### A.3.2 Exemples de résultats

Nous présentons dans cette section quelques résultats d'estimation des paramètres dans la cadre du modèle de champs de Markov couple convolutif, introduit dans le chapitre 5. Les paramètres à estimer sont les suivants :

— le paramètre $\alpha$ régulant la distribution de champ de Markov ;
— la moyenne spectrale $\boldsymbol{\mu}$ ;
— l'écart-type $\sigma^2$ ;
— les paramètres de corrélation spectrale $\rho_1$ et $\rho_2$.

Nous nous plaçons dans la même configuration expérimentale que dans la section 5.3.1. Nous effectuons 100 estimations des paramètres avec des réalisations $\mathbf{Y} = \mathbf{y}$ différentes, sous $\mathrm{RSB} = -14$ dB. Pour mesurer les performances de l'estimation, nous mesurons pour les paramètres scalaires l'écart relatif entre les valeurs réelles et mesurées, et l'erreur quadratique moyenne (EQM) normalisée pour le paramètre vectoriel $\boldsymbol{\mu}$. Pour $\alpha$, la valeur estimée sur la vérité terrain $\mathbf{x}$ sert de valeur de référence. Les valeurs des paramètres sont :

— $\alpha = 1{,}299$ ;
— $\sigma = 1{,}273$ ;
— $\rho_1 = 0{,}081$ et $\rho_2 = 0{,}016$ ;
— $\boldsymbol{\mu}$ a trois coefficients non nuls, comme illustré en figure 5.2e.

La figure A.3 reporte les résultats obtenus. Plusieurs constats peuvent être tirés de cet exemple :

— l'algorithme atteint très rapidement (1 à deux itérations) un état stable, dans lequel les paramètres évoluent très peu ;
— pour les paramètres $\sigma$ et $\rho_1$, l'estimateur semble sans biais et de variance constante

selon les itérations ;

— un biais est en revanche observable sur les estimations des paramètres $\alpha$ et $\rho_2$. Le paramètre $\rho_2$ étant de faible valeur, un faible écart induit un grand écart relatif, ce qui explique la grande variance relative de l'estimateur. Pour $\alpha$, cela peut être imputé aux limites de l'estimateur des moindres carrés de [Derin et Elliott, 1987]. Plusieurs études ont spécifiquement traité le problème de l'estimation de ce paramètre, par des modèles MCMC [Pereyra et al., 2013; Descombes et al., 1999] notamment.

Notons enfin que dans l'implémentation des méthodes, les segmentations sont très peu influencées par les légers apparaissant dans l'estimation des paramètres.

# B

# Résultats additionnels



## B.1 Champs de Markov triplets orientés : textures de Brodatz

Nous reportons ici des résultats complémentaires de ceux présentés dans le chapitre 6. Ces résultats concernent la segmentation d'images issues des textures de Brodatz [Brodatz, 1966; nor, 2017], et sont reportés dans les figures B.1 et B.2.

La légende de ces figures et la suivante :

— ligne 1 : observation $\mathbf{Y} = \mathbf{y}$.
— ligne 2 : résultats de segmentation à trois classes par le MPM dans le cadre d'un modèle de champ de Markov caché à bruit indépendant ;
— ligne 3 : résultat de segmentation à trois classes au sens du MPM dans le modèle CMTO ;
— ligne 4 : incertitudes $\mathbf{u}^x$ associées à cette segmentation. Nous pouvons constater, comme dans le chapitre 6, qu'elles traduisent bien les changement de stationnarité au sein de $\hat{\mathbf{x}}$ ;
— ligne 5 : segmentation des orientations avec $\Omega_v = \{\pi/12, 3\pi/12, \ldots, 11\pi/12\}$, au sens du MPM dans le modèle CMTO. Elles sont représentées par leur valeur et par les courbes tangentes en chaque point (en rouge). Nous pouvons constater, avec les images 0 et 2 par exemple, que cette segmentation suit bien les orientations apparentes dans $\mathbf{y}$ ;
— ligne 6 : incertitudes associées à $\hat{\mathbf{v}}$. Elles traduisent là aussi les changement de stationnarité dans cette segmentation, et notamment au sein des régions ne présentant pas de direction privilégiée apparente.

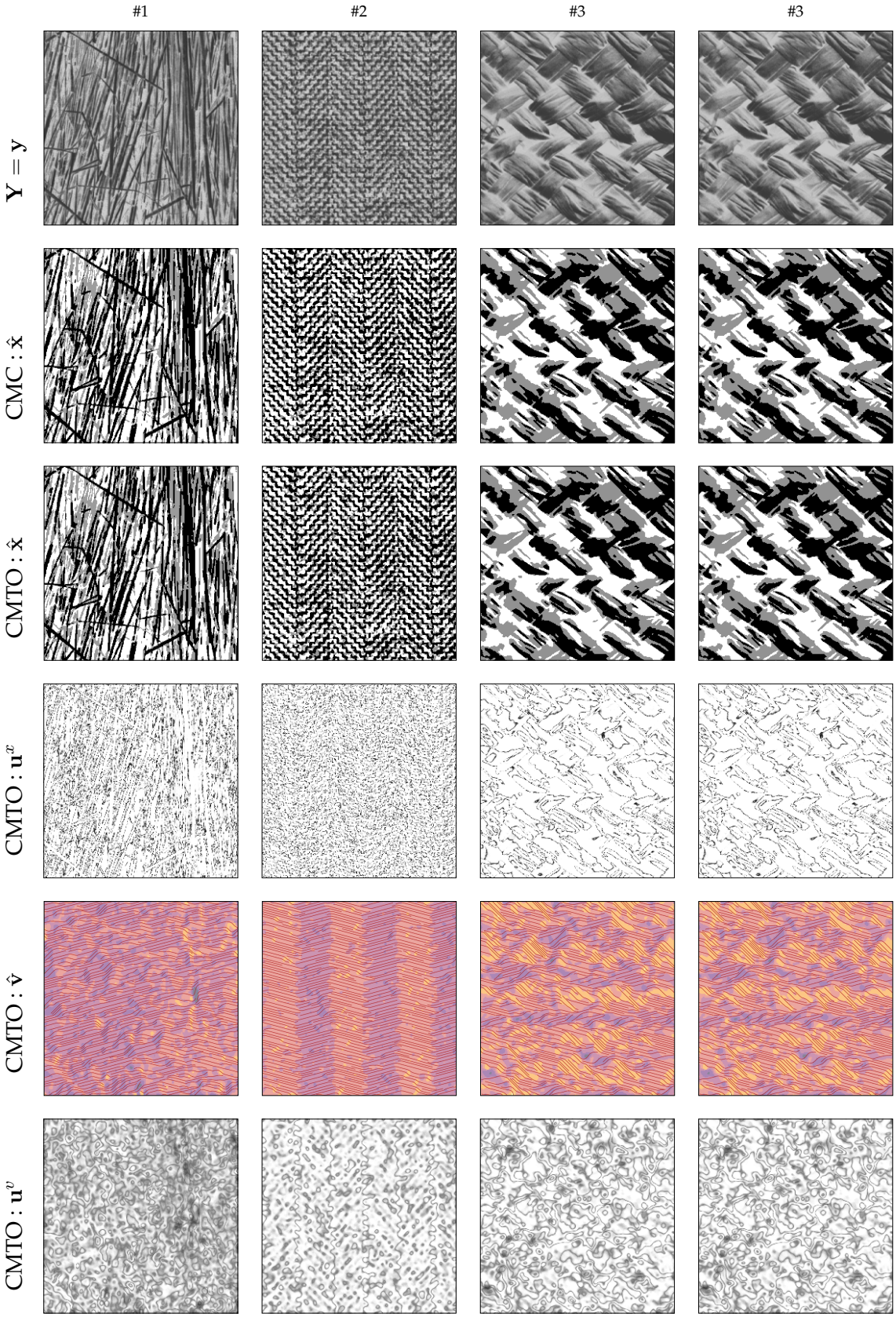



Figure B.1 – Résultats de segmentation sur 8 textures de Brodatz (1/2).

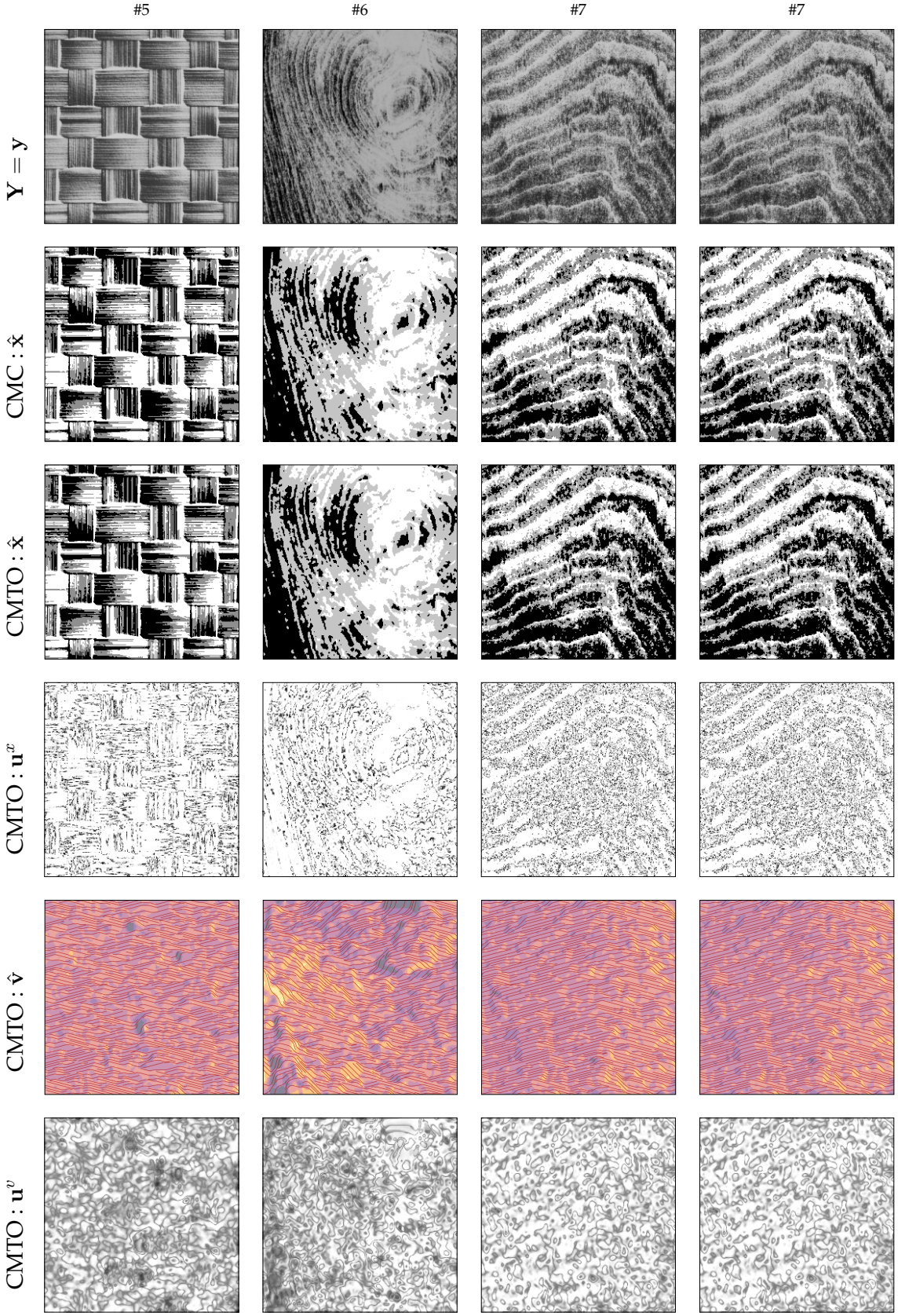



Figure B.2 – Résultats de segmentation sur 8 textures de Brodatz (2/2) .

## B.2 Images hyperspectrales MUSE

Ces résultats complètent ceux présentés dans le chapitre 10 sur la détection de sources étendues dans les images hyperspectrales MUSE.

### B.2.1 Objets individuels

Les résultats sont reportés dans les figures B.3 (image *udf10*), B.4, B.5 et B.6 (image *mosaic*). La légende de ces figures est la suivante :

— colonne 1 : moyenne spectrale de l'image hyperspectrale $\mathbf{Y} = \mathbf{y}$ de taille $64 \times 64$ spectres (inverse vidéo) ;
— colonnes 2-3 : résultats de détection ;
— colonne 4 : incertitudes associes à la méthode CMco (blanc : très certain, noir : très incertain) ;
— colonne 5 : moyenne des spectres de $\mathbf{y}$ de la région détectée pour les deux méthodes. Ces spectres sont exprimés exprimées en $10^{-20}\mathrm{erg\ s^{-1}cm^{-2}\AA^{-1}}$ en fonction de la longueur d'onde (Å). Lorsque la méthode ne produit pas de détection, aucun spectre n'est affiché.

Les objets sont de plus repérés par leur identifiant (« ID ») et leur redshift, en ordonnée dans la colonne 1. Le cercle rouge affiché en chaque image illustre la lageur à mi-hauteur de la FSF à la longueur d'onde d'émission de Lyman-alpha de l'objet.

### B.2.2 Objets doubles

Les résultats de détection d'objets doubles de l'image *mosaic* sont reportés dans la figure B.7. La légende de cette figure est la suivante :

— colonne 1 : moyenne spectrale de l'image $\mathbf{Y} = \mathbf{y}$ de taille $128 \times 128$ spectres, et localisation (en rouge) des objets individuels ;
— colonne 2-3 : résultats de détection. Les orientations sont représentées par leur valeur et par les courbes tangentes en chaque point ;
— colonne 4 : incertitudes associes à la méthode CMTO (blanc : très certain, noir : très incertain) ;
— colonne 5 : moyenne des spectres de $\mathbf{y}$ des régions détectées par les deux méthodes, exprimés en $10^{-20}\mathrm{erg\ s^{-1}cm^{-2}\AA^{-1}}$ en fonction de la longueur d'onde (Å).

Les objets sont repérés par leur identifiant (« ID ») et leur redshift dans la colonne 1, et les cercles rouges représentent la largeur à mi-hauteur de la FSF pour la longueur d'onde de Lyman-alpha des objets.

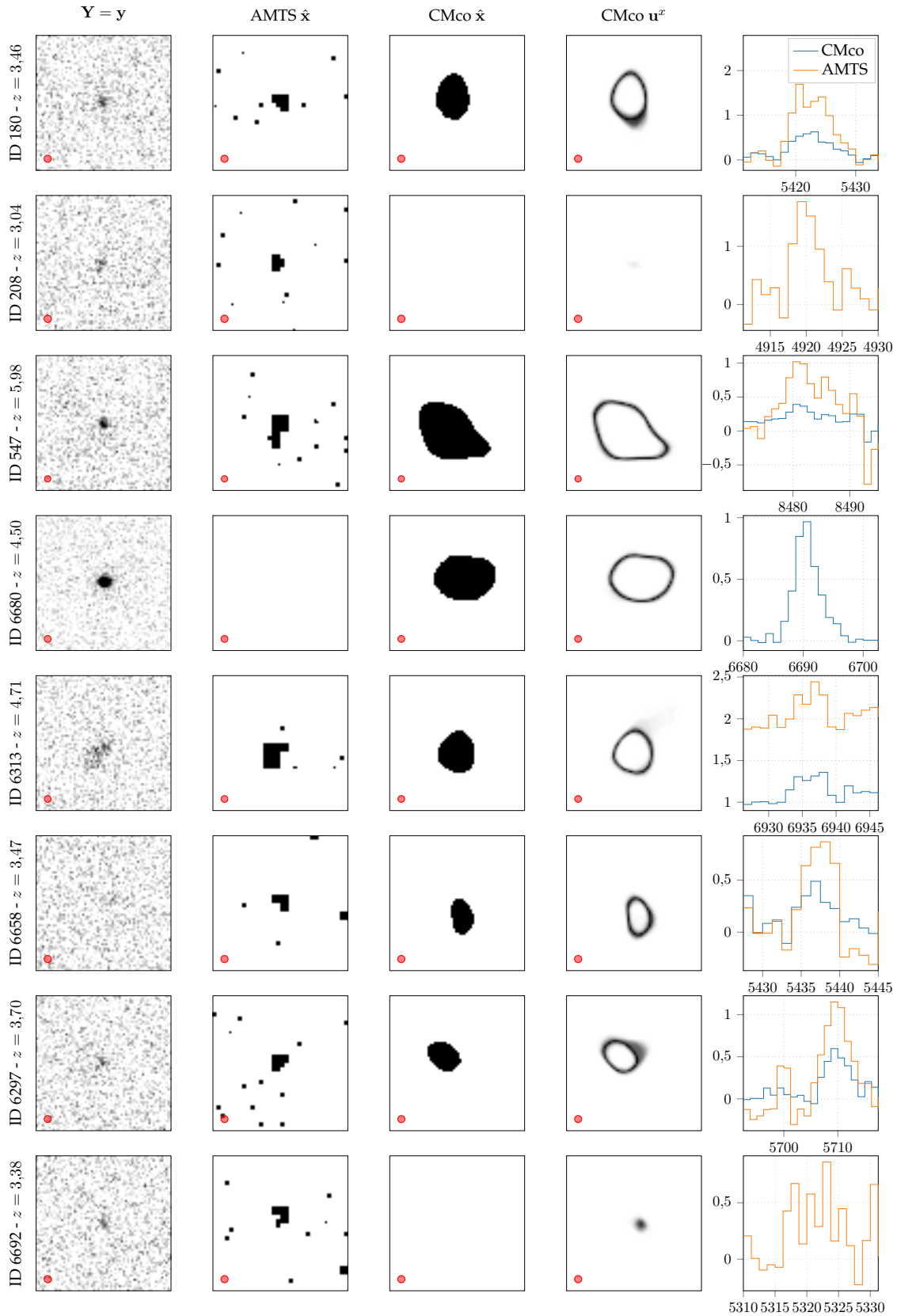



Figure B.3 – Résultats de segmentation sur 8 objets des images *udf10*.

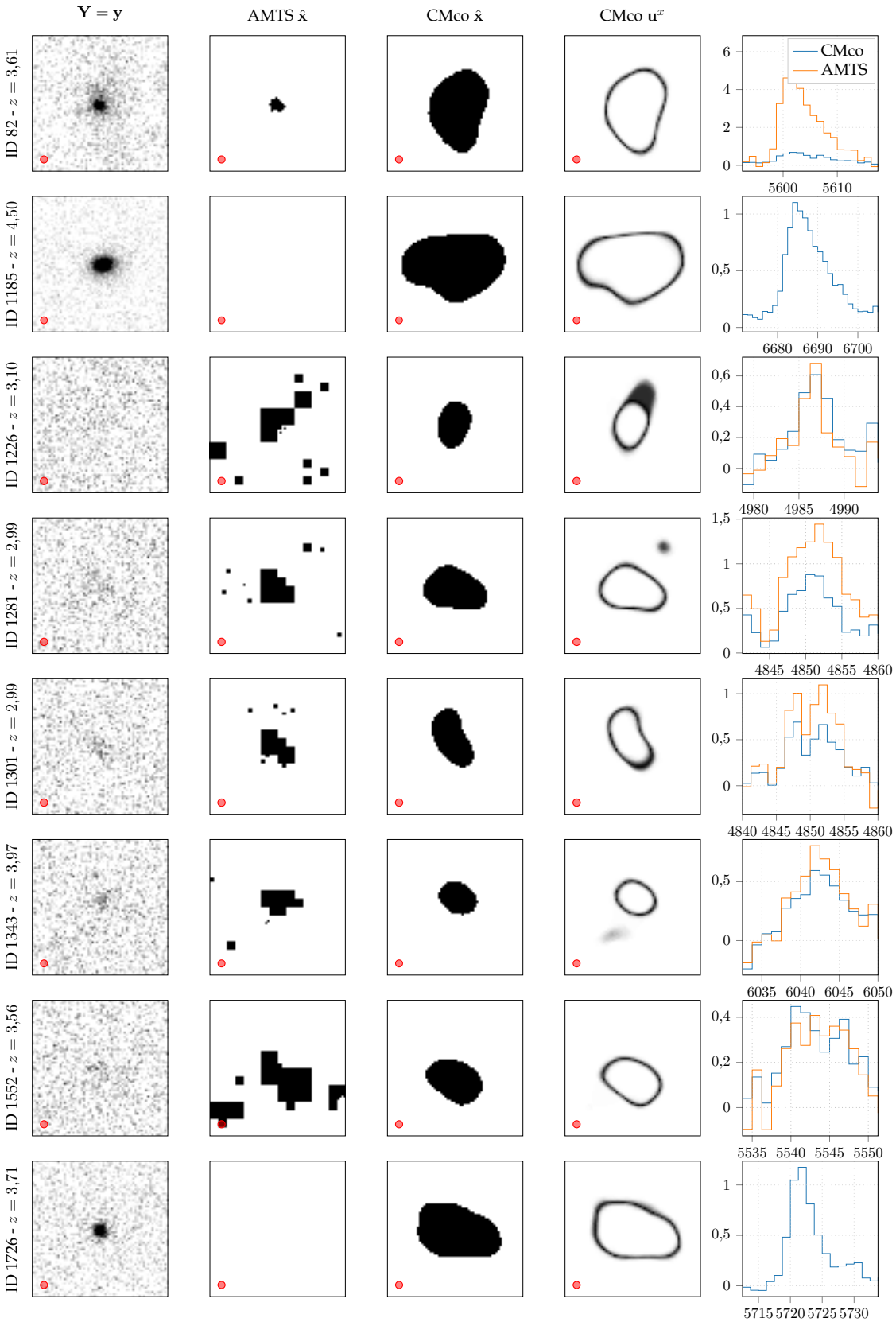



Figure B.4 – Résultats de segmentation sur 24 objets de l'image *mosaic* (1/3).

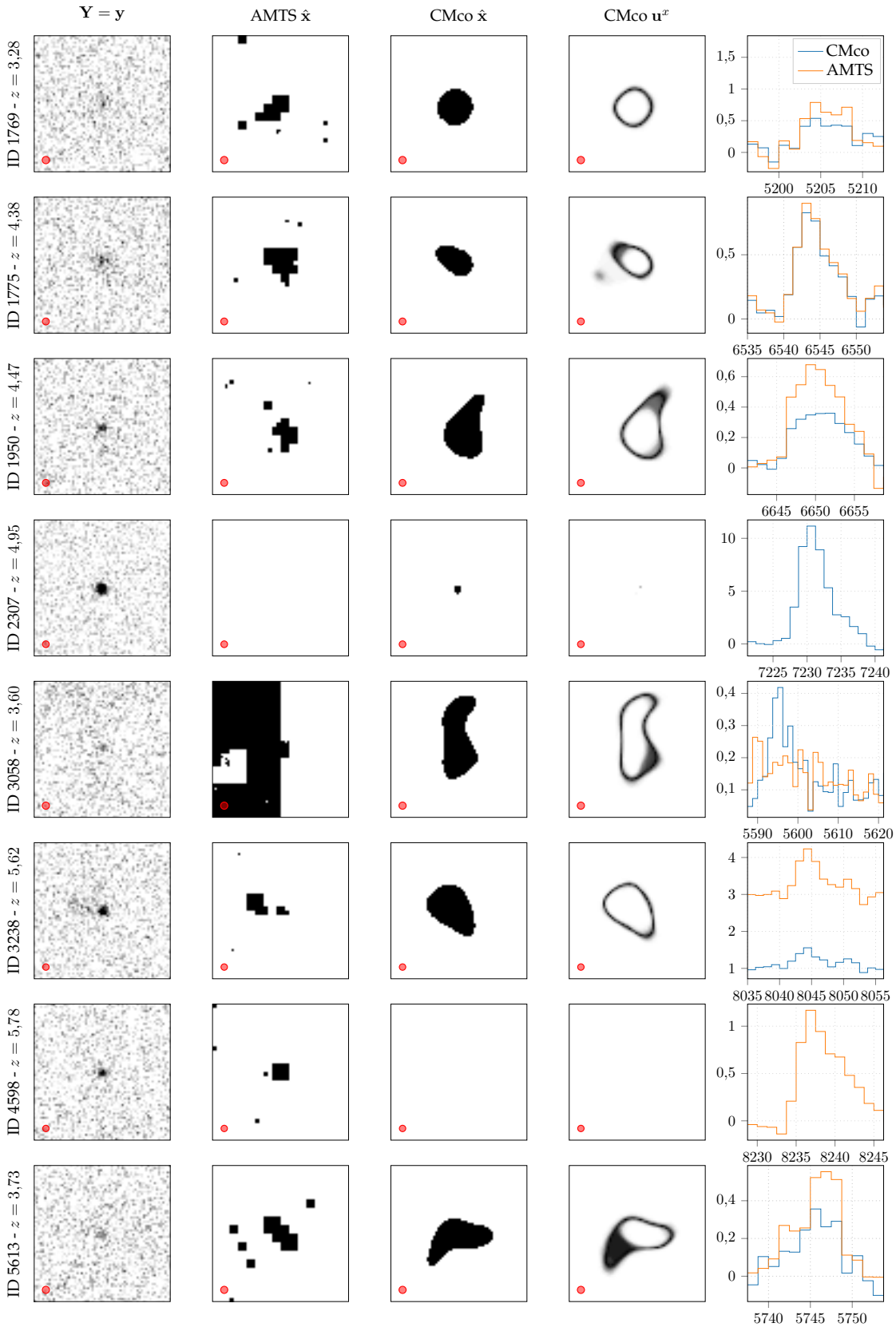



Figure B.5 – Résultats de segmentation sur 24 objets de l'image *mosaic* (2/3).

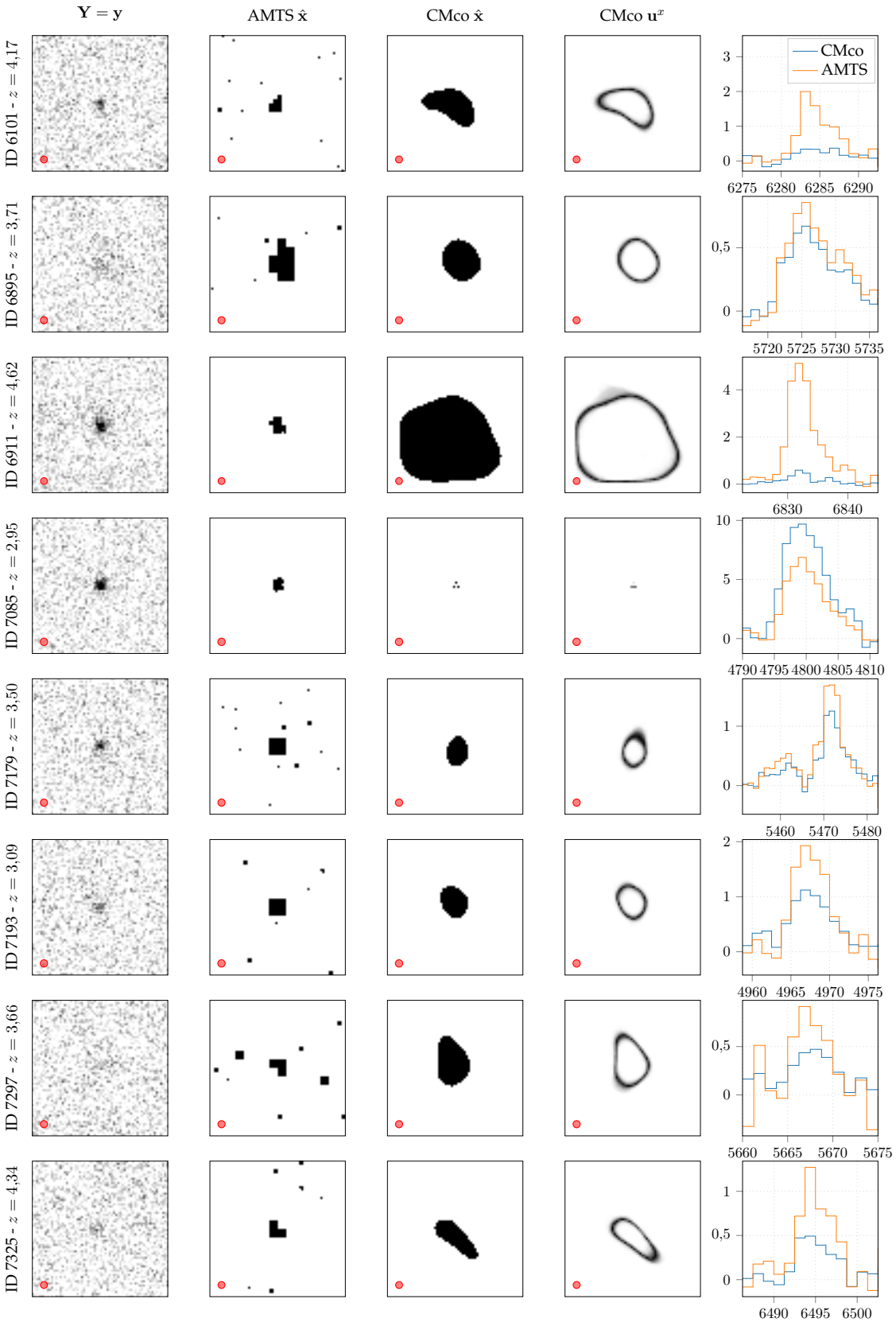



FIGURE B.6 – Résultats de segmentation sur 24 objets de l'image *mosaic* (3/3).

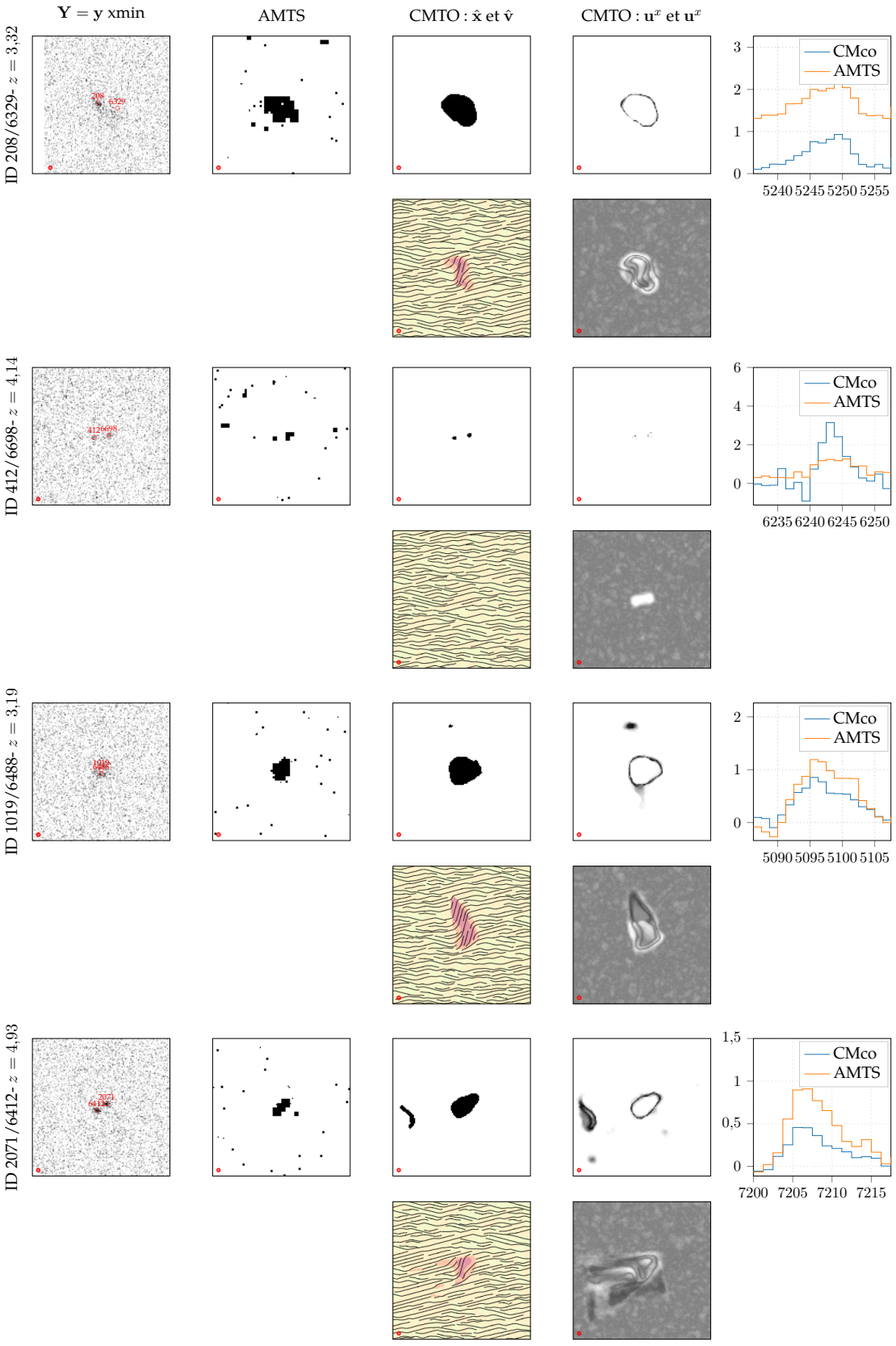



Figure B.7 – Détections obtenues pour 4 paires d'objets.

# Bibliographie

(2017). AVIRIS website. https ://aviris.jpl.nasa.gov/. Cité en page 9.

(2017). MBT database. http ://multibandtexture.recherche.usherbrooke.ca/. Cité en pages 90 et 163.

(2017). MUSE science website. http://muse-vlt.eu/science/. Cité en pages 10 et 12.

(2017). Pléiade website. https://pleiades.cnes.fr/. Cité en pages 90 et 91.

Ahmad, O., Collet, C., et Salzenstein, F. (2015). Spatio-spectral Gaussian random field modeling approach for target detection on hyperspectral data obtained in very low SNR. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2090–2094. IEEE. Cité en pages 30, 48, et 50.

Ashton, E. A. (1998). Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 36(2) :506–517. Cité en page 25.

August, J. et Zucker, S. W. (2003). Sketches with curvature : The curve indicator random field and Markov processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4) :387–400. Cité en page 79.

Bacher, R., Chatelain, F., et Michel, O. (2017a). Global error control procedure for spatially structured targets. In *Signal Processing Conference (EUSIPCO), 2017 25rd European*. IEEE. Cité en pages 67, 68, 74, 76, 77, 78, 149, et 150.

Bacher, R., Meillier, C., Chatelain, F., et Michel, O. (2017b). Robust Control of Varying Weak Hyperspectral Target Detection With Sparse Nonnegative Representation. *IEEE Transactions on Signal Processing*, 65(13) :3538–3550. Cité en pages 74, 149, et 150.

Bacon, R., Accardo, M., Adjali, L., Anwand, H., Bauer, S., Biswas, I., Blaizot, J., Boudon, D., Brau-Nogue, S., Brinchmann, J., et al. (2010). The MUSE second-generation VLT instrument. In *SPIE Astronomical Telescopes+ Instrumentation*. SPIE. Cité en pages 115 et 116.

Bacon, R., Brinchmann, J., Richard, J., Contini, T., Drake, A., Franx, M., Tacchella, S., Vernet, J., Wisotzki, L., Blaizot, J., et al. (2015). The MUSE 3D view of the hubble deep field south. *Astronomy & Astrophysics*, 575 :A75. Cité en pages 43, 49, 51, et 116.

Bacon, R., Conseil, S., Mary, D., Brinchmann, J., Shepherd, M., Akhlaghi, M., Weilbacher, P., Piqueras, L., Wisotzki, L., Lagattuta, D., Epinat, B., Guerou, A., Inami, H., Cantalupo, S., Clastres, C., Courbot, J.-B., Contini, T., Rochard, J., Maseda, M., Bouwens, R., BouchÃ©, N., Kollatschny, W., Schaye, J., Anna Marino, R., Pello, R., Herenz, C., Guiderdoni, B., et Carollo, M. (2017). The MUSE Hubble Ultra Deep Field Survey : I. Survey description,

data reduction and source detection. *Astronomy & Astrophysics*. Cité en pages 4, 10, 38, 116, 117, 118, 119, 141, et 143.

Bacon, R., Piqueras, L., Conseil, S., Richard, J., et Shepherd, M. (2016). MPDAF : MUSE Python Data Analysis Framework. Cité en page 119.

Bali, N. et Mohammad-Djafari, A. (2008). Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis. *IEEE Transactions on Image Processing*, 17(2) :217–225. Cité en page 12.

Basu, S. (1996). Bayesian hypotheses testing using posterior density ratios. *Statistics & probability letters*, 30(1) :79–86. Cité en page 26.

Bauer, C., Pock, T., Sorantin, E., Bischof, H., et Beichel, R. (2010). Segmentation of interwoven 3d tubular tree structures utilizing shape priors and graph cuts. *Medical Image Analysis*, 14(2) :172–184. Cité en page 80.

Beckwith, S. V., Stiavelli, M., Koekemoer, A. M., Caldwell, J. A., Ferguson, H. C., Hook, R., Lucas, R. A., Bergeron, L. E., Corbin, M., Jogee, S., et al. (2006). The Hubble ultra deep field. *The Astronomical Journal*, 132(5) :1729. Cité en pages 116 et 141.

Bellman, R. (1956). Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences*, 42(10) :767–769. Cité en page 8.

Benboudjema, D. et Pieczynski, W. (2005). Unsupervised image segmentation using triplet Markov fields. *Computer Vision and Image Understanding*, 99(3) :476–498. Cité en pages 59 et 65.

Benboudjema, D. et Pieczynski, W. (2007). Unsupervised statistical segmentation of nonstationary images using triplet Markov fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8) :1367–1378. Cité en pages 59 et 65.

Benediktsson, J. A., Palmason, J. A., et Sveinsson, J. R. (2005). Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3) :480–491. Cité en page 12.

Benmansour, F. et Cohen, L. D. (2011). Tubular structure segmentation based on minimal path method and anisotropic enhancement. *International Journal of Computer Vision*, 92(2) :192–210. Cité en page 80.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236. Cité en pages 55 et 80.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, pages 179–195. Cité en page 85.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 259–302. Cité en pages 59, 62, 81, et 83.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer. Cité en page 12.

Borges, J. S., Bioucas-Dias, J. M., et Marcal, A. R. (2011). Bayesian hyperspectral image segmentation with discriminative class learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6) :2151–2164. Cité en page 12.

Brewer, L. N., Ohlhausen, J. A., Kotula, P. G., et Michael, J. R. (2008). Forensic analysis of bioagents by X-ray and TOF-SIMS hyperspectral imaging. *Forensic science international*, 179(2) :98–106. Cité en page 9.

Bricq, S., Collet, C., et Armspach, J.-P. (2006). Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas. In *Biomedical Imaging : Nano to Macro, 2006. 3rd IEEE International Symposium on*, pages 386–389. IEEE. Cité en pages 56, 57, et 65.

Broadwater, J. B. et Chellappa, R. (2010). Adaptive threshold estimation via extreme value theory. *IEEE Transactions on Signal Processing*, 58(2) :490–500. Cité en page 25.

Brodatz, P. (1966). *Textures : a photographic album for artists and designers*. Dover Pubns. Cité en pages 90 et 163.

Broniatowski, M., Celeux, G., et Diebolt, J. (1983). Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste. *Data analysis and informatics*, 3 :359–373. Cité en page 63.

Camps-Valls, G. et Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(6) :1351–1362. Cité en page 12.

Camps-Valls, G., Tuia, D., Bruzzone, L., et Benediktsson, J. A. (2014). Advances in hyperspectral image classification : Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1) :45–54. Cité en page 12.

Carlotto, M. J. (2005). A cluster-based approach for detecting man-made objects and changes in imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(2) :374–387. Cité en page 25.

Casella, G. et Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA. Cité en page 126.

Celeux, G., Chauveau, D., Diebolt, J., Lavergne, C., Bosson, J.-L., et Robert, C. (1995). On stochastic versions of the EM algorithm. *Rapports de recherche- INRIA*. Cité en page 64.

Celeux, G. et Diebolt, J. (1986). L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. *Revue de statistique appliquée*, 34(2) :35–52. Cité en pages 63 et 74.

Celeux, G. et Diebolt, J. (1992). A stochastic approximation type EM algorithm for the mixture problem. *Stochastics : An International Journal of Probability and Stochastic Processes*, 41(1-2) :119–134. Cité en pages 64 et 85.

Chang, C.-I. (2003). *Hyperspectral imaging : techniques for spectral detection and classification*, volume 1. Springer Science & Business Media. Cité en page 24.

Chang, C.-I. (2005). Orthogonal subspace projection (OSP) revisited : A comprehensive study and analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(3) :502–518. Cité en page 23.

Chen, J. Y. et Reed, I. S. (1987). A detection algorithm for optical targets in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, (1) :46–59. Cité en page 25.

Chiang, S.-S., Chang, C.-I., et Ginsberg, I. W. (2001). Unsupervised target detection in hyperspectral images using projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7) :1380–1391. Cité en page 25.

Courbot, J.-B., Mazet, V., Monfrini, E., et Collet, C. (2016a). Detection of faint extended sources in hyperspectral data and application to HDF-S MUSE observations. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1891–1895. IEEE. Cité en pages 2 et 5.

Courbot, J.-B., Mazet, V., Monfrini, E., et Collet, C. (2017a). Extended faint source detection in astronomical hyperspectral images. *Signal Processing*, 135 :274–283. Cité en pages 2, 4, 5, 67, 68, 76, 77, et 150.

Courbot, J.-B., Monfrini, E., Mazet, V., et Collet, C. (2016b). Oriented Triplet Markov fields for Hyperspectral Image Segmentation. In *2016 8th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)*. IEEE. Cité en pages 3 et 5.

Courbot, J.-B., Monfrini, E., Mazet, V., et Collet, C. (2017b). Arbres de markov triplets pour la segmentation d'images. In *Colloque GRETSI*. Cité en pages 3 et 5.

Courbot, J.-B., Monfrini, E., Mazet, V., et Collet, C. (2017c). Oriented triplet Markov fields. *Soumis aux Pattern Recognition Letters*. Cité en pages 3, 4, et 5.

Courbot, J.-B., Rust, E., Monfrini, E., et Collet, C. (2016c). Vertebra segmentation based on two-step refinement. *Journal of Computational Surgery*, 4(1) :1. Cité en page 56.

Dalla Mura, M., Villa, A., Benediktsson, J. A., Chanussot, J., et Bruzzone, L. (2011). Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3) :542–546. Cité en page 12.

Dass, S. C. (2004). Markov random field models for directional field and singularity extraction in fingerprint images. *IEEE Transactions on Image Processing*, 13(10) :1358–1367. Cité en pages 59, 62, et 79.

Dempster, A. P., Laird, N. M., et Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38. Cité en page 63.

Derin, H. et Elliott, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1) :39–55. Cité en pages 59, 62, 72, 74, 85, 102, et 161.

Descombes, X., Mangin, J.-F., Pechersky, E., et Sigelle, M. (1995). Fine structures preserving Markov model for image processing. In *Proceedings of the 9th Scandinavian Conference on Image Analysis*. Cité en pages 59, 62, et 79.

Descombes, X., Morris, R., Zerubia, J., et Berthod, M. (1999). Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood. *IEEE Transactions on Image Processing*, 8(7) :954–963. Cité en page 161.

Dias, J. G. et Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing*, 14(4) :323–332. Cité en page 64.

Dobigeon, N., Tourneret, J.-Y., Richard, C., Bermudez, J. C. M., McLaughlin, S., et Hero, A. O. (2014). Nonlinear unmixing of hyperspectral images : Models and algorithms. *IEEE Signal Processing Magazine*, 31(1) :82–94. Cité en page 11.

Donoho, D. et Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics*, pages 962–994. Cité en page 26.

Du, Q. et Kopriva, I. (2008). Automated target detection and discrimination using constrained kurtosis maximization. *IEEE Geoscience and Remote Sensing Letters*, 5(1) :38–42. Cité en page 25.

Eches, O., Benediktsson, J. A., Dobigeon, N., et Tourneret, J.-Y. (2013). Adaptive Markov random fields for joint unmixing and segmentation of hyperspectral images. *IEEE Transactions on Image Processing*, 22(1) :5–16. Cité en page 67.

Edelman, G., Gaston, E., Van Leeuwen, T., Cullen, P., et Aalders, M. (2012). Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic science international*, 223(1) :28–39. Cité en page 9.

Erdinç, A. et Aksoy, S. (2015). Anomaly detection with sparse unmixing and gaussian mixture modeling of hyperspectral images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, pages 5035–5038. IEEE. Cité en page 24.

Ethier, S. N. et Kurtz, T. G. (2009). *Markov processes : characterization and convergence*, volume 282. John Wiley & Sons. Cité en page 56.

Fjortoft, R., Delignon, Y., Pieczynski, W., Sigelle, M., et Tupin, F. (2003). Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 41(3) :675–686. Cité en pages 59 et 62.

Frontera-Pons, J., Veganzones, M. A., Pascal, F., et Ovarlez, J.-P. (2016). Hyperspectral anomaly detectors using robust estimators. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2) :720–731. Cité en page 24.

Fruchter, A. et Hook, R. N. (1997). Novel image-reconstruction method applied to deep Hubble Space Telescope images. In *Optical Science, Engineering and Instrumentation'97*, pages 120–125. International Society for Optics and Photonics. Cité en page 118.

Geman, S. et Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6) :721–741. Cité en pages 55, 59, 62, 69, 81, 83, 84, et 157.

Giordana, N. et Pieczynski, W. (1997). Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5) :465–475. Cité en pages 56 et 62.

Gonzalez, J., Low, Y., Gretton, A., et Guestrin, C. (2011). Parallel Gibbs sampling : from colored fields to thin junction trees. In *International Conference on Artificial Intelligence and Statistics*, pages 324–332. Cité en pages 84 et 156.

Hayes, B. et al. (2013). First links in the Markov chain. *American Scientist*, 101(2) :252. Cité en page 55.

Hidding, J., Shandarin, S. F., et van de Weygaert, R. (2013). The Zel'dovich approximation : key to understanding cosmic web complexity. *Monthly Notices of the Royal Astronomical Society*, page stt2142. Cité en page 112.

Huck, A. et Guillaume, M. (2010). Asymptotically CFAR-unsupervised target detection and discrimination in hyperspectral images with anomalous-component pursuit. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(11) :3980–3991. Cité en pages 20 et 25.

Inami, H., Bacon, R., Brinchmann, J., Richard, J., et al. (2017). The MUSE Hubble Ultra Deep Field Survey : II. Spectroscopic Redshift and Comparisons to Color Selections of High- z Galaxies. *Astronomy&Astrophysics*, Soumis. Cité en page 143.

Kay, S. M. (1998). *Fundamentals of statistical signal processing : Detection theory, vol. 2*. Prentice-Hall PTR. Cité en pages 18, 19, 23, 24, et 25.

Kindermann, R. et Snell, L. (1980). *Markov random fields and their applications*. Cité en page 55.

Kraut, S., Scharf, L. L., et McWhorter, L. T. (2001). Adaptive subspace detectors. *IEEE Transactions on Signal Processing*, 49(1) :1–16. Cité en pages 22, 23, et 48.

Laferté, J.-M., Pérez, P., et Heitz, F. (2000). Discrete Markov image modeling and inference on the quadtree. *IEEE Transactions on Image Processing*, 9(3) :390–404. Cité en pages 60, 62, 95, 97, et 103.

Lanchantin, P., Lapuyade-Lahorgue, J., et Pieczynski, W. (2008). Unsupervised segmentation of triplet Markov chains hidden with long-memory noise. *Signal Processing*, 88(5) :1134–1151. Cité en page 65.

Leclercq, F., Bacon, R., Wisotzki, L., et al. (2017). The MUSE Hubble Ultra Deep Field Survey : VIII. Extended Lyman $\alpha$ haloes around high-$z$ star-forming galaxies. *Astronomy&Astrophysics*, Soumis. Cité en pages 142 et 143.

Lehmann, E. L. et Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media. Cité en page 126.

Li, J., Bioucas-Dias, J. M., et Plaza, A. (2012). Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(3) :809–823. Cité en page 67.

Li, S. Z. (2009). *Markov random field modeling in image analysis*. Springer Science & Business Media. Cité en page 62.

Li, W., Prasad, S., et Fowler, J. E. (2014). Hyperspectral image classification using Gaussian mixture models and Markov random fields. *Geoscience and Remote Sensing Letters, IEEE*, 11(1) :153–157. Cité en page 67.

Liddle, A. (2015). *An introduction to modern cosmology*. John Wiley & Sons. Cité en pages 111 et 113.

Lin, T. et Bourennane, S. (2013). Survey of hyperspectral image denoising methods based on tensor decompositions. *EURASIP journal on Advances in Signal Processing*, 2013(1) :186. Cité en page 11.

Liu, C. et Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica*, pages 19–39. Cité en pages 129 et 131.

Loncan, L., de Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A., Simoes, M., et al. (2015). Hyperspectral pansharpening : A review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3) :27–46. Cité en page 11.

Lu, G., Halig, L., Wang, D., Qin, X., Chen, Z. G., et Fei, B. (2014). Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging. *Journal of Biomedical Optics*, 19(10) :106004–106004. Cité en page 9.

Lunga, D., Prasad, S., Crawford, M. M., et Ersoy, O. (2014). Manifold-learning-based feature extraction for classification of hyperspectral data : A review of advances in manifold learning. *IEEE Signal Processing Magazine*, 31(1) :55–66. Cité en page 12.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA. Cité en page 158.

Malpica, J. A., Rejas, J. G., et Alonso, M. C. (2008). A projection pursuit algorithm for anomaly detection in hyperspectral imagery. *Pattern Recognition*, 41(11) :3313–3327. Cité en page 25.

Manolakis, D., Lockwood, R., Cooley, T., et Jacobson, J. (2009). Is there a best hyperspectral detection algorithm ? In *SPIE Defense, Security, and Sensing*, pages 733402–733402. International Society for Optics and Photonics. Cité en pages 23 et 24.

Manolakis, D., Siracusa, C., et Shaw, G. (2001). Hyperspectral subpixel target detection using the linear mixing model. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7) :1392–1409. Cité en page 24.

Markov, A. A. (1913). An example of statistical investigation of the text Eugene Onegin concerning the connection of samples in chains. *Science in Context (traduction 2006)*, 19(04) :591–600. Cité en page 55.

Marroquin, J., Mitter, S., et Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical association*, 82(397) :76–89. Cité en pages 56, 59, 62, 63, 69, 81, 83, 103, et 156.

Matteoli, S., Diani, M., et Corsini, G. (2010). A tutorial overview of anomaly detection in hyperspectral images. *Aerospace and Electronic Systems Magazine, IEEE*, 25(7) :5–28. Cité en page 25.

Meillier, C. (2015). *Détection de sources quasi-ponctuelles dans des champs de données massifs*. PhD thesis, Université Grenoble Alpes. Cité en page 20.

Meillier, C., Bacher, R., Chatelain, F., et Michel, O. (2017). Méthode de sigma-clipping par point fixe pour l'estimation de la distribution sous H0 dans le cadre de tests multiples. In *Colloque GRETSI*. Cité en page 149.

Meillier, C., Chatelain, F., Michel, O., et Ayasso, H. (2015). Nonparametric Bayesian extraction of object configurations in massive data. *IEEE Transactions on Signal Processing*, 63(8) :1911–1924. Cité en page 67.

Mercier, G., Derrode, S., et Lennon, M. (2003). Hyperspectral image segmentation with Markov chain model. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, volume 6, pages 3766–3768. IEEE. Cité en pages 56 et 62.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., et Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6) :1087–1092. Cité en page 59.

Mignotte, M., Collet, C., Perez, P., et Bouthemy, P. (2000). Sonar image segmentation using an unsupervised hierarchical MRF model. *IEEE Transactions on Image Processing*, 9(7) :1216–1231. Cité en page 96.

Mittelman, R., Dobigeon, N., et Hero, A. O. (2012). Hyperspectral image unmixing using a multiresolution sticky HDP. *IEEE Transactions on Signal Processing*, 60(4) :1656–1671. Cité en page 11.

Moffat, A. (1969). A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry. *Astronomy & Astrophysics*, 3 :455. Cité en page 43.

Monfrini, E., Ledru, T., Vaie, E., et Pieczynski, W. (1999). Segmentation non supervisée d'images par arbres de Markov cachés. In *17ème Colloque GRETSI*. Cité en pages 60, 62, et 97.

Monfrini, E. et Pieczynski, W. (2005). Estimation de mélanges généralisés dans les arbres de Markov cachés, application à la segmentation des images de cartons d'orgue de barbarie. *Traitement du Signal*, 22(2). Cité en pages 64, 95, et 151.

Mountrakis, G., Im, J., et Ogole, C. (2011). Support vector machines in remote sensing : A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3) :247–259. Cité en page 12.

Neyman, J. et Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231 :pp. 289–337. Cité en page 19.

Parente, M. et Plaza, A. (2010). Survey of geometric and statistical unmixing algorithms for hyperspectral images. In *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pages 1–4. IEEE. Cité en page 11.

Paris, S. (2013). *Méthodes de détection parcimonieuses pour signaux faibles dans du bruit : application à des données hyperspectrales de type astrophysique*. PhD thesis, Nice. Cité en page 26.

Paris, S., Mary, D., et Ferrari, A. (2011). PDR and LRMAP detection tests applied to massive hyperspectral data. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on*, pages 93–96. IEEE. Cité en pages 25, 26, et 35.

Paris, S., Mary, D., et Ferrari, A. (2013a). Detection tests using sparse models, with application to hyperspectral data. *IEEE Transactions on Signal Processing*, 61(6) :1481–1494. Cité en page 67.

Paris, S., Suleiman, R. F. R., Mary, D., et Ferrari, A. (2013b). Constrained likelihood ratios for detecting sparse signals in highly noisy 3D data. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3947–3951. IEEE. Cité en pages 26, 27, 31, 32, 33, et 38.

Pati, Y. C., Rezaiifar, R., et Krishnaprasad, P. S. (1993). Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE. Cité en page 35.

Pereyra, M., Dobigeon, N., Batatia, H., et Tourneret, J.-Y. (2013). Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm. *IEEE Transactions on Image Processing*, 22(6) :2385–2397. Cité en page 161.

Petremand, M., Jalobeanu, A., et Collet, C. (2012). Optimal bayesian fusion of large hyperspectral astronomical observations. *Statistical Methodology*, 9(1) :44–54. Cité en pages 9 et 11.

Pieczynski, W. (1992). Statistical image segmentation. *Machine Graphics and Vision*, 1(1) :261–268. Cité en page 64.

Pieczynski, W. (1994). Champs de markov cachés et estimation conditionnelle itérative. *Traitement du signal*, 11(2) :141–153. Cité en page 64.

Pieczynski, W. (2002). Arbres de markov couple. *Comptes Rendus Mathématique*, 335(1) :79–82. Cité en pages 60 et 65.

Pieczynski, W. (2003a). Arbres de Markov triplet et fusion de Dempster–Shafer. *Comptes Rendus Mathématique*, 336(10) :869–872. Cité en pages 60, 65, 96, et 98.

Pieczynski, W. (2003b). Pairwise Markov chains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5) :634–639. Cité en page 65.

Pieczynski, W., Hulard, C., et Veit, T. (2003). Triplet Markov chains in hidden signal restoration. In *International Symposium on Remote Sensing*, pages 58–68. International Society for Optics and Photonics. Cité en page 65.

Pieczynski, W. et Tebbache, A.-N. (2000). Pairwise Markov random fields and segmentation of textured images. *Machine Graphics and Vision*, 9(3) :705–718. Cité en pages 59, 65, et 68.

Pieper, M., Manolakis, D., Lockwood, R., Cooley, T., Armstrong, P., et Jacobson, J. (2011). Hyperspectral detection and discrimination using the ACE algorithm. In *Proc. SPIE*, volume 8158, pages 815807–815807. Cité en pages 22 et 23.

Pike, R., Lu, G., Wang, D., Chen, Z. G., et Fei, B. (2016). A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging. *IEEE Transactions on Biomedical Engineering*, 63(3) :653–663. Cité en page 9.

Provost, J.-N., Collet, C., Rostaing, P., Pérez, P., et Bouthemy, P. (2004). Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps. *Computer Vision and Image Understanding*, 93(2) :155–174. Cité en pages 60 et 62.

Rasti, B., Sveinsson, J. R., et Ulfarsson, M. O. (2014). Wavelet-based sparse reduced-rank regression for hyperspectral image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10) :6688–6698. Cité en page 11.

Reed, I., Mallett, J., et Brennan, L. (1974). Rapid convergence rate in adaptive arrays. *Aerospace and Electronic Systems, IEEE Transactions on*, (6) :853–863. Cité en pages 22, 23, et 48.

Reed, I. et Xiaoli, Y. (1990). Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(10) :1760–1770. Cité en pages 24, 25, 30, et 43.

Rellier, G., Descombes, X., Falzon, F., et Zerubia, J. (2004). Texture feature analysis using a Gauss-Markov model in hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(7) :1543–1551. Cité en page 67.

Richmond, C. D. (2000). Performance of a class of adaptive detection algorithms in nonhomogeneous environments. *IEEE Transactions on Signal Processing*, 48(5) :1248–1262. Cité en page 24.

Rigamonti, R. et Lepetit, V. (2012). Accurate and efficient linear structure segmentation by leveraging ad hoc features with learned filters. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 189–197. Springer. Cité en page 79.

Robert, C. et Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media. Cité en page 84.

Rodarmel, C. et Shan, J. (2002). Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science*, 62(2) :115. Cité en page 12.

Sagan, H. (1994). *Space-filling curves*, volume 18. Springer. Cité en page 56.

Salzenstein, F. et Collet, C. (2006). Fuzzy Markov random fields versus chains for multispectral image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11) :1753–1767. Cité en pages 67 et 151.

Salzenstein, F. et Pieczynski, W. (1997). Parameter estimation in hidden fuzzy Markov random fields and image segmentation. *Graphical models and image processing*, 59(4) :205–220. Cité en pages 59 et 62.

Schweizer, S. M. et Moura, J. M. (2000). Hyperspectral imagery : Clutter adaptation in anomaly detection. *IEEE Transactions on Information Theory*, 46(5) :1855–1871. Cité en pages 23 et 67.

Seneta, E. (1996). Markov and the birth of chain dependence theory. *International Statistical Review*, pages 255–263. Cité en page 55.

Serre, D., Villeneuve, E., Carfantan, H., Jolissaint, L., Mazet, V., Bourguignon, S., et Jarno, A. (2010). Modeling the spatial PSF at the VLT focal plane for MUSE WFM data analysis purpose. In *SPIE Astronomical Telescopes+ Instrumentation*. SPIE. Cité en pages 43 et 116.

Shapiro, S. S. et Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4) :591–611. Cité en page 130.

Smits, P. C. et Dellepiane, S. G. (1997). Synthetic aperture radar image segmentation by a detail preserving Markov random field approach. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4) :844–857. Cité en page 79.

Soto, K. T., Lilly, S. J., Bacon, R., Richard, J., et Conseil, S. (2016). ZAP–enhanced PCA sky subtraction for integral field spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 458(3) :3210–3220. Cité en page 118.

Soulez, F., Bongard, S., Thiébaut, É., et Bacon, R. (2011). Restoration of hyperspectral astronomical data from integral field spectrograph. In *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pages 1–4. IEEE. Cité en page 9.

Soulez, F., Thiébaut, É., et Denis, L. (2013). Restoration of hyperspectral astronomical data with spectrally varying blur. *EAS Publications Series*, 59 :403–416. Cité en page 11.

Spitzer, F. (1971). Markov random fields and Gibbs ensembles. *The American Mathematical Monthly*, 78(2) :142–154. Cité en page 55.

Stein, D. W., Beaven, S. G., Hoff, L. E., Winter, E. M., Schaum, A. P., et Stocker, A. D. (2002). Anomaly detection from hyperspectral imagery. *IEEE Signal Processing Magazine*, 19(1) :58–69. Cité en page 25.

Varshney, P. K. et Arora, M. K. (2004). *Advanced image processing techniques for remotely sensed hyperspectral data*. Springer Science & Business Media. Cité en page 9.

Villeneuve, E. et Carfantan, H. (2014). Nonlinear deconvolution of hyperspectral data with MCMC for studying the kinematics of galaxies. *IEEE Transactions on Image Processing*, 23(10) :4322–4335. Cité en page 11.

Villeneuve, E., Carfantan, H., et Serre, D. (2011). PSF estimation of hyperspectral data acquisition system for ground-based astrophysical observations. In *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pages 1–4. IEEE. Cité en page 116.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., et Hernquist, L. (2014). Introducing the Illustris Project : simulating the coevolution of dark and visible matter in the Universe. *Monthly Notices of the Royal Astronomical Society*, 444(2) :1518–1547. Cité en pages 13, 112, et 114.

Vollmer, B., Perret, B., Petremand, M., Lavigne, F., Collet, C., Van Driel, W., Bonnarel, F., Louys, M., Sabatini, S., et MacArthur, L. (2013). Simultaneous Multi-band Detection of Low Surface Brightness Galaxies with Markovian Modeling. *The Astronomical Journal*, 145(2) :36. Cité en page 67.

Wei, G. C. et Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411) :699–704. Cité en page 64.

Weilbacher, P., Streicher, O., Urrutia, T., Jarno, A., Pécontal-Rousset, A., Bacon, R., et Böhm, P. (2012). Design and capabilities of the MUSE data reduction software and pipeline. In *SPIE Astronomical Telescopes+ Instrumentation*. SPIE. Cité en page 117.

Wisotzki, L., Bacon, R., Blaizot, J., Brinchmann, J., Herenz, E. C., Schaye, J., Bouché, N., Cantalupo, S., Contini, T., Carollo, C. M., Caruana, J., Courbot, J.-B., Emsellem, E., Kamann, S., Kerutt, J., Leclercq, F., Lilly, S. J., Patrício, V., Sandin, C., Steinmetz, M., Straka, L. A., Urrutia, T., Verhamme, A., Weilbacher, P. M., et Wendt, M. (2016). Extended Lyman $\alpha$ haloes around individual high-redshift galaxies revealed by MUSE. *Astronomy & Astrophysics*, 587 :A98. Cité en pages 4, 49, et 142.

Wright, E. L. (2006). A cosmology calculator for the World Wide Web. *Publications of the Astronomical Society of the Pacific*, 118(850) :1711. Cité en page 113.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, pages 95–103. Cité en page 63.

Wu, D. et Sun, D.-W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment : A review—Part I : Fundamentals. *Innovative Food Science & Emerging Technologies*, 19 :1–14. Cité en page 9.

Wu, Y., Li, M., Zhang, P., Zong, H., Xiao, P., et Liu, C. (2011). Unsupervised multi-class segmentation of SAR images using triplet Markov fields models based on edge penalty. *Pattern Recognition Letters*, 32(11) :1532–1540. Cité en pages 59 et 65.

Xia, J., Chanussot, J., Du, P., et He, X. (2015). Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 53(5) :2532–2546. Cité en page 67.

Yahiaoui, M., Monfrini, E., et Dorizzi, B. (2016). Markov Chains for unsupervised segmentation of degraded NIR iris images for person recognition. *Pattern Recognition Letters*, 82 :116–123. Cité en page 56.

Yuan, Q., Zhang, L., et Shen, H. (2012). Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10) :3660–3677. Cité en page 11.

Zhang, P., Li, M., Wu, Y., Gan, L., Liu, M., Wang, F., et Liu, G. (2012). Unsupervised multi-class segmentation of SAR images using fuzzy triplet Markov fields model. *Pattern Recognition*, 45(11) :4018–4033. Cité en pages 59 et 65.

Zhang, Q., Liu, Y., Blum, R. S., Han, J., et Tao, D. (2017). Sparse Representation based Multi-sensor Image Fusion : A Review. *arXiv preprint arXiv :1702.03515*. Cité en page 11.

Zhang, R., Sheng, W., et Ma, X. (2013). Improved switching CFAR detector for non-homogeneous environments. *Signal Processing*, 93(1) :35–48. Cité en page 25.

Zhang, Y., Brady, M., et Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1) :45–57. Cité en pages 59 et 62.

Jean-Baptiste COURBOT

# Traitement statistique d'images hyperspectrales pour la détection d'objets diffus : application aux données astronomiques du spectro-imageur MUSE

**Résumé** – Nous étudions le problème de la détection et de la segmentation dans des images extrêmement bruitées. L'application est la détection, dans les données hyperspectrales astronomiques de l'instrument MUSE, de halos (localisés et homogènes dans les images) et de filaments (structures anisotropes à grande échelle). Dans un premier temps, nous étudions le problème de détection par tests d'hypothèses dans des images hyperspectrales en nous appuyant sur des contraintes de formes spatiales, spectrales et de similarité entre spectres. Nous introduisons ensuite un modèle de champ de Markov couple convolutif, qui permet de poser le problème de détection comme le cas particulier d'un problème de segmentation, tout en apportant un *a priori* markovien sur la classification recherchée. Ensuite, afin de modéliser les structures orientées dans les images, nous introduisons un modèle de champ de Markov triplet permettant la segmentation simultanée des orientations et des classes. Dans le but de modéliser des structures à grande échelle dans les images, nous introduisons également un modèle d'arbre de Markov triplet permettant la prise en compte simultanée de composantes hiérarchiques inter-résolution et d'homogénéité au sein d'une résolution. Chaque modèle a été validé et comparé à l'état de l'art, puis tous ont été comparés sur des données synthétiques dans le contexte de la détection dans des images hyperspectrales astronomiques. Le manuscrit présente enfin l'analyse des résultats obtenus sur des données réelles issues de l'instrument MUSE.

**Mots-clés** : segmentation, détection, modèles markoviens, images hyperspectrales astronomiques

**Abstract** – We study the detection and segmentation problems in extremely noised images. The main application of these works is the detection of large-scale structures in MUSE astronomical hyperspectral images, namely haloes (localized and homogenous in images) and filaments (anisotropic large-scale structures). First, we study the hypothesis-testing detection in hyperspectral images, based on spatial and spectral shape constraints as well as similarity constraints. Then, we introduce a pairwise Markov field model which allows the formulation of the detection problem as a special case of the segmentation problem, while introducing a Markovian prior on the result. Next, in order to model oriented structures in images, we propose a triplet Markov field model allowing the joint segmentation of orientations and classes in images. Finally, we study the modelling of large-scale structures in images by introducing a triplet Markov tree model handling inter-resolution dependancy jointly with homogeneity within resolutions. The two latter models were introduced in the general framework of image segmentation. Each model was validated with respect to its alternatives, then all models were compared on synthetic data in the context of detection within astronomical hyperspectral images. Finally, this document presents the analysis of the results on real MUSE images.

**Keywords** : segmentation, detection, Markovian modeling, astronomical hyperspectral images