



HAL
open science

Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de Markov

Farid Flitti

► **To cite this version:**

Farid Flitti. Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de Markov. Traitement du signal et de l'image [eess.SP]. Université Louis Pasteur - Strasbourg I, 2005. Français. NNT: . tel-00156963

HAL Id: tel-00156963

<https://theses.hal.science/tel-00156963v1>

Submitted on 25 Jun 2007

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

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



N° d'ordre :

THÈSE

présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur - Strasbourg I

École doctorale : Mathématiques, sciences de l'information et de l'ingénieur
Discipline : Électronique, électrotechnique, automatique
Spécialité : Traitement d'images et vision par ordinateur

Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de Markov

Soutenue publiquement
le 08 décembre 2005
par

Farid FLITTI

Membres du jury :

Président :	M. Albert BIJAOUI	Astronome	OCA, Nice
Rapporteur interne :	M. Olivier BIENAYMÉ	Astronome	ULP, Strasbourg
Rapporteur externe :	M. Henri MAITRE	Professeur	ENST, Paris
Rapporteur externe :	M. Patrick PÉREZ	Direct. de Rech.	IRISA Rennes
Directeur de thèse :	M. Christophe COLLET	Professeur	ULP, Strasbourg
Examineur :	M. Wojciech PIECZYNSKI	Professeur	INT, Evry

Travail effectué au sein du Laboratoire des Sciences
de l'Image, de l'Informatique et de la Télédétection, UMR - 7005 CNRS - ULP.

Remerciements

Je remercie Messieurs Olivier Bienaymé, Henri Maître et Patrick Pérez pour l'intérêt qu'ils ont porté à mes travaux en acceptant la tâche fastidieuse de rapporteurs. Je remercie également Messieurs Albert Bijaoui et Wojciech Pieczynski pour avoir accepté de participer au jury.

Je remercie mon directeur de thèse Monsieur Christophe Collet pour la confiance et le soutien qu'il m'a accordés ainsi que pour la justesse des conseils qu'il m'a prodigués durant ces trois années de thèse. Je lui suis reconnaissant de m'avoir fait partager son enthousiasme et son goût pour la recherche.

Ce travail a été réalisé au LSIIT (Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection), dans l'équipe MIV (Modèles, Images et Vision) en collaboration avec le CDS (Centre de Données astronomiques de Strasbourg) dans le cadre des ACI IDHA (Images Distribuées Hétérogènes pour l'Astronomie) et MDA (Masses de Données en Astronomie). J'exprime ma gratitude à Madame Françoise Genova, directrice du CDS, et à Monsieur Fabrice Heitz, directeur du LSIIT, pour leur accueil dans leur laboratoire respectif.

Je tiens à remercier spécialement François Bonnarel et Mireille Louys pour leur gentillesse, leur aide, leurs conseils, les multiples discussions et surtout pour m'avoir accueilli dans leur bureau pendant une grande partie de ma thèse.

Je remercie également Bernd Vollmer, Eric Slezak et Annabelle Joannic avec qui j'ai collaboré durant différentes périodes de ma thèse.

Merci à tous les membres des équipes MIV et CDS, pour leur gentillesse et leur soutien. Plus particulièrement, je remercie Monsieur Ernest Hirsch pour l'aide qu'il m'a prodigué pour mes activités pédagogiques.

Je remercie mes collègues doctorants : Aïcha, Matthieu, Redouane, Sylvain, Torbjörn et Vincent, pour l'ambiance agréable qu'ils ont créée. Je salue également les nouveaux arrivants de l'équipe : André, les 2 Mathieu, Michael et Stéphanie.

Ma pensée la plus émue est adressée à ma famille, particulièrement à mes parents qui m'ont toujours soutenu. Je leur témoigne ici ma profonde gratitude. Nul remerciement ne pourra exprimer ma reconnaissance envers celle que j'ai choisie pour femme et qui m'a supporté durant ces longues années de travail.

Enfin, je remercie tous ceux qui ont contribué à ma formation.

Merci à Dieu qui m'a donné la force d'aller jusqu'au bout de cette thèse

Table des matières

Introduction générale	1
1 Modèles Markoviens	3
1.1 Introduction	4
1.2 Éléments de l'inférence bayésienne	4
1.3 Modèles graphiques	6
1.3.1 Graphe de dépendance non orienté	7
1.3.2 Graphe de dépendance orienté	9
1.3.3 Manipulation des graphes non orientés	10
1.3.4 Tâches génériques sur les modèles graphiques	12
1.3.5 Méthodes exactes d'inférence	12
1.3.6 Méthodes approximatives d'inférence	15
1.4 Modèles Markoviens pour les images	17
1.4.1 Champs de Markov	17
1.4.2 Chaîne de Markov	18
1.4.3 Approches à hiérarchie induite	19
1.4.4 Approches à hiérarchie explicite	21
1.5 Inférence sur le quadarbre de Markov caché	25
1.6 Inférence sur le quadarbre de Markov couple	28
1.7 Conclusion	31
2 Réduction de dimension et lois non gaussiennes multivariées	33
2.1 Introduction	34
2.2 Nécessité de réduire le nombre de bandes	34
2.3 Quelques techniques de réduction de dimension	35
2.3.1 L'analyse en composantes principales	35
2.3.2 L'analyse en composantes principales probabiliste	36
2.3.3 L'analyse en composantes indépendantes	37
2.3.4 Le mélange d'analyseurs en composantes principales probabilistes	40
2.4 Schémas de réduction adoptés	41
2.4.1 Regroupement de bandes et projections locales	41
2.4.2 Réduction des cubes radio par mélange de gaussiennes	44
2.4.3 Régularisation du mélange d'analyseurs en composantes principales probabilistes	46
2.5 Modélisation de l'attache aux données	48
2.5.1 Quelques densités de probabilités monovariées	49

2.5.2	Loi jointe de variables aléatoires indépendantes	51
2.5.3	Loi jointe de variables aléatoires gaussiennes corrélées	51
2.5.4	Loi jointe de variables aléatoires corrélées : technique de décorrélation	52
2.5.5	Loi jointe de variables aléatoires corrélées : théorie des copules	53
2.6	Conclusion	55
3	Estimation de paramètres	57
3.1	Introduction	57
3.2	Algorithmes d'estimations	57
3.2.1	Algorithme EM	57
3.2.2	Variante de l'EM	60
3.2.3	Algorithme ECI	61
3.2.4	L'algorithme K-moyennes segmental	62
3.2.5	L'algorithme de segmentation courante	62
3.3	Le quadarbre de Markov caché	63
3.3.1	Le cas d'une attache aux données gaussienne	63
3.3.2	Le cas d'une attache aux données gaussienne généralisée	65
3.3.3	Résultats sur des images synthétiques	67
3.4	Le quadarbre de markov couple	69
3.4.1	Estimation des paramètres de l' <i>a priori</i>	72
3.4.2	Estimation des paramètres de l'attache aux données : cas gaussien	72
3.4.3	Estimation des paramètres de l'attache aux données : cas gaussien généralisé	73
3.4.4	Résultats sur les images synthétiques	75
3.5	Régularisation de l'algorithme MACPP par le quadarbre de markov caché	75
3.5.1	Estimation des paramètres	75
3.5.2	Résultats sur des images synthétiques	78
3.6	Conclusion	82
4	Application à l'imagerie astronomique	85
4.1	Introduction	85
4.2	Images astronomiques	85
4.3	Réduction/segmentation	87
4.3.1	Détection de galaxie lointaines à faible rayonnement	87
4.3.2	Segmentation du cube radio de la raie HI de la galaxie spirale NGC4254	93
4.3.3	Réduction/Segmentation jointe par la MACPP régularisée	95
4.4	Arbres de Markov et analyse multirésolution	102
4.4.1	Analyse en pyramide avec une seule ondelette	102
4.4.2	Modélisation par arbres de Markov	103
4.5	Débruitage des images avec arbres de Markov	104
4.5.1	Débruitage par sélection des coefficients pertinents	104
4.5.2	Débruitage par rétrécissement des coefficients de détails	105
4.6	Fusion d'images sur l'arbre de Markov caché	108
4.7	Conclusion	114
	Conclusion générale	115

Bibliographie 121

Introduction générale

Depuis toujours l'homme a été fasciné par l'observation du ciel. Son observation a évolué au cours du temps du simple émerveillement poétique à la poursuite des trajectoires des astres et constellations et jusqu'aux études astrophysiques modernes de l'Univers et les théories de son évolution. Ce développement spectaculaire est dû en grande partie à l'évolution des capteurs astronomiques orientés vers le ciel, à savoir les télescopes optiques, les radio télescopes et le télescope spatial. Ces instruments permettent d'avoir une vision multi-longueurs d'ondes de l'univers, offrant une caractérisation spectrale de chaque zone observée. L'analyse humaine des images multispectrales est fastidieuse et requière l'examen individuel des images aux différentes longueurs d'ondes. Il est de ce point de vue souhaitable de disposer de chaînes de traitements automatiques fournissant un résumé efficace des images étudiées et permettant aux astronomes de se concentrer sur des régions d'intérêt qui apparaissent plus clairement. Cette thèse s'inscrit dans cette optique en ayant comme but principal la segmentation d'images multispectrales astronomiques et accessoirement la restauration et la fusion de telles images.

La première tâche revient donc à classer les vecteur-pixels en tenant compte de leur voisinage spatial. De façon générale, et quelle que soit la méthode utilisée, la classification de données multidimensionnelles reste à ce jour un problème difficile. Les algorithmes d'apprentissage requièrent généralement un nombre d'échantillons suffisant pour couvrir tous les aspects du comportement du processus étudié. En effet, le nombre d'échantillons requis augmente avec la dimension de telle sorte que l'apprentissage devient rapidement inefficace en pratique. C'est le problème connu sous le nom de *malédiction de la dimension* ou *phénomène de Hughes*. Pour une image de taille $H \times W$ pixels par d bandes spectrales, une longueur d'onde supplémentaire, ne changeant pas le nombre de pixels (échantillons), augmente la complexité de l'estimation puisque celle ci est alors effectuée dans \mathbb{R}^{d+1} . À nombre d'échantillons fixe, l'augmentation de la dimension rend l'espace de plus en plus vide. Ce problème, bien connu en télédétection, peut être abordé en effectuant une étape préliminaire de réduction de l'espace de représentation. En effet, plusieurs techniques ont été proposées en analyse d'images. Elles cherchent toutes une projection linéaire ou non sur un espace de plus faible dimension en maximisant un critère donné. Comme techniques linéaires on peut citer l'Analyse en Composantes Principales (ACP), l'Analyse Factorielle (AF), l'Analyse en Composantes Indépendantes (ACI) et l'Analyse Discriminante de Fisher (ADF). Pour les techniques non linéaires, on peut citer les deux méthodes de poursuite de projections proposées par Jimenez et *al.* et Rellier et *al.*, l'ACP non linéaire et le mélange de méthodes linéaires locales. Nous pouvons aussi, dans le cas où le nombre de bandes est suffisant, considérer chaque vecteur-pixel comme une courbe 1D qu'on approxime avec un modèle à nombre de paramètres réduit. Les paramètres du

modèle en chaque vecteur-pixel codent efficacement la courbe associée donnant ainsi un ensemble d'images paramètres qu'on segmentera à la place de l'image originale.

Dans ce cadre nous avons proposé trois schémas de traitement pour les images astronomiques, à savoir le regroupement de bandes suivis de projections locales [49, 50], la réduction des cubes radio par mélange de gaussiennes [50, 55, 54] et la régularisation du mélange d'analyseurs en composantes principales probabilistes [35]. Cette dernière technique réalise les opérations de réduction et classification simultanément. Nous avons repris le modèle de mélange d'analyseurs en composantes principales probabilistes introduit par Tipping *et al.* dans lequel nous avons changé l'*a priori* qui ne prenait pas en compte les voisinages des pixels, par un *a priori* markovien. Nous pouvons ainsi utiliser les champs de Markov, les chaînes de Markov avec parcours fractal sur l'image ou l'arbre de Markov caché. Cette méthode peut être étendue à d'autres modèles de mélanges de projections locales.

Pour la tâche de segmentation proprement dite, nous avons opté pour des modèles Markoviens sur l'arbre, permettant une prise en compte hiérarchique du voisinage spatial entre pixels sans la lourdeur que peut entraîner l'utilisation des champs de Markov. Dans ce cadre, nous avons étudié le quadarbre de Markov caché, introduit par Laferté *et al.* et Crouse *et al.* et repris par Provost *et al.* dans le cadre mutispectral, en modélisant différemment le terme d'attache aux données multidimensionnel dans le cas non gaussien. Nous avons ainsi pu calculer des lois gaussiennes généralisées multidimensionnelles en utilisant la théorie de copules [51].

D'un autre côté, nous avons étudié l'arbre de Markov couple introduit par Pieczynski et adapté par Monfrini *et al.* dans le cas d'un bruit gaussien pour la restauration d'images. Nous avons étendu ce modèle au cas non gaussien en nous servant de la théorie des copules. Les tests sur des images de synthèse se sont révélés très prometteurs (papier prêt à soumission).

Les modèles de l'arbre de Markov caché et couple par leurs conceptions sont bien adaptés aux données multirésolutions. Ceci nous a poussé à les appliquer dans le domaine des ondelettes. L'application directe des arbres de Markov cachés sur des transformées en ondelettes adaptées aux images astronomiques nous a permis de faire une sélection automatique des coefficients pertinents permettant ainsi la restauration de telles images. Plus loin, nous avons développé un schéma de fusion offrant ainsi une synthèse du contenu des différentes bandes [52, 53].

Le manuscrit est organisé en quatre chapitres. Le premier débute par le rappel de la théorie bayésienne et enchaîne sur la présentation des modèles graphiques et les modèles markoviens en imagerie. Nous présentons à la fin de ce chapitre les modèles du quadarbre de Markov caché et couple et le calcul du critère MPM sur chacun des deux. Le deuxième chapitre est consacré à la résolution du problème de dimensionnalité et au calcul de l'attache aux données. Nous y présentons quelques techniques classiques de réduction de dimensionnalité ainsi que les trois schémas proposés. De plus, nous y exposons les différentes techniques de calcul des lois multidimensionnelles dans les cas gaussien et non

gaussien. Le troisième chapitre traite l'estimation des paramètres. Nous y présentons les versions complètes non supervisées des différents algorithmes avec quelques résultats sur des images synthétiques. Le dernier chapitre est réservé aux applications des modèles étudiés pour la classification, la restauration et la fusion d'images astronomiques.

1 Modèles Markoviens

Notations

Notation	Signification
X	Champ des variables cachées, étiquettes dans le cas discret
Y	Champ des observations
x	réalisation du champ des variables cachées
y	réalisation du champ des observations
S	Ensemble de sites sur une grille, ou ensemble de nœuds dans un graphe
T	Ensemble de sites sur une grille
L	Ensemble d'arêtes dans un graphe
G	Graphe
X_a	l'ensemble des variables aléatoires $\{X_t; t \in a\}$ avec $a \subset S$
x_s	Réalisation de la variable aléatoire cachée au site $s \in S$
y_s	Observation au site $s \in S$
Ω_x	Ensemble des valeurs possibles de x_s
Ω_y	Ensemble des valeurs possibles de y_s
Λ_x	Ensemble des configurations possibles de X
Λ_y	Ensemble des configurations possibles de Y
$\mathcal{L}(\cdot, \cdot)$	Fonction coût
$p(x)$	Probabilité de x
$\mathbf{E}(x)$	Espérance mathématique de x
ν_s	Ensemble des voisins du site s
\mathcal{N}	Système de voisinage
c	Clique relative à un système de voisinage
\mathcal{C}	Ensemble de cliques
$\Psi_c(x)$	fonction locale associée à la clique c
$V_c(x)$	potentiel local associé à la clique c : $V_c(x) = -\ln \Psi_c(x)$
$m_{ij}(\cdot)$	message passé du site i au site j
S_{ij}	Séparateur entre deux états i et j dans un arbre de jonction
$U(x)$	Énergie associée aux champ x
s^-	Parent du site s
s^+	Enfant du site s
$\geq s$	Descendants du site s , s inclus
$\leq s$	Ancêtres du site s , s inclus
$> s$	Descendants du site s , s non inclus
$< s$	Ancêtres du site s , s non inclus
$ S $	Cardinal de l'ensemble S

1.1 Introduction

Les modèles de Markov cachés, dont l'origine remonte aux travaux du mathématicien russe Markov, ne cessent de susciter l'intérêt de la communauté du traitement de l'information depuis leurs premiers succès en reconnaissance automatique de la parole au début des années 70. De nos jours, leurs domaines d'application couvrent le traitement des images, la biologie (analyse de séquences de gènes), la météorologie, l'analyse financière... etc. Récemment, la théorie des modèles graphiques s'est présentée comme une généralisation des modèles de Markov. Ces approches permettent de modéliser notre connaissance sur les processus étudiés, étape nécessaire dans le cadre de la théorie de décision basée sur l'inférence bayésienne. Ce chapitre est dédié à l'exposé de ces modèles. Nous commencerons par rappeler les éléments de l'inférence bayésienne, et nous enchaînerons par la présentation d'une synthèse sur les modèles graphiques qui nous permettra d'introduire les modèles markoviens en traitement d'image et plus précisément les arbres de Markov, caché et couple, sur lesquels se base ce travail.

1.2 Éléments de l'inférence bayésienne

De nombreuses applications de traitement d'images se basent sur la recherche d'informations cachées X , discrètes ou continues, à partir d'observations Y . Cette recherche s'appuie généralement sur l'établissement d'un modèle *direct* intégrant toute la connaissance disponible sur la formation des Y à partir des X . Retrouver les informations cachées à partir des observations consiste alors à *inverser* ce modèle direct. Ceci n'est pas une tâche simple car les problèmes inverses en traitement d'images sont généralement des problèmes mal posés. Un problème est dit bien posé s'il existe une solution unique dépendante continûment des observations. La dernière propriété signifie que lorsque l'erreur sur Y tend vers zéro, l'erreur induite sur la solution X tend aussi vers zéro ce qui est lié à la stabilité de la solution et la robustesse de la méthode. Pour les problèmes inverses en imagerie, le bruit d'observation dégrade significativement la solution qui peut alors être très éloignée de celle espérée. Une solution est d'ajouter la connaissance disponible *a priori* sur la solution pour réduire l'espace de recherche.

Dans ce cadre, l'inférence bayésienne [128] présente une alternative très attractive, fournissant une formulation mathématique souple et riche au problème d'inversion. Elle se base sur la détermination de la probabilité *a posteriori* des informations cachées sachant les observations $p(X = x/Y = y)$. Cette dernière peut être calculée avec la règle de Bayes comme suit :

$$\begin{aligned} p(X = x/Y = y) &= \frac{p(X = x, Y = y)}{p(Y = y)} \\ &= \frac{p(Y = y/X = x)p(X = x)}{p(Y = y)} \end{aligned} \quad (1.1)$$

où $p(X = x, Y = y)$ est la probabilité jointe¹, $p(Y = y/X = x)$ est la vraisemblance des observations conditionnellement aux informations cachées, $p(X = x)$ est l'*a priori*

¹Pour des raisons de simplification nous noterons $p(X = x) = p_X(x)$ par $p(x)$.

sur X et $p(Y = y)$ est une constante de normalisation. Il est clair que l'*a posteriori* offre un degré de liberté supplémentaire sur la simple vraisemblance, permettant ainsi de prendre en compte la connaissance disponible en amont sur X , les deux quantités étant similaires dans le cas d'un *a priori* non informatif ne favorisant aucune partie de l'espace des solutions. Plus loin encore, les estimateurs bayésiens permettent d'associer à chaque estimation \hat{x} de x un coût $\mathcal{L}(x, \hat{x})$ mesurant le prix à payer en choisissant \hat{x} sachant que la vraie solution est x . Cette pénalisation appelée fonction coût est très pratique dans la mesure où certaines décisions ont plus d'impact que les autres. L'estimateur optimal au sens de Bayes est celui qui minimise le coût moyen, *i.e.*, risque de Bayes :

$$\hat{x}_{opt}(y) = \arg \min_{\hat{x}} \mathbf{E}(\mathcal{L}(x, \hat{x})/y) \quad (1.2)$$

Dans la suite, nous présenterons trois estimateurs bayésiens classiques. Avant cela, introduisons quelques notations.

Soit le champ aléatoire des observations $Y = \{Y_s, s \in T\}$ tel que $\forall s \in T : Y_s \in \Omega_y$ et $Y \in \Lambda_y = \Omega_y^{|T|}$ l'ensemble de toutes les configurations possibles de Y . De même, soit le champ aléatoire des informations cachées $X = \{X_s, s \in S\}$ tel que $\forall s \in S : X_s \in \Omega_x$ et $X \in \Lambda_x = \Omega_x^{|S|}$ l'ensemble de toutes les configurations possibles de X . S et T sont deux grilles qui peuvent être identiques ou non et s représente un élément de l'une d'elles.

L'estimateur du MAP

L'estimateur du Maximum *a Posteriori* MAP pénalise de manière identique toutes les estimations différentes de la vraie solution x . Ainsi, deux estimations qui diffèrent respectivement d'un site et de la moitié des sites de la solution exacte auront la même pénalisation. La fonction coût est dans ce cas :

$$\mathcal{L}(x, \hat{x}) = 1 - \delta(x, \hat{x}) \quad \text{avec} \quad \delta(a, b) = \begin{cases} 1 & \text{si } a = b \\ 0 & \text{sinon} \end{cases} \quad (1.3)$$

le symbole de Kronecker.

L'estimateur résultant est donné par :

$$\hat{x}_{MAP}(y) = \arg \max_x p(x/y) \quad (1.4)$$

L'estimateur du MPM

L'estimateur du Mode des Marginales *a Posteriori* MPM pénalise une solution proportionnellement au nombre de sites erronés. La fonction coût qui lui est associée est :

$$\mathcal{L}(x, \hat{x}) = \sum_{s \in S} 1 - \delta(x_s, \hat{x}_s) \quad (1.5)$$

On obtient alors l'estimateur suivant :

$$\forall s \in S \quad \hat{x}_{sMPM}(y) = \arg \max_{x_s} p(x_s/y) \quad (1.6)$$

L'estimateur du champ moyen

Cet estimateur comptabilise l'erreur quadratique en chaque site. Sa fonction de coût est de la forme :

$$\mathcal{L}(x, \hat{x}) = \sum_{s \in S} (x_s - \hat{x}_s)^2 \quad (1.7)$$

ce qui donne l'estimateur :

$$\forall s \in S \quad \hat{x}_{sCM}(y) = \mathbf{E}(x_s/y) \quad (1.8)$$

qui représente l'espérance conditionnelle de X_s sachant les observations y . Il est à noter que dans le cas où les étiquettes X prennent leurs valeurs dans l'ensemble discret Λ_x , l'estimée du champ moyen n'est pas forcément dans le même ensemble.

1.3 Modèles graphiques

Comme nous l'avons vu dans la section précédente, les observations et les informations cachées sont considérées comme des variables aléatoires dont on veut modéliser les dépendances statistiques qui se trouvent efficacement résumées par la loi jointe $p(X, Y)$. Récemment, la modélisation graphique des processus stochastiques a suscité un vif intérêt. En effet, un graphe de dépendance représente de manière visuelle et compacte les dépendances existantes entre les variables aléatoires. Ceci permet d'utiliser efficacement les algorithmes existants sur les graphes pour réaliser différentes opérations dont le calcul de la loi jointe fait partie.

Définition 1 (Graphe de dépendance)

Un graphe de dépendance $G(S, L)$, orienté ou non, est composé d'un ensemble de nœuds (sites) S , représentant des variables aléatoires, et un ensemble d'arêtes L tel que l'absence d'une arête entre les sites s et t représente une relation d'indépendance conditionnelle entre les variables aléatoires associées à ces deux sites.

Soit le processus aléatoire Z indexé sur le graphe $G(S, L)$, et soit l'ensemble des nœuds $a \subset S$, on note Z_a l'ensemble des variables aléatoires $\{Z_t; t \in a\}$.

Définition 2 (Indépendance conditionnelle)

Soit \mathbf{A}, \mathbf{B} et \mathbf{C} trois ensembles de variables aléatoires. \mathbf{A} et \mathbf{B} sont indépendants conditionnellement à \mathbf{C} si et seulement si :

$$p(\mathbf{A}, \mathbf{B}/\mathbf{C}) = p(\mathbf{A}/\mathbf{C}) p(\mathbf{B}/\mathbf{C}) \quad (1.9)$$

et on note $\mathbf{A} \perp \mathbf{B}/\mathbf{C}$.

Définition 3 (Système de voisinage)

Soit S un ensemble de sites. À chaque site $s \in S$ est associé un ensemble $\nu_s \subset S$, appelé voisinage de s , vérifiant :

$$\begin{aligned} -s &\notin \nu_s \\ -t \in \nu_s &\iff s \in \nu_t \end{aligned}$$

$\mathcal{N} \triangleq \{\nu_s, s \in S\}$ est appelé système de voisinage.

Définition 4 (Clique)

Une clique c est soit un singleton soit un ensemble de sites mutuellement voisins tel que pour toute paire $\{s, t\} \in c$ nous avons $t \in \nu_s$. On note \mathcal{C} l'ensemble de toutes les cliques associé à un système de voisinage \mathcal{N} . Une clique est dite maximale si aucune clique ne la contient.

1.3.1 Graphe de dépendance non orienté

Un graphe de dépendance non orienté est un graphe dont toutes les arêtes sont non dirigées. Un processus aléatoire Z associé au graphe non orienté $G(S, L)$ est un champ de Markov. Le voisinage ν_s de tout site $s \in S$ est l'ensemble des sites $\{t \in S, (t, s) \in L\}$. Le processus Z vérifie [112, 89] :

(MP) la propriété de Markov par paire si pour toute paire de sites $\{s, t\} \in S$ tels que $s \notin \nu_t$, les variables aléatoires Z_s et Z_t sont indépendantes conditionnellement à toutes les autres, *i.e.*, :

$$\forall \{s, t\} \in S : (s, t) \notin L \Rightarrow Z_s \perp Z_t / Z_{S \setminus \{s, t\}};$$

(ML) la propriété de Markov locale si conditionnellement à son voisinage ν_s , la variable aléatoire Z_s est indépendante du reste du graphe, *i.e.*, :

$$\forall s \in S : p(Z_s / Z_{S \setminus \{s\}}) = p(Z_s / Z_{\nu_s});$$

(MG) la propriété de Markov globale si pour tous trois sous-ensembles de sites non vides et disjoints inclus dans S , \mathbf{a}, \mathbf{b} et \mathbf{c} , tels que \mathbf{c} sépare² \mathbf{a} et \mathbf{b} , $Z_{\mathbf{a}}$ et $Z_{\mathbf{b}}$ sont indépendants conditionnellement à $Z_{\mathbf{c}}$. On note $Z_{\mathbf{a}} \perp Z_{\mathbf{b}} / Z_{\mathbf{c}}$ (voir exemple de la Fig. 1.1).

La propriété (ML) est souvent utilisée pour spécifier un champ de Markov [92]. Les trois propriétés de Markov ne sont pas nécessairement équivalentes. Généralement, nous avons [89] :

$$(\text{MG}) \Rightarrow (\text{ML}) \Rightarrow (\text{MP})$$

Si la propriété suivante est vérifiée :

$$\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \text{ sous-ensembles disjoints de } S,$$

$$\text{Si } Z_{\mathbf{a}} \perp Z_{\mathbf{b}} / Z_{\mathbf{c}} \cup Z_{\mathbf{d}} \text{ et } Z_{\mathbf{a}} \perp Z_{\mathbf{c}} / Z_{\mathbf{b}} \cup Z_{\mathbf{d}} \text{ alors } Z_{\mathbf{a}} \perp Z_{\mathbf{b}} \cup Z_{\mathbf{c}} / Z_{\mathbf{d}}$$

nous avons alors l'équivalence [89] :

$$(\text{MG}) \Leftrightarrow (\text{ML}) \Leftrightarrow (\text{MP})$$

²tout chemin d'un site de \mathbf{a} vers un site de \mathbf{b} passe par au moins un site de \mathbf{c}

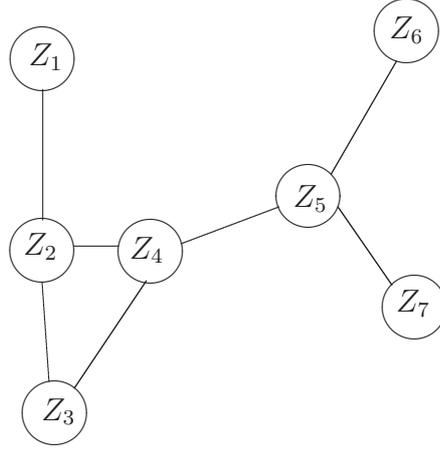


FIG. 1.1 – Propriété de Markov global sur un graphe non orienté : nous avons par exemple $\{Z_1, Z_2, Z_3\} \perp \{Z_6, Z_7\} / \{Z_4, Z_5\}$ et $\{Z_1, Z_2, Z_3, Z_4\} \perp \{Z_6, Z_7\} / Z_5$ et $\{Z_1, Z_2, Z_3\} \perp \{Z_5, Z_6, Z_7\} / Z_4$

Définition 5 (Champ de Gibbs)

Un processus aléatoire Z indexé par un ensemble de site S est un champ de Gibbs relativement au système de voisinage \mathcal{N} si sa distribution jointe est une distribution de Gibbs :

$$p(Z = z) = \frac{1}{\zeta} \exp - \sum_{c \in \mathcal{C}} V_c(z) \quad (1.10)$$

où ζ est une constante de normalisation et V_c le potentiel correspondant aux cliques $c \in \mathcal{C}$, avec \mathcal{C} l'ensemble de cliques associé au système de voisinage \mathcal{N} .

Théorème 1 (Hammersley-Clifford)

Soit Z un processus aléatoire indexé par un ensemble de site S , à valeurs dans Ω_z , sur lequel est défini le système de voisinage \mathcal{N} .

Z est un champ de Markov avec $\forall z \in \Lambda_z p(Z = z) > 0 \Leftrightarrow Z$ est un champ de Gibbs

On peut déduire [9], que la probabilité conditionnelle exprimée par la propriété de Markov locale peut s'écrire :

$$\forall s \in S : p(Z_s / Z_{V_s}) \propto \exp - \sum_{c \in \mathcal{C} : s \in c} V_c(z) \quad (1.11)$$

D'autre part, si on pose $V_c = -\ln \Psi_c$ dans l'Eq. 1.10, on retrouve la forme factorisée de la loi jointe d'un champ de Gibbs Z :

$$p(Z = z) \propto \prod_{c \in \mathcal{C}} \Psi_c(z) \quad (1.12)$$

où $\Psi_c(z) > 0, \forall z \in \Lambda_z$, sont des fonctions locales définies sur l'ensemble des cliques \mathcal{C} .

On peut donc conclure que tout processus aléatoire défini par une loi jointe qui peut être factorisée de la même façon que celle de l'Eq.1.12, suivant un ensemble de cliques \mathcal{C} , peut être représenté par un graphe non orienté avec le système de voisinage \mathcal{N} associé à l'ensemble \mathcal{C} .

1.3.2 Graphe de dépendance orienté

Un graphe de dépendance orienté $G = (S, L)$, également appelé réseau bayésien, est un graphe dont les arêtes sont orientées. Les cycles sont absents dans ce genre de graphe en prenant en compte l'orientation des arêtes [106]. On définit pour chaque site s ses parents s^- (l'ensemble des sites qui pointent vers s), ses enfants s^+ (l'ensemble des sites vers lesquels il pointe), ses ancêtres $< s$ (l'ensemble de ses parents, les parents de ses parents... etc) et ses descendants $> s$ (l'ensemble de ses enfants, les enfants de ses enfants... etc).

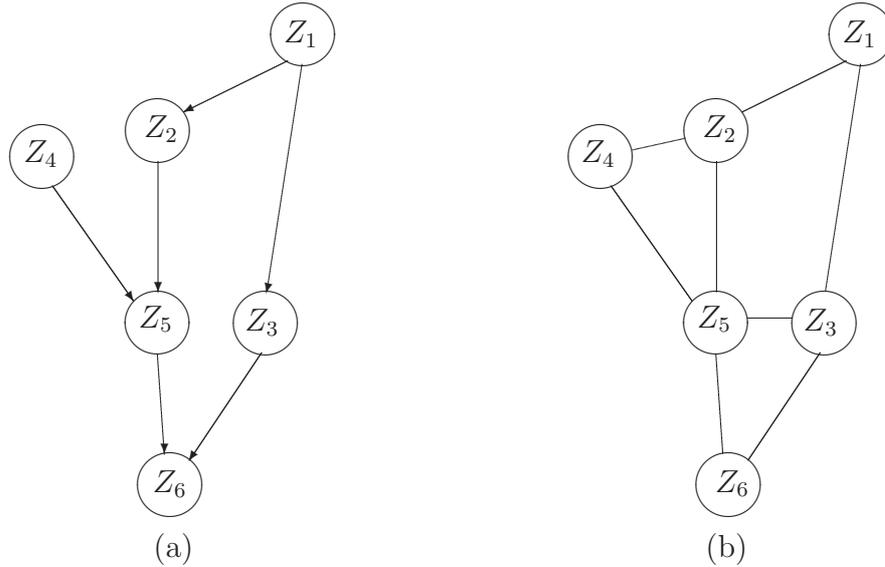


FIG. 1.2 – Graphes orientés : (a) Exemple d'un graphe orienté (b) Transformation du graphe (a) en un graphe non orienté équivalent (*i.e.*, moralisation)

Pour un processus aléatoire Z associé au graphe orienté $G = (S, L)$, les relations d'indépendances conditionnelles sont plus complexes que pour le modèle non orienté. La plus simple relation d'indépendance codée par un graphe orienté est que chaque site est indépendant de ses ancêtres sachant ses parents [106] :

$$p(Z_s/Z_{<s}) = p(Z_s/Z_{s^-})$$

Pour le graphe de la Fig. 1.2 par exemple, la variable aléatoire Z_5 possède deux parents Z_2 et Z_4 et donc :

$$p(Z_5/Z_1, Z_2, Z_4) = p(Z_5/Z_2, Z_4)$$

La loi jointe du processus aléatoire Z peut alors se mettre sous la forme factorisée suivante :

$$p(Z = z) \propto \prod_{s \in S} p(z_s/z_{s^-}) \quad (1.13)$$

En comparant cette expression avec l'Eq. 1.12 et posant $\Psi_{\{s\} \cup s^-}(z) = p(z_s/z_{s^-})$, on peut voir qu'il est possible de construire un graphe non orienté $G^m = (S, L^m)$ équivalent au graphe orienté $G = (S, L)$ [89]. G^m , appelé graphe morale, est obtenu à partir de G

en deux étapes : 1) transformer toutes les arêtes orientées en non orientées 2) ajouter des arêtes entre les parents initialement non liés et qui ont des enfants en commun. Cette opération de conversion est appelée moralisation.

Chaque site est indépendant de tout le graphe sachant sa couverture de Markov (Markov blanket en anglais). La couverture de Markov étant les parents, les enfants et autres parents des enfants. Ceci donne :

$$p(Z_s/Z_{\{t,t \neq s\}}) = p(Z_s/Z_{s^-}, Z_{s^+}, Z_{\{i^-:i \in s^+ \text{ et } i^- \neq s\}})$$

En appliquant cette règle à la variable Z_5 du graphe de la Fig. 1.2, on peut écrire :

$$p(Z_5/Z_{\{t,t \neq 5\}}) = p(Z_5/Z_2, Z_3, Z_4, Z_6)$$

Pour l'exemple de la Fig. 1.2, le graphe moral (à droite) obtenu à partir du graphe orienté (à gauche) montre bien que Z_5 n'est pas conditionnellement indépendante de Z_3 , dépendance qui n'est pas immédiate dans le graphe de gauche.

1.3.3 Manipulation des graphes non orientés

La manipulation d'une loi jointe d'un processus aléatoire Z associé à un graphe de dépendance non orienté $G = (S, L)$ fait appel à trois mécanismes différents : le conditionnement, la marginalisation et le partitionnement.

Le conditionnement

Le conditionnement du graphe $G = (S, L)$ par rapport à un sous ensemble $\mathbf{a} \subset S$ revient à figer les variables aléatoires $Z_{\mathbf{a}}$ et considérer la loi jointe des variables restantes. Le graphe correspondant à la loi conditionnelle $p(Z_{S \setminus \mathbf{a}}/Z_{\mathbf{a}})$ est le graphe $\hat{G} = (S \setminus \mathbf{a}, \hat{L})$ tel que $\hat{L} = \{(s, t) \in L : \{s, t\} \subset S \setminus \mathbf{a}\}$.

La marginalisation

La marginalisation du graphe $G = (S, L)$ par rapport à un sous ensemble $\mathbf{a} \subset S$ revient à intégrer par rapport aux variables aléatoires $Z_{\mathbf{a}}$ pour obtenir la loi marginale des variables restantes. Le graphe correspondant à la loi marginale $p(Z_{S \setminus \mathbf{a}})$ est le graphe $\hat{G} = (S \setminus \mathbf{a}, \hat{L})$ tel que $\hat{L} = \{(s, t) \subset (S \setminus \mathbf{a})^2 : ((s, t) \in L) \text{ ou } (\exists \text{ chaîne } \subset \mathbf{a} \text{ joignant } s \text{ et } t)\}$.

Le partitionnement

Le partitionnement du graphe $G = (S, L)$ signifie le regroupement de certaines variables aléatoires pour former des "méta-variables" dont l'interaction globale peut s'avérer pertinente. On définit ainsi une partition d'ensembles non vides $(\mathbf{a}_k)_{k=1}^p$ de S . Le graphe correspondant au vecteur aléatoire $(Z_{\mathbf{a}_k})_k$ résultat de la partition de Z est le graphe $\hat{G} = (\hat{S}, \hat{L})$ tel que $\hat{S} = \{1, \dots, p\}$ et $\hat{L} = \{(k, l) \subset \hat{S}^2 : \exists (s, t) \in L, s \in \mathbf{a}_k \text{ et } t \in \mathbf{a}_l\}$.

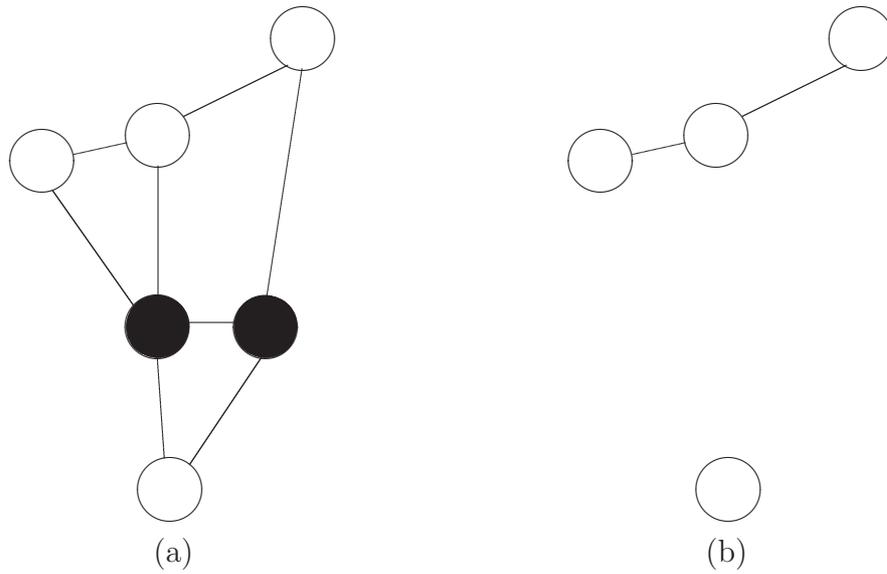


FIG. 1.3 – Conditionnement d'un graphe non orienté : (a) Exemple d'un graphe non orienté (b) Conditionnement du graphe (a) par rapport aux variables en noir

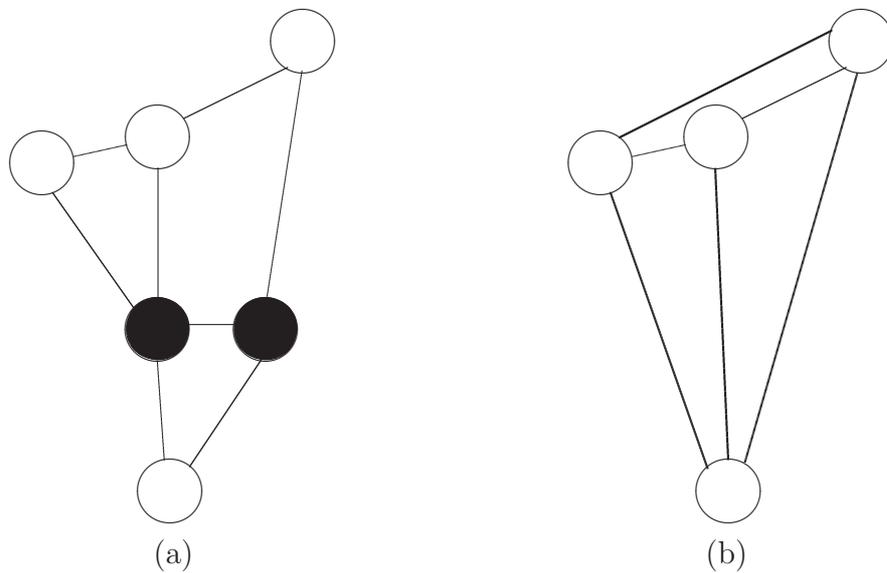


FIG. 1.4 – Marginalisation dans un graphe non orienté : (a) Exemple d'un graphe non orienté (b) Marginalisation du graphe (a) par rapport aux variables en noir

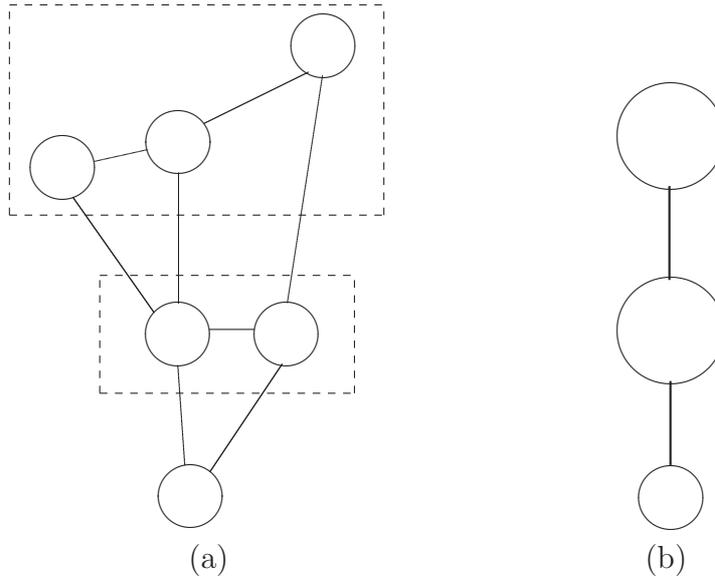


FIG. 1.5 – Partitionnement d’un graphe non orienté : (a) Exemple d’un graphe non orienté (b) Partitionnement du graphe (a) en trois ”méta-variables”

1.3.4 Tâches génériques sur les modèles graphiques

Pour tout processus aléatoire Z , constitué d’une partie observée Y et d’une partie cachée X pouvant être décrit par une ou plusieurs classes de lois factorisées, trois tâches génériques peuvent être rencontrées :

- **Sélection de modèle** Il s’agit de choisir une classe \mathcal{M} de lois (modèle) pour décrire au mieux les données observées. L’approche bayésienne se base sur le calcul de probabilité *a posteriori* du modèle sachant les observations $p(\mathcal{M}/Y)$. Cette tâche, généralement complexe, n’est que rarement abordée en imagerie et ne sera pas reprise dans le cadre de ce travail.
- **Apprentissage** Une fois le modèle \mathcal{M} choisi, l’apprentissage consiste à estimer l’ensemble des paramètres Θ associés à ce modèle étant données les observations disponibles. Les principaux algorithmes utilisés pour résoudre cette tâche seront décrits dans le chapitre 3.
- **Inférence** L’inférence a pour but d’estimer les variables non observées d’après les observations et la classe de lois retenue pour laquelle les paramètres sont déjà estimés. Les principales techniques de résolution exactes et approximatives des problèmes d’inférence seront décrites ci-après.

1.3.5 Méthodes exactes d’inférence

À la base de la résolution du problème d’inférence on retrouve deux opérations qui sont la marginalisation et la maximisation. Dans ce qui suit nous nous intéresserons à la marginalisation, la maximisation pouvant être conduite de façon similaire en remplaçant les sommations par des maximisations.

Technique des éliminations successives

Nous introduirons cet algorithme par un exemple [101]. Soit le processus aléatoire décrit par le graphe de la Fig. 1.6. Calculons la marginale $p(z_1)$.

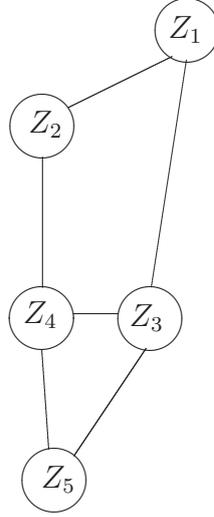


FIG. 1.6 – Exemple d'un graphe non orienté

$$\begin{aligned}
 p(z_1) &= \sum_{z_2} \sum_{z_3} \sum_{z_4} \sum_{z_5} \frac{1}{\zeta} \Psi(z_1, z_2) \Psi(z_1, z_3) \Psi(z_2, z_4) \Psi(z_3, z_4, z_5) \\
 &= \frac{1}{\zeta} \sum_{z_2} \Psi(z_1, z_2) \sum_{z_3} \Psi(z_1, z_3) \sum_{z_4} \Psi(z_2, z_4) \sum_{z_5} \Psi(z_3, z_4, z_5) \\
 &= \frac{1}{\zeta} \sum_{z_2} \Psi(z_1, z_2) \sum_{z_3} \Psi(z_1, z_3) \sum_{z_4} \Psi(z_2, z_4) m_5(z_3, z_4) \\
 &= \frac{1}{\zeta} \sum_{z_2} \Psi(z_1, z_2) \sum_{z_3} \Psi(z_1, z_3) m_4(z_2, z_3) \\
 &= \frac{1}{\zeta} \sum_{z_2} \Psi(z_1, z_2) m_3(z_2, z_1) \\
 &= \frac{1}{\zeta} m_2(z_1)
 \end{aligned} \tag{1.14}$$

La constante de normalisation ζ peut être obtenue par une dernière sommation suivant z_1 . Chaque sommation apparaissant dans le calcul contient au plus trois variables, d'où une réduction de la complexité de $|\Omega_z|^5$ à $|\Omega_z|^3$. On peut réaliser la sommation précédente dans n'importe quel ordre. Il est cependant judicieux de choisir un ordre de sommation pour lequel la complexité soit minimale. En effet, comme nous l'avons vu plus haut, les marginalisations successives induisent une série de transformations sur le graphe initial introduisant à chaque fois de nouvelles liaisons entre les sites, et donc différents ordres de sommation induisent différentes suites de graphes. La clique c_{tw} de plus grand cardinal associé à l'ordre de sommation de plus faible complexité définit ce qu'on appelle la *largeur*

de l'arbre (*treewidth* en anglais) du graphe. Par convention, *largeur de l'arbre* est égale au cardinal de c_{tw} moins 1. Le problème est donc de trouver l'ordre de marginalisation qui réalise la *largeur de l'arbre*.

La procédure par éliminations successives fournit une marginale à la fois. Il faut donc relancer le processus autant de fois que le nombre de marginales que l'on veut calculer. L'algorithme décrit ci-après permet de mieux gérer cette complexité de calcul.

Algorithme du passage de message

Cet algorithme est utilisé lorsque le graphe $G = (S, L)$ est un arbre. Un arbre est un graphe non orienté, sans cycle, dans lequel chaque noeud a au plus un parent. Il y a un chemin unique entre deux sites s et t de S . Pour calculer la marginale en un site quelconque s , considérons le comme racine de l'arbre. Nous aurons ainsi un nouvel arbre où nous choisissons un ordre d'élimination tel que les enfants sont éliminés avant leur parents. L'étape élémentaire de l'algorithme d'élimination, décrite auparavant, devient :

$$m_{ji}(z_i) = \sum_{z_j} \left(\Psi(z_i, z_j) \prod_{k \in \nu_j \setminus i} m_{kj}(z_j) \right) \quad (1.15)$$

où ν_j est le voisinage de j et $m_{ji}(z_i)$, appelé message envoyé par j à i , est le terme résultant de l'élimination de j . La marginale désirée est donnée par :

$$p(z_s) \propto \prod_{k \in \nu_s} m_{ks}(z_s) \quad (1.16)$$

En pratique, un arbre possède une racine naturelle r et un ensemble de terminaisons S^0 appelées feuilles. Tout site $s \in S \setminus (\{r\} \cup S^0)$ possède un parent s^- et des enfants s^+ , la racine n'ayant pas de parent et les feuilles ne possédant pas d'enfants. On calcule alors tous les messages possibles sur le graphe en deux passes sur l'arbre : *feuilles* \rightarrow *racine* et *racine* \rightarrow *feuilles*. On est ainsi assuré lors du calcul du message $m_{ji}(z_i)$ que tous les messages $m_{kj}(z_j)$, $k \in \nu_j \setminus i$, sont déjà calculés. Les marginales peuvent être ensuite calculées grâce à l'Eq. 1.16.

Cet algorithme appelé aussi "*sum-product algorithm*" permet d'avoir une estimation au sens du MPM. Pour le critère MAP les sommes sont remplacés par des maximisations d'où l'appellation "*max-product algorithm*".

Algorithme de l'arbre de jonction

Cet algorithme est une généralisation de l'algorithme de passage du message pour un graphe quelconque. Il consiste à transformer le graphe en un hyper-arbre où chaque noeud est une clique. Aussi, on vérifie que les noeuds se trouvant sur l'unique chemin entre C_i et C_j contiennent les variables $C_i \cap C_j$. On associe à l'arête séparant C_i et C_j un séparateur S_{ij} . L'arbre de cliques ainsi construit est appelé arbre de jonctions. Les cliques constituant les noeuds de l'arbre ne sont pas celles du graphe original, mais celles d'un graphe augmenté obtenu comme suit. On choisit un ordre d'élimination et on réalise

les opérations graphiques correspondantes à toutes ces marginalisations. On note toutes les nouvelles liaisons ainsi créées et on les rajoute au graphe original (voir Fig. 1.7). Le

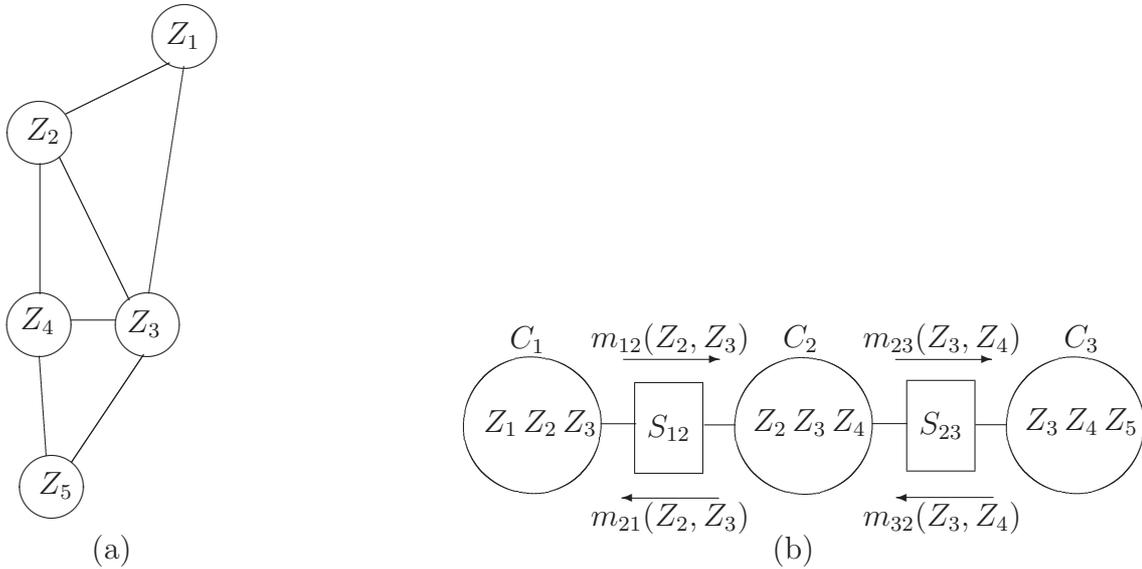


FIG. 1.7 – Exemple d’un arbre de jonction (a) Graphe augmenté obtenu à partir du graphe de la Fig. 1.6. (b) Arbre de jonction correspondant ($S_{ij} = C_i \cap C_j$)

message passé entre les cliques i et j à travers S_{ij} dans ce cas est :

$$m_{ij}(z_{S_{ij}}) = \sum_{C_i \setminus S_{ij}} \left(\Psi(z_{C_i}) \prod_{k \in \nu_i \setminus j} m_{ki}(z_{S_{ki}}) \right) \quad (1.17)$$

Il faut calculer tous les messages en respectant le même protocole que dans l’algorithme du passage de message : une clique ne peut envoyer de message à son voisin que si elle a reçu tout les message des autres voisins. En pratique on choisit les deux passes sur l’arbre : *feuilles* \rightarrow *racine* et *racine* \rightarrow *feuilles*. Une fois ceci réalisé, les marginales sont données par :

$$p(z_{C_i}) \propto \prod_{k \in \nu_i} m_{ki}(z_{S_{ki}}) \quad (1.18)$$

Cet algorithme n’est plus pratique dès que la *largeur de l’arbre* devient importante. Dans ce cas on a recours à des méthodes approximatives.

1.3.6 Méthodes approximatives d’inférence

Dans le cas où marginalisation et maximisation ne peuvent pas être menées d’une façon exacte, les quantités d’intérêt sont approchées par des méthodes dites de Monte Carlo. L’idée de base est la suivante : lorsque le calcul de moyennes suivant une loi $p(z)$ n’est pas possible, on tire des échantillons suivant cette loi ou une loi assez ressemblante, tel que moyennes et marginales puissent être approximées en se basant sur les moyennes empiriques calculées à partir des échantillons tirés [101].

Échantillonnage pondéré (Importance sampling)

Cet algorithme suggère d'approximer la moyenne de la fonction $f(z)$ suivant la loi $p(z)$ en utilisant M échantillons $z^{(m)}$, $m = 1, \dots, M$, tirés suivant une loi $q(z)$, comme suit [101, 112] :

$$\begin{aligned} \mathbf{E}[f(z)] &= \sum_z p(z) f(z) \\ &= \sum_z q(z) \left(\frac{p(z)}{q(z)} f(z) \right) \\ &\approx \frac{1}{M} \sum_{m=1}^M \frac{p(z^{(m)})}{q(z^{(m)})} f(z^{(m)}) \end{aligned} \quad (1.19)$$

La qualité de l'approximation dépend de M et du degré de ressemblance entre p et q . Dans les modèles graphiques un choix serait d'éliminer certaines arêtes du graphe de p . Ceci reste néanmoins difficile pour des graphes complexes, ce qui conduit généralement à l'utilisation des méthodes d'échantillonnage basées sur les chaînes de Markov.

Échantillonnage par chaîne de Markov

Les algorithmes de Monte Carlo par chaîne de Markov (MCMC) sont basés sur l'échantillonnage suivant une chaîne de Markov dont l'état stationnaire donnerait la distribution désirée [101]. L'échantillonneur de Gibbs est un exemple d'algorithme MCMC. Pour un processus Z sur un graphe $G = (S, L)$, cet algorithme opère comme suit :

- sélectionner la variable z_s ;
- calculer $p(Z_s = l / Z_{S \setminus \{s\}})$ pour toutes les valeurs possibles de $l \in \Omega_z$;
- choisir une valeur \hat{l} suivant cette loi ;
- remplacer z_s par \hat{l} .

La complexité de calcul de la quantité $p(z_s / Z_{S \setminus \{s\}})$ est réduite par la propriété de Markov local, puisqu'alors :

$$p(z_s / Z_{S \setminus \{s\}}) = p(z_s / z_{\nu_s}) \propto \exp - \sum_{c \in \mathcal{C}: s \in c} V_c(z)$$

Lorsqu'il n'est pas possible de calculer $p(Z_s = l / S \setminus \{s\})$, on utilise l'algorithme de Metropolis-Hastings [101] qui opère comme suit :

- sélectionner la variable z_s ;
- tirer suivant une loi uniforme sur Ω_z une valeur \tilde{z}_s ;
- accepter \tilde{z}_s avec la probabilité $\alpha = \min(1, \exp(-\frac{1}{T} \Delta V_c))$, où $\Delta V_c = \sum_{c \in \mathcal{C}: s \in c} V_c(\tilde{z}) - \sum_{c \in \mathcal{C}: s \in c} V_c(z)$, avec \tilde{z} ne différant de z qu'au niveau du site s où z_s est changée par \tilde{z}_s et T un paramètre de "température" pris égal à 1.

Les algorithmes cités ci-dessus permettent d'obtenir une estimation au sens du MPM du champ Z . Pour l'estimation au sens du MAP, il existe d'autres algorithmes tel que le recuit simulé [80] ou l'ICM (Iterated Conditional Modes) [9, 11]. Le premier est basé sur la diminution progressive de la température autorisant ainsi des valeurs diminuant le terme

$\exp(-\frac{1}{T}\Delta V_c)$ aux premières itérations pour éviter de rester piégé dans des maxima locaux. En réduisant la température, le recuit simulé tend vers un comportement déterministe interdisant toute diminution de $\exp(-\frac{1}{T}\Delta V_c)$ et donc favorisant des valeurs proches de la solution courante. La convergence n'étant garantie qu'à l'infini, cet algorithme est très gourmand en temps de calcul. L'ICM permet de palier à cet inconvénient puisqu'il maximise en chaque site la quantité $\exp(-\sum_{c \in \mathcal{C}: s \in c} V_c(z))$. La convergence vers un maximum, qui n'est pas forcément global, est obtenue après un nombre fini d'itérations.

1.4 Modèles Markoviens pour les images

1.4.1 Champs de Markov

Une image est statistiquement perçue comme une réalisation y d'un champ aléatoire d'observations noté Y indexé sur une grille régulière S et prenant ses valeurs dans $\Omega_y = \mathbb{R}$. On associe au champ aléatoire des observations Y un champ aléatoire de variables cachées (étiquettes) noté X , indexé généralement aussi sur S et prenant ses valeurs dans Ω_x . Généralement, le nœud représentant y_s est lié à celui la variable cachée x_s pour tout $s \in S$. Les états cachés voisins sont liés entre eux pour tenir compte de la corrélation spatiale entre les pixels voisins dans l'image. Deux types de voisinages sont très utilisés : le voisinage en 4-connexité et le voisinage 8-connexité (voir Fig. 1.8). Suivant cette

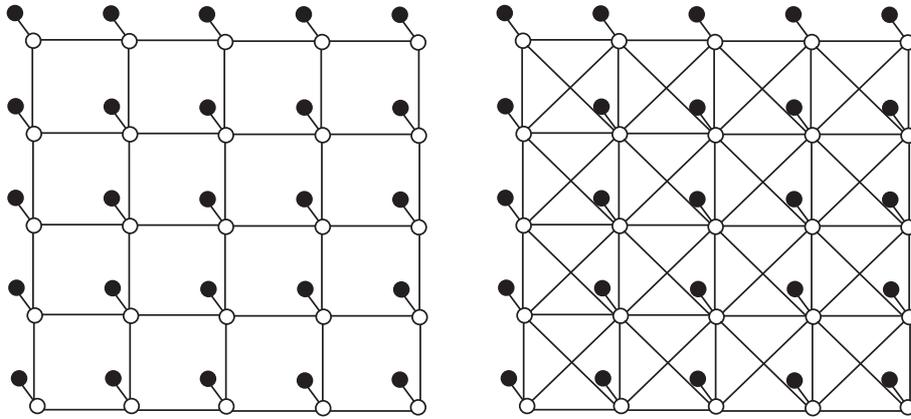


FIG. 1.8 – Graphes de dépendance d'un champ de Markov : voisinage 4-connexité à gauche et voisinage 8-connexité à droite. Les cercles blancs représentent le états cachés et cercles noirs les observations qui peuvent être vectorielles

modélisation chaque observation est indépendante du reste du graphe étant donné son état caché. On considère en plus le terme d'attache aux données strictement positif, *i.e.*, $\forall s \in S, p(y_s/x_s) > 0$. Ceci nous permet d'écrire la loi jointe des observations et des étiquettes :

$$p(x_s, y_s) \propto \exp \left(- \sum_{c \in \mathcal{C}} V_c(x) - \sum_{s \in S} l_s(x_s, y_s) \right) \quad (1.20)$$

où $l_s(x_s, y_s) = -\ln p(y_s/x_s)$ et $V_c(x)$ est le potentiel associé aux cliques c . Ce dernier est souvent calculé suivant le modèle de Potts, ce qui donne :

$$V_c(x) = \beta \left[1 - \sum_{\{s,t\} \in c: c \in \mathcal{C}} \delta(x_s, x_t) \right] \quad (1.21)$$

où $\delta(\cdot, \cdot)$ est le symbole de Kronecker et $\beta > 0$ est un paramètre favorisant la similitude entre deux sites voisins. On ne considère ainsi que les paires horizontales, verticales et parfois diagonales de pixels voisins.

Le critère MPM sur les champs de Markov consiste à générer une suite de M réalisations du champ X selon la loi *a posteriori* $p(X/Y)$, puis à choisir pour chaque site l'étiquette qui apparaît le plus souvent après élimination des p premières jugées trop dépendantes de l'initialisation. Pour le critère MAP, l'algorithme du recuit simulé et l'ICM sont largement utilisés dans la littérature [92].

Dans tous les cas, les algorithmes basés sur les champs de Markov sont fort demandeurs en temps de calcul. Une solution consiste à négliger certains voisinages et modéliser le processus caché X par une simple chaîne de Markov avec un parcours fractal de l'image.

1.4.2 Chaîne de Markov

Chaque variable cachée est liée au plus à deux voisins dans la grille S choisis selon un parcours fractal [47] pour tenir compte le mieux possible du voisinage spatial. Dans [8, 61] le parcours de Hilbert-Peano a été utilisé (voir Fig. 1.9). Deux voisins dans la chaîne sont voisins dans la grille, mais deux voisins dans la grille ne le sont pas forcément dans la chaîne. Les cliques associées à une chaîne de Markov sont les singletons et les paires (Fig. 1.10). En choisissant un sens de parcours, on peut ordonner les étiquettes et

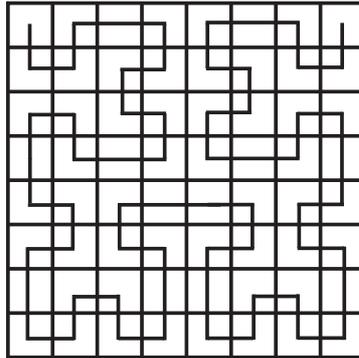


FIG. 1.9 – Parcours fractal de Hilbert-Peano sur une grille 8×8

les observations X_i et Y_i , $i = 1, \dots, |S|$.

Deux hypothèses d'indépendance sont considérées :

- Un état caché est indépendant de son passé sachant son parent :

$$p(x_i/x_1, \dots, x_{i-1}) = p(x_i/x_{i-1}) \quad (1.22)$$

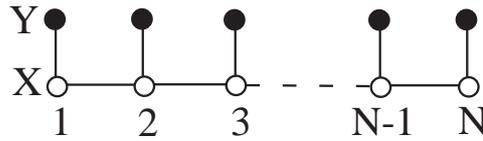


FIG. 1.10 – Graphe de dépendance d'une chaîne de Markov

- Une observation est indépendante de la chaîne et des autres observations sachant son état caché :

$$p(y_i/x, y \setminus \{y_i\}) = p(y_i/x_i) \quad (1.23)$$

Lorsque la probabilité de transition $p(x_i/x_{i-1})$ est indépendante de i , la chaîne est dite homogène. Les deux hypothèses précédentes permettent d'écrire la loi jointe sous la forme factorisée :

$$p(x, y) = p(x_1) \prod_{i=2}^{|S|} p(x_i/x_{i-1}) \prod_{i=1}^{|S|} p(y_i/x_i) \quad (1.24)$$

L'algorithme forward-backward de Baum et Welch [6, 43] est utilisé pour résoudre le problème d'inférence dans le cas du MPM, et l'algorithme de Viterbi [56] est utilisé dans le cas du MAP. Ces algorithmes sont des versions particulières des algorithmes "sum-product" et "max-product" et donc appartiennent à la famille des algorithmes de passage de message.

D'autres alternatives pour contourner les problèmes de la lourdeur des champs de Markov sont les approches hiérarchiques. De plus, c'est parfois la nature même des données qui impose le recours à des modèles hiérarchiques, un exemple étant les représentations multirésolutions des images. On distinguera donc les approches dites à *hiérarchie induite* où c'est la stratégie de résolution qui est hiérarchique, des approches dites à *hiérarchie explicite* où les modèles eux-mêmes sont hiérarchiques.

1.4.3 Approches à hiérarchie induite

Les algorithmes multiéchelles reposent sur 1) la définition d'un certain nombre de processus cachés de moins en moins complexes à partir du processus initial X ; 2) la construction des fonctions d'énergie correspondant à ces processus à partir de celles du processus X ; 3) la résolution du problème d'inférence suivant une stratégie descendante (coarse to fine) illustrée à la Fig. 1.11. On présente brièvement dans ce qui suit le principe de deux techniques qui sont l'approche par groupe de renormalisation et les algorithmes multiéchelles.

Groupe de renormalisation

Ces techniques sont inspirées de la physique statistique et appliquées initialement en traitement d'images par Gidas [59]. Le principe est le suivant : à partir du champ de Markov $X = X^0$ défini sur la grille $S = S^0$, on construit un ensemble de champs X^1, \dots, X^R définis respectivement sur des sous-grilles S^1, \dots, S^R telles que $|S^0| > |S^1| >$

$\dots > |S^R|$, en spécifiant les distributions conditionnelles $p(x^{i+1}/x^i)$. On arrive ainsi à déduire par récurrence les distributions des champs X^i , $i = 1, \dots, R$:

$$p(x^{i+1}, y) = \sum_{x^i} p(x^{i+1}/x^i) p(x^i, y) \quad (1.25)$$

Cette dernière peut se mettre sous la forme suivante :

$$p(x^i, y) \propto \exp -U^i(x^i, y) \quad (1.26)$$

avec

$$U^i(x^i, y) = -\ln \left(\sum_{x^{i-1}} p(x^i/x^{i-1}) \exp -U^{i-1}(x^{i-1}, y) \right) \quad (1.27)$$

Une fois ces énergies définies, on utilise une approche d'optimisation descendante où l'estimation du champ X^i est projetée sur le niveau $i - 1$, le champ correspondant X^{i-1} est optimisé à son tour et on ré-itére la descente jusqu'à atteindre le champ X^0 sur la grille X^0 . La difficulté avec cette technique réside dans la spécification des énergies

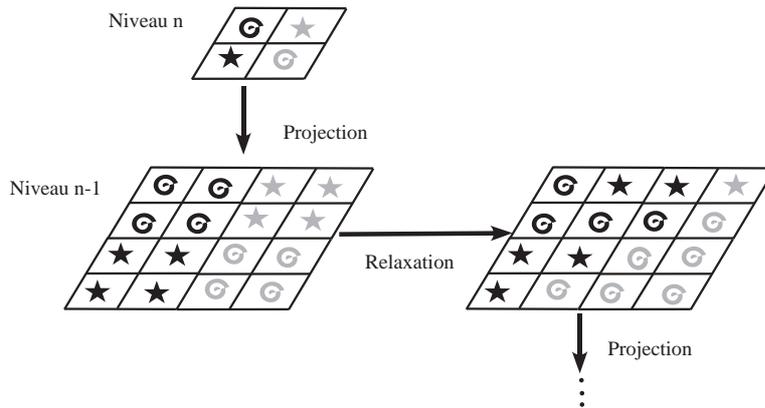


FIG. 1.11 – Stratégie d'optimisation descendante (*coarse to fine*)

U^i , $i = 1, \dots, N$, qui est analytiquement inaccessible sauf pour des cas simples [59, 42]. Un autre inconvénient est que les champs X^i , $i = 1, \dots, R$ ne sont pas généralement Markoviens et donc les algorithmes développés pour les champs de Markov ne peuvent pas être utilisés directement.

Algorithmes multiéchelles

L'idée de ces approches [18, 65, 119, 110] est de construire une suite d'espaces emboîtés $\Lambda_x^R \subset \Lambda_x^{R-1} \subset \dots \subset \Lambda_x^0 = \Lambda_x$. On définit ainsi une suite de partitions hiérarchisées de $S^i = \bigcup_{k=1, \dots, N_i} S_k^i$ de S telles que chaque bloc S_k^i est formé par un ensemble de blocs du niveau plus fin S^{i-1} (Fig. 1.12). En associant à chaque bloc S_k^i un état caché unique

X_k^i , on définit un ensemble de champs cachés X^i correspondants aux différentes échelles $i=1, \dots, R$. Il est possible alors de définir un nouveau système de voisinage et construire les fonctions d'énergie $U^i(x^i, y)$ à partir de l'énergie à pleine résolution $U(x, y)$. On minimise alors la fonction $U^i(x^i, y)$ et on projette le résultat sur le niveau $i - 1$ de proche en proche jusqu'à la résolution la plus fine. L'intérêt de cette méthode est la conservation du caractère Markovien des champs aux différentes échelles.

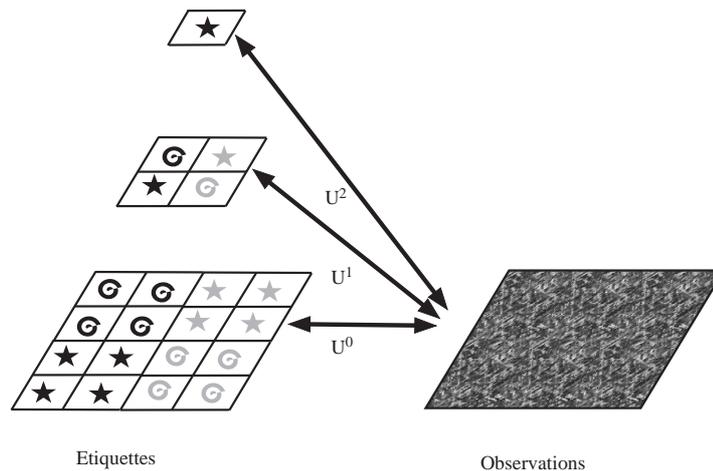


FIG. 1.12 – Illustration des approches multiéchelles

1.4.4 Approches à hiérarchie explicite

Approche multirésolution

Les approches multirésolutions reposent sur l'existence d'une pyramide d'observations répartie en $R + 1$ échelles, Y^i , $i = 0, \dots, R$. Chaque échelle Y^i est une version grossière de l'échelle Y^{i-1} qui se trouve immédiatement en dessous, Y^0 étant la plus fine (Fig. 1.13). Associé à ces données est définie une pyramide de variables cachées où chaque échelle, X^i , $i = 0, \dots, R$, constitue un champ de Markov. Il est possible de définir un système de voisinage différent pour chaque champ X^i , approche *multi-modèles* [1, 124, 34], ou bien garder le même système de voisinage, approche *mono-modèles* [4, 3, 15, 29, 17, 64]. Une stratégie d'optimisation descendante est ensuite utilisée : le résultat de segmentation du champ de Markov X^i est projeté sur l'échelle $i - 1$ pour servir d'initialisation à la segmentation du champ X^{i-1} . Ceci permet de gagner en temps de calcul par rapport à l'utilisation d'un champ unique à la résolution la plus fine.

Modèles continus sur l'arbre

Il s'agit ici de graphes intrinsèquement hiérarchiques. Le modèle le plus simple et le plus célèbre est l'arbre qui est un graphe non orienté $G = (S, L)$, sans cycle, dans lequel chaque noeud a au plus un parent. Il y a un chemin unique entre deux sites s et t de S . L'absence de cycle dans ce graphe offre la possibilité d'utilisation d'algorithmes non

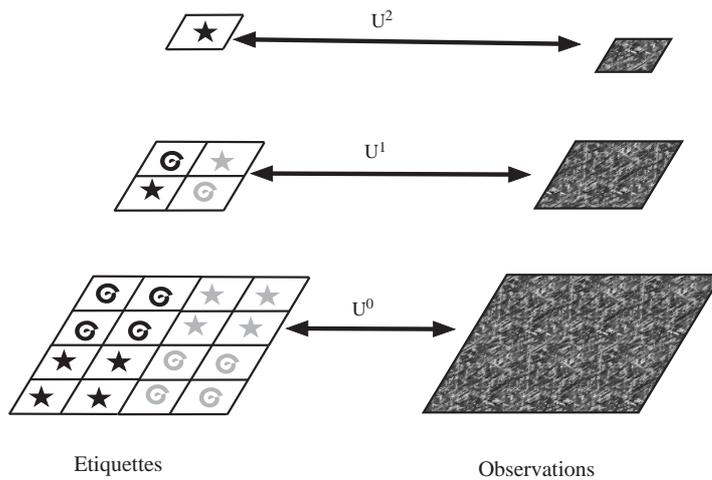


FIG. 1.13 – Illustration des approches multirésolutions

itératifs relativement rapides. Ce modèle a été développé par Chou et *al.* [32] et Basseville et *al.* [5] dans le cas des signaux 1D et puis adapté par Luetzgen et *al.* [96] pour les signaux 2D. Le modèle de Luetzgen est basé sur deux équations linéaires. Une équation d'état liant les états cachés (à valeurs continues) d'un enfant et son parent, et une équation de mesure mettant en relation l'observation et son état caché. L'algorithme est non itératif et consiste en deux passes sur l'arbre : *feuilles* \rightarrow *racine* et *racine* \rightarrow *feuilles*. Cet algorithme a été appliqué avec succès pour l'estimation de mouvement [97] et l'estimation de la bathymétrie [48].

Modèles discrets sur l'arbre

Ce modèle à variables cachées discrètes a été introduit par Bouman et *al.* [19]. Le processus caché est un quadarbre (*i.e.*, chaque site à part les feuilles dispose de 4 enfants). Les observations ne sont introduites qu'au niveau des feuilles en les liant indépendamment une par une à leurs états cachés respectifs. Jugeant le critère MAP inadapté aux structures hiérarchiques car il pénalise de la même façon l'erreur à toutes les échelles, Bouman et *al.* ont introduit l'algorithme SMAP (Sequential MAP) dont le principe est de pénaliser plus sévèrement les erreurs aux échelles grossières. Cet estimateur n'est cependant pas calculé de manière exacte [20].

Crouse et *al.* [38] et Laferté et *al.* [83] ont étendu ce modèle à des données multirésolutions avec des algorithmes pour le calcul exact des critères MAP et MPM. Provost et *al.* [121] l'ont généralisé pour le cas multispectral-multirésolution. Ce modèle va être largement développé dans la suite.

Récemment, Pieczynski [116] a introduit l'arbre de Markov couple où il considère un processus Z indexé sur un arbre, et défini à partir des deux processus X et Y , représentant respectivement les étiquettes et les observations, tel qu'en chaque site s , la variable aléatoire Z_s est donnée par $Z_s = (X_s, Y_s)$. L'observation dépend, en plus de

son état caché, des états cachés et des observations associées à ses enfants et son parent. Ce modèle sera détaillé plus loin dans ce chapitre.

On peut également citer l'arbre de Markov triplet de Pieczynski [117] où un processus auxiliaire caché U est introduit de sorte que le processus Z indexé sur un arbre est donné par $Z_s = (X_s, U_s, Y_s)$ pour tout site $s \in S$. Ce modèle généralise le modèle précédent et peut se montrer bien adapté à des processus non stationnaires [85].

Effet de bloc et solutions

L'inconvénient majeur du quadarbre est l'existence d'un "effet bloc" sur les résultats qui peut être parfois assez gênant. Plusieurs travaux ont été proposés pour contourner cette difficulté. Bouman et *al.* [20] ont introduit des liens supplémentaires dans le quadarbre tels que les sites de chaque bloc de 2×2 ont en plus de leur parent commun deux autres parents (Fig. 1.14). Le graphe n'est plus un arbre et les algorithmes itératifs sont

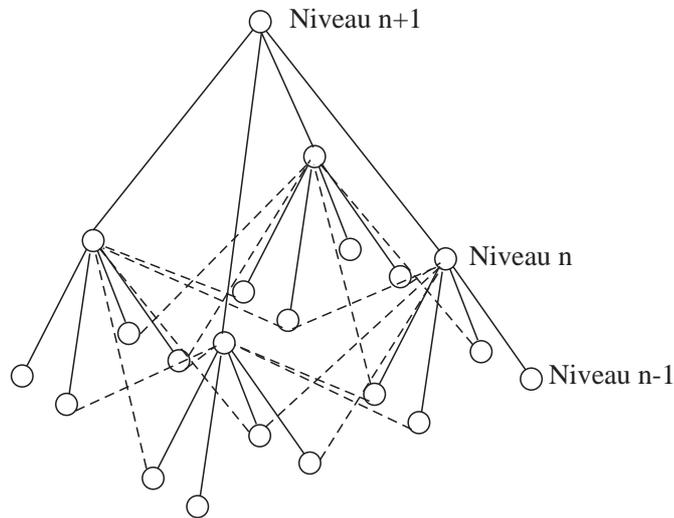


FIG. 1.14 – Quadarbre augmenté de Bouman et *al.*. Les liens supplémentaires sont mis en lignes interrompues

inévitables. Le modèle de Kato et *al.* [79] est un graphe construit à la base d'un arbre. Il est basé sur le système de voisinage classique de l'arbre augmenté d'un voisinage à 4-connexités (Fig. 1.15). Ce choix rend ce modèle beaucoup plus complexe que celui de Bouman et *al.*. Les auteurs ont utilisé un recuit simulé avec une température décroissante de la racine vers les feuilles. On peut citer aussi la solution apportée par Monfrini et *al.* [105] qui consiste à doubler le nombre de niveaux et utiliser un diarbre avec 2 enfants au lieu de 4 (Fig. 1.16). Une autre approche consiste en l'utilisation d'un arbre dynamique [142, 125]. Les liaisons entre parent et enfant ne sont plus figées mais évoluent au cours des itérations pour s'adapter au mieux aux données, le résultat étant toujours un arbre (Fig. 1.17). Dans Chardin et *al.* [30] les auteurs ont utilisé un quadarbre tronqué dans lequel la racine n'est plus un site unique mais une grille (Fig. 1.18). Ils ont abouti ainsi à un algorithme semi-itératif.

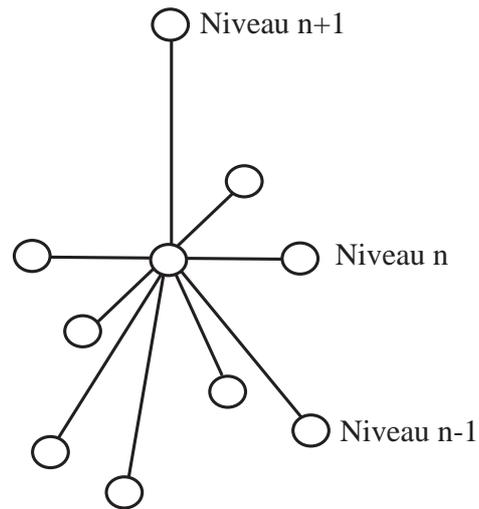


FIG. 1.15 – Système de voisinage dans le modèle de Kato et *al.*

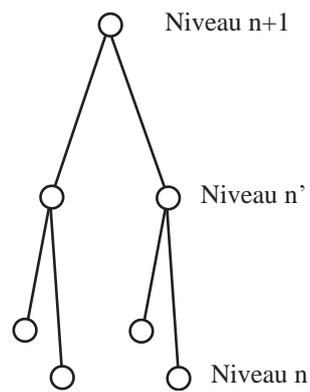


FIG. 1.16 – Diarbre de Monfrini et *al.*

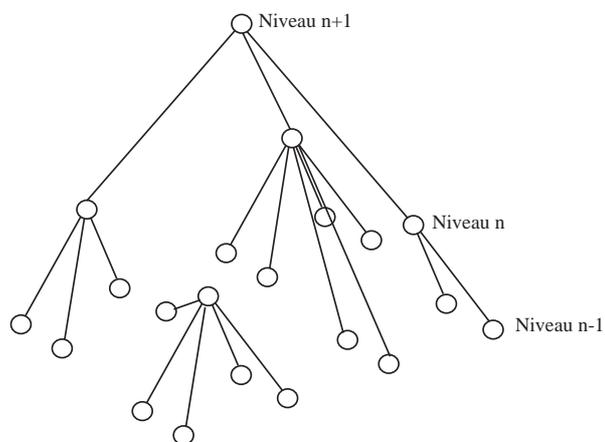
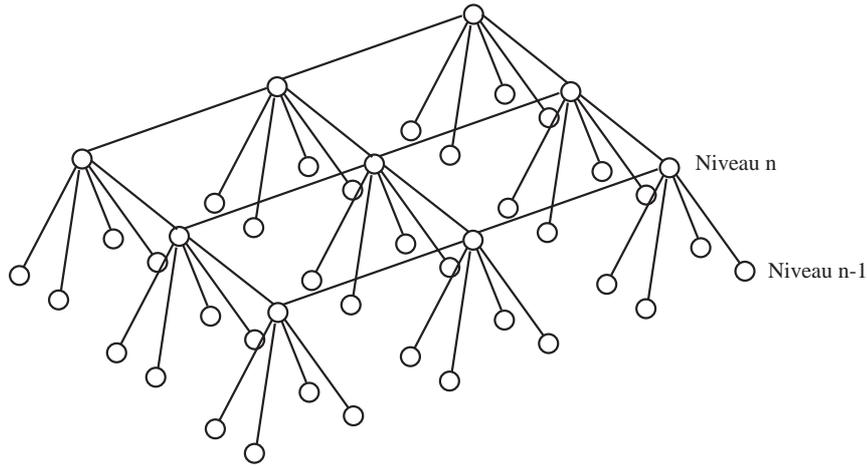


FIG. 1.17 – Exemple d'un graphe résultat de l'algorithme de l'arbre dynamique

FIG. 1.18 – Arbre tronqué de Chardin et *al.*

1.5 Inférence sur le quadarbre de Markov caché

Un quadarbre $G = (S, L)$ (Fig. 1.19) est un arbre dans lequel chaque nœud s , mis à part la racine r , a un unique prédécesseur, son *parent* s^- . De même, mis à part les nœuds terminaux (les *feuilles*), chaque nœud possède 4 *enfants* s^+ . L'ensemble des sites S peut être décomposé en *échelles*, $S = S^0 \cup S^1 \dots \cup S^R$, suivant le chemin partant des *feuilles* vers la racine. Ainsi, $S^R = \{r\}$, S^n regroupe 4^{R-n} sites, et S^0 est l'échelle la plus fine formée par les feuilles.

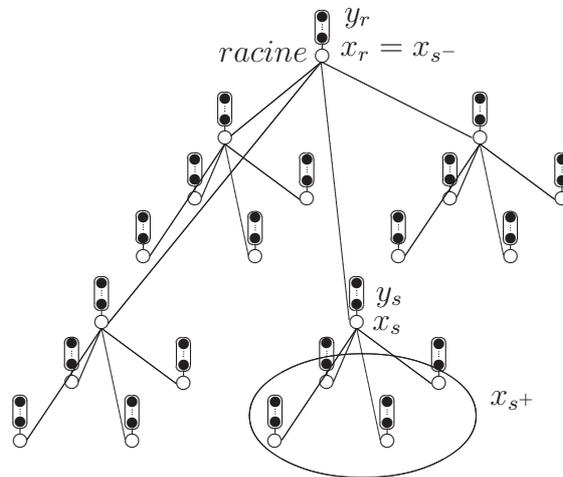


FIG. 1.19 – Exemple d'un quadarbre correspondant à une grille de 4×4 . Les cercles blancs représentent les étiquettes X_s et les noirs les observations Y_s qui peuvent être multispectrales. Chaque nœud s a un parent s^- unique et quatre enfants s^+ .

On considère les deux processus stochastiques $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ indexés sur S et correspondant respectivement aux étiquettes et aux observations. Chaque X_s prend ses valeurs dans un ensemble discret fini d'étiquettes $\Omega_x = \{\omega_1, \dots, \omega_K\}$ et Y_s dans \mathbb{R}^N .

X est supposé markovien en échelle, ce qui implique que les étiquettes x^n de l'échelle n sont conjointement indépendantes des étiquettes des échelles supérieures, *i.e.*, x^k , $k > n$, sachant les étiquettes x^{n+1} de l'échelle $n + 1$:

$$p(x^n/x^k, k > n) = p(x^n/x^{n+1}) \quad (1.28)$$

De plus, étant donné son parent, chaque état caché (étiquette) est supposé indépendant de tous ses ancêtres (son parent, le parent de son parent...). La probabilité de transition inter-échelle peut être écrite sous la forme factorisée suivante [95, 83] :

$$p(x^n|x^{n+1}) = \prod_{s \in S^n} p(x_s|x_{s-}) \quad (1.29)$$

Dans le cadre de cette thèse on supposera l'homogénéité en échelle, *i.e.*, la probabilité de transition (parent/enfant) ne dépend pas de l'échelle :

$$\forall n, \forall s \in S^n : P(x_s = \omega_j/x_{s-} = \omega_i) = a_{ij}$$

Chaque observation Y_s est indépendante de tous les autres nœuds conditionnellement à son état caché X_s . Ainsi la vraisemblance peut être exprimée sous la forme suivante :

$$p(y|x) = \prod_{n=0}^R p(y^n/x^n) = \prod_{n=0}^R \prod_{s \in S^n} p(y_s/x_s) \quad (1.30)$$

où

$$\forall s \in S^n, \forall n \in \{0, \dots, R\}, p(y_s|x_s = \omega_i) \triangleq f_i^n(y_s) \quad (1.31)$$

représente la vraisemblance de l'observation y_s . Les différentes modélisations possibles pour ce terme dans le cas multispectral seront discutées au chapitre 2.

En se basant sur ces hypothèses, la distribution jointe $p(x, y)$ peut être factorisée comme suit [83] :

$$p(x, y) = p(x_r) \prod_{s \neq r} p(x_s/x_{s-}) \prod_{s \in S} p(y_s/x_s) \quad (1.32)$$

où $p(x_r = \omega_i) = \pi_i$ est la probabilité *a priori* pour le champ d'étiquettes, et $P(x_s = \omega_j/x_{s-} = \omega_i) = a_{ij}$ est la probabilité de transition (parent/enfant).

Les paramètres de ce modèles sont :

1. les paramètres Θ_x du modèle *a priori* :
 - la probabilité *a priori* $\{\pi_i\}_{i=1, \dots, K}$
 - les probabilités de transitions parent/enfant, $\{a_{ij}\}_{i, j=1, \dots, K}$;
2. les paramètres Θ_y des vraisemblances qui dépendent du modèle choisi.

Le critère MPM sur le quadarbre de Markov caché

Le critère MPM consiste à choisir en chaque site s l'étiquette qui maximise la marginale *a posteriori* $p(x_s/y)$. Nous décrivons ici l'algorithme de Laferté et *al.* [83]. La passe montante consiste à calculer les probabilités $p(x_s/y_{\geq s})$, où $\geq s$ sont les descendants de s . Nous avons :

$$\begin{aligned}
 p(x_s/y_{\geq s}) &= \frac{p(y_{\geq s}/x_s) p(x_s)}{p(y_{\geq s})} \\
 &= \frac{p(y_s/x_s) p(x_s) \prod_{t \in s^+} p(y_{\geq t}/x_s)}{p(y_{\geq s})} \\
 &= \frac{p(y_s/x_s) p(x_s) \prod_{t \in s^+} \frac{p(x_s/y_{\geq t}) p(y_{\geq t})}{p(x_s)}}{p(y_{\geq s})} \\
 &\propto p(y_s/x_s) p(x_s)^{-3} \prod_{t \in s^+} p(x_s/y_{\geq t})
 \end{aligned} \tag{1.33}$$

D'un autre côté :

$$\begin{aligned}
 p(x_s/y_{\geq t}) &= \sum_{x_t \in \Omega_x} p(x_s, x_t/y_{\geq t}) \\
 &= \sum_{x_t \in \Omega_x} p(x_s/x_t) p(x_t/y_{\geq t})
 \end{aligned} \tag{1.34}$$

Ceci permet d'avoir la récurrence suivante :

$$p(x_s/y_{\geq s}) \propto p(y_s/x_s) p(x_s)^{-3} \prod_{t \in s^+} \sum_{x_t \in \Omega_x} p(x_s/x_t) p(x_t/y_{\geq t}) \tag{1.35}$$

Cette récurrence nécessite le calcul de la loi *a priori* $p(x_s)$ en chaque site ainsi que la probabilité de transition inverse (enfant/parent). La première est obtenue de proche en proche à partir de la racine :

$$p(x_s) = \sum_{x_{s^-} \in \Omega_x} p(x_s/x_{s^-}) p(x_{s^-}) \tag{1.36}$$

La deuxième s'obtient en utilisant l'*a priori* en chaque site et la probabilité de transition directe de la manière suivante :

$$p(x_{s^-}/x_s) = \frac{p(x_s/x_{s^-}) p(x_{s^-})}{p(x_s)} \tag{1.37}$$

À l'échelle la plus fine, $\forall s \in S^0$:

$$\begin{aligned}
 p(x_s/y_{\geq s}) &= p(x_s/y_s) \\
 &\propto p(y_s/x_s) p(x_s)
 \end{aligned} \tag{1.38}$$

À la fin de la passe montante on obtient la marginale à la racine $p(x_r/y)$. Cette quantité constitue le point de départ de la passe descendante. En effet, les marginales pour les échelles en dessous de la racine s'obtiennent par :

$$p(x_s/y) = \sum_{x_{s^-} \in \Omega_x} p(x_s, x_{s^-}/y) \tag{1.39}$$

où

$$\begin{aligned}
p(x_s, x_{s^-}/y) &= p(x_s/x_{s^-}, y) p(x_{s^-}/y) \\
&= p(x_s/x_{s^-}, y_{\geq s}) p(x_{s^-}/y) \\
&= \frac{p(x_s, x_{s^-}/y_{\geq s})}{p(x_{s^-}/y_{\geq s})} p(x_{s^-}/y) \\
&= \frac{p(x_{s^-}/x_s) p(x_s/y_{\geq s})}{\sum_{x_s \in \Omega_x} p(x_{s^-}/x_s) p(x_s/y_{\geq s})} p(x_{s^-}/y) \tag{1.40}
\end{aligned}$$

En remplaçant par l'expression de la probabilité de transition inverse (Eq. 1.37) on obtient :

$$p(x_s, x_{s^-}/y) = \frac{p(x_s/x_{s^-}) p(x_{s^-}) p(x_s/y_{\geq s})}{p(x_s) \sum_{x_s \in \Omega_x} p(x_{s^-}/x_s) p(x_s/y_{\geq s})} p(x_{s^-}/y) \tag{1.41}$$

Une fois les lois marginales en chaque site obtenues, les étiquettes sont choisies comme suit :

$$\forall s \in S \quad \hat{x}(y) = \arg \max_{x_s \in \Omega_x} p(x_s/y) \tag{1.42}$$

1.6 Inférence sur le quadarbre de Markov couple

Soit un quadarbre $G = (S, L)$ défini de la même manière que le quadarbre de Markov caché Fig. 1.20, et soient $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ deux processus stochastiques indexés

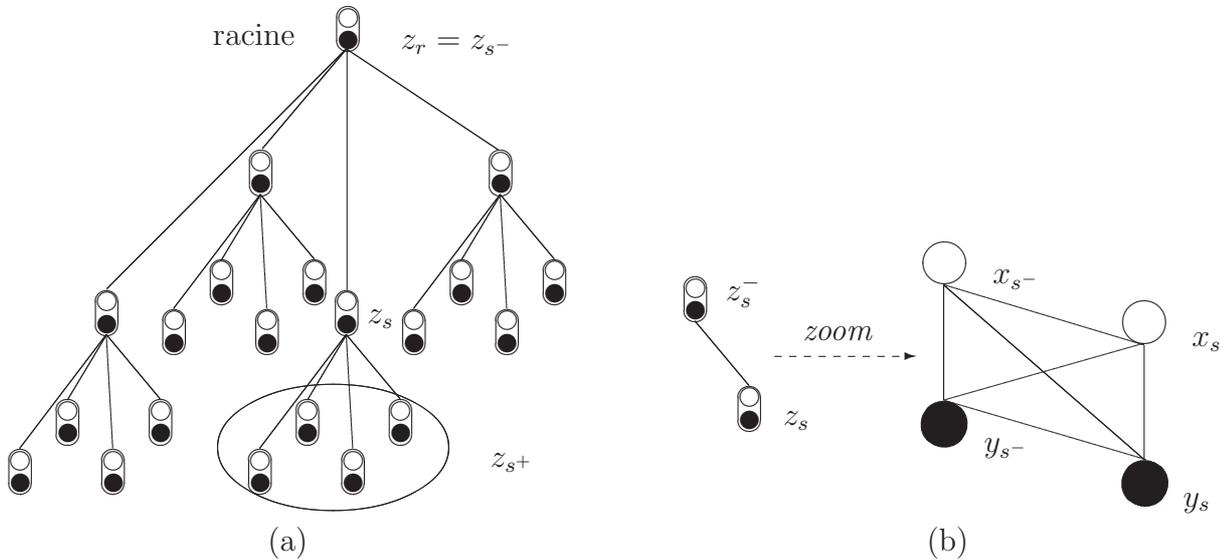


FIG. 1.20 – Quadarbre de Markov couple : (a) Graphe de dépendance illustré pour une grille 4×4 . Les cercles blancs représentent les étiquettes et les noirs les observations, respectivement X_s et Y_s , $s \in S$, tel que le processus $Z_s = (X_s, Y_s)_{s \in S}$ est défini sur un arbre. (b) Détails des liens la dépendance entre z_s et z_{s^-} .

sur S correspondant respectivement aux étiquettes³ et aux observations. On considère le

³ $\Omega_x = \{\omega_i\}_{i=1, \dots, K}$ est l'ensemble des étiquettes

processus stochastique $Z = (Z_s)_{s \in S}$ tel que $Z_s = (X_s, Y_s)$ prenant ses valeurs dans $\Omega_x \times \mathbb{R}$. Z est un arbre de Markov couple si et seulement si :

$$p(z) = p(z_r) \prod_{i=0}^{R-1} \prod_{s \in S^i} p(z_s/z_{s-}) \quad (1.43)$$

Sachant z_s , l'ensemble des descendants $\{z_t\}_{t \in > s}$ sont indépendants du reste du graphe (*i.e.* ; l'ensemble $\bar{s} = S \setminus \geq s$). De plus, Il faut noter que pris séparément, les processus X et Y ne sont pas forcément des arbres de Markov (*i.e.* ; leurs lois jointes respectives ne peuvent ne pas vérifier l'Eq. 1.43) [116, 104].

Il est clair que le quadarbre de Markov couple est une généralisation du quadarbre de Markov caché dans lequel seul le processus caché X est un arbre de Markov (Fig. 1.21). Ainsi le modèle couple enrichit les liaisons parent/enfant : l'observation y_s dépend en plus de son état caché x_s de l'état caché de son parent x_{s-} , des états cachés de ses enfants x_{s+} , de l'observation y_{s-} liée au parent et des observations y_{s+} liées à ce enfants. La probabilité de transition $p(z_s/z_{s-})$ se transforme en utilisant les hypothèses d'indépendance du quadarbre de Markov caché comme suit :

$$\begin{aligned} p(z_s/z_{s-}) &= p(y_s, x_s/y_{s-}, x_{s-}) \\ &= p(y_s/x_s, y_{s-}, x_{s-})p(x_s/y_{s-}, x_{s-}) \\ &= p(y_s/x_s)p(x_s/x_{s-}) \end{aligned} \quad (1.44)$$

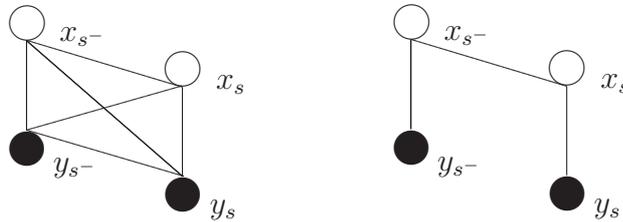


FIG. 1.21 – Différence entre l'arbre de Markov couple à gauche et l'arbre de Markov caché à droite. Dans l'arbre couple, l'observation y_s ne dépend plus que de son état caché x_s . Elle est également liée aux états cachés et aux observations du parent et des enfants.

Le critère MPM sur le quadarbre de Markov couple

L'algorithme consiste en deux passes sur le quadarbre. La passe montante a pour but le calcul des messages $\beta(x_s) = p(y_{>s}/z_s)$. Nous avons [116] :

$$\begin{aligned}
\beta(x_s) &= p(y_{>s}/z_s) \\
&= \sum_{x_t \in \Omega_x} \prod_{t \in s^+} p(x_t, y_t, y_{>t}/z_s) \\
&= \sum_{x_t \in \Omega_x} \prod_{t \in s^+} p(z_t, y_{>t}/z_s) \\
&= \sum_{x_t \in \Omega_x} \prod_{t \in s^+} p(y_{>t}/z_t, z_s) p(z_t/z_s) \\
&= \sum_{x_t \in \Omega_x} \prod_{t \in s^+} p(y_{>t}/z_t) p(z_t/z_s) \\
&= \prod_{t \in s^+} \sum_{x_t \in \Omega_x} \beta(x_t) p(z_t/z_s)
\end{aligned} \tag{1.45}$$

La probabilité de transition $p(z_s/z_{s-})$ (voir Fig. 1.20-b) est calculée en utilisant la règle de Bayes comme suit :

$$p(z_s/z_{s-}) = \frac{p(y_s, y_{s-}/x_s, x_{s-}) p(x_s, x_{s-})}{\sum_{x_s \in \Omega_x} p(y_{s-}/x_s, x_{s-}) p(x_s, x_{s-})} \tag{1.46}$$

où

$$\forall s \in S^n, \forall n \in \{0, \dots, R\}, p(y_s, y_{s-}/x_s = \omega_j, x_{s-} = \omega_i) \triangleq f_{i,j}^n(y_s, y_{s-}) \tag{1.47}$$

et

$$\forall s \in S^n, \forall n \in \{0, \dots, R\}, p(y_{s-}/x_s = \omega_j, x_{s-} = \omega_i) \triangleq f_{i,j}^n(y_{s-}) \tag{1.48}$$

représentent la vraisemblance bidimensionnelle et sa marginale requises pour cette passe. Une solution immédiate consiste à utiliser la distribution gaussienne bidimensionnelle, entièrement définie par son vecteur moyenne et sa matrice de covariance. Pour le cas de marginales non gaussiennes, l'expression analytique de la densité bidimensionnelle n'est le plus souvent pas disponible. Nous exposerons dans le chapitre 2 différentes manières de résoudre ce problème.

Ainsi la passe montante est donnée par :

$$\begin{aligned}
\beta(x_s) &= 1, s \in S^0, \\
\beta(x_s) &= \prod_{t \in s^+} \sum_{x_t \in \Omega_x} \beta(x_t) p(z_t/z_s); s \in (S - S^0)
\end{aligned} \tag{1.49}$$

La probabilité *a posteriori* à la racine $p(x_r/y)$ est obtenue à la fin de la passe montante :

$$p(x_r/y) = \frac{\beta(x_r) p(z_r)}{\sum_{x_r \in \Omega_x} \beta(x_r) p(z_r)} \tag{1.50}$$

avec $p(z_r) = p(x_r) p(y_r/x_r)$ la probabilité à la racine, $\pi_i = p(x_r = \omega_i)$ étant la probabilité *a priori* des étiquettes.

En utilisant les hypothèses d'indépendance associées au quadarbre de Markov couple, la marginale *a posteriori*, pour les échelles en dessous de la racine, peut être écrite sous la forme [116] :

$$\begin{aligned}
\forall s \in (S - S^R) & : \\
p(x_s/y) &= \sum_{x_{s^-} \in \Omega_x} p(x_s, x_{s^-}/y) \\
&= \sum_{x_{s^-} \in \Omega_x} p(x_{s^-}/y) p(x_s/x_{s^-}, y)
\end{aligned} \tag{1.51}$$

Cette récurrence nécessite le calcul de la probabilité de transition *a posteriori* $p(x_s/x_{s^-}, y)$. Nous avons :

$$\begin{aligned}
p(x_s/x_{s^-}, y) &= p(x_s/z_{s^-}, y_s, y_{>s}, y_{S/>s} \cup \{s, s^-\}) \\
&\propto p(z_s, y_{>s}/z_{s^-}, y_{S/>s} \cup \{s, s^-\}) \\
&\propto p(y_{>s}, z_s/z_{s^-}) \\
&\propto p(y_{>s}/z_s) p(z_s/z_{s^-})
\end{aligned} \tag{1.52}$$

ce qui donne :

$$p(x_s/x_{s^-}, y) = \frac{\beta(x_s) p(z_s/z_{s^-})}{\sum_{x_s \in \Omega_x} \beta(x_s) p(x_s, y_s/z_{s^-})} \tag{1.53}$$

Il est ainsi possible de calculer les lois marginales en chaque site, ce qui permet d'obtenir \hat{x} suivant le critère MPM. Les paramètres de l'arbre de Markov couple sont :

1. les paramètres Θ_x de l'*a priori* :
 - la probabilité *a priori* $\pi_i = P(x_r = \omega_i)$, $i = 1, \dots, K$
 - les probabilités jointes $p(i, j) = p(x_s = \omega_j, x_{s^-} = \omega_i)$
2. les paramètres Θ_y des vraisemblances qui dépendent du modèle choisi.

1.7 Conclusion

Une synthèse des modèles graphiques et des modèles markoviens en imagerie a été présentée, ce qui permet de situer et d'introduire les modèles utilisés par la suite, *i.e.*, le quadarbre de Markov caché et quadarbre de Markov couple. Ces modèles offrent intrinsèquement la possibilité de traitements de données multispectrales et/ou multirésolutions telles que celles issues d'une analyse en ondelettes. Nous avons montré comment il est possible de calculer les marginales nécessaires pour l'utilisation du critère MPM en deux passes sur le quadarbre. Cependant, nous avons passé sous silence à chaque fois l'expression analytique du terme d'attache aux données, ainsi que l'estimation des hyperparamètres du modèle. Les différentes possibilités de modélisation pour ce terme feront l'objet du prochain chapitre.

2 Réduction de dimension et lois non gaussiennes multivariées

Notations

Notation	Signification
X	Champ des étiquettes
Y	Champ des observations multispectrales
\mathbf{y}	Vecteur d'observations
\mathbf{t}	Vecteur de variables cachées (continues) associé à \mathbf{y} dans l'espace de projection
S	Ensemble de sites sur une grille, ou ensemble de nœuds dans un graphe
x_s	Réalisation de la variable aléatoire cachée au site $s \in S$
\mathbf{y}_s	Vecteur observations au site $s \in S$
\mathbf{t}_s	Vecteur de variables cachées au site $s \in S$ associé à \mathbf{y}_s dans l'espace de projection
W	Matrice de projection
A	Matrice de formation des données à partir des variables cachées
d	Dimension de l'espace original
r	Dimension de l'espace réduit
Σ_y	Matrice de covariance des vecteurs \mathbf{y}_i
λ_i	$i^{\text{ème}}$ valeurs propres d'une matrice
Λ	Matrice diagonale des valeurs propres
U	Matrice des vecteurs propres
\mathbf{w}_i	Colonne i de la matrice W
μ	Moyenne des vecteurs \mathbf{y}_i
σ	Écart type
π_i	Probabilité <i>a priori</i> de la composante i dans un mélange de lois
R_{ni}	La responsabilité <i>a posteriori</i> pour la composante i d'avoir généré le vecteur \mathbf{x}_n
$f(x)$	Densité de x
$\mathbf{E}(x)$	Espérance de x
$H(x)$	Entropie de x
C_i	$i^{\text{ème}}$ cluster d'images
d_{ij}	Distance entre les clusters i et j

2.1 Introduction

En fin du précédent chapitre nous avons introduit les modèles des arbres de Markov cachés et couples sans parler du terme d'attache aux données (*i.e.*, la vraisemblance). Dans le cadre de l'imagerie multispectrale, deux problèmes peuvent être rencontrés lors de la manipulation de ce terme. Le premier est lié au phénomène de Hughes qui se manifeste par la perte de précision sur l'estimation des paramètres avec l'augmentation de la dimension. Le second est la modélisation de lois de probabilités multivariées dans le cas non gaussien où les expressions analytiques ne sont généralement pas disponibles.

Ce chapitre est constitué de deux parties, la première expose les techniques de réduction de dimension et la deuxième les techniques permettant le calcul de lois multidimensionnelles non gaussiennes. Dans la première partie, on commencera par justifier la réduction de données, on présentera ensuite quelques techniques classiques et enfin trois schémas de réduction que nous avons développés dans le cadre de l'imagerie astronomique. Dans la deuxième partie, on exposera quelque lois monodimensionnelles et les techniques de calcul des lois multidimensionnelles non gaussiennes corrélées.

2.2 Nécessité de réduire le nombre de bandes

La classification d'images multidimensionnelles se heurte, lorsque le nombre de bandes spectrales est important, au problème connu sous le nom de *malédiction de la dimension* ou *phénomène de Hughes* [68]. En effet, à nombre de pixels fixe, l'ajout d'une bande spectrale s'accompagne par un accroissement du nombre de paramètres du modèle et donc inévitablement une baisse de précision de leur estimation entraînant une dégradation de la qualité des décisions prises. L'espace de plus en plus vide avec l'augmentation de la dimension (voir Fig. 2.1).

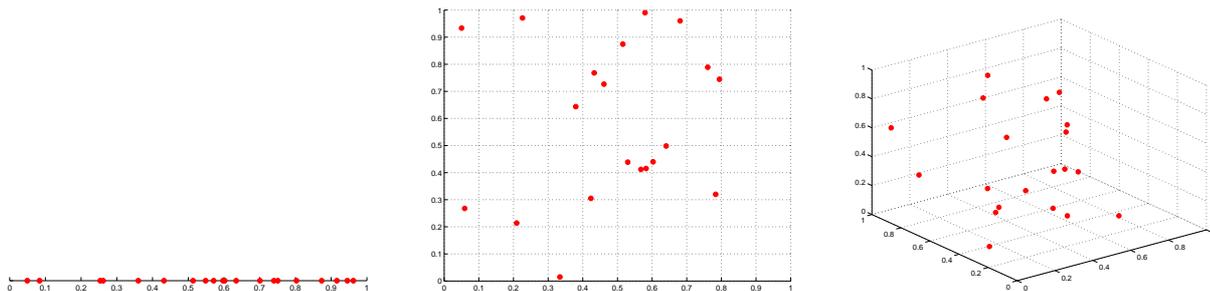


FIG. 2.1 – Illustration du problème de la malédiction de dimension. De gauche à droite, nous avons tiré suivant une loi uniforme 20 échantillons de dimension respectivement 1, 2 et en 3. En augmentant la dimension, à nombre d'échantillons fixe, l'espace présente des échantillons de plus en plus clairsemés

Plusieurs techniques de réduction de dimensionnalité ont été proposées en imagerie hyperspectrale de télédétection pour contourner ce problème[86, 2, 126, 91]. Le but est

donc de rechercher une projection vers un espace de plus faible dimension en optimisant un critère donné. Comme exemples de techniques des projections linéaires, on peut évoquer l'Analyse en Composantes Principales (ACP), l'Analyse Factorielle (AF), l'Analyse en Composantes Indépendantes (ACI) et l'Analyse Discriminante de Fisher (ADF) [45]. Pour les techniques non linéaires, on peut citer les deux méthodes de poursuite de projections proposée par Jimenez et *al.* [74] et Rellier et *al.* [127], l'ACP non linéaire [2] et le mélange de méthodes linéaires locales [78, 137, 58, 90].

Dans le cas où le nombre de bandes est suffisant (images hyperspectrales), une solution consiste à considérer chaque vecteur-pixel comme une courbe 1D qu'on approxime avec un modèle à paramètres réduits [103, 54]. Les paramètres du modèle en chaque vecteur-pixel codent efficacement la courbe associée donnant ainsi un ensemble d'images paramètres qu'on segmentera au lieu de l'image originale.

Nous présenterons dans les sections suivantes les techniques de réduction que nous avons utilisées dans les schémas de réduction des images astronomiques.

2.3 Quelques techniques de réduction de dimension

2.3.1 L'analyse en composantes principales

L'analyse en Composantes Principales (ACP) est sans doute la technique la plus utilisée en analyse de données multidimensionnelles [75]. Ayant un ensemble de vecteurs \mathbf{y}_i , $i = 1, \dots, N$, de dimension d , la méthode consiste à chercher les axes de projection orthogonaux suivant lesquels la variance est maximisée. L'approximation optimale, au sens de l'erreur quadratique moyenne, d'un vecteur \mathbf{y}_i par un vecteur $\hat{\mathbf{t}}_i$ de dimension $r < d$ est donnée par :

$$\hat{\mathbf{t}}_i = W_r^t (\mathbf{y}_i - \mu) \quad (2.1)$$

où μ est la moyenne des \mathbf{y}_i et W_r est la matrice de projection composée des r premiers vecteurs propres de la matrice de covariance des données Σ_y , correspondant aux r plus grandes valeurs propres données dans un ordre descendant $(\lambda_i)_{i=1, \dots, r}$. La matrice de covariance des données réduites est diagonale d'éléments $(\lambda_i)_{i=1, \dots, r}$.

L'erreur quadratique de l'approximation est donnée par la somme des valeurs propres écartées :

$$e^2 = \sum_{i=r+1}^d \lambda_i \quad (2.2)$$

Le choix de r peut être basé sur l'équation (2.2), ou d'une manière équivalente sur le choix d'un seuil p entre 0 et 1 tel que :

$$\frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^d \lambda_i} \geq p \quad (2.3)$$

On peut noter deux inconvénients majeurs de l'ACP. Le premier est l'absence d'un modèle génératif des données et d'une densité de probabilité associée (*i.e.*, vraisemblance). Le second est son caractère global qui suppose implicitement que la distribution des données

est un hyperellipsoïde caractérisé par sa moyenne et sa matrice de covariance globale [74]. Ainsi, elle peut provoquer la perte définitive d'informations caractérisant d'éventuelles structures locales des données.

Pour remédier au premier problème, Tipping et *al.* [138] ont introduit l'ACP probabiliste (ACPP) qui fera l'objet de la prochaine section. La solution au second problème est obtenue par le modèle de mélange d'ACPP qui sera abordé plus loin.

2.3.2 L'analyse en composantes principales probabiliste

L'ACP probabiliste (ACPP) se dérive d'un modèle de variables cachées, avec les hypothèses d'un bruit isotrope et un *a priori* gaussien [138]. Le vecteur observé \mathbf{y} de dimension d est généré à partir du vecteur caché \mathbf{t} de dimension r suivant l'expression :

$$\mathbf{y} = A\mathbf{t} + \mu + \epsilon \quad (2.4)$$

où A est une matrice de dimension $d \times r$, μ est la moyenne des données et ϵ est un bruit gaussien de moyenne nulle et de matrice de covariance $\sigma^2 I$, I étant la matrice identité $d \times d$.

La vraisemblance est alors donnée par :

$$p(\mathbf{y}/\mathbf{t}) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y} - A\mathbf{t} - \mu\|^2\right\} \quad (2.5)$$

Un *a priori* gaussien est choisi pour \mathbf{t} :

$$p(\mathbf{t}) = (2\pi)^{-\frac{r}{2}} \exp\left\{-\frac{1}{2}\mathbf{t}^t\mathbf{t}\right\} \quad (2.6)$$

ce qui donne la probabilité gaussienne du vecteur \mathbf{y} :

$$p(\mathbf{y}) = (2\pi)^{-\frac{d}{2}} |C|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu)^t C^{-1}(\mathbf{y} - \mu)\right\} \quad (2.7)$$

avec $C = \sigma^2 I + AA^t$ une matrice $d \times d$.

La probabilité *a posteriori* est donnée par :

$$p(\mathbf{t}/\mathbf{y}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{r}{2}} |M|^{\frac{1}{2}} \exp\left[\frac{-1}{2}\{\mathbf{t} - M^{-1}A^t(\mathbf{y} - \mu)\}^t (\sigma^{-2}M) \{\mathbf{t} - M^{-1}A^t(\mathbf{y} - \mu)\}\right] \quad (2.8)$$

avec :

$$M = \sigma^2 I + A^t A \quad (2.9)$$

une matrice $r \times r$.

La log-vraisemblance des données par rapport au modèle est :

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^N \ln\{p(\mathbf{y}_i)\} \\ &= -\frac{N}{2} \{d \ln(2\pi) + \ln |C| + \text{tr}(C^{-1}\Sigma_y)\} \end{aligned} \quad (2.10)$$

où $\Sigma_y = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n - \mu)(\mathbf{y}_n - \mu)^t$ est la matrice de covariance des données observées.

La maximisation de \mathcal{L} donne les estimations suivantes des paramètres [138] :

$$\begin{aligned}\mu_{ML} &= \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \\ \sigma_{ML}^2 &= \frac{1}{d-r} \sum_{j=r+1}^d \lambda_j \\ A_{ML} &= U_r (\Lambda_r - \sigma_{ML}^2 I)^{\frac{1}{2}} \mathfrak{R}\end{aligned}\tag{2.11}$$

où les λ_i sont les valeurs propres de Σ_y ordonnées en valeurs décroissantes ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$), Λ_r est la matrice diagonale des r premières valeurs propres, U_r est la matrice des r premiers vecteurs propres correspondants et \mathfrak{R} est une matrice de rotation orthogonale $r \times r$ quelconque.

Avec l'ACPP, le vecteur d'observation \mathbf{y}_i n'est plus caractérisé par un vecteur caché unique sur l'espace réduit, comme c'est le cas de l'ACP classique Eq. 2.1, mais par une distribution gaussienne *a posteriori* Eq. 2.8. Un estimateur du vecteur caché associé à \mathbf{y}_i est donc la moyenne de la loi *a posteriori* :

$$\hat{\mathbf{t}}_i = M_{ML}^{-1} A_{ML}^t (\mathbf{y}_i - \mu)\tag{2.12}$$

avec $M_{ML} = \sigma^2 I + A_{ML}^t A_{ML}$ (Eq. 2.9).

Nous disposons ainsi d'un modèle de génération des observations (Eq. 2.4) et une densité associée (Eq. 2.5) qui sert à mesurer l'adéquation d'une observation à ce dernier. Néanmoins, malgré cet apport le second inconvénient de l'ACP (*i.e.*, son caractère global) persiste toujours.

2.3.3 L'analyse en composantes indépendantes

L'analyse en composantes indépendantes (ACI) est souvent illustrée par le problème de la "cocktail party" : d personnes tiennent conversation dans un salon dans lequel d microphones d'enregistrement sont installés. Nous avons d signaux vocaux (sources) que l'on suppose indépendants et que l'on veut retrouver à partir des d enregistrements. La formulation mathématique du problème est la suivante [36, 70] :

$$\mathbf{y} = A\mathbf{t}\tag{2.13}$$

où \mathbf{y} est le vecteur des observations, A est la matrice de mélange à définir et \mathbf{t} sont les sources à déterminer, supposées indépendantes et non gaussiennes [36, 70]. Il faut noter deux ambiguïtés associées à ce modèle sur les puissances des sources et leur ordre.

Le théorème de la limite centrale affirme que la somme de variables aléatoires indépendantes tend, sous certaines conditions, vers la distribution gaussienne. De plus, la somme de deux

variables aléatoires indépendantes est généralement plus proche de la gaussienne que n'importe laquelle des deux variables [72]. Les observations qui sont des combinaisons linéaires des sources sont alors davantage gaussiennes que ces dernières, et l'estimation de $W = A^{-1}$ peut être basée sur la maximisation de la non gaussianité des sources. Une mesure de cette non gaussianité s'impose donc. La plus classique est le Kurtosis qui est le cumulants normalisé d'ordre 4 [70] :

$$Kurt(\mathbf{y}) = E\{\mathbf{y}^4\} - 3E\{\mathbf{y}^2\}^2 \quad (2.14)$$

Des valeurs strictement négatives du Kurtosis sont obtenues avec des distributions sous-gaussiennes (plus aplaties que la gaussienne) tandis que les distributions sur-gaussiennes (plus piquées que la gaussienne) donnent des valeurs strictement positives. Malgré sa simplicité, le Kurtosis a l'inconvénient d'être très sensible aux données aberrantes [67].

La seconde mesure trouve ses bases dans la théorie de l'information. Il est établi que pour un ensemble de variables aléatoire de même variance, celle qui suit la loi gaussienne maximise l'entropie donnée par :

$$H(\mathbf{y}) = - \int f_{\mathbf{y}}(\mathbf{u}) \log f_{\mathbf{y}}(\mathbf{u}) d\mathbf{u} \quad (2.15)$$

où $f_{\mathbf{y}}(\mathbf{y})$ est la densité de \mathbf{y} . Cette remarque permet d'introduire une mesure de la non gaussianité appelée négentropie [72] définie par :

$$J(\mathbf{y}) = H(\mathbf{y}_g) - H(\mathbf{y}) \quad (2.16)$$

où \mathbf{y}_g est un vecteur aléatoire gaussien de même covariance que \mathbf{y} . La négentropie est égale à la divergence de Kullback-Leibler entre $f_{\mathbf{y}}(\mathbf{y})$ et $f_{\mathbf{y}_g}(\mathbf{y}_g)$ [36]. Etant donnée la difficulté d'évaluer J d'après l'Eq. 2.16, cette quantité est en pratique approximée. Une méthode classique d'approximation est [76] :

$$J(\mathbf{y}) = \frac{1}{12} E\{\mathbf{y}^3\}^2 - \frac{1}{48} Kurt(\mathbf{y})^2 \quad (2.17)$$

Cet estimateur basé sur le kurtosis demeure peu robuste. Hyvärinen [71] a proposé une approximation robuste de la négentropie :

$$J(\mathbf{y}) = [E\{G(\mathbf{y})\} - E\{G(\mathbf{y}_g)\}]^2 \quad (2.18)$$

où G est une fonction non quadratique¹ puisqu'alors J serait trivialement nul pour toutes les distributions. La maximisation de J permet de retrouver la matrice de projection

¹Les choix suivants pour G sont avérés très pertinents [70] :

$$G_1(\mathbf{u}) = \frac{1}{a_1} \log \cosh(a_1 \mathbf{u}) ; 1 \leq a_1 \leq 2 \quad \text{et} \quad G_2(\mathbf{u}) = -\exp\left(-\frac{\mathbf{u}^2}{2}\right)$$

Leurs dérivées respectives sont :

$$g_1(\mathbf{u}) = \frac{1}{a_1} \tanh(a_1 \mathbf{u}) ; 1 \leq a_1 \leq 2 \quad \text{et} \quad g_2(\mathbf{u}) = \mathbf{u} \exp\left(-\frac{\mathbf{u}^2}{2}\right)$$

$$W = A^{-1}.$$

Comme l'indépendance implique la non corrélation, l'espace de recherche est restreint aux composantes décorréliées. Cela est mis en pratique grâce à une procédure de blanchiment des observations préalable à l'optimisation de J . Cette opération n'est autre qu'une ACP sous la contrainte d'une matrice de covariance identité pour les vecteurs transformés. Elle est donnée par :

$$\tau = U\Lambda^{-\frac{1}{2}}U^t\mathbf{y} \quad (2.19)$$

où τ est le vecteur blanchi, Λ est la matrice des valeurs propres de Σ_y et U est la matrice des vecteurs propres de Σ_y .

Hyvärinen [69] a proposé un algorithme robuste (Algo. 2.1, p. 39.), baptisé fastICA, pour l'optimisation des vecteur colonnes \mathbf{w}_i de W .

Algo. 2.1 Algorithme FastICA

- Centrage des données
- Blanchiment (Eq.2.19) pour obtenir τ
- Choix initial de \mathbf{w}_i unitaire (choix aléatoire par exemple)

Répéter

- Mise à jour de \mathbf{w}_i :

$$\mathbf{w}_i \leftarrow E\{\tau g(\mathbf{w}_i^t \tau)\} - E\{g'(\mathbf{w}_i^t \tau)\}\mathbf{w}_i$$

où g' est la dérivée de g .

- Normalisation de \mathbf{w}_i : $\mathbf{w}_i \leftarrow \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$

Jusqu'à convergence de \mathbf{w}_i

Afin d'empêcher les \mathbf{w}_i d'être estimés d'une façon identique, ces derniers doivent être décorréliés à chaque itération. Cela peut se faire de deux manières : la décorrélation symétrique ou la décorrélation déflationniste. La première consiste à faire une itération pour tous les \mathbf{w}_i en "parallèle" puis les décorréler simultanément. La seconde est basée sur la procédure d'orthogonalisation de Gram-Schmidt [37]. Les vecteurs \mathbf{w}_i sont estimés l'un après l'autre. Pour chaque vecteur \mathbf{w}_i , $i = 2, \dots, d$ une étape supplémentaire est ajoutée dans l'algorithme fastICA qui consiste à son orthogonalisation par rapport aux vecteurs précédemment estimés. Cela est réalisé par :

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \sum_{j=1}^{i-1} (\mathbf{w}_i^t \mathbf{w}_j) \mathbf{w}_j \quad (2.20)$$

L'algorithme fastICA avec décorrélation déflationniste est décrit par l'Algo. 2.2, p. 40. Cette manière de faire privilégie les premiers vecteurs estimés. Pour des fins de réduction, c'est cette méthode qui est utilisée plutôt que la première car on est ainsi assuré de garder les axes qui maximisent le plus la néguentropie. De ce point de vue cet algorithme peut être classé comme un algorithme de poursuite de projection très performant, dont l'index de projection est la néguentropie.

Algo. 2.2 Algorithme fastICA avec décorrélation déflationniste

- Centrage des données
 - Blanchiment (Eq.2.19) pour obtenir τ
- Pour tout** $i = 1, \dots, r$ **Faire**
- Choix initial de \mathbf{w}_i aléatoirement par exemple

Répéter

- Mise à jour de \mathbf{w}_i :

$$\mathbf{w}_i \leftarrow E\{\tau g(\mathbf{w}_i^t \tau)\} - E\{g'(\mathbf{w}_i^t \tau)\} \mathbf{w}_i$$

où g' est la dérivée de g .

- Si $i > 1$, faire l'orthogonalisation suivant l'Eq. 2.20

- Normalisation de \mathbf{w}_i : $\mathbf{w}_i \leftarrow \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$

Jusqu'à convergence de \mathbf{w}_i

Fin Pour

Néanmoins, il est légitime de s'interroger sur l'utilité de la maximisation de la négentropie. En effet la distribution normale est souvent jugée la moins intéressante, et les axes de projections les plus intéressants sont ceux qui montrent le moins cette distribution. Un premier argument est que les distributions multimodales montrent des structures de clusters. Un autre est que la gaussienne maximise l'entropie qui est une mesure du désordre ou du manque de structure des données analysées, chose qu'on cherche à éviter [70].

2.3.4 Le mélange d'analyseurs en composantes principales probabilistes

L'ACPP introduite par Tipping et *al.* [138] ne parvient à palier que le premier inconvénient de l'ACP, *i.e.*; l'absence de modèle statistique associé. Le deuxième (*i.e.*, son caractère global) restant toujours posé, Tipping et *al.* ont considéré un mélange d'ACPP locales, chacune d'elles modélisant efficacement une partie de l'espace (un cluster) de sorte que la génération de chaque observation soit partagée par tous les modèles locaux. On aboutit ainsi à un modèle global non linéaire. La vraisemblance de chaque observation est donnée par :

$$p(\mathbf{y}_n) = \sum_{i=1}^P \pi_i p(\mathbf{y}_n/i) \quad (2.21)$$

où P est le nombre de composantes dans le mélange, π_i l'*a priori* de chaque composante et $P(\mathbf{y}_n/i)$ une ACPP associée à cette composante caractérisée par la moyenne μ_i , la variance du bruit σ_i^2 et la matrice A_i .

La responsabilité *a posteriori* de la composante i d'avoir généré le vecteur \mathbf{y}_n est définie par :

$$R_{ni} = p(i/\mathbf{y}_n) = \frac{p(\mathbf{y}_n/i)\pi_i}{p(\mathbf{y}_n)} \quad (2.22)$$

L'algorithme EM est utilisé pour l'estimation des différents paramètres de ce modèle

[137], ce qui donne les mises à jour suivantes :

$$\pi_i^{[q+1]} = \frac{1}{d} \sum_{n=1}^N R_{ni}^{[q]} \quad (2.23)$$

$$\mu_i^{[q+1]} = \frac{\sum_{n=1}^N R_{ni}^{[q]} \mathbf{y}_n}{\sum_{n=1}^N R_{ni}^{[q]}} \quad (2.24)$$

$A_i^{[q+1]}$ et $(\sigma_i^{[q+1]})^2$ sont obtenues grâce à l'Eq. 2.11 par décomposition en valeurs propres de la matrice de covariance pondérée :

$$(\Sigma_y)_i^{[q+1]} = \frac{1}{\pi_i^{[q+1]} N} \sum_{n=1}^N R_{ni}^{[q]} (\mathbf{y}_n - \hat{\mu}_i)(\mathbf{y}_n - \hat{\mu}_i)^t \quad (2.25)$$

Cette méthode présente l'avantage d'estimer la partition de l'espace en même temps que les ACP locales les modélisant. Cependant son utilisation directe en imagerie multispectrale n'est pas bien adaptée car elle ne tient pas compte de l'information de voisinage entre pixels qui est très importante pour la régularisation des résultats de traitement. Nous verrons plus loin comment cela peut être réalisé dans le cadre de la segmentation d'images multispectrales.

2.4 Schémas de réduction adoptés

2.4.1 Regroupement de bandes et projections locales

Les bandes spectrales dont les longueurs d'ondes sont proches, sont généralement très corrélées, et l'information qu'elles apportent est redondante. En se basant sur ce constat, nous avons adopté une technique [49] qui consiste à 1) regrouper les images (bandes spectrales) *similaires* formant ainsi un certain nombre de groupes d'images ; 2) réaliser une projection locale au sein de chaque groupe (Fig. 2.2).

Le groupement des bandes spectrales se base sur un algorithme de coalescence de bas-en-haut (*cf.* Fig. 2.3), démarrant avec un nombre de groupes égal au nombre de bandes, et groupant à chaque itération les deux bandes les plus proches au sens d'une mesure de similarité multirésolution basée sur les histogrammes normalisés [31] combinés avec les moments d'inerties d'ordre un (barycentres) [62]. L'histogramme normalisé h_n (appelé aussi probabilité heuristique) d'une image est déduit de l'histogramme classique h par division par le nombre de pixels dans l'image. La distance entre deux images 1 et 2 est calculée suivant l'algorithme Algo. 2.3, p. 42.

A chaque fois que deux clusters C_i et C_j sont regroupés pour former le cluster k , la distance entre celui-ci et tous les autres clusters C_h , $h \in \{1, \dots, k-1\} - \{i, j\}$, est calculée par [31] :

$$d_{hk} = d_{hi} + d_{hj} - d_{ij} \quad (2.27)$$

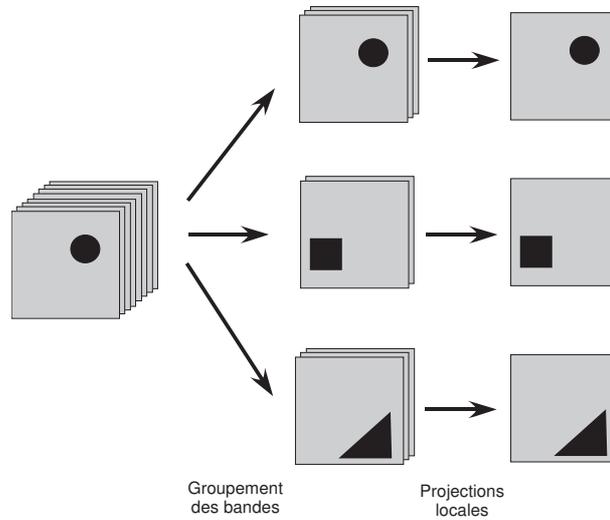


FIG. 2.2 – Regroupement de bandes et projections locales : les images correspondantes aux différentes bandes spectrales sont regroupées suivant une mesure de similarité multirésolution. Les groupes ainsi définis sont réduits avec des projections par ACI ou ACP.

Algo. 2.3 Calcul de la distance entre deux images

- Décomposer les deux images en échelle par filtrage passe bas et décimation

Pour tout niveau de résolution i **Faire**

- Calculer les histogrammes normalisés h_1^i et h_2^i pour les deux images
- Calculer les barycentres g_1^i et g_2^i pour les deux images
- Calculer la divergence entre h_1^i et h_2^i par [31] :

$$D_{12}^i = (h_1^i - h_2^i) \log \frac{h_1^i}{h_2^i} \quad (2.26)$$

- Calculer la distance euclidienne l_{12}^i entre g_1^i et g_2^i

Fin Pour

- Sommer D_{12}^i et l_{12}^i sur tous les niveaux de résolution pour obtenir D_{12} et l_{12}
 - Normaliser D_{12} et l_{12} par leurs maximums respectifs obtenus sur toutes les paires d'images. On obtient ainsi D_{n12} et l_{n12}
 - Sommer D_{n12} et l_{n12} pour avoir la distance d_{12} entre les images 1 et 2
-

où d_{hi} est la distance entre les clusters C_h et C_i . Il existe plusieurs façons de calculer d_{hk} , celle retenue (Eq. 2.27) donne la structure en arbre² la plus équilibrée pour les données [84].

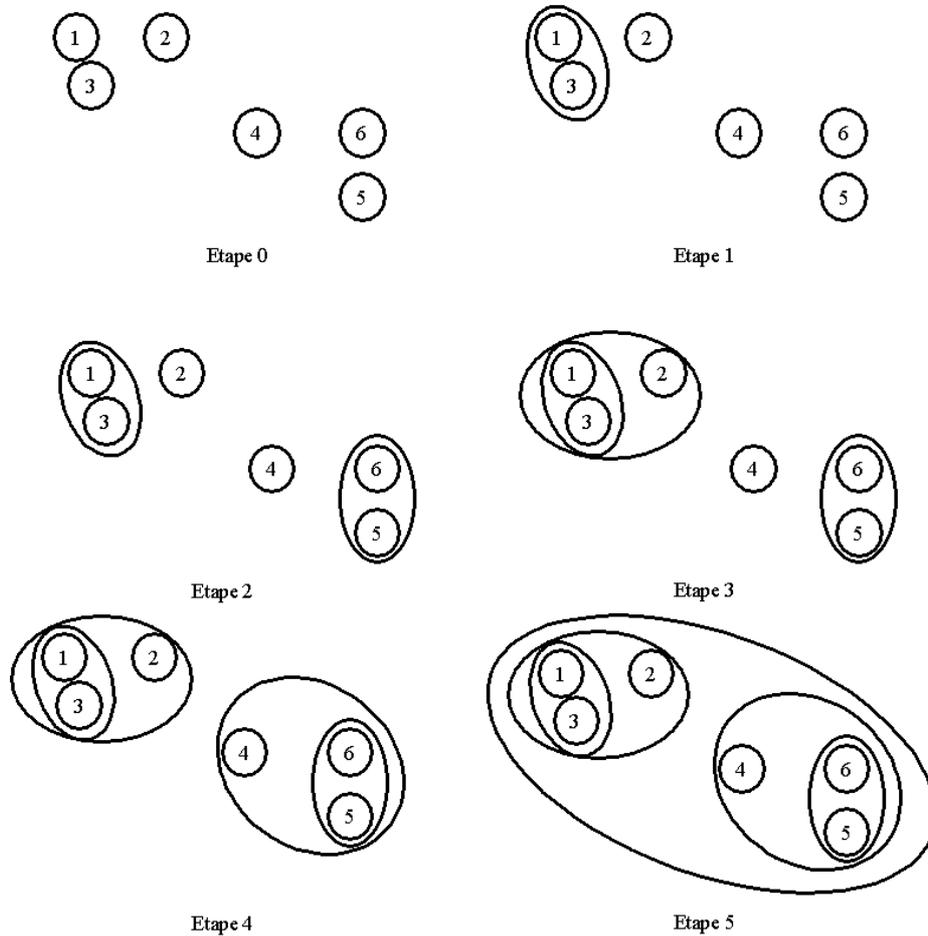


FIG. 2.3 – Illustration de l’algorithme de coalescence de bas en haut.

Une fois les groupes de bandes établis, nous projetons les images de chaque groupe en utilisant une technique linéaire (ACP ou ACI). Nous obtenons ainsi une projection globale non linéaire.

Deux paramètres sont à déterminer par l’utilisateur, à savoir le nombre de groupes et le nombre d’images gardées après réduction au sein de chaque groupe. Le premier peut être fixé par la contrainte sur le classifieur, *i.e.*, le nombre de bandes supportées par ce dernier, ou en imposant une limite supérieure de la distance utilisée lors du groupement. Le deuxième peut être estimé à partir de l’Eq. 2.2 utilisée localement sur chaque groupe. Cette technique sera utilisée pour la réduction d’un cube de 12 bandes spectrales pour la

²Il s’agit de l’arbre construit par l’algorithme de coalescence, les feuilles étant les images aux différentes bandes.

détection de galaxies à faible rayonnement (Sect. 4.3.1, p. 87).

2.4.2 Réduction des cubes radio par mélange de gaussiennes

Les techniques de réduction se heurtent souvent au problème de la signification physique des images réduites. En effet l'application d'une ACP sur un cube d'images (image multi-composantes), ressort un cube réduit maximisant la variance. Cependant prise une par une, les images de ce cube n'ont pas forcément de signification physique. Plus flagrant, le cas de l'ACI où le cube réduit est sensé représenter des images sources dont la combinaison linéaire a fourni le cube observé. Or en pratique, l'existence de telles images sources est difficilement vérifiée, et même dans le cas où des sources existent dans la scène, le modèle doit être local et non pas global pour tous les sites de l'image [109]. Pour les cubes de données radioastronomiques, les astronomes utilisent d'habitude dans leurs analyses quelques paramètres du spectre en chaque site, à savoir la carte d'intensité (moment 0), le champs de vitesse (moment 1) et la distribution de la dispersion de la vitesse (moment 2). En effet, ce type de cubes correspond le plus souvent à une observation des variations d'une raie unique (e.g., une raie de Co ou HI), autour de sa longueur d'onde de référence λ_r , causées par le mouvement radial du gaz observé (effet Doppler). La relation entre la longueur d'onde observée, λ_o , de la raie et sa vitesse est calculée comme suit :

$$v = \frac{\lambda_o - \lambda_r}{\lambda_r} c \quad (2.28)$$

où c est la vitesse de la lumière. Le but est alors d'étudier le champ de vitesses. L'utilisation des trois moments revient à modéliser chaque spectre par une gaussienne de moyenne, amplitude et variance égales respectivement aux moments d'ordre 0, 1 et 2. Nous avons généralisé ce concept en modélisant chaque spectre par un mélange de P composantes gaussiennes. Quelques paramètres seulement, $3 \times P$, caractérisent le spectre représenté initialement avec un vecteur de $d \gg P$ éléments.

Soit une image multispectrale avec d bandes définie sur une grille régulière S de taille $H \times W$ pixels. Cette image peut être vue comme $H \times W$ spectres \mathbf{y}_s , $s \in S$, de taille d . Chaque spectre est modélisé par un mélange de P -composantes gaussiennes comme suit :

$$\hat{\mathbf{y}}_s(j) = \sum_{k=1}^P \frac{a_k(s)}{\sqrt{2\pi\sigma_k^2(s)}} \exp\left(-\frac{1}{2\sigma_k^2(s)}(j - \mu_k(s))^2\right) \quad (2.29)$$

où $a_k(s)$ est le poids, $\sigma_k^2(s)$ la variance, $\mu_k(s)$ la moyenne associés à la $k^{ième}$ composante gaussienne du $s^{ième}$ spectre, et j est le numéro de la bande spectrale.

L'estimation des paramètres de ce modèle est réalisée grâce à une adaptation de l'algorithme EM (Expectation Maximation). L'EM est une méthode itérative pour l'estimation des paramètres par maximisation de la vraisemblance. Il a de nombreuses applications [99, 14], mais parmi toutes, l'estimation des paramètres du mélange de densités gaussiennes est sans doute la plus largement utilisée dans la communauté de la reconnaissance de formes par méthodes statistiques. Cette application consiste à approximer une densité de probabilité de données observées par un mélange de P lois gaussiennes [40, 14, 99].

D'habitude, l'algorithme EM utilise des réalisations d'une certaine densité. Dans notre cas on suppose que le spectre lui même représente une densité. Ainsi, une petite adaptation de l'EM s'impose. Pour satisfaire aux hypothèses, chaque spectre \mathbf{y}_s est normalisé pour avoir la forme d'une densité, *i.e.* $\mathbf{y}_s(j) \geq 0, \forall j$ et $\sum_{j=1}^d \mathbf{y}_s(j) = 1$.

L'analogie avec l'EM classique, nous permet de trouver les mises à jour des paramètres du modèle de mélange :

$$a_k^{[q+1]}(s) = \frac{\sum_{j=1}^d \mathbf{y}_s(j) \mathcal{R}^{[q]}(k, s, j)}{\sum_{j=1}^d \mathbf{y}_s(j)} \quad (2.30)$$

$$\mu_k^{[q+1]}(s) = \frac{\sum_{j=1}^d j \mathbf{y}_s(j) \mathcal{R}^{[q]}(k, s, j)}{\sum_{j=1}^d \mathbf{y}_s(j) \mathcal{R}^{[q]}(k, s, j)} \quad (2.31)$$

$$\sigma_k^{2[q+1]}(s) = \frac{\sum_{j=1}^d (j - \mu_k^{[q+1]}(s))^2 \mathbf{y}_s(j) \mathcal{R}^{[q]}(k, s, j)}{\sum_{j=1}^d \mathbf{y}_s(j) \mathcal{R}^{[q]}(k, s, j)} \quad (2.32)$$

où la contribution de la $k^{\text{ième}}$ composante dans le mélange est donnée par : $\mathcal{R}^{[q]}(k, s, j)$, est calculée comme suit :

$$\mathcal{R}^{[q]}(k, s, j) = \frac{\frac{a_k^{[q]}(s)}{\sqrt{2\pi\sigma_k^{2[q]}(s)}} \exp\left(\frac{-1}{2\sigma_k^{2[q]}(s)}(j - \mu_k^{[q]}(s))^2\right)}{\sum_{k=1}^P \frac{a_k^{[q]}(s)}{\sqrt{2\pi\sigma_k^{2[q]}(s)}} \exp\left(\frac{-1}{2\sigma_k^{2[q]}(s)}(j - \mu_k^{[q]}(s))^2\right)}, \quad (2.33)$$

q indique le numéro de l'itération courante (Fig. 2.4-a). Cette méthode suppose que les poids $a_k(s)$ sont tous positifs.

La modélisation avec l'Eq. 2.29 requiert 3 paramètres pour chaque composante gaussienne. On aboutit ainsi en chaque pixel à $3 \times P$ paramètres au lieu du vecteur original à d échantillons spectraux. Cependant, à cause de la limitation du nombre de bandes liée au phénomène de Hughes, nous ne pouvons pas dépasser une dizaine d'images paramètres ce qui donnerait une modélisation avec un mélange de 3 composantes au plus. Ceci risque de conduire à une approximation trop grossière du spectre. Pour contourner cette difficulté, nous avons opté pour l'utilisation d'un nombre important de gaussiennes choisies au préalable avec des moyennes et des variances fixes (*indépendants de s*). Nous n'aurons ainsi qu'à optimiser les poids, ce qui revient à tripler le nombre de composantes dans le mélange pour le même nombre de paramètres. Le choix des gaussiennes est cependant délicat. Une possibilité est de sélectionner dans l'ensemble des gaussiennes $\Xi = \{(\mu_k(s), \sigma_k^2(s)); k = 1, \dots, P; s \in S\}$, obtenues à l'issue de l'application de l'algorithme EM sur le modèle de l'Eq. 2.29, celles qui se révèlent les plus représentatives. Pour cette tâche nous avons utilisé l'algorithme des K-moyennes [45] sur l'ensemble Ξ pour extraire P clusters et prendre les centroïdes associés comme gaussiennes les plus pertinentes (Fig. 2.4-(b)). Le modèle de l'Eq. 2.29 devient :

$$\hat{\mathbf{y}}_s(j) = \sum_{k=1}^P \frac{\tilde{a}_k(s)}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-1}{2\sigma_k^2}(j - \mu_k)^2\right) \quad (2.34)$$

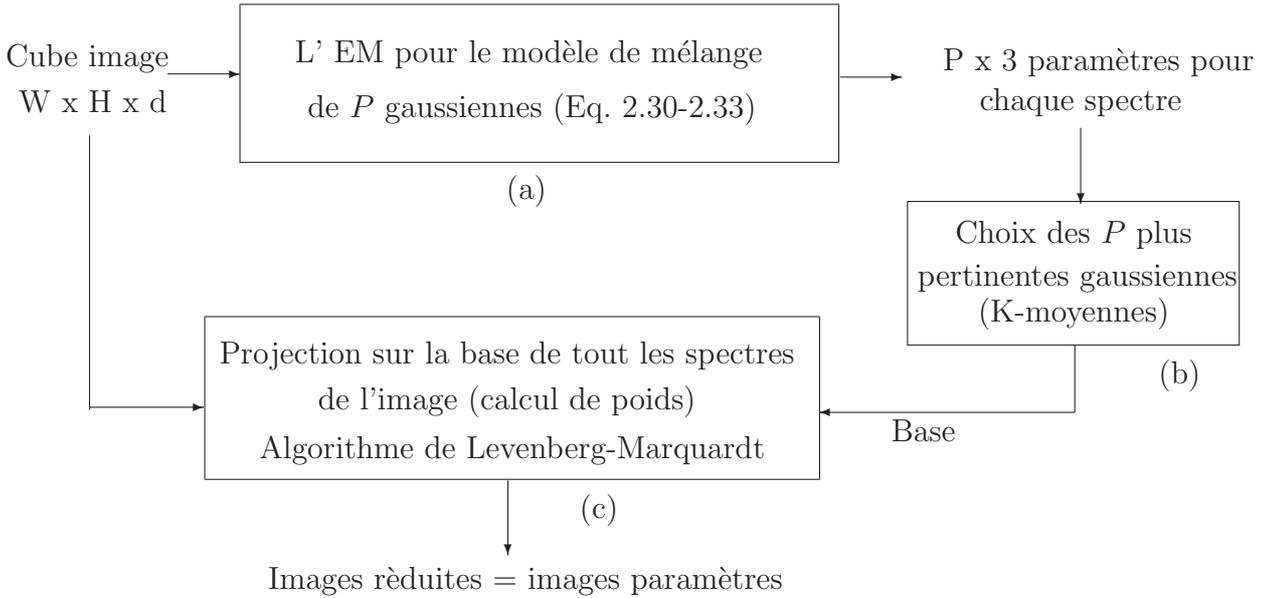


FIG. 2.4 – Réduction par mélange de gaussiennes des cubes radio

où μ_k et σ_k^2 ne dépendent plus de la position s . Ceci est équivalent à la projection des spectres sur le sous espace généré par la base $\left\{ \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-1}{2\sigma_k^2}(j - \mu_k)^2\right) \right\}_{k=1, \dots, P}$.

Cette projection, *i.e.*, le calcul des $(\tilde{a}_k(s))_{k=1, \dots, P; s \in S}$, peut être réalisée grâce à l'algorithme de Levenberg-Marquardt ou les moindres carrés [140] (Fig. 2.4-(c)). Ainsi nous obtenons enfin une image multi-composantes \mathcal{A} avec P composantes de taille $H \times W$ où chaque pixel s est représenté par un vecteur de P éléments $\mathcal{A}_s = (\tilde{a}_k(s))_{k=1, \dots, P}$. Cette technique sera utilisée pour la réduction du cube radio à 48 bandes de la raie HI de la galaxie spirale NGC4254 Sect. 4.3.2, p. 93.

2.4.3 Régularisation du mélange d'analyseurs en composantes principales probabilistes

Le mélange d'ACPP (Sect. 2.3.4) est un outil puissant et bien formulé mathématiquement pour l'analyse des données complexes présentant plusieurs structures locales dans l'espace à d dimensions. Cependant, son utilisation directe en imagerie multi-composantes revient à traiter les pixels de l'image sans tenir compte de leurs positions sur la grille. Les clusters établis présentent une solution non lisse au problème de segmentation. En effet l'*a priori* utilisé dans l'Eq. 2.21 ne prend pas en compte la position du vecteur (pixel). En modèles graphiques, ceci est équivalent à avoir un processus caché X indexé sur la grille S de l'image tel que chaque vecteur (pixel) \mathbf{y}_s est lié uniquement à son état caché x_s sans aucune liaison entre les variables cachées. Nous proposons de régulariser la MACPP par la liaison des états cachés suivant un modèle adapté à l'image. Nous pouvons ainsi utiliser les chaînes de Markov sur un parcours fractal, les champs de Markov ou les modèles hiérarchiques tels que l'arbre de Markov caché.

La probabilité de chaque observation \mathbf{y}_s devient :

$$p(\mathbf{y}_s) = \sum_{i=1}^K p(x_s = \omega_i) p(\mathbf{y}_s/x_s = \omega_i) \quad (2.35)$$

où chaque classe w_i est représentée par une ACPD locale, et X un processus de Markov caché modélisant les dépendances entre pixels voisins et permettant ainsi de régulariser la solution du mélange d'ACPD.

Nous pouvons alors écrire la probabilité jointe des états cachés et des observations comme suit :

$$p(x, \mathbf{y}) \propto \prod_{c \in \mathcal{C}} \Psi_c(x) \prod_{s \in \mathcal{S}} p(\mathbf{y}_s/x_s) \quad (2.36)$$

avec $p(\mathbf{y}_s/x_s)$ est le terme d'attache aux données au site s et $\Psi_c(x)$ est la fonction locale associée à la clique c . D'une manière équivalente, nous pouvons écrire :

$$p(x, \mathbf{y}) \propto \exp \left(- \sum_{c \in \mathcal{C}} V_c(x) - \sum_{s \in \mathcal{S}} l_s(x_s, \mathbf{y}_s) \right) \quad (2.37)$$

où $l_s(x, \mathbf{y}_s) = -\ln p(\mathbf{y}_s/x_s)$ et $V_c(x) = -\ln \Psi_c(x)$ est le potentiel associé à la clique c . Ce dernier est calculé suivant le modèle de Potts pour un champs de Markov et correspond aux probabilités de transitions pour un arbre ou une chaîne.

Les paramètres de la MACPD régularisée se répartissent en deux ensembles à savoir les paramètres de l'*a priori* Θ_x qui dépendent du modèle markovien choisi, et les paramètres de l'attache aux données Θ_y qui sont ici les caractéristiques des ACPD locales associées aux classes. Il est clair d'après l'Eq. 2.37 que l'optimisation de ces deux ensembles de paramètres peut être découplée. Ainsi, il est possible d'adapter aisément l'algorithme EM de Tipping et al. [137] en changeant les probabilités *a priori* par les probabilités *a posteriori* en chaque site pour l'estimation des Θ_y . L'estimation des paramètres de l'*a priori* Θ_x dépend du modèle Markovien choisi (pour le cas de l'arbre Markov caché, voir le chapitre 3).

La probabilité $p(\mathbf{y}_s/x_s = \omega_i)$ est calculée comme la vraisemblance de \mathbf{y}_s par rapport à l'ACPD locale correspondant à la classe ω_i . Comme l'Eq. 2.7 le suggère, cette vraisemblance est donnée par :

$$P(\mathbf{y}_s/x_s = \omega_i) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y}_s - \mu_i)^t C_i^{-1} (\mathbf{y}_s - \mu_i) \right\} \quad (2.38)$$

La matrice $C_i = \sigma_i^2 I + A_i A_i^t$ est obtenue de manière analogue à l'Eq. 2.11 par décomposition en valeurs propres de la matrice de covariance pondérée :

$$\Sigma_i = \frac{\sum_{s \in \mathcal{S}} P(x_s = \omega_i/\mathbf{y}) (\mathbf{y}_s - \hat{\mu}_i) (\mathbf{y}_s - \hat{\mu}_i)^t}{\sum_{s \in \mathcal{S}} P(x_s = \omega_i/\mathbf{y})} \quad (2.39)$$

avec

$$\hat{\mu}_i = \frac{\sum_{s \in S} P(x_s = \omega_i / \mathbf{y}) \mathbf{y}_s}{\sum_{s \in S} P(x_s = \omega_i / \mathbf{y})} \quad (2.40)$$

Il est vrai que le modèle de la MACPP régularisée revient à utiliser une gaussienne multidimensionnelle pour le terme d'attache aux données (Eq.2.38). Cependant, il présente deux avantages par rapport à la modélisation gaussienne classique :

1. **Covariance à nombre de paramètres réduit** : La matrice de covariance d'une loi gaussienne d -dimensionnelle requière en général $d(d+1)/2$ paramètres indépendants, tandis que pour la matrice de covariance $C_i = \sigma_i^2 I + A_i A_i^t$, associée à la densité gaussienne multidimensionnelle issue de l'ACPP de la classe ω_i (Eq. 2.38), le nombre de paramètres indépendants est $d r_i + 1 - r_i(r_i - 1)/2$ où r_i est la dimension réduite de l'ACCP et le terme $r_i(r_i - 1)/2$ correspond au nombre de paramètres nécessaire pour spécifier la rotation \mathfrak{R}_i associée à A_i (Eq. 2.11) . Ce nombre dépend donc du choix de r_i : on retrouve ainsi le cas gaussien classique pour $r_i = d - 1$ et le cas d'une gaussienne isotrope pour $r_i = 0$ [137].
2. **Richesse en information** : En effet, la MACPP présente deux types de variables cachées qui sont le processus discret X des étiquettes et les vecteurs cachés \mathbf{t}_s^i qui représentent les projections des observations \mathbf{Y}_s dans l'espace de dimension réduite r_i (Eq. 2.4) grâce à l'ACPP de la classe ω_i . En se basant sur l'Eq. 2.2 appliquée pour chaque ACCP locale, nous pouvons déterminer la dimension locale de chaque classe. La somme de ces dimensions locales donne la dimension réduite globale. Nous aboutissons ainsi à un ensemble d'images réduites représentatives du cube initial. Plus loin encore, dans le cas où le cube initial représente une image multispectrale, nous pouvons récupérer les spectres propres de chaque classe ce qui constitue une signature de celle-ci. Ceci est très utile dans le cas où les classes correspondent à des réalités physiques telles que des objets astronomiques, des types de sol ou des étendues d'eau.

Notons enfin que cette technique peut être étendue à d'autres mélanges de projections locales dès qu'une modélisation statistique est disponible. On peut ainsi avoir le mélange d'analyseurs factoriaux régularisé basé sur le travail de Ghahramani et *al.* [58], ou le mélange d'analyseurs en composantes indépendantes basé sur le travail de Roberts et *al.* [129]. Nous détaillerons dans le chapitre 3 le cas de la MACPP régularisée par un quadarbre de Markov caché Sect. 3.5, p. 75.

2.5 Modélisation de l'attache aux données

Le choix du terme d'attache aux données est un élément important dans tout modèle Markovien car il modélise la dispersion des données observées dans chaque classe. Dans le cas de lois monodimensionnelles, le large éventail de lois existantes dans la littérature offre une grande liberté dans la sélection de la loi appropriée. Cependant dans le cas multidimensionnel, mis à part quelques cas spéciaux, nous ne disposons pas d'expressions analytiques des densités de probabilités. Ceci explique en partie la popularité de la loi gaussienne multivariée qui dispose d'une formule analytique élégante et d'estimateurs aux sens du maximum de vraisemblance assez simple à mettre en œuvre. Malheureusement, le modèle gaussien n'est pas toujours valable ce qui implique le passage par des techniques

particulières de calcul pour retrouver les densités. Après avoir présenté deux lois monodimensionnelles qui seront utilisées par la suite, nous énumérerons différentes possibilités de modélisations de lois multivariées.

2.5.1 Quelques densités de probabilités monovariées

Loi gaussienne

La loi gaussienne monodimensionnelle est de loin la densité la plus utilisée à cause de ces propriétés analytiques très intéressantes d'une part, et du théorème de la limite centrale d'autre part. Une variable aléatoire gaussienne $l \in \mathbb{R}$ a pour densité :

$$f(l) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left(-\frac{(l-\mu)^2}{2\sigma^2}\right) \quad (2.41)$$

Elle est complètement déterminée par ses deux paramètres que sont la moyenne μ et la variance σ^2 . Ils sont estimés au sens du maximum de vraisemblance à partir de N échantillons $l_i, i = 1, \dots, N$, comme suit :

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N l_i \quad (2.42)$$

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (l_i - \hat{\mu})^2 \quad (2.43)$$

Malgré la facilité de sa manipulation, la loi gaussienne n'est pas adaptée pour de nombreuses applications. Il est alors utile de disposer d'autres densités de probabilités. La gaussienne généralisée, qui comme son nom l'indique est une généralisation de la loi gaussienne, sera abordée à la section suivante.

Loi gaussienne généralisée

Il s'agit d'une famille de lois qui constitue une extension de la loi gaussienne. Elle est caractérisée par trois paramètres : la moyenne μ , la variance σ^2 et le paramètre de forme γ . C'est ce dernier paramètre qui permet de couvrir en plus de la loi gaussienne ($\gamma = 2$), des lois piquées dites sur-gaussiennes ($\gamma < 2$) et des lois aplaties dites sous-gaussiennes ($\gamma > 2$) (Fig. 2.5).

Une variable aléatoire gaussienne généralisée $l \in \mathbb{R}$ a pour densité [139, 121] :

$$f(l) = \frac{\eta(\gamma)\gamma}{[2\Gamma(1/\gamma)]} \exp[-(\eta(\gamma)|l-\mu|)^\gamma] \quad (2.44)$$

avec $\eta(\gamma) = \left[\frac{\Gamma(3/\gamma)}{\sigma^2\Gamma(1/\gamma)}\right]^{\frac{1}{2}}$ et $\Gamma(\gamma) = \int_0^\infty z^{\gamma-1} \exp(-z) dz$ est la fonction Gamma.

L'estimation des paramètres de la gaussienne généralisée n'est pas aussi simple que dans le cas gaussien [139, 121]. L'utilisation de la méthode des moments, avec N échantillons $l_i, i = 1, \dots, N$, donne les mêmes expressions que celles des Eqs. 2.42 et 2.43 respectivement

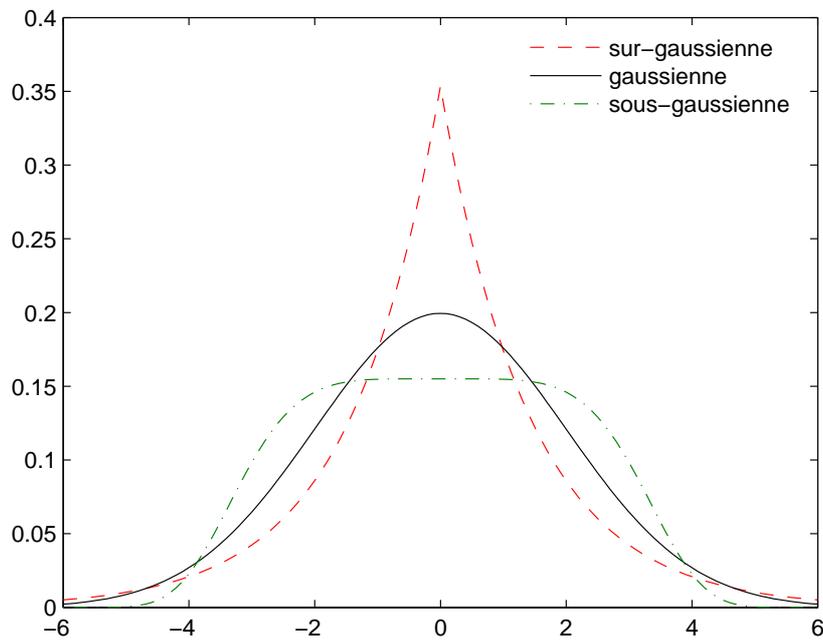


FIG. 2.5 – Tracé de la densité de probabilité gaussienne généralisée de moyenne nulle et de variance 2 pour différentes valeurs du paramètre de forme : sur-gaussienne ($\gamma = 1$), gaussienne ($\gamma = 2$) et sous-gaussienne ($\gamma = 5$).

pour la moyenne et la variance. Le paramètre de forme peut être retrouvé en résolvant numériquement l'équation ci-après obtenue avec le moment d'ordre 4 [139, 121] :

$$\hat{m}^{(4)} = \frac{1}{N} \sum_{i=1}^N (l_i - \hat{\mu})^4 = \sigma^4 \frac{\Gamma(\frac{5}{\hat{\gamma}}) \Gamma(\frac{1}{\hat{\gamma}})}{\Gamma^2(\frac{3}{\hat{\gamma}})} \quad (2.45)$$

Néanmoins, cette méthode n'est pas efficace [139]. Provost et *al.* ont proposé dans [121] une méthode d'estimation hybride consistant à estimer le paramètre de forme et la variance par maximum de vraisemblance et la moyenne par la méthode des moments. Les estimateurs correspondants sont les suivants :

$$\begin{aligned} \hat{\mu} &= \frac{1}{N} \sum_{i=1}^N l_i \\ \hat{\gamma} = \gamma &: \gamma + \Psi(1/\gamma) + \log(\gamma/M) + \log G_\gamma - \gamma \frac{G'_\gamma}{G_\gamma} = 0 \\ \hat{\sigma}^2 &= \frac{\Gamma(3/\hat{\gamma})}{\Gamma(1/\hat{\gamma})} \left[\left(\frac{\hat{\gamma}}{N} \right)^{\frac{1}{\hat{\gamma}}} G_{\hat{\gamma}} \right]^2 \end{aligned} \quad (2.46)$$

où Ψ est la fonction digamma : $\Psi(l) = \frac{\partial}{\partial l} \int_0^\infty z^{l-1} \exp(-z) dz$, G_γ est la γ -norme à la puissance N : $G_\gamma = \sum_{i=1}^N |l_i - \hat{\mu}|^\gamma$ et G'_γ sa dérivée par rapport à γ : $G'_\gamma = \frac{\partial G_\gamma}{\partial \gamma}$.

Cette technique d'estimation s'est avérée performante pour la classification non supervisée en imagerie SPOT [120].

2.5.2 Loi jointe de variables aléatoires indépendantes

Soient les vecteurs d'observations $\mathbf{y}_s = (\mathbf{y}_s^1, \dots, \mathbf{y}_s^d)^t$. Si nous considérons que les éléments $\mathbf{y}_s^1, \dots, \mathbf{y}_s^d$ du vecteur \mathbf{y}_s sont indépendants les uns des autres, alors la densité jointe $f(\mathbf{y}_s)$ est donnée par le produit simple des lois monodimensionnelles $f^i(\mathbf{y}_s^i)$, $i = 1, \dots, d$ [111]. Ceci s'écrit :

$$\begin{aligned} f(\mathbf{y}_s) &= f(\mathbf{y}_s^1, \dots, \mathbf{y}_s^d) \\ &= \prod_{i=1}^d f^i(\mathbf{y}_s^i) \end{aligned} \quad (2.47)$$

Nous pouvons choisir le même type de loi pour toutes les variables \mathbf{y}_s^i , $i = 1, \dots, d$, ou des types de lois différentes pour chacune d'elles suivant l'application. Les paramètres associés à ce modèle sont l'ensemble des paramètres des lois monodimensionnelles. La littérature statistique est riche de telles lois ainsi que des estimateurs appropriés. Toutefois, il est rare de trouver dans la pratique des applications où les observations vectorielles vérifient l'hypothèse d'indépendance. L'utilisation de l'Eq. 2.47 n'est alors plus justifiée d'où la nécessité de recourir à d'autres formulations.

2.5.3 Loi jointe de variables aléatoires gaussiennes corrélées

La loi gaussienne multivariée est sans doute la plus utilisée en analyse de données multidimensionnelles. Elle constitue le meilleur modèle dans le cas où la dispersion des

vecteurs d'observations $\mathbf{y}_s = (\mathbf{y}_s^1, \dots, \mathbf{y}_s^d)^t$, $s \in S$, dans l'espace \mathbb{R}^d est un super-ellipsoïde. La loi jointe $f(\mathbf{y}_s)$ dans ce cas est donnée par :

$$f(\mathbf{y}_s) = \frac{1}{(2\pi)^{\frac{d}{2}}} \det(\Sigma)^{-\frac{1}{2}} \exp(-0.5 (\mathbf{y}_s - \mu)^t \Sigma^{-1} (\mathbf{y}_s - \mu)) \quad (2.48)$$

avec $\mu = (\mu^1, \dots, \mu^d)^t$ est le vecteur moyenne et Σ est la matrice de covariance des données qui constituent les deux paramètres de cette loi.

Il existe des estimateurs au sens du maximum de vraisemblance, très simples à mettre en œuvre pour la moyenne et la matrice de covariance. Il sont donnés par les expressions suivantes :

$$\hat{\mu} = \frac{1}{|S|} \sum_{s=1}^{|S|} \mathbf{y}_s \quad (2.49)$$

pour la moyenne, et

$$\hat{\Sigma} = \frac{1}{|S|} \sum_{s=1}^{|S|} (\mathbf{y}_s - \hat{\mu}) (\mathbf{y}_s - \hat{\mu})^t \quad (2.50)$$

pour la matrice de covariance.

2.5.4 Loi jointe de variables aléatoires corrélées : technique de décorrélation

Cette méthode est basée sur la technique de changement de variables aléatoires, bien connue en théorie de probabilités, et généralise la méthode de calcul de la Sect. 2.5.2. En effet, il faut trouver une transformation adéquate telle que les variables résultantes soient indépendantes ou au moins décorrélées. Soient les vecteurs d'observations $\mathbf{y}_s = (\mathbf{y}_s^1, \dots, \mathbf{y}_s^N)^t$, $s \in S$. Nous cherchons une transformation linéaire :

$$\mathcal{Y}_s = W \mathbf{y}_s \quad (2.51)$$

telle que les variables \mathcal{Y}_s^i , $i = 1, \dots, N$, soient indépendantes, ce qui permet d'écrire :

$$f_{\mathcal{Y}}(\mathcal{Y}_s) = \prod_{i=1}^N f_{\mathcal{Y}}^i(\mathcal{Y}_s^i) \quad (2.52)$$

Dans Pieczynski et *al.* [118] les auteurs se sont contentés d'une décorrélation (*i.e.*, ACP Sect. 2.3.1) tandis que Derrode et *al.* [41] ont opéré une analyse en composantes indépendantes (ACI Sect. 2.3.3).

Quelle que soit la transformation W , la densité de probabilité des observations originales \mathbf{y}_s , $s \in S$, se dérive de celle des donnée transformées \mathcal{Y}_s , $s \in S$, comme suit [63] :

$$f_{\mathbf{y}}(\mathbf{y}_s) = |\det(W)| f_{\mathcal{Y}}(\mathcal{Y}_s) \quad (2.53)$$

$$= |\det(W)| \prod_{i=1}^N f_{\mathcal{Y}}^i(\mathcal{Y}_s^i) \quad (2.54)$$

Nous pouvons ainsi calculer la loi jointe de tout ensemble de variables aléatoires corrélées. Pour les lois marginales des variables transformées, il est possible d'utiliser tout l'éventail de lois monodimensionnelles disponibles dans la littérature, et par exemple la loi gaussienne généralisée introduite à la Sect. 2.5.1.

2.5.5 Loi jointe de variables aléatoires corrélées : théorie des copules

Les statisticiens se sont depuis toujours intéressés à la relation entre les lois multidimensionnelles et leurs marginales. On peut citer les travaux de M. Fréchet [57] dans les années cinquantes. A. Sklar [134] a introduit en 1959 une nouvelle classe de fonctions qu'il a nommé copules, permettant de lier une fonction de répartition bidimensionnelle à ses marginales. La théorie des copules a ensuite été exploitée pour le développement d'espaces métriques probabilistes et des mesures non paramétriques d'indépendance. Le livre de R.B. Nelsen [108] représente une bonne introduction à cette théorie. Dans ce qui suit nous décrirons succinctement cette théorie et comment elle peut être utilisée pour le calcul de densités multidimensionnelles.

Théorème 2 (Sklar)

Soient les variables aléatoires $\mathbf{y}^1, \dots, \mathbf{y}^d$ dont la fonction de répartition jointe est F et les fonctions de répartition marginales sont respectivement F^1, \dots, F^d . Alors il existe une fonction C , dite d -copule, telle que :

$$F(\mathbf{y}^1, \dots, \mathbf{y}^d) = C(F^1(\mathbf{y}^1), \dots, F^d(\mathbf{y}^d)) \quad (2.55)$$

Si F^1, \dots, F^d sont continues, alors C est unique. Sinon, C est déterminée uniquement sur $Im(F^1) \times \dots \times Im(F^d)$. Inversement, si C est une d -copule et F^1, \dots, F^d sont des fonctions de répartition des variables aléatoires $\mathbf{y}^1, \dots, \mathbf{y}^d$, alors la fonction F définie par l'Eq. 2.55 est une fonction de répartition jointe de ces variables.

Ce théorème établit clairement la relation entre la fonction de répartition jointe et ses marginales. La d -copule peut être vue comme une fonction de répartition d -dimensionnelle dont les marginales sont uniformes sur l'intervalle $[0, 1]$ [114]. Ceci peut être facilement établi en effectuant le changement de variables $u_i = F^i(\mathbf{y}^i)$ dans l'Eq. 2.55. D'une manière plus rigoureuse, une d -copule est donnée par la définie comme suit :

Définition 6 (Copule d -dimensionnelle)

Une copule d -dimensionnelle (ou d -copule) est une fonction C de $\mathbf{I}^d = [0, 1]^d$ vers \mathbf{I} avec les propriétés suivantes :

1. Pour chaque \mathbf{u} dans \mathbf{I}^d ,

$$C(\mathbf{u}) = 0 \text{ si } \exists k, 1 \leq k \leq d : \mathbf{u}_k = 0$$

et

$$C(\mathbf{u}) = \mathbf{u}_k \text{ si } \forall m \neq k, 1 \leq m \leq d : \mathbf{u}_m = 1$$

2. Pour chaque \mathbf{u} et \mathbf{v} dans \mathbf{I}^d tel que $\mathbf{u} \leq \mathbf{v}$,

$$V_c([\mathbf{u}, \mathbf{v}]) = \Delta_{\mathbf{u}}^{\mathbf{v}} C(\mathbf{t}) = \Delta_{\mathbf{u}_d}^{\mathbf{v}_d} \Delta_{\mathbf{u}_{d-1}}^{\mathbf{v}_{d-1}} \cdots \Delta_{\mathbf{u}_2}^{\mathbf{v}_2} \Delta_{\mathbf{u}_1}^{\mathbf{v}_1} C(\mathbf{t}) \geq 0$$

où $V_c([\mathbf{u}, \mathbf{v}])$ est le C -volume de $[\mathbf{u}, \mathbf{v}]$ et :

$$\Delta_{\mathbf{u}_k}^{\mathbf{v}_k} C(\mathbf{t}) = C(t_1, \dots, t_{k-1}, \mathbf{v}_k, t_{k+1}, \dots, t_d) - C(t_1, \dots, t_{k-1}, \mathbf{u}_k, t_{k+1}, \dots, t_d) \quad (2.56)$$

Les copules disposent de deux bornes dites de Fréchet-Hoeffding (Fig. 2.6) :

– une borne supérieure :

$$\forall(\mathbf{u}_1, \dots, \mathbf{u}_d) \in \mathbf{I}^d : M^d(\mathbf{u}_1, \dots, \mathbf{u}_d) = \min(\mathbf{u}_1, \dots, \mathbf{u}_d) \quad (2.57)$$

– une borne inférieure :

$$\forall(\mathbf{u}_1, \dots, \mathbf{u}_d) \in \mathbf{I}^d : W^d(\mathbf{u}_1, \dots, \mathbf{u}_d) = \max(\mathbf{u}_1 + \dots + \mathbf{u}_d + d - 1, 0) \quad (2.58)$$

tel que pour toute d -copule C :

$$\forall(\mathbf{u}_1, \dots, \mathbf{u}_d) \in \mathbf{I}^d : W^d(\mathbf{u}_1, \dots, \mathbf{u}_d) \leq C(\mathbf{u}_1, \dots, \mathbf{u}_d) \leq M^d(\mathbf{u}_1, \dots, \mathbf{u}_d) \quad (2.59)$$

Il faut noter que W^d est une copule pour tout $d \geq 2$ tandis que M^d n'est une copule que pour le cas $d = 2$.

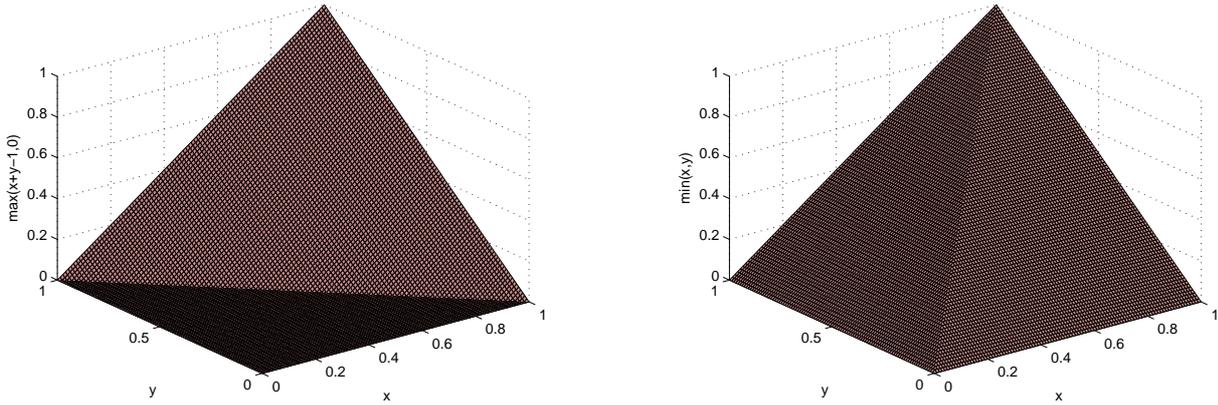


FIG. 2.6 – Bornes inférieure (à gauche) et supérieure (à droite) de Fréchet-Hoeffding d'une 2-copule

Densité d'une copule et lien entre une densité jointe et ses marginales

Dans le cas où C est différentiable, il est possible de définir une relation entre la densité jointe et ses marginales en dérivant l'Eq. 2.55 [108] :

$$f(\mathbf{y}^1, \dots, \mathbf{y}^d) = f^1(\mathbf{y}^1) \times \cdots \times f^d(\mathbf{y}^d) c(F^1(\mathbf{y}^1), \dots, F^d(\mathbf{y}^d)) \quad (2.60)$$

où f^m est la densité de probabilité correspondant à F^m et

$$c(\mathbf{u}_1, \dots, \mathbf{u}_d) = \frac{\partial^d C}{\partial \mathbf{u}_1 \times \cdots \times \partial \mathbf{u}_d}$$

est la densité de la copule.

Densité de la copule gaussienne

Considérons maintenant que le vecteur d'observations $\mathbf{y} = (\mathbf{y}^1, \dots, \mathbf{y}^d)^t$ suit une loi gaussienne d -dimensionnelle de moyenne nulle et de matrice de corrélation R , et que chaque éléments de ce vecteur suit une loi gaussienne centrée réduite. La densité jointe pour ce vecteur est alors :

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{N}{2}}} \det(R)^{-\frac{1}{2}} \exp(-0.5 \mathbf{y}^t R^{-1} \mathbf{y}) \quad (2.61)$$

Sachant que $f^i(\mathbf{y}^i) = \frac{1}{\sqrt{2\pi}} \exp(-0.5 (\mathbf{y}^i)^2)$, et en utilisant l'Eq.2.60, on peut déduire la densité de la copule gaussienne c_g :

$$\forall \mathbf{u} = (u_1, \dots, u_N)^t : c_g(\mathbf{u}) = |R|^{-\frac{1}{2}} \exp \left[-\frac{\tilde{\mathbf{u}}^t (R^{-1} - I) \tilde{\mathbf{u}}}{2} \right] \quad (2.62)$$

où $\tilde{\mathbf{u}} = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))^T$ avec $\Phi(\cdot)$ la fonction de répartition de la gaussienne monodimensionnelle centrée réduite, et I la matrice identité $d \times d$.

Il est ainsi possible de modéliser n'importe quelle densité jointe de variables corrélées à partir des densités marginales et la matrice de corrélation R , définie non négative, en utilisant la densité de copule gaussienne. Les paramètres d'un tel modèle sont la matrice de corrélation et les paramètres des densités marginales. Ces derniers sont estimés suivant les lois marginales choisies. Tandis que la matrice de corrélation peut être estimée de manière classique [108, 21, 22, 51] ou avec la méthode décrite dans [114].

Plusieurs autres types de copules ont été introduites dans la littérature en statistique parmi lesquelles : Student, Archimedean, Gumbel, Joe... etc [108]. Nous nous restreindrons dans cette thèse à la copule gaussienne qui malgré sa simplicité reste assez répandue.

2.6 Conclusion

Dans la première partie de ce chapitre, nous avons évoqué le problème de la dimensionnalité, présenté quelques techniques classiques pour le résoudre et trois schémas de réduction que nous avons adoptés dans le cadre de ce travail. Nous avons également décrit les différentes manières de modéliser le terme d'attache aux données dans le cas multispectral. Nous avons présenté le cas classique de la modélisation gaussienne ainsi que des techniques permettant le calcul de lois multidimensionnelles corrélées de marginales quelconques grâce à la théorie des copules. De cette façon, les éléments des modèles markoviens que nous avons présentés au chapitre 1 ont été complétés. Il nous reste maintenant une étape importante pour aboutir à des algorithmes autonomes (*i.e.*, non supervisés) qui est l'apprentissage des paramètres et qui fait l'objet du prochain chapitre.

3 Estimation de paramètres

3.1 Introduction

Nous avons présenté dans le chapitre 1 les quadarbres de Markov caché et couple. Ces modèles ont été complétés dans le chapitre 2 par le détail du terme d'attache aux données. Cependant, le calcul pratique des paramètres de ces modèles reste à définir. L'estimation des paramètres correspond à la seconde tâche générique sur les modèles graphiques. Dans le cas où l'on dispose des données complètes, la technique la plus utilisée est le maximum de vraisemblance. Quand celui-ci est difficile à obtenir, le maximum de pseudo vraisemblance est utilisé. Cependant lorsque les données sont incomplètes, la partie inconnue est souvent intégrée hors du problème d'estimation. Ceci correspond à la solution fournie par les algorithmes de type EM (Expectation-Maximization). Une autre approche consiste à fixer la partie cachée, en choisissant une réalisation suivant un critère donné, et optimiser ensuite par rapport aux paramètres comme dans le cas des données complètes. Cette dernière technique correspond à la segmentation courante.

La structure du chapitre est la suivante. Nous commencerons par une courte synthèse des algorithmes d'estimation. Nous détaillerons ensuite les procédures d'estimation utilisées dans le cas du quadarbre de Markov caché avec un terme d'attache aux données multidimensionnel gaussien ou gaussien généralisé, exposerons la procédure de segmentation non supervisée dans chaque cas, et présenterons quelques résultats sur des images synthétiques. Après cela, nous développerons de la même façon le quadarbre de Markov couple. Enfin, l'estimation des paramètres et la segmentation non supervisée avec l'approche MACPP régularisée seront abordées avec quelques résultats sur des images synthétiques.

3.2 Algorithmes d'estimations

3.2.1 Algorithme EM

L'EM, est une méthode d'optimisation itérative pour l'estimation de paramètres au sens du maximum de vraisemblance. Son idée de base est qu'au lieu d'utiliser les seules observations Y et faire des maximisations complexes ou des simulations, on augmente les observations avec des variables cachées X pour simplifier les calculs et réaliser une série de maximisations simples [136, 99]. Les observations Y sont donc considérées comme des données incomplètes, auxquelles on rajoute les données manquantes X pour aboutir aux données complètes (X, Y) . En effet, le fait que les variables cachées X contiennent des

informations pertinentes pour l'estimation des paramètres Θ , et Θ en retour, permettent de retrouver vraisemblablement les valeurs de X , suggère la stratégie suivante pour l'estimation des paramètres Θ à partir des seules observations Y : retrouver les variables cachées à partir d'une estimée initiale de Θ , ré-estimer Θ en se basant sur les observations Y et les variables cachées évaluées X et réitérer jusqu'à convergence des estimés [132]. Des applications spécifiques basées sur cette idée astucieuse, sont apparues dans la littérature dès les années 1920 [93]. Dempster et *al.* [40] ont formalisé comment retrouver les X à chaque étape et ont présenté l'algorithme dans sa forme générale en l'appelant Expectation-Maximization (EM).

La distribution des données complètes est donnée par :

$$p(X, Y/\Theta) = p(Y/\Theta) p(X/Y, \Theta) \quad (3.1)$$

Le passage au logarithme donne la log-vraisemblance des données complètes :

$$\log p(X, Y/\Theta) = L(\Theta) + \log p(X/Y, \Theta) \quad (3.2)$$

où $L(\Theta) = \log p(Y/\Theta)$ est la log-vraisemblance des observations et $\log p(X/Y, \Theta)$, appelée distribution prédictive conditionnelle [136], est un terme capturant la dépendance entre X et Θ et jouant un rôle important dans la procédure EM (Sect. 3.2.1).

Le calcul du second terme de la partie droite de l'Eq. 3.2 n'est pas possible à cause de la non disponibilité des données manquantes. On calcule alors la moyenne suivant la distribution prédictive $p(X/Y, \Theta^{[q]})$, avec $\Theta^{[q]}$ une estimée préliminaire de Θ . Ceci donne :

$$\mathcal{Q}(\Theta, \Theta^{[q]}) = L(\Theta) + \mathcal{H}(\Theta, \Theta^{[q]}) \quad (3.3)$$

avec

$$\mathcal{Q}(\Theta, \Theta^{[q]}) = \int_{\mathcal{X}} \log p(X, Y/\Theta) p(X/Y, \Theta^{[q]}) dX \quad (3.4)$$

et

$$\mathcal{H}(\Theta, \Theta^{[q]}) = \int_{\mathcal{X}} \log p(X/Y, \Theta) p(X/Y, \Theta^{[q]}) dX \quad (3.5)$$

Dempster et *al.* [40] ont proposé un algorithme avec des itérations à deux étapes (Algo. 3.1, p. 59) et ont montré que le choix des paramètres $\Theta^{[q+1]}$, qui maximisent la fonction auxiliaire $\mathcal{Q}(\Theta, \Theta^{[q]})$, donnerait une valeur $L(\Theta^{[q+1]})$ de la log-vraisemblance supérieure ou égale à $L(\Theta^{[q]})$. En effet :

$$\begin{aligned} L(\Theta^{[q+1]}) - L(\Theta^{[q]}) &= \mathcal{Q}(\Theta^{[q+1]}, \Theta^{[q]}) - \mathcal{Q}(\Theta^{[q]}, \Theta^{[q]}) \\ &\quad + \mathcal{H}(\Theta^{[q]}, \Theta^{[q]}) - \mathcal{H}(\Theta^{[q+1]}, \Theta^{[q]}) \end{aligned} \quad (3.6)$$

Le premier terme $\mathcal{Q}(\Theta^{[q+1]}, \Theta^{[q]}) - \mathcal{Q}(\Theta^{[q]}, \Theta^{[q]})$ est supérieur ou égal à 0 puisque

$$\Theta^{[q+1]} = \arg \max_{\Theta} \mathcal{Q}(\Theta, \Theta^{[q]})$$

et le second terme peut être réécrit sous la forme :

$$\mathcal{H}(\Theta^{[q]}, \Theta^{[q]}) - \mathcal{H}(\Theta^{[q+1]}, \Theta^{[q]}) = \int_{\mathcal{X}} \log \frac{p(X/Y, \Theta^{[q]})}{p(X/Y, \Theta^{[q+1]})} p(X/Y, \Theta^{[q]}) dX \quad (3.9)$$

Algo. 3.1 EM

- Choix des paramètres initiaux $\Theta^{[0]}$, $q = 0$.

Répéter

- Étape E :

$$\mathcal{Q}(\Theta, \Theta^{[q]}) = E(\log P(X, Y/\Theta)/Y, \Theta^{[q]}) \quad (3.7)$$

- Étape M :

$$\Theta^{[q+1]} = \arg \max_{\Theta} \mathcal{Q}(\Theta, \Theta^{[q]}) \quad (3.8)$$

$q=q+1$

Jusqu'à convergence

En utilisant l'inégalité de Jensen¹, on peut montrer que cette différence est supérieure ou égale à 0 pour tout $\Theta \neq \Theta^{[q]}$. Il est clair alors qu'il suffit de trouver $\Theta^{[q+1]}$ telle que $\mathcal{Q}(\Theta^{[q+1]}, \Theta^{[q]}) \geq \mathcal{Q}(\Theta^{[q]}, \Theta^{[q]})$ pour avoir une augmentation de la log-vraisemblance des observations. C'est l'algorithme EM généralisé (Algo. 3.2, p. 59).

Algo. 3.2 EM généralisé

- Choix des paramètres initiaux $\Theta^{[0]}$, $q = 0$.

Répéter

- Étape E :

$$\mathcal{Q}(\Theta, \Theta^{[q]}) = E(\log P(X, Y/\Theta)/Y, \Theta^{[q]}) \quad (3.10)$$

- Étape M :

$$\Theta^{[q+1]} : \text{tel que } \mathcal{Q}(\Theta^{[q+1]}, \Theta^{[q]}) \geq \mathcal{Q}(\Theta^{[q]}, \Theta^{[q]}) \quad (3.11)$$

$q=q+1$

Jusqu'à convergence

L'algorithme EM pour maximiser l'*a posteriori*

L'algorithme EM présenté généralement comme méthode de maximisation de la vraisemblance, peut être utilisé pour maximiser la loi *a posteriori* [40, 136]. En effet le logarithme de la distribution *a posteriori* des paramètres est donné par :

$$\log p(\Theta/Y) = \log p(\Theta/X, Y) - \log p(X/Y, \Theta) + \log p(X/Y) \quad (3.12)$$

Le calcul des moyennes par rapport à la distribution prédictive conditionnelle $p(X/Y, \Theta^{[q]})$ donne :

$$\log p(\Theta/Y) = \mathcal{Q}^*(\Theta, \Theta^{[q]}) - \mathcal{H}(\Theta, \Theta^{[q]}) + \mathcal{K}(\Theta^{[q]}) \quad (3.13)$$

où

$$\mathcal{Q}^*(\Theta, \Theta^{[q]}) = \mathcal{Q}(\Theta, \Theta^{[q]}) - \log(\pi(\Theta))$$

¹L'inégalité de Jensen affirme que pour toute variable aléatoire X et toute fonction convexe $\phi(\cdot)$, on a l'inégalité suivante $\mathbf{E}[\phi(X)] \geq \phi[\mathbf{E}(X)]$

avec $\pi(\Theta)$ la distribution *a priori* des paramètres, et

$$\mathcal{K}(\Theta^{[q]}) = \int_X \log P(X/Y) p(X/Y, \Theta^{[q]}) dX \quad (3.14)$$

qui ne dépend pas de $\Theta^{[q+1]}$.

Suivant le même raisonnement que pour l'EM classique, on montre que le choix d'un paramètre $\Theta^{[q+1]}$ qui maximise $\mathcal{Q}^*(\Theta, \Theta^{[q]})$, ou au moins tel que $\mathcal{Q}^*(\Theta^{[q+1]}, \Theta^{[q]}) \geq \mathcal{Q}^*(\Theta^{[q]}, \Theta^{[q]})$, entraîne une augmentation de l'*a posteriori* d'une quantité positive ou nulle [40, 136].

Propriétés de l'EM

La convergence de l'EM est garantie vers un point stationnaire (maximum ou point de selle) de la log-vraisemblance. Lorsque celle-ci en possède plusieurs, la convergence vers le maximum global n'est pas garantie. Dépendant de son initialisation, l'algorithme peut donc rester piégé par un maximum local ou un point de selle [40, 24].

Par ailleurs, la convergence de l'EM est linéaire et dépend du taux d'information sur Θ apporté par les variables cachées X [132] :

$$\mathcal{D} = \frac{- \left[\frac{\partial^2}{\partial \Theta^2} \log p(X/Y, \Theta) \right]_{\Theta=\hat{\Theta}}}{- \left[\frac{\partial^2}{\partial \Theta^2} \log p(X, Y/\Theta) \right]_{\Theta=\hat{\Theta}}}$$

Ainsi, lorsque ce rapport est grand la convergence peut être très lente. Néanmoins, à l'opposé, lorsque l'apport des variables cachées est nul, l'algorithme converge en une itération.

3.2.2 Variantes de l'EM

Plusieurs variantes de l'EM ont été présentées pour remédier aux inconvénients de ce dernier et/ou faciliter les calculs de ses deux étapes. Pour s'affranchir du problème des minima locaux, Celeux et *al.* [25] ont introduit une version stochastique de l'EM baptisée SEM. La phase stochastique de cette technique consiste à générer une réalisation $x^{[q]}$ suivant la loi $P(X/Y, \Theta^{[q]})$ et à calculer la fonction auxiliaire $\mathcal{Q}(\Theta, \Theta^{[q]}) = \log P(x^{[q]}, Y/\Theta)$ à maximiser. Dans le même esprit, Celeux et *al.* [26] ont également proposé un autre algorithme, appelé SAEM pour Stochastic Approximation EM, introduisant une séquence de paramètres (v_q) , équivalente à la température dans le recuit simulé, de telle sorte que l'estimateur se comporte comme un SEM au départ et un EM vers la fin. L'estimée à l'itération q est $\Theta^{[q]} = (1 - v_q)\Theta_{EM}^{[q]} + v_q\Theta_{SEM}^{[q]}$. Une technique assez proche, qui a été introduite pour d'autres raisons est le Monte Carlo EM (MCEM) de Wie et *al.* [141]. Son but est de faciliter le calcul de la fonction auxiliaire à l'étape E et éviter le calcul analytique de la moyenne suivant la distribution prédictive conditionnelle. Elle consiste à tirer M réalisations $(x_i^{[q]})_{i=1, \dots, M}$ suivant la loi $P(X/Y, \Theta^{[q]})$ et calculer la fonction auxiliaire $\mathcal{Q}(\Theta, \Theta^{[q]}) = \frac{1}{M} \sum_{i=1}^M \log P(x_i^{[q]}, Y/\Theta)$ à maximiser. Quand $M = 1$, on retrouve le SEM, et quand M est grand cet algorithme tend vers l'EM. Il est recommandé de démarrer avec

une valeur faible de M et l'augmenter lorsqu'on s'approche de la valeur de l'optimum [136].

Les algorithmes SEM, SAEM et MCEM correspondent à une perturbation aléatoire du système dynamique discret généré par l'EM, ce qui permet en pratique de ne pas rester piégé dans des minima locaux.

Toujours dans l'optique de simplification du calcul de la vraisemblance des données complètes, nécessaire à l'étape E, Chalmond [27, 28] et Zhang et *al.* [144] ont proposé d'utiliser la pseudo-vraisemblance introduite par Besag [9, 10] dans le cadre des données complètes. Plus précisément dans le cas d'un champs de Markov où chaque observation ne dépend que de son état caché, $\log p(X, Y, /\Theta)$ est remplacé par $\sum_{s \in S} \log p(X_s / X_{\nu_s}, \Theta) + \sum_{s \in S} \log p(Y_s / X_s, \Theta)$. Le calcul de l'espérance se fait de la même façon que pour le MCEM.

D'autres travaux ont été proposés pour remédier au problème de lenteur de la convergence. On peut citer le turbo EM de Louis [94] basé sur la méthode de Newton-Raphson, le travail de Jamshidian et *al.* [73] basé sur gradient conjugué et les travaux de Lange [87, 88] basés sur le gradient et le quasi-Newton.

Lorsque la maximisation de la fonction auxiliaire est trop complexe pour faire l'objet d'un calcul analytique, la version ECM (Expectation Conditional Maximization) introduite par Meng et *al.* [100] peut être utilisée. Elle consiste à remplacer à l'itération q la maximisation de $\mathcal{Q}(\Theta, \Theta^{[q]})$, où $\Theta = (\Theta_1, \Theta_2)$, par deux maximisations partielles. Dans la première, Θ_2 est fixé à $\Theta_2^{[q]}$, et on maximise par rapport à Θ_1 ; tandis que dans la deuxième Θ_1 est fixé à $\Theta_1^{[q+1]}$ et on maximise par rapport à Θ_2 [136]. Cet algorithme est un EM généralisé puisque l'augmentation de la vraisemblance à chaque itération est garantie, bien que $\mathcal{Q}(\Theta^{[q+1]}, \Theta^{[q]})$, obtenue par les deux maximisations partielles, n'est généralement pas le maximum de $\mathcal{Q}(\Theta, \Theta^{[q]})$.

3.2.3 Algorithme ECI

L'idée de l'algorithme ECI (Estimation Conditionnelle Itérative) introduit par Pieczynski [115] est basée sur l'hypothèse de l'existence d'un estimateur $\hat{\Theta}(X, Y)$. On recherche alors à l'approcher au mieux dans l'ensemble de variables aléatoires, fonctions des observations Y . La meilleure approximation au sens de l'erreur quadratique moyenne est l'espérance conditionnelle $\mathbf{E}(\hat{\Theta}(X, Y)/Y)$ qui dépend de Θ . Elle est donc approchée itérativement partant d'une estimée initiale $\Theta^{[0]}$, suivant l'expression :

$$\Theta^{[q+1]} = \mathbf{E}(\hat{\Theta}(X, Y)/Y, \Theta^{[q]})$$

La procédure ECI ne trouve à notre connaissance aucune preuve théorique de sa convergence. Cependant différentes expériences et applications montrent son efficacité [7, 60, 82, 102, 120]. Il est à noter que lorsque la moyenne ne peut pas être calculée, on a recours au tirage d'échantillons $(x_i^{[q]})_{i=1, \dots, M}$, suivant la loi $p(X/Y, \Theta^{[q]})$ et l'approximation de la moyenne par $\frac{1}{M} \sum_i \hat{\Theta}(x_i^{[q]}, Y)$.

Enfin, il faut préciser qu'aucune contrainte n'est faite sur l'estimateur utilisé avec cette technique. Cependant, lorsque le maximum de vraisemblance est utilisé, l'ECI ne diffère

de l'EM que par l'ordre des opérations de maximum et de l'espérance conditionnelle. Lorsque cet ordre peut être inversé on aura équivalence entre les deux algorithmes [39].

3.2.4 L'algorithme K-moyennes segmental

L'algorithme K-moyennes segmental (KMS) a été introduit par Rabiner et *al.* [122, 123, 77]. Son idée s'intègre dans la stratégie globale d'utilisation des variables cachées X pour l'estimation. On cherche $\hat{\Theta}$ telle que :

$$\hat{\Theta} = \underset{\Theta}{\operatorname{arg\,max}} \left\{ \max_{x \in \Omega_x} p(X, Y / \Theta) \right\} \quad (3.15)$$

On fixe ainsi les variables cachées X au lieu de les sommer hors du problème, par le biais du calcul de la moyenne par rapport à la distribution prédictive, comme le suggère l'EM. Dans le cas d'une chaîne de Markov, l'algorithme KMS revient à optimiser les paramètres sur le chemin le plus vraisemblable, tandis que l'EM le fait sur tous les chemins possibles dans le treillis engendré par toutes les transitions possibles. Chaque itération de l'algorithme comporte donc deux étapes fondamentales : 1) segmentation et 2) optimisation. On démarre avec un ensemble de paramètres initiaux $\Theta^{[0]}$, et on alterne les deux étapes suivant Eq. 3.15 jusqu'à la convergence [77].

3.2.5 L'algorithme de segmentation courante

La formulation du KMS suggère l'utilisation du MAP pour l'étape de segmentation et le maximum de vraisemblance pour l'optimisation. Lorsque d'autres critères sont utilisés, on retrouve une généralisation du KMS appelée méthode de segmentation courante [92, 102]. La mise à jour des paramètres s'écrit alors :

$$\Theta^{[q+1]} = \hat{\Theta}(\hat{X}(\Theta^{[q]}, Y), Y) \quad (3.16)$$

où $\hat{\Theta}$ est un estimateur donné et \hat{X} une méthode de segmentation.

Il faut noter la ressemblance entre la segmentation courante et l'ECI. Dans les deux techniques, on recherche des réalisations des observations cachées X en se basant sur une estimation courante $\Theta^{[q]}$. Ceci est réalisé par segmentation dans le cas de la segmentation courante et par tirages aléatoires suivant la loi $p(X/Y, \Theta^{[q]})$ dans le cas de l'ECI.

De même que pour l'ECI, La segmentation courante ne trouve aucune preuve théorique de convergence malgré son succès pratique [92]. Cette technique sera utilisée par la suite pour l'estimation des paramètres du quadarbre de markov couple et du quadarbre de markov caché dans le cas non gaussien. La raison est l'apparition de la fonction de répartition de la loi gaussienne généralisé dans l'expression du terme d'attache aux donnée, formulée grâce à la théorie des copules, ce qui rend la dérivation de ce terme complexe.

3.3 Le quadarbre de Markov caché

3.3.1 Le cas d'une attache aux données gaussienne

Il s'agit du quadarbre de Markov caché introduit à la Sect. 1.5 où la vraisemblance $f_i^k(\cdot)$ de la classe ω_i à l'échelle k est une loi gaussienne multivariée (Sect. 2.5.3).

L'algorithme EM pour l'estimation des paramètres

Les hypothèses d'indépendance sur le quadarbre nous permettent d'écrire la probabilité jointe des données complètes sous la forme :

$$p(x, y/\Theta) = p(x_r/\Theta_x) \prod_{s \neq r} p(x_s/x_{s-}, \Theta_x) \prod_{s \in S} p(\mathbf{y}_s/x_s, \Theta_y) \quad (3.17)$$

où Θ_x et Θ_y sont respectivement les paramètres du modèle *a priori* et de l'attache aux données. La fonction auxiliaire de l'EM s'écrit alors :

$$\mathcal{Q}(\Theta, \Theta^{[q]}) = \mathbf{E}(\log p(x_r/\Theta_x)) + \sum_{s \neq r} \mathbf{E}(\log p(x_s/x_{s-}, \Theta_x)) + \sum_{s \in S} \mathbf{E}(\log p(\mathbf{y}_s/x_s, \Theta_y)) \quad (3.18)$$

Rappelons les notations suivantes :

- $\pi_i = p(x_r = \omega_i)$
- $a_{ij} = p(x_s = \omega_j/x_{s-} = \omega_i)$
- $h_s(i, j) = p(x_s = \omega_j, x_{s-} = \omega_i/Y)$
- $\xi_s(i) = p(x_s = \omega_i/Y)$
- $\forall s \in S^k : f_i^k(\mathbf{y}_s) = p(\mathbf{y}_s/x_s = \omega_i)$ et $\mathbf{y}_s \in \mathbb{R}^d$

Nous avons alors :

$$\begin{aligned} \mathcal{Q}(\Theta, \Theta^{[q]}) &= \sum_i \xi_r^{[q]}(i) \log \pi_i + \sum_{s \neq r} \sum_i \sum_j h_s^{[q]}(i, j) \log a_{ij} \\ &+ \sum_{k=0}^R \sum_{s \in S^k} \sum_i \xi_s^{[q]}(i) \log f_i^k(\mathbf{y}_s) \end{aligned} \quad (3.19)$$

Cette quantité est à maximiser sous les contraintes suivantes $\sum_i \pi_i = 1$ et $\sum_j a_{ij} = 1, \forall i$.

En utilisant la technique du multiplicateur de Lagrange, on obtient le critère *Crit* à maximiser :

$$Crit = \mathcal{Q}(\Theta, \Theta^{[q]}) - \lambda \left[\sum_i \pi_i - 1 \right] - \sum_i \varsigma_i \left[\sum_j a_{ij} - 1 \right] \quad (3.20)$$

Calcul des probabilités *a priori*

La dérivée du critère de l'Eq. 3.20 par rapport à la probabilité π_i est $\frac{\partial Crit}{\partial a_{ij}} = \frac{\xi_r^{[q]}(i)}{\pi_i} - \lambda$. Son annulation donne :

$$\frac{\partial Crit}{\partial a_{ij}} = 0 \Rightarrow \pi_i = \frac{\xi_r^{[q]}(i)}{\lambda} \quad (3.21)$$

En utilisant le fait que $\sum_i \pi_i = 1$, on trouve facilement que $\lambda = 1$, d'où la mise à jour de l'*a priori* :

$$\pi_i^{[q+1]} = \xi_r^{[q]}(i) \quad (3.22)$$

Calcul des probabilités de transition

La dérivation du critère de l'Eq. 3.20 par rapport à la probabilité a_{ij} , et l'annulation du résultat donne :

$$a_{ij} = \frac{\sum_{s \neq r} h_s^{[q]}(i, j)}{\varsigma_i} \quad (3.23)$$

Par ailleurs :

$$\sum_j a_{ij} = 1 = \sum_j \frac{\sum_{s \neq r} h_s^{[q]}(i, j)}{\varsigma_i} = \frac{\sum_{s \neq r} \xi_{s^-}^{[q]}(i)}{\varsigma_i} \quad (3.24)$$

Nous avons enfin :

$$a_{ij}^{[q+1]} = \frac{\sum_{s \neq r} h_s^{[q]}(i, j)}{\sum_{s \neq r} \xi_{s^-}^{[q]}(i)} \quad (3.25)$$

Calcul des paramètres de l'attache aux données

Dans le cas gaussien le terme d'attache aux données s'exprime sous la forme :

$$\forall s \in S^k \quad : \quad f_i^k(\mathbf{y}_s) = \frac{1}{(2\pi)^{\frac{N}{2}}} \det(\Sigma_i^k)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{y}_s - \boldsymbol{\mu}_i^k)^t (\Sigma_i^k)^{-1} (\mathbf{y}_s - \boldsymbol{\mu}_i^k)\right) \quad (3.26)$$

En remplaçant cette expression dans l'Eq. 3.19, le terme associé à l'attache aux données devient :

$$\begin{aligned} & \sum_{k=0}^R \sum_{s \in S^k} \sum_i \xi_s^{[q]}(i) \log f_i^k(\mathbf{y}_s) \\ &= \sum_{k=0}^R \sum_{s \in S^k} \sum_i \xi_s^{[q]}(i) \left(-\frac{1}{2} \log \det(\Sigma_i^k) - \frac{1}{2} (\mathbf{y}_s - \boldsymbol{\mu}_i^k)^t (\Sigma_i^k)^{-1} (\mathbf{y}_s - \boldsymbol{\mu}_i^k) \right) + C^{ste} \end{aligned} \quad (3.27)$$

Ainsi en dérivant par rapport à la moyenne $\boldsymbol{\mu}_i^k$ et en égalant le résultat à 0, on obtient :

$$\sum_{s \in S^k} (\Sigma_i^k)^{-1} (\mathbf{y}_s - \boldsymbol{\mu}_i^k) \xi_s^{[q]}(i) = 0 \quad (3.28)$$

d'où la mise à jour suivante de la moyenne :

$$\boldsymbol{\mu}_i^{k[q+1]} = \frac{\sum_{s \in S^k} \mathbf{y}_s \xi_s^{[q]}(i)}{\sum_{s \in S^k} \xi_s^{[q]}(i)} \quad (3.29)$$

Avant de dériver par rapport à la matrice de covariance, on réécrit l'Eq. 3.27 sous la forme :

$$\begin{aligned} & \sum_{k=0}^R \sum_{s \in S^k} \sum_i \xi_s^{[q]}(i) \log f_i^k(\mathbf{y}_s) \\ = & \sum_{k=0}^R \sum_i \left[\frac{1}{2} \log \det((\Sigma_i^k)^{-1}) \sum_{s \in S^k} \xi_s^{[q]}(i) - \frac{1}{2} \sum_{s \in S^k} \xi_s^{[q]}(i) \text{tr}((\Sigma_i^k)^{-1} \Upsilon_{ki}) \right] + C^{ste} \end{aligned} \quad (3.30)$$

où $\Upsilon_{ki} = (\mathbf{y}_s - \boldsymbol{\mu}_i^k)(\mathbf{y}_s - \boldsymbol{\mu}_i^k)^t$. En utilisant les règles du calcul matriciel², on obtient la dérivée par rapport à $(\Sigma_i^k)^{-1}$:

$$\begin{aligned} \frac{\partial Crit}{\partial (\Sigma_i^k)^{-1}} &= \frac{1}{2} \sum_{s \in S^k} \xi_s^{[q]}(i) (2\Sigma_i^k - \text{diag}(\Sigma_i^k)) - \frac{1}{2} \sum_{s \in S^k} \xi_s^{[q]}(i) (2\Upsilon_{ki} - \text{diag}(\Upsilon_{ki})) \\ &= 2\Xi - \text{diag}(\Xi) \end{aligned} \quad (3.31)$$

avec $\Xi = \sum_{s \in S^k} \xi_s^{[q]}(i) (\Sigma_i^k - \Upsilon_{ki})$. En annulant cette dérivée on trouve $\Xi = 0$, ce qui donne la mise à jour suivante de la matrice de covariance :

$$\Sigma_i^{k[q+1]} = \frac{\sum_{s \in S^k} \xi_s^{[q]}(i) (\mathbf{y}_s - \boldsymbol{\mu}_i^{k[q+1]})(\mathbf{y}_s - \boldsymbol{\mu}_i^{k[q+1]})^t}{\sum_{s \in S^k} \xi_s^{[q]}(i)} \quad (3.32)$$

Segmentation non supervisée

Nous avons maintenant rassemblé tous les éléments pour présenter l'algorithme de segmentation non supervisée sur le quadarbre dans le cas d'un terme d'attache aux données modélisé par une loi normale multidimensionnelle. La procédure débute avec une estimée initiale³ des paramètres $\Theta^{[0]}$ et répète les deux passes sur le quadarbre, pour le calcul des probabilités *a posteriori* $h_s(i, j)$ et $\xi_s(i)$, suivies des mises à jour décrites dans les sections précédentes, jusqu'à la convergence. La carte de segmentation est alors obtenue en utilisant le critère MPM (Algo. 3.3, p. 66.).

3.3.2 Le cas d'une attache aux données gaussienne généralisée

Dans le cas où les bruits dans les différentes bandes de l'image multispectrale sont non gaussiens, l'utilisation de la loi normale n'est plus justifiée. Nous utilisons alors la théorie des copules pour le calcul de la loi multidimensionnelle avec des marginales gaussiennes généralisées corrélées [51]. Cela nous permet de couvrir un large ensemble de lois symétriques sur-gaussiennes, gaussiennes et sous-gaussiennes. L'attache aux données de la classe i à l'échelle k est donc :

$$f_i^k(\mathbf{y}_s) = c g_i^k(\mathfrak{G}_{i1}^k(\mathbf{y}_s^1), \dots, \mathfrak{G}_{iN}^k(\mathbf{y}_s^N)) \prod_{j=1}^N \mathfrak{g}(\boldsymbol{\mu}_{ij}^k, \sigma_{ij}^k, \gamma_{ij}^k; \mathbf{y}_s^j) \quad (3.33)$$

² $\frac{\partial \log \det(A)}{\partial A} = 2A^{-1} - \text{diag}(A^{-1})$ et $\frac{\partial \text{tr}(AB)}{\partial A} = B + B^t - \text{diag}(B)$, avec $\text{diag}(B)$ la matrice diagonale dont les éléments représentent la diagonale de B

³ Dans tous nos algorithmes, nous avons utilisé une initialisation par l'algorithme des k-moyennes [45].

Algo. 3.3 Segmentation non supervisée sur le quadarbre de Markov caché : cas gaussien

Initialisation : Estimée initiale $\Theta^{[0]}$

Répéter

- Passe montante sur le quadarbre : calcul des probabilités $p(x_s/\mathbf{y}_{\geq s})$ [Eq. 1.35]
- Passe descendante sur le quadarbre : calcul des probabilités *a posteriori* $h_s(i, j)$ et $\xi_s(i)$ [Eqs. 1.39 et 1.41]
- Mise à jour des paramètres *a priori* Θ_x [Eqs. 3.22 et 3.25]
- Mise à jour des paramètres de l'attache aux données Θ_y [Eqs. 3.29 et 3.32]

Jusqu'à convergence

- Obtention de la carte de segmentation $\hat{x}_s, s \in S$ suivant le critère MPM [Eq. 1.42]

Sortie : Carte de segmentation $\hat{x}_s, s \in S$

où cg_i^k est la densité de la copule gaussienne définie par la matrice de corrélation R_i^k (Eq. 2.62), $\mathbf{g}(\boldsymbol{\mu}_{ij}^k, \sigma_{ij}^k, \gamma_{ij}^k; \mathbf{y}_s^j)$ est la $j^{\text{ième}}$ marginale gaussienne généralisée définie par $\boldsymbol{\mu}_{ij}^k, \sigma_{ij}^k, \gamma_{ij}^k$ respectivement sa moyenne, sa variance et son paramètre de forme (Eq. 2.44), \mathcal{G}_{i1}^k est la fonction de répartition marginale correspondante, et \mathbf{y}_s^l l'observation au site s de la bande l .

L'utilisation de l'EM pour l'estimation des paramètres, donnerait une fonction auxiliaire de même allure que pour le cas gaussien (Eq. 3.19). L'optimisation par rapport aux paramètres du modèle *a priori* est la même que pour le cas gaussien. Cependant, pour les paramètres de l'attache aux données la dérivation de l'Eq. 3.33 est complexe et conduit à un calcul inextricable. C'est la raison pour laquelle nous utilisons la technique de la segmentation courante.

Segmentation courante pour l'estimation de paramètres

La segmentation courante est une technique assez simple à mettre en œuvre. Disposant d'estimées $\Theta^{[0]}$ des paramètres, on effectue les deux passes classiques sur le quadarbre pour le calcul des lois *a posteriori* $p(x_s/Y)$ qui permettent l'utilisation du critère MPM pour avoir une carte de segmentation intermédiaire $\hat{x}^{[q]}$. L'estimation des paramètres *a priori*

revient alors à faire de simples comptages :

$$\pi_i^{[q+1]} = \frac{\sum_{s \in S} \delta(\omega_i, \hat{x}_s^{[q]})}{|S|} \quad (3.34)$$

$$a_{i,j}^{[q+1]} = \frac{\sum_{s \neq r} \delta(\omega_j, \hat{x}_s^{[q]}) \delta(\omega_i, \hat{x}_{s^-}^{[q]})}{\sum_{s \neq r} \delta(\omega_i, \hat{x}_{s^-}^{[q]})} \quad (3.35)$$

où $\delta(\cdot, \cdot)$ est le symbole de Kronecker.

Pour les paramètres de la densité f_i^k , on définit l'ensemble des observations à l'échelle k appartenant à la classe ω_i :

$$O_i^k = \{\mathbf{y}_s : s \in S^k \text{ et } \hat{x}_s^{[q]} = \omega_i\}$$

On estime alors les paramètres $\boldsymbol{\mu}_{ij}^k, \sigma_{ij}^k, \gamma_{ij}^k$ de la $j^{\text{ième}}$ marginale gaussienne généralisée de la classe i à l'échelle k en utilisant les observations de l'ensemble O_i^k suivant la technique hybride [120] décrite au chapitre précédent (Eq. 2.46). La matrice de corrélation R_i^k est estimée de manière classique à partir de l'ensemble :

$$\tilde{O}_i^k = \{\tilde{\mathbf{t}}_s = (\tilde{\mathbf{t}}_s^1, \dots, \tilde{\mathbf{t}}_s^N)^t : \forall m = 1, \dots, N, \tilde{\mathbf{t}}_s^m = \Phi^{-1}(F_i^k(\mathbf{y}_s^m)) \text{ avec } \mathbf{y}_s \in O_i^k\}$$

où $\Phi(\cdot)$ est la fonction de répartition de la gaussienne monodimensionnelle centrée réduite et F_i^k la fonction de répartition de la marginale f_i^k . La mise à jour de R_i^k s'écrit alors :

$$R_i^{k[q+1]} = \sum_{s: \tilde{\mathbf{t}}_s \in \tilde{O}_i^k} \tilde{\mathbf{t}}_s \tilde{\mathbf{t}}_s^t \quad (3.36)$$

L'algorithme Algo. 3.4, p. 68 résume la procédure de segmentation non supervisée sur le quadarbre de Markov caché dans le cas gaussien généralisé.

3.3.3 Résultats sur des images synthétiques

Pour tester et comparer les algorithmes sur le quadarbre gaussien et gaussien généralisé, nous avons généré deux ensembles d'images, chacun est constitué de trois images tri-bandes. Le premier ensemble (Fig.3.1) représente deux classes gaussiennes avec les paramètres donnés dans la Tab. 3.1 et un coefficient de corrélation inter-bande unique pour chaque image de l'ensemble, de valeurs : (a) $\rho = 0$, (b) $\rho = 0.3$ and (c) $\rho = 0.8$. Le deuxième ensemble (Fig.3.2) représente deux classes gaussiennes généralisées avec les moyennes et variances données dans la Tab. 3.1, et un paramètre de forme égal à 1, 2, 3 respectivement pour les bandes 1, 2, 3. Comme pour le cas gaussien, nous considérons trois valeurs du facteur de corrélation inter-bandes : (d) $\rho = 0$, (e) $\rho = 0.3$ and (f) $\rho = 0.8$. Ces deux ensembles sont segmentés en deux classes avec le quadarbre de Markov caché en utilisant les deux modélisations de la vraisemblance citées ci-haut, *i.e.*, la loi gaussienne multivariée et la loi gaussienne généralisée multidimensionnelle obtenue grâce à la théorie des copules.

Les taux d'erreurs obtenus sont résumés à la Tab. 3.2. Dans le cas gaussien Fig. 3.1, on constate que le quadarbre caché avec la loi gaussienne multivariée dépasse légèrement le

Algo. 3.4 Segmentation non supervisée sur le quadarbre de Markov caché : cas gaussien généralisé

Initialisation : Estimée initiale $\Theta^{[0]}$

Répéter

- Passe montante sur le quadarbre : calcul des probabilités $p(x_s/\mathbf{y}_{\geq s})$ [Eq. 1.35]
- Passe descendante sur le quadarbre : calcul des probabilités *a posteriori* $h_s(i, j)$ et $\xi_s(i)$ [Eqs. 1.39-1.41]
- Obtention de la carte de segmentation intermédiaire $\hat{x}^{[q]}$ en maximisant $\xi_s(i)$ en chaque site
- Mise à jour des paramètres *a priori* Θ_x [Eqs. 3.34-3.35]
- Mise à jour des paramètres de la densité copule [Eq. 3.36] sur les ensembles \tilde{O}_i^k
- Mise à jour des paramètres des marginales gaussiennes généralisées [Eq. 2.46] sur les ensembles O_i^k

Jusqu'à convergence

- Obtention de la carte de segmentation $\hat{x}_s, s \in S$ suivant le critère MPM [Eq. 1.42]

Sortie : Carte de segmentation $\hat{x}_s, s \in S$

quadarbre caché avec la loi gaussienne généralisée obtenue grâce à la théorie des copules et cela malgré la difficulté supplémentaire due à l'estimation du paramètre de forme. Cependant dans le cas des classes gaussiennes généralisées Fig.3.2, la tendance est inversée et le modèle avec la gaussienne généralisée se révèle plus performant. De plus, on peut observer que le taux d'erreur minimum est à chaque fois obtenu avec le taux de corrélation le plus élevé. Ceci peut s'expliquer en partie par le fait que l'ellipsoïde formé par la représentation des pixels dans l'espace 3D (pour ces images tribandes), devient de plus en plus excentrique et orientée avec l'augmentation de la valeur du coefficient de corrélation, de telle sorte que dans certains cas la séparation entre deux classes devient plus prononcée (Fig. 3.3).

Les résultats obtenus avec le quadarbre de Markov caché, avec la nouvelle modélisation du terme d'attache aux données dans le gaussien généralisé multidimensionnel, sur les images synthétique sont satisfaisants. L'aptitude de traitement de ce modèle de données multirésolution et multispectrale, nous a poussé à l'utiliser pour la fusion d'images multispectrale dans le domaine d'ondelettes. Les détails de la technique et des résultats sur des images astronomiques seront présentés dans le prochain chapitre (Sect. 4.6, p. 108.).

	classe 1		classe 2	
	μ_1	σ_1	μ_2	σ_2
Bande 1	120	16	136	16
Bande 2	120	16	136	16
Bande 3	128	16	128	16

TAB. 3.1 – Moyennes et variances des deux classes pour les trois bandes

	Gaussien			Gaussien Généralisé		
	(a)	(b)	(c)	(d)	(e)	(f)
GM	1.71	1.95	1.07	2.08	2.38	1.16
C	1.77	2.59	1.19	0.87	1.58	1.09

TAB. 3.2 – Taux d'erreur, en %, obtenus avec les deux modèles de vraisemblance. (GM) Gaussienne Multivariée . (C) Gaussienne multivariée généralisée obtenue avec la théorie des Copules. Classes gaussiennes avec une corrélation interbande : (a) $\rho = 0$, (b) $\rho = 0.3$ and (c) $\rho = 0.8$. Classes gaussiennes généralisées avec une corrélation interbande : (d) $\rho = 0$, (e) $\rho = 0.3$ and (f) $\rho = 0.8$

3.4 Le quadarbre de markov couple

Dans la formulation du quadarbre de Markov couple introduite dans la Sect. 1.6 nous n'avons pas spécifié les termes d'attaches aux données requis pour le calcul de la probabilité de transition $p(z_s/z_{s-})$. En effet, selon l'Eq. 1.46, ce terme requiert le calcul des probabilités $f_{i,j}^k(y_s, y_{s-})$ et sa marginale $f_{i,j}^k(y_{s-})$ pour chaque échelle k . Nous utilisons deux modélisations différentes selon la présence ou non de bruit gaussien. En effet, dans le cas où le bruit aux différentes échelles est non gaussien, il est plus judicieux d'utiliser la

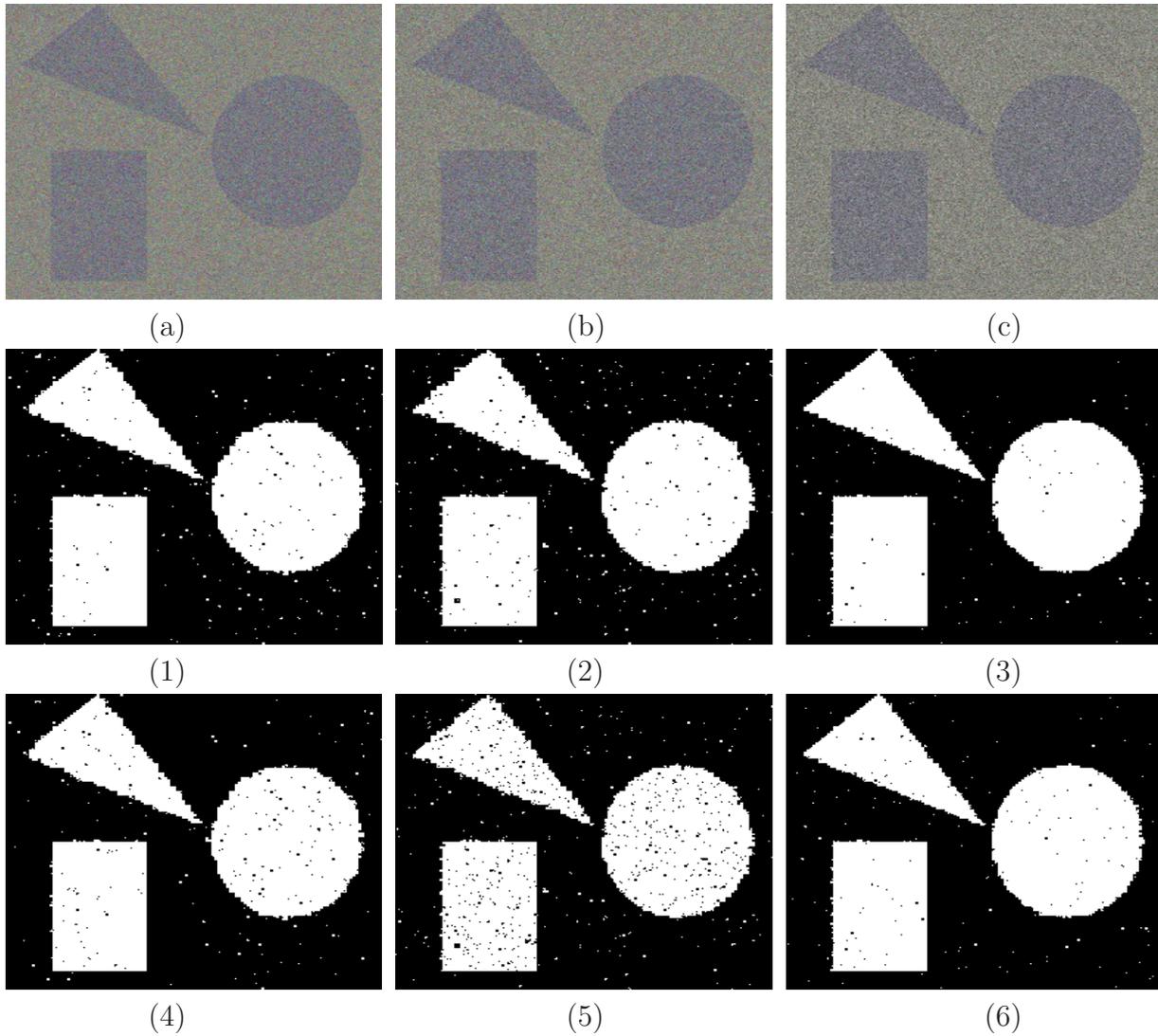


FIG. 3.1 – Quadarbre de Markov caché, cas gaussien : images originales tribandes (visualisées en RGB) avec chacune un coefficient de corrélation interbande unique : (a) $\rho = 0$, (b) $\rho = 0.3$ and (c) $\rho = 0.8$; cartes de segmentations 2 classes obtenues avec le quadarbre de Markov caché et une vraisemblance gaussienne multivariée (images 1,2 et 3), et celles obtenues avec le quadarbre de Markov caché et une vraisemblance gaussienne généralisée multivariée (images 4,5 et 6).

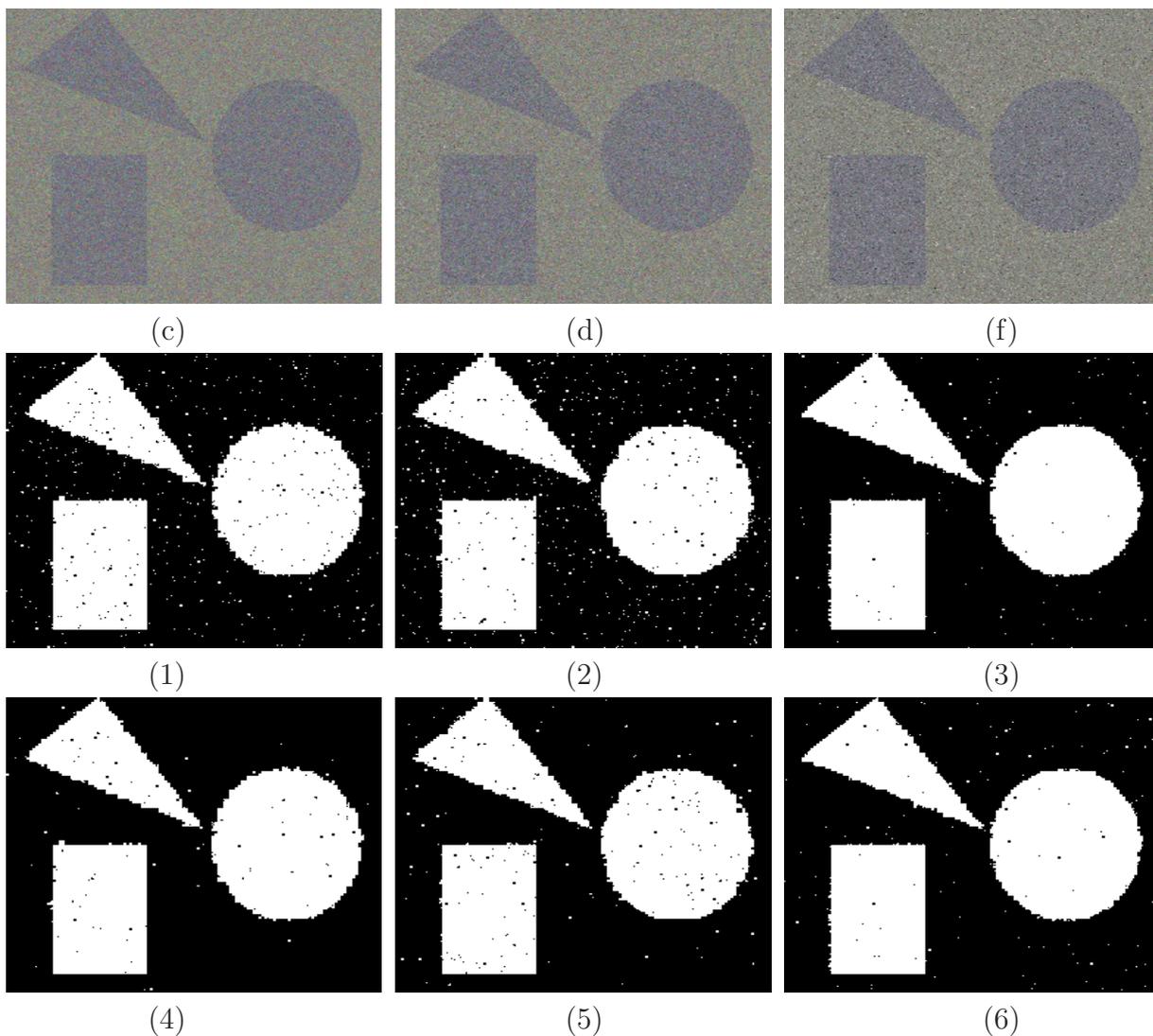


FIG. 3.2 – Quadarbre de Markov caché, cas gaussien généralisé : images originales tri-bandes (visualisées en RGB) avec chacune un coefficient de corrélation interbande unique : (c) $\rho = 0$, (d) $\rho = 0.3$ and (f) $\rho = 0.8$; cartes de segmentations 2 classes obtenues avec le quadarbre de Markov caché et une vraisemblance gaussienne multivariée (images 1,2 et 3), et ceux obtenues avec le quadarbre de Markov caché et une vraisemblance gaussienne généralisée multivariée (images 4,5 et 6).

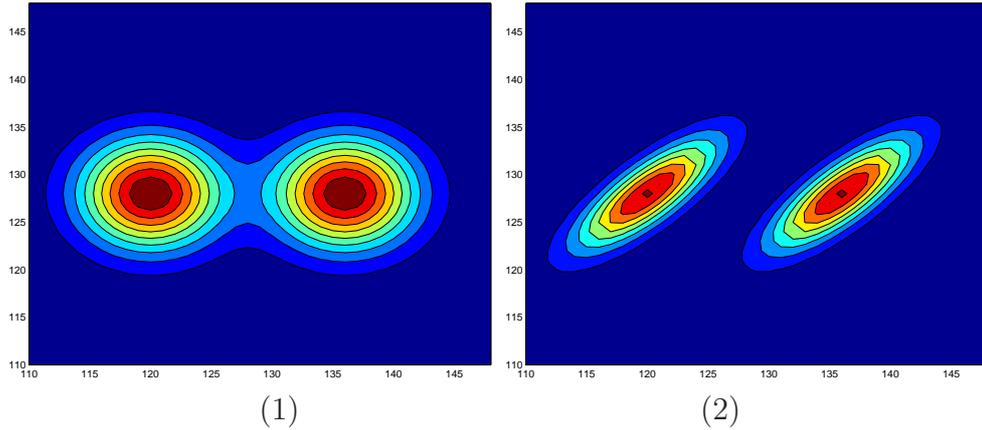


FIG. 3.3 – Exemple d'iso-contours de la densité de probabilité d'une variable bidimensionnelle gaussienne avec deux classes. (1) non-corrélée (2) corrélée de coefficient $\rho = 0.8$

gaussienne généralisée pour couvrir un large éventail de lois. Dans les deux cas nous avons utilisé la méthode de segmentation courante. L'estimation des paramètres du modèle *a priori*, semblables dans les deux cas, va donc être abordée en premier.

3.4.1 Estimation des paramètres de l'*a priori*

A partir des paramètres initiaux $\Theta^{[0]}$, on effectue deux passes sur le quadarbre couple pour le calcul des marginales *a posteriori* $p(x_s/Y)$ (Eqs. 1.49 et 1.51). Ceci permet l'utilisation du critère MPM pour avoir une carte segmentation intermédiaire $\hat{x}^{[q]}$. La mise à jour de la probabilité *a priori* est :

$$\pi_i^{[q+1]} = \frac{\sum_{s \in S} \delta(\omega_i, \hat{x}_s^{[q]})}{|S|} \quad (3.37)$$

La mise à jour des probabilités jointes est alors donnée par :

$$\begin{aligned} p(i, j)^{[q+1]} &= p(x_s^{[q]} = \omega_j, x_{s^-}^{[q]} = \omega_i) \\ &= \frac{\sum_{s \neq r} \delta(\omega_j, \hat{x}_s^{[q]}) \delta(\omega_i, \hat{x}_{s^-}^{[q]})}{|S| - 1} \end{aligned} \quad (3.38)$$

3.4.2 Estimation des paramètres de l'attache aux données : cas gaussien

L'estimation des paramètres des vraisemblances est faite à partir des ensembles O_{ij}^k définis comme suit :

$$O_{ij}^k = \{(y_s, y_{s^-}) : s \in S^k \text{ et } \hat{x}_s^{[q]} = \omega_j \text{ et } \hat{x}_{s^-}^{[q]} = \omega_i\} \quad (3.39)$$

Dans le cas gaussien, les paramètres de la densité gaussienne bivariable $f_{i,j}^k(y_s, y_{s-})$ sont :

$$\boldsymbol{\mu}_{ij}^k = \frac{1}{|O_{ij}^k|} \sum_{(y_s, y_{s-}) \in O_{ij}^k} \begin{pmatrix} y_s \\ y_{s-} \end{pmatrix} \quad (3.40)$$

$$\Sigma_{ij}^k = \frac{\sum_{(y_s, y_{s-}) \in O_{ij}^k} \left(\begin{pmatrix} y_s \\ y_{s-} \end{pmatrix} - \boldsymbol{\mu}_{ij}^k \right) \left(\begin{pmatrix} y_s \\ y_{s-} \end{pmatrix} - \boldsymbol{\mu}_{ij}^k \right)^T}{|O_{ij}^k|} \quad (3.41)$$

Les paramètres de la densité $f_{i,j}^k(y_{s-})$ se déduisent directement de celle de $f_{i,j}^k(y_s, y_{s-})$. L'algorithme Algo. 3.5, p. 73 synthétise les tâches requises pour la segmentation non supervisée sur le quadarbre de Markov couple dans le cas gaussien.

Algo. 3.5 Segmentation non supervisée sur le quadarbre de Markov couple : cas gaussien

Initialisation : Estimée initiale $\Theta^{[0]}$

Répéter

- Passe montante sur le quadarbre : calcul des probabilités $\beta(x_s) = p(y_{>s}/z_s)$ [Eq. 1.49 et 1.50]
- Passe descendante sur le quadarbre : calcul des probabilités *a posteriori* $p(x_s/y)$ [Eqs. 1.51 et 1.53]
- Obtention de la carte de segmentation intermédiaire $\hat{x}^{[q]}$ suivant le critère MPM
- Mise à jour des paramètres *a priori* Θ_x [Eqs. 3.37 et 3.38]
- Mise à jour des paramètres de l'attache aux données [Eqs. 3.40 et 3.41] sur les ensembles O_{ij}^k

Jusqu'à convergence

- Obtention de la carte de segmentation $\hat{x}_s, s \in S$ suivant le critère MPM [Eq. 1.42]

Sortie : Carte de segmentation $\hat{x}_s, s \in S$

3.4.3 Estimation des paramètres de l'attache aux données : cas gaussien généralisé

L'estimation dans le cas gaussien généralisé, se base toujours sur les ensembles O_{ij}^k . Les paramètres de la densité gaussienne généralisée bivariable $f_{i,j}^k(y_s, y_{s-})$ sont les paramètres

des marginales gaussiennes généralisés $f_{i,j}^k(y_s)$ et $f_{i,j}^k(y_{s-})$ et la matrice de corrélation. Pour chaque distribution marginale, on utilise les éléments correspondants des paires de l'ensemble O_{ij}^k avec l'estimation hybride de l'Eq. 2.46. La matrice de corrélation R_{ij}^k est estimée de manière classique à partir de l'ensemble :

$$\tilde{O}_{ij}^k = \{\tilde{\mathbf{t}}_s = (\tilde{\mathbf{t}}_s^1, \tilde{\mathbf{t}}_s^2)^t : \forall m = 1, 2, \tilde{\mathbf{t}}_s^m = \Phi^{-1}(F_i^k(\mathbf{y}_s^m)) \text{ avec } \mathbf{y}_s \in O_{ij}^k\}$$

où $\Phi(\cdot)$ est la fonction de répartition de la gaussienne monodimensionnelle centrée réduite et F_i^k la fonction de répartition de la marginale f_i^k . La mise à jour de R_i^k s'écrit alors :

$$R_i^{k[q+1]} = \sum_{s: \tilde{\mathbf{t}}_s \in \tilde{O}_{ij}^k} \tilde{\mathbf{t}}_s \tilde{\mathbf{t}}_s^t \quad (3.42)$$

L'algorithme 3.6, p. 74 récapitule la procédure de segmentation non supervisée sur le quadarbre de Markov couple dans le cas gaussien généralisé.

Algo. 3.6 Segmentation non supervisée sur le quadarbre de Markov couple : cas gaussien généralisé

Initialisation : Estimée initiale $\Theta^{[0]}$

Répéter

- Passe montante sur le quadarbre : calcul des probabilités $\beta(x_s) = p(y_{>s}/z_s)$ [Eq. 1.49 et 1.50]
- Passe descendante sur le quadarbre : calcul des probabilités *a posteriori* $p(x_s/y)$ [Eqs. 1.51 et 1.53]
- Obtention de la carte de segmentation intermédiaire $\hat{x}^{[q]}$ suivant le critère MPM
- Mise à jour des paramètres *a priori* Θ_x [Eqs. 3.37 et 3.38]
- Mise à jour des paramètres de la densité copule [Eq. 3.42] sur les ensembles \tilde{O}_{ij}^k
- Mise à jour des paramètres des marginales gaussiennes généralisée [Eq. 2.46] sur les ensembles O_{ij}^k

Jusqu'à convergence

- Obtention de la carte de segmentation $\hat{x}_s, s \in S$ suivant le critère MPM [Eq. 1.42]

Sortie : Carte de segmentation $\hat{x}_s, s \in S$

3.4.4 Résultats sur les images synthétiques

Nous avons généré deux images de tailles 256×256 chacune pour tester les performances du quadarbre couple. La première représente deux classes gaussiennes de paramètres donnés à la Tab. 3.3-a, et la deuxième correspond à deux classes gaussiennes généralisées de paramètres donnés à la Tab. 3.3-b (Fig. 3.4 ligne 1). Pour donner un sens à notre test, les observations doivent être injectés sur deux échelles au moins dans le quadarbre couple. Nous avons donc généré 2 échelles supplémentaires par filtrage passe-bas et décimation. Les données multirésolutions sur trois échelles ainsi obtenues sont alors segmentées en deux classes par le quadarbre de Markov caché avec la vraisemblance gaussienne (G-HMT), le quadarbre de Markov couple avec des termes d'attache aux données gaussiens (G-PMT) et la quadarbre de Markov couple avec des termes d'attache aux données gaussiens généralisés (GG-PMT). Les résultats sont présentés à la Fig. 3.4. On constate que le modèle couple (Fig.3.4 lignes 3 and 4) surpasse le modèle caché (Fig.3.4 ligne 2) dans tous les cas. De plus, pour le modèle couple il est clair que la modélisation gaussienne généralisée apporte une nette amélioration.

Les résultats obtenus sur les images synthétiques montrent l'efficacité du modèle du quadarbre de Markov couple. De plus, son adaptation aux données multirésolution suggère son utilisation dans le domaine ondelettes. Nous avons utilisé ce modèle pour le débruitage d'images basé sur une transformée multirésolution. La Sect. 4.5, p. 104, du prochain chapitre est consacrée à cette application.

	μ	σ
Classe 1	124	16
Classe 2	132	16

(a)

	μ	σ	γ
Classe 1	124	16	1.5
Classe 2	132	16	2.5

(b)

TAB. 3.3 – Paramètres des classes : (a) cas gaussien (b) cas gaussien généralisé

3.5 Régularisation de l'algorithme MACPP par le quadarbre de markov caché

3.5.1 Estimation des paramètres

Dans le chapitre précédent nous avons présenté le cadre général de la régularisation de la MACPP pour la segmentation d'images multispectrales Sect. 2.4.3. Dans cette section, on considère le quadarbre de Markov caché comme modèle pour le champ des étiquettes X . La probabilité jointe des étiquettes s'écrit alors :

$$p(x) = p(x_r) \prod_{s \neq r} p(x_s/x_{s-}) \tag{3.43}$$

De plus, à l'échelle la plus fine (bas du quadarbre), chaque état caché de x_s , $s \in S^0$, est lié à un vecteur d'observation \mathbf{y}_s . Les corrélations entre pixels voisins sont ainsi prises en

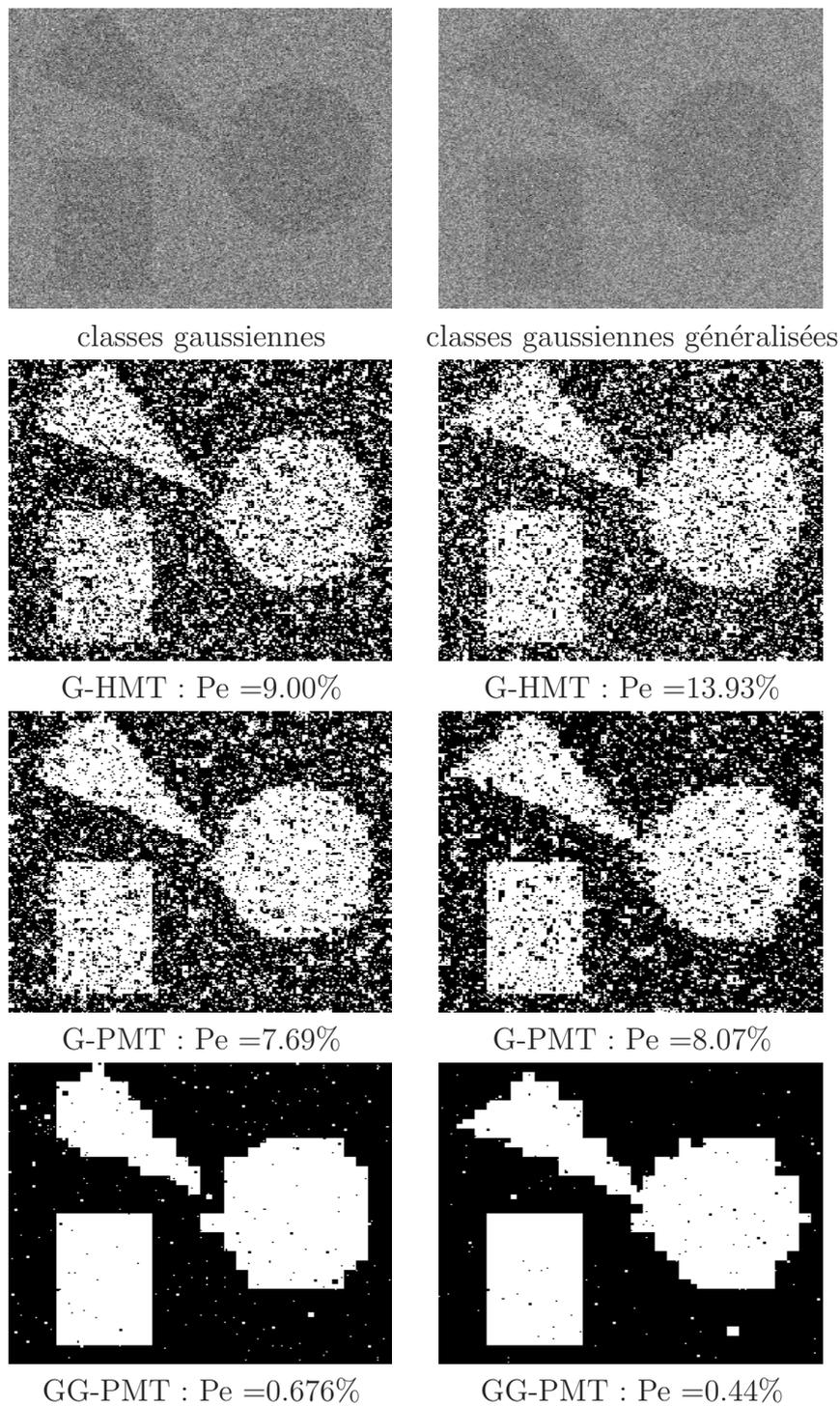


FIG. 3.4 – Résultats de Segmentation : dans l'ordre lexicographique, les images simulées avec les classes gaussiennes et gaussiennes généralisées, la segmentation avec le quadarbre de Markov caché avec une vraisemblance gaussienne (G-HMT), la segmentation avec le quadarbre de Markov couple avec une vraisemblance gaussienne (G-PMT) et segmentation avec le quadarbre de Markov couple avec une vraisemblance gaussienne généralisée (GG-PMT). "Pe" est le taux d'erreurs.

compte d'une manière hiérarchique. L'expression de la probabilité jointe est donnée par :

$$p(x, y/\Theta) = p(x_r/\Theta_x) \prod_{s \neq r} p(x_s/x_{s-}, \Theta_x) \prod_{s \in S^0} p(\mathbf{y}_s/x_s, \Theta_y) \quad (3.44)$$

avec le terme d'attache aux données fourni par :

$$P(\mathbf{y}_s/x_s = \omega_i, \Theta_y) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp\left\{\frac{-1}{2}(\mathbf{y}_s - \boldsymbol{\mu}_i)^t C_i^{-1}(\mathbf{y}_s - \boldsymbol{\mu}_i)\right\} \quad (3.45)$$

où $C_i = \sigma_i^2 I + A_i A_i^t$, A_i étant la matrice de transformation de dimension $d \times r_i$ associée à l'ACPP de la $i^{\text{ème}}$ classe.

La fonction auxiliaire correspondante a la même forme que l'Eq. 3.19. L'optimisation des paramètres du modèle *a priori*, étant la même que dans le quadarbre de Markov caché, fournit les mises à jour semblables aux Eqs. 3.22 et 3.25. Pour les paramètres de l'attache aux données, caractérisant les ACPP locales, nous suivons la démarche de Tipping et *al.* [137]. On réécrit le terme correspondant de la fonction auxiliaire comme suit :

$$\begin{aligned} & \sum_{s \in S^0} \sum_i \xi_s^{[q]}(i) \log f_i^k(\mathbf{y}_s) \\ = & \sum_i -\frac{\sum_{s \in S^0} \xi_s^{[q]}(i)}{2} [N \log(2\pi) + \log \det(C_i) + \text{tr}(C_i^{-1} \Upsilon_i)] \end{aligned} \quad (3.46)$$

avec

$$\Upsilon_i = \frac{1}{\sum_{s \in S^0} \xi_s^{[q]}(i)} \sum_{s \in S^0} \xi_s^{[q]}(i) (\mathbf{y}_s - \boldsymbol{\mu}_i)(\mathbf{y}_s - \boldsymbol{\mu}_i)^t \quad (3.47)$$

La dérivation de ce terme par rapport à $\boldsymbol{\mu}_i$ donne l'estimateur classique :

$$\boldsymbol{\mu}_i^{[q+1]} = \frac{\sum_{s \in S^0} \mathbf{y}_s \xi_s^{[q]}(i)}{\sum_{s \in S^0} \xi_s^{[q]}(i)} \quad (3.48)$$

De même, la dérivation par rapport à A_i donne [81] :

$$\frac{\partial \left[\sum_{s \in S^0} \sum_i \xi_s^{[q]}(i) \log f_i^k(\mathbf{y}_s) \right]}{\partial A_i} = \sum_{s \in S^0} \xi_s^{[q]}(i) (C_i^{-1} \Upsilon_i C_i^{-1} A_i - C_i^{-1} A_i) \quad (3.49)$$

Il est montré dans [137] que la seule solution non nulle à cette équation est donnée par :

$$A_i^{[q+1]} = U_{r_i} (\Lambda_{r_i} - \sigma_i^2 I)^{\frac{1}{2}} \Phi_i \quad (3.50)$$

avec

$$\sigma_i^2 = \frac{1}{N - r_i} \sum_{j=r_i+1}^N \lambda_j \quad (3.51)$$

où les λ_j sont les valeurs propres de Υ_i ordonnées en valeurs décroissantes ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$), Λ_{r_i} est la matrice diagonale des r_i premières valeurs propres, U_{r_i} est la matrice des r_i premiers vecteurs propres correspondants et Φ_i est une matrice de rotation orthogonale $r_i \times r_i$ quelconque.

L'algorithme Algo. 3.7, p. 78 résume la procédure de segmentation non supervisée avec l'ACPP régularisée.

Algo. 3.7 Segmentation non supervisée avec la MACPP régularisée par le quadarbre de Markov caché : cas gaussien

Initialisation : Estimée initiale $\Theta^{[0]}$

Répéter

- Passe montante sur le quadarbre : calcul des probabilités $p(x_s/\mathbf{y}_{\geq s})$ [Eq. 1.35]
- Passe descendante sur le quadarbre : calcul des probabilités *a posteriori* $h_s(i, j)$ et $\xi_s(i)$ [Eqs. 1.39-1.41]
- Mise à jour des paramètres *a priori* Θ_x [Eqs. 3.22-3.25]
- Mise à jour des paramètres de l'attache aux données Θ_y [Eqs. 3.48-3.51]

Jusqu'à convergence

- Obtention de la carte de segmentation $\hat{x}_s, s \in S$ suivant le critère MPM [Eq. 1.42]

Sortie : Carte de segmentation $\hat{x}_s, s \in S$

3.5.2 Résultats sur des images synthétiques

Images non corrélées

Pour tester notre approche, nous avons généré 3 ensembles de trois images non corrélées. Chaque ensemble représente deux classes gaussiennes correspondant à un objet géométrique plus le fond avec les paramètres de la Tab. 3.4. On obtient ainsi les 9 images en haut de la Fig. 3.5. La carte de segmentation en 4 classes avec le maximum de vraisemblance basée sur la MACPP et MACPP régularisée sont données en bas de de la Fig. 3.5. Le dimension de chaque ACPP est fixée à 1. Il est très facile de noter l'effet de l'ajout de la régularisation markovienne où 95 % des pixels sont correctement classés.

	classe 1		classe 2	
	μ_1	σ_1	μ_2	σ_2
Bande 1	7.75	1	8.25	1
Bande 2	7.75	1	8.25	1
Bande 3	8	1	8	1

TAB. 3.4 – Paramètres des classes gaussiennes. Dans la troisième bande de chaque ensemble les classes sont complètement confondues

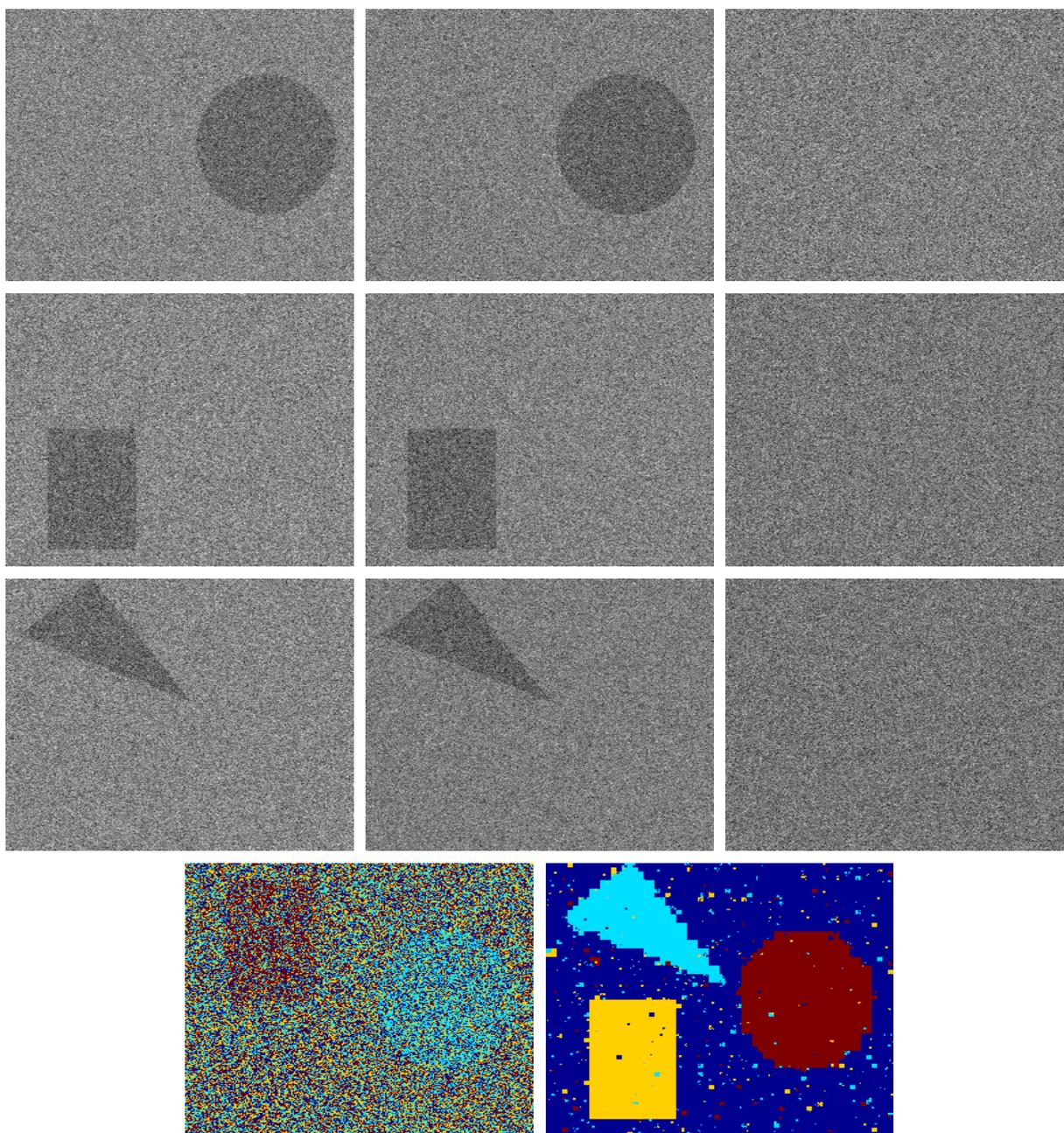


FIG. 3.5 – Résultats de segmentation des images non corrélées. Les images synthétiques en haut. Chaque trois images de la même ligne sont générées avec les paramètres de la Tab. 3.4. En bas, la carte de segmentation obtenue avec le maximum de vraisemblance basé sur l'ACPP (à gauche) est très chahutée, tandis que celle obtenue avec l'ACPP régularisée (à droite) présente un taux d'erreur de 5.39 %

Images corrélées

De la même façon que pour le cas non corrélé, nous avons généré avec le paramètre de la Tab. 3.4 trois ensembles de trois image corrélées avec un coefficient de corrélation 0.8. On obtient ainsi les 9 images en haut de la Fig. 3.6. Les cartes de segmentation en quatre classes avec le maximum de vraisemblance basé sur la MACPP et MACPP régularisée sont données en bas de de la Fig. 3.6 où l'on constate que 98.9 % des pixels sont correctement classés avec l'ACPP régularisée. Des résultats supplémentaires sur des images réelles en télédétection (région de Strasbourg) et en astronomie (32 bandes dans le domaine du visible autour la raie de d'oxygène OIII (489nm à 506nm) de l'étoile Eta Carinae) seront présentés au prochain chapitre (Sect. 4.3.3, p. 95).

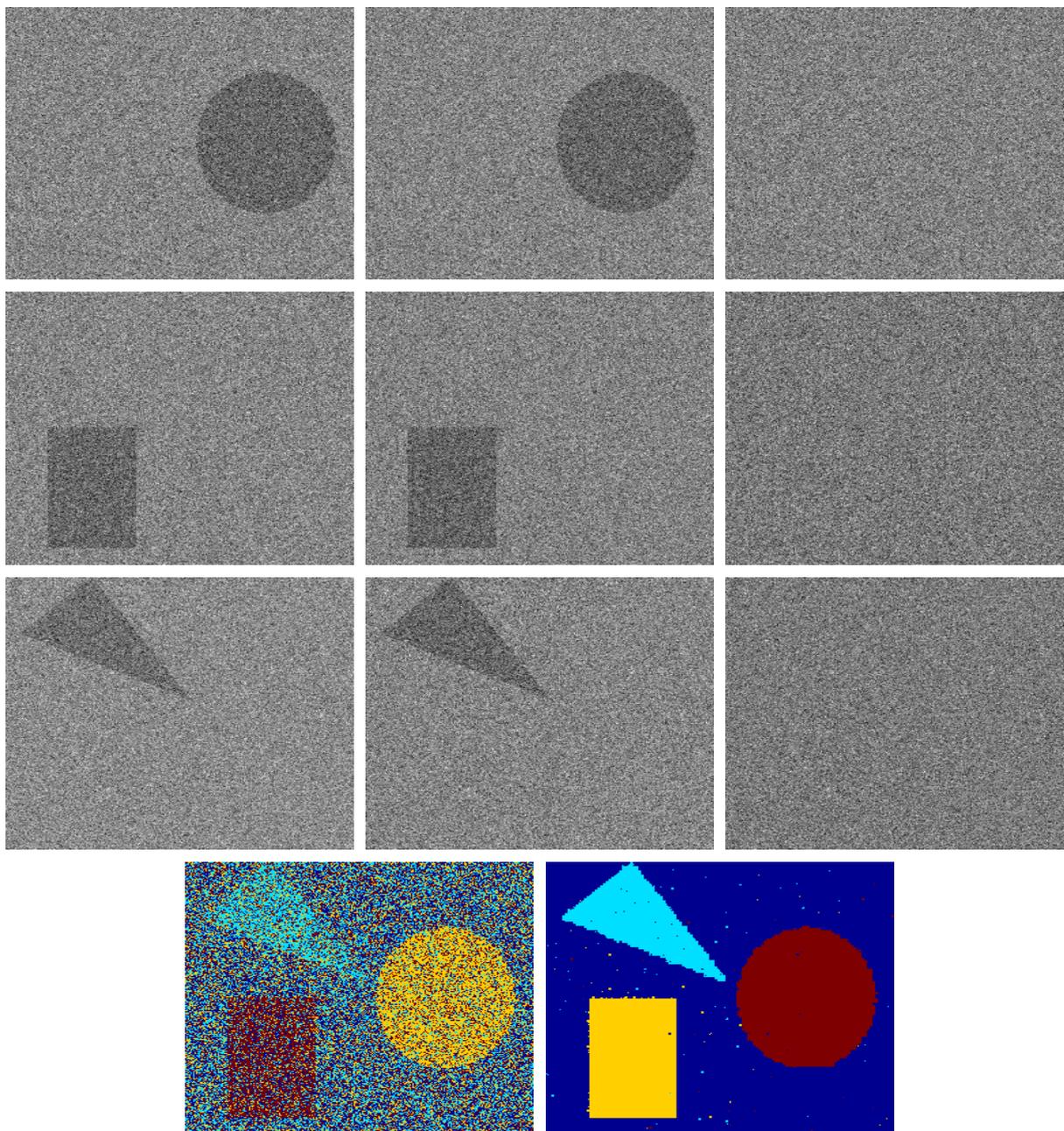


FIG. 3.6 – Résultats de segmentation des images corrélées (coefficient de corrélation interbandes de 0.8). Les images synthétiques en haut. Chaque trois images de la même ligne sont générées avec les paramètres de la Tab. 3.4. En bas, la carte de segmentation obtenue avec le maximum de vraisemblance basé sur l'ACPP (à gauche) est très chahutée, tandis que celle obtenue avec l'ACPP régularisée (à droite) présente un taux d'erreur de 0.012 %

3.6 Conclusion

Dans ce chapitre nous avons présenté une courte synthèse des algorithmes d'estimation de paramètres. Ceci nous a permis d'introduire les procédures d'estimation de paramètres associées aux modèles du quadarbres de Markov caché et couple, avec les vraisemblance gaussiennes et gaussiennes généralisées, ainsi que la l'ACPP régularisée. La présentation des algorithmes de segmentation non supervisés avec tous ces modèles, ont naturellement suivi. Nous avons présenté dans chaque cas des résultats sur des images synthétiques qui ont démontré la pertinence de ces méthodes. Le prochain chapitre sera consacré à quelques applications en imagerie astronomiques.

4 Application à l'imagerie astronomique

4.1 Introduction

Dans les chapitres précédents, nous avons décrit en détails les algorithmes de segmentation non supervisée pour les arbres de Markov cachés et couplés avec le critère MPM. Nous avons évoqué le problème de la dimension, présenté quelques techniques de réduction que nous avons utilisées et détaillé la modélisation du terme d'attache aux données dans le cas multidimensionnel gaussien et gaussien généralisé. Ces procédures ont été validées sur des images synthétiques. Dans ce chapitre nous mettrons en pratique ces modèles pour la segmentation, le débruitage et la fusion d'images astronomiques.

Nous commencerons par la présentation de quelques spécificités des images astronomique et de l'intérêt des outils de traitement d'images dans ce domaine. Les applications décrites dans ce chapitre se répartissent en deux grandes parties. La première concerne des applications de réduction/segmentation d'images multispectrales où les schémas de réduction que nous avons décrits à la Sect. 2.4 sont couplés avec le quadarbre de markov caché. La deuxième partie traite l'application des quadarbres de Markov aux données multirésolutions. Nous rappellerons brièvement la théorie de l'analyse utilisée puis nous présenterons l'intérêt de la modélisation par les arbres de Markov. Ceci nous permettra de présenter les deux techniques de débruitage d'images par sélection des coefficients pertinents et par rétrécissement des coefficients de détails. Enfin nous exposerons la fusion d'images multispectrales par arbres de Markov et analyse multirésolution.

4.2 Images astronomiques

L'observation multi-longueur d'onde du ciel offre aux astronomes la possibilité de voir les objets astronomiques sous différents aspects, de détecter certaines de leurs caractéristiques telles que le rayonnement émis ou la présence d'éléments chimiques, et d'appréhender leurs mouvements. D. Egret [46] donne un petit exemple de l'utilité de telles observations. Par exemple, l'étude d'étoiles en cours de formation dans la constellation de l'Aigle par le télescope Antu de l'observatoire européen austral (ESO) est gênée par la présence de cocons de poussières les entourant. Il est alors possible de percer ce voile en utilisant deux autres instruments : le télescope spatial européen ISO, car ses détecteurs sont sensibles aux rayonnements infrarouges qui traversent les nuages de poussières, et Chandra, le satellite de la NASA qui capte les rayons X émis par les astres les plus jeunes

de notre Galaxie. Il faut cependant remarquer que si Chandra est encore en activité, ISO n'est plus qu'un débris dans l'espace circumterrestre. Malgré cela, ses observations du ciel sont bien archivées et mises à disposition des astronomes par quelques centres de données astronomiques dont fait parti le CDS à Strasbourg. La coopération entre ces centres, les laboratoires de recherche en astrophysique, les laboratoires de traitement de données et les centres de documentation fait l'objet d'un effort mondial pour concevoir un outil permettant une meilleure exploitation des données présentes et passées de tous les détecteurs astronomiques du monde, ainsi que toute la documentation déjà existante, c'est ce que l'on appelle l'Observatoire Virtuel (VO). Comme son nom l'indique, ce projet se veut une fenêtre virtuelle sur le ciel qui permet à un utilisateur d'accéder depuis son bureau, sur une simple requête, et de manière transparente à toutes les observations, les catalogues et les articles de journaux disponibles dans tous les centres astronomiques de la planète grâce à la technologie GRID¹.

Le besoin au VO s'accroît d'autant que les télescopes, au sol ou dans l'espace, couvrent enfin l'ensemble du spectre électromagnétique, des ondes radio aux rayons gamma très énergétiques, avec toute l'accumulation des données qui en découle. Les prévisions pour les nouveaux télescopes en construction durant les deux années à venir parlent d'une quantité énorme de données à étudier, jusqu'à 5 000 milliards d'octets chaque nuit. Une partie importante de cette avalanche de données risque de rester inexploitée. De ce point de vue l'apport attendu des outils de traitements automatiques des données est donc crucial. Ces traitements ne visent pas à remplacer le travail des astronomes mais plutôt à le faciliter en synthétisant l'information des grands cubes de données et/ou en facilitant la détection de zones d'intérêt. Cela concerne l'intégration des méthodes classiquement utilisées par les astronomes ainsi que celles développées dans le domaine du traitement de l'information et adaptées aux images astronomiques.

En effet les images astronomiques présentent quelque caractéristiques générales [107] qui sont : la présence du bruit, des sources ponctuelles (étoiles ou objets ressemblants), des objets plus ou moins étendus (galaxies et nébuleuse résolues, planètes et autres objets de système solaire), des objets mats (faibles rayonnement), des structures diffuses et des objets superposés. Ajouté à cela l'aspect multi-longueur d'ondes, les outils d'analyse en imagerie astronomique sont donc confrontés à une grande variété de problèmes parmi lesquels la réduction de dimensionnalité, la restauration, la segmentation, la détection, la classification d'objets, la fusion, la reconstruction 3D et la compression.

Notons enfin que le développement d'outils originaux et adaptés aux besoins de l'astronomie nécessite la collaboration des spécialistes du traitement de l'information et des astronomes pour intégrer le plus possible les connaissances disponibles *a priori* dans les modèles de traitement, et valider les techniques développées d'un point de vue astrophysique.

¹Globalisation des Ressources Informatiques et des Données

4.3 Réduction/segmentation

4.3.1 Détection de galaxie lointaines à faible rayonnement

Nous avons appliqué la technique de réduction par regroupement de bandes et projections locales, décrite à la Sect. 2.4.1, p. 41, aux images de galaxies lointaines situées dans la zone "Chandra Deep Field south". Il s'agit de 12 bandes spectrales (Fig. 4.1) de galaxies à faible rayonnement. Ces données très hétérogènes car issues de trois capteurs différents à savoir le WFI², le HST-ACS³ et ISAAC⁴ dont les longueurs d'ondes d'observation vont de l'infrarouge à l'ultraviolet. Elles ont été recalées et rééchantillonnées à la résolution la plus fine, en se basant sur les correspondances entre les positions dans le ciel et les coordonnées dans les images, et des outils d'interpolation comme les fonctions B-splines. Le but est d'isoler les galaxies du fond.

Bande	nom	capteur	long. d'onde en microns	résol. init. en arsec/pixel
1	U38	WFI	0.344 - 0.382	0.24
2	B99	WFI	0.406 - 0.565	0.24
3	V89	WFI	0.495 - 0.585	0.24
4	RC162	WFI	0.57 - 0.73	0.24
5	ICLWP	WFI	0.7 - 1.1	0.24
6	F435W	HST ACS	0.435	0.05
7	F606W	HST ACS	0.606	0.05
8	F775W	HST ACS	0.775	0.05
9	F850LP	HST ACS	0.850	0.05
10	J	ISAAC	1.11 - 1.36	0.14
11	H	ISAAC	1.5 - 1.8	0.14
12	KS	ISAAC	2.0 - 2.32	0.14

TAB. 4.1 – Détail des 12 bandes observées

Groupes	1	2	3	4
Bandes	2,9	6,11,5,8,10,3	7,8	4,12

TAB. 4.2 – Groupement des 12 bandes de la Fig. 4.1 en 4 clusters.

Le nombre de classes a été estimé à 4 par l'astronome. Nous avons donc choisi un nombre de clusters égal au nombre de classes pour l'étape de regroupement. En utilisant la technique de coalescence décrite à la Sect. 2.4.1, p. 41, nous avons abouti à la partition de la Tab. 4.2.

Les projections locales sur chaque groupe d'images avec l'ACP ou l'ACI ont donné les images résultats des Figs. 4.2 et 4.3. La segmentation des 4 images résultantes de la réduction par le quadarbre de Markov caché avec un terme d'attache aux données gaussien montrent dans le cas de l'ACI une meilleure détection des zones d'intérêt. Cependant le fond reste assez sur-segmenté. Néanmoins, en augmentant le nombre de clusters à 5 (Tab.

²Wide Field Imager de l'observatoire européen installé au Chili

³Advanced Camera for Surveys installée sur le Hubble Space Telescope

⁴Infrared Spectrometer And Array Camera

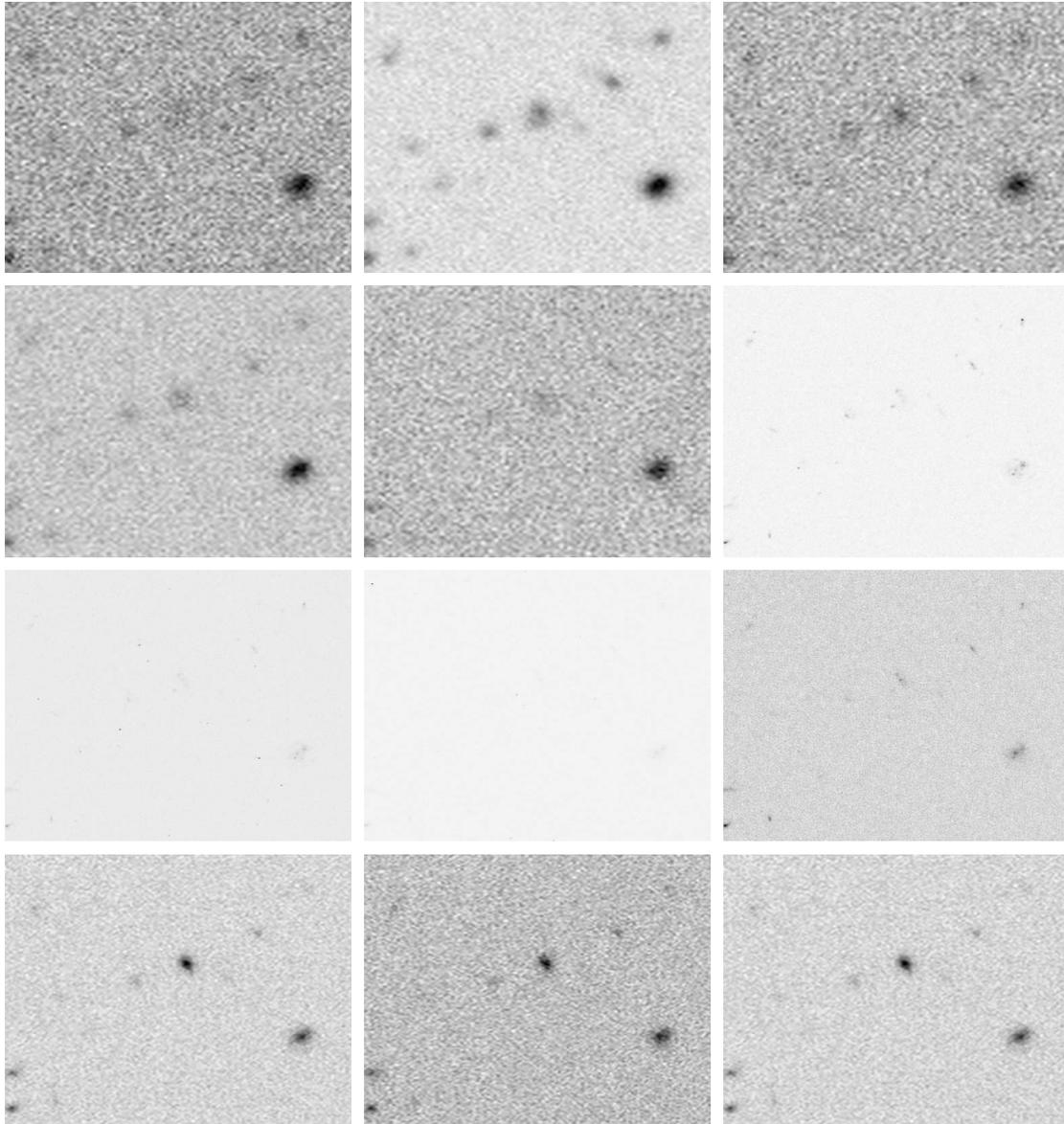


FIG. 4.1 – 12 bandes spectrales de galaxies lointaines situées dans la zone Chandra Deep Field South.

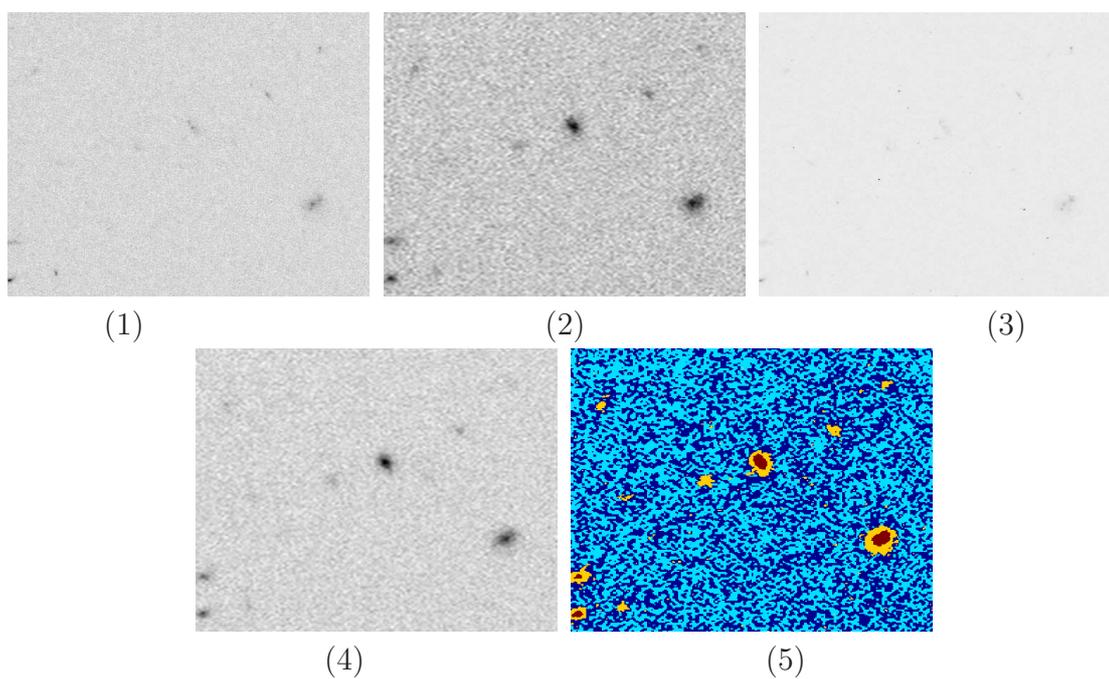


FIG. 4.2 – Résultats de la réduction avec le regroupement en 4 clusters (Tab. 4.2) et projection sur le premier axe de l'ACP (images 1,2,3,4) et carte de segmentation à 4 classes obtenue avec le quadarbre de markov caché avec une attache aux données gaussienne multidimensionnelle (image 5)

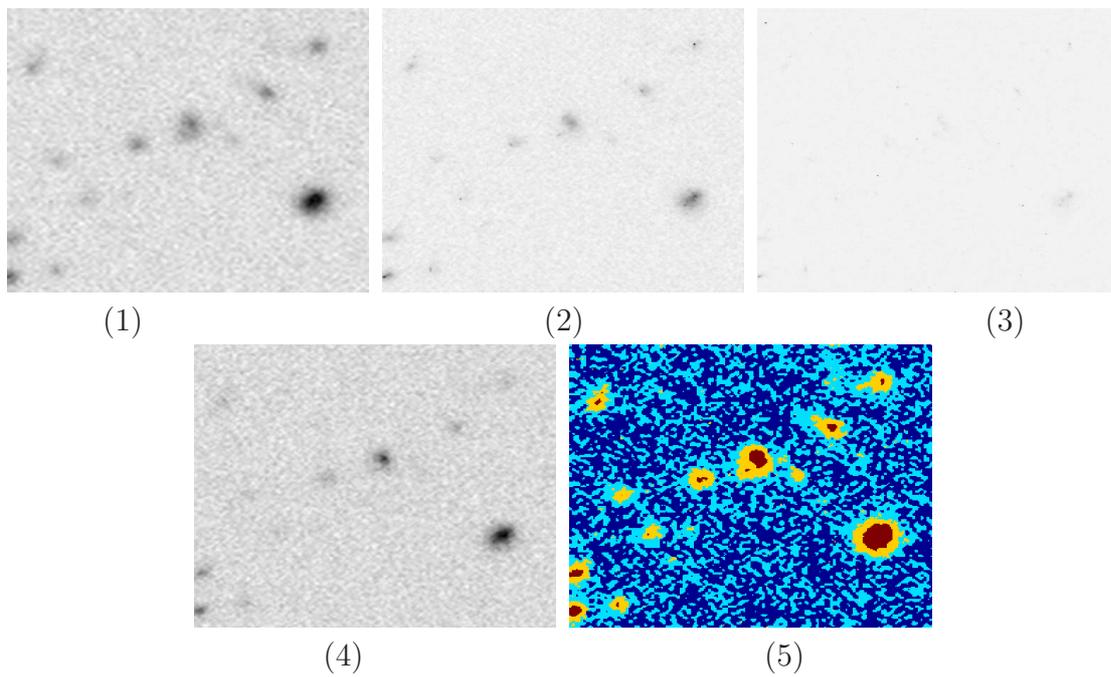


FIG. 4.3 – Résultats de la réduction avec le regroupement en 4 clusters (Tab. 4.2) et projection par ACI avec l'algorithme FastICA (images 1,2,3,4) et carte de segmentation à 4 classes obtenue avec le quadarbre de markov caché avec une attache aux données gaussienne multidimensionnelle (image 5)

4.3), nous constatons une amélioration sur la segmentation avec l'ACP (Figs. 4.4 et 4.5). Ceci est dû au fait que l'augmentation du nombre de clusters a permis d'isoler la bande 2 qui paraît la plus informative entre toutes et la moins bruitée.

Groupes	1	2	3	4	5
Bandes	6,11,5,8,10,3	7,8	4,12	9	2

TAB. 4.3 – Groupement des 12 bandes de la Fig. 4.1 en 5 clusters.

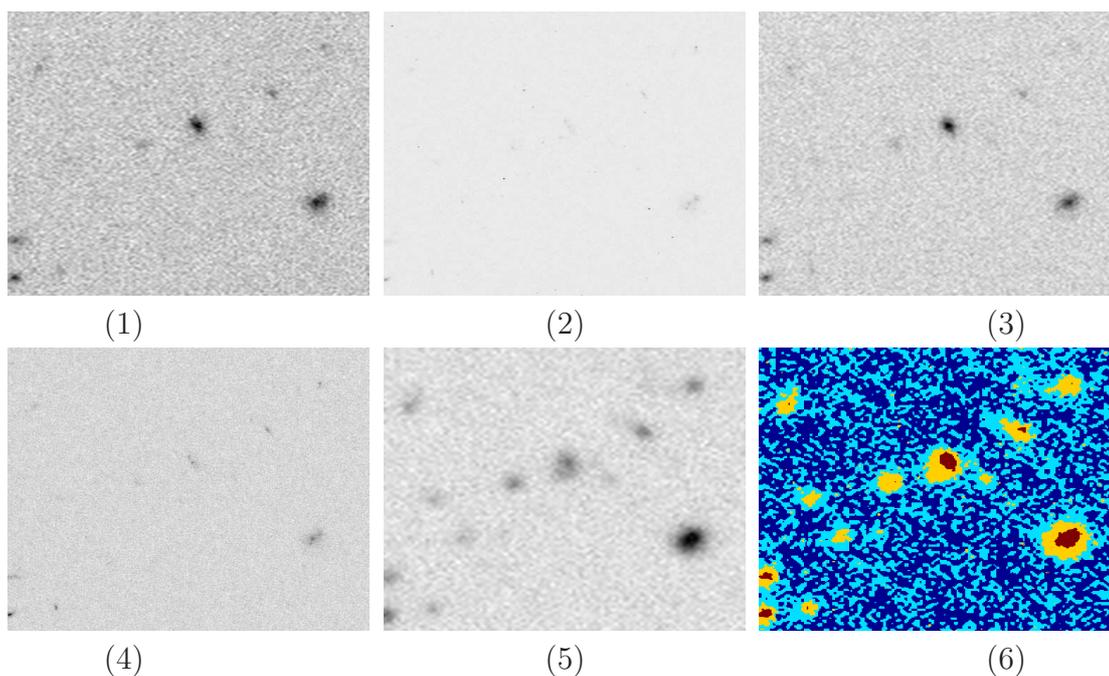


FIG. 4.4 – Résultats de la réduction avec le regroupement en 5 clusters (Tab. 4.3) et projection sur le premier axe de l'ACP (images 1,2,3,4,5) et carte de segmentation à 4 classes obtenue avec le quadarbre de markov caché avec une attache aux données gaussienne multidimensionnelle (image 6)

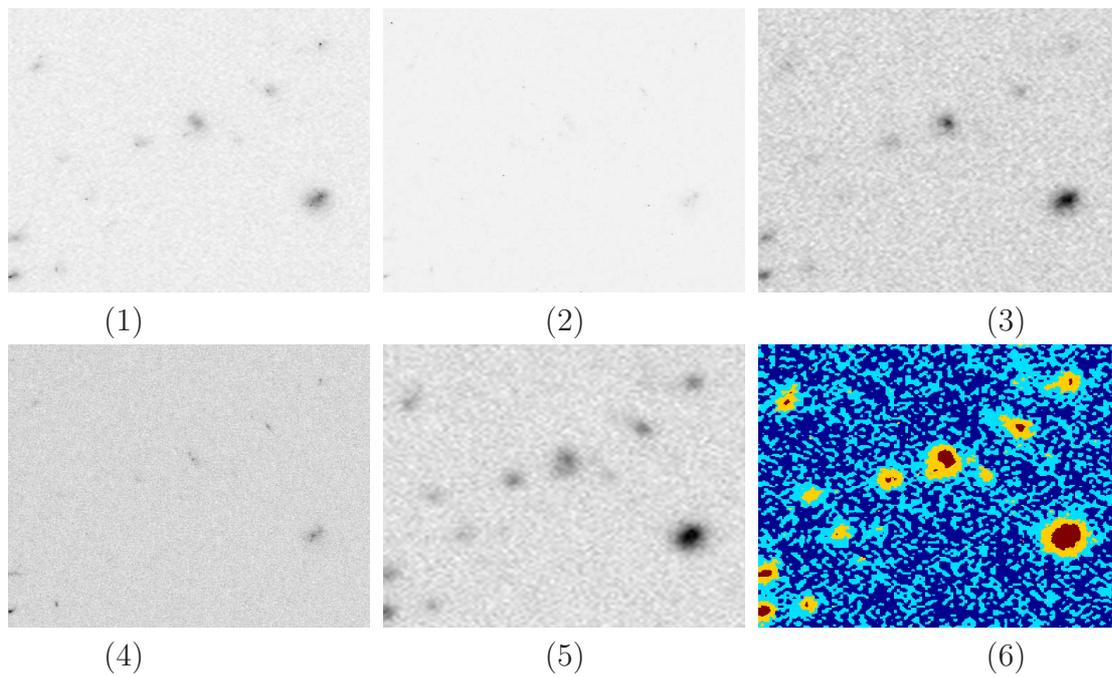


FIG. 4.5 – Résultats de la réduction avec le regroupement en 5 clusters (Tab. 4.3) et projection par ACI avec l'algorithme FastICA (images 1,2,3,4,5) et carte de segmentation à 4 classes obtenue avec le quadarbre de markov caché avec une attache aux données gaussienne multidimensionnelle (image 6)

4.3.2 Segmentation du cube radio de la raie HI de la galaxie spirale NGC4254

NGC4254 est une galaxie spirale située dans le cluster de la Vierge. L'observation de cette galaxie sur 42 bandes autour de la raie HI à 21cm est faite par le VLA (Very Large Array) [113]. La vitesse centrale du cube est $v = 2408 \text{ km s}^{-1}$ correspondant à la bande numéro 22 et la résolution de vitesse est de 10 km s^{-1} par bande. La Fig. 4.6 montre les trois premiers moments calculés sur l'intégralité du cube. La distribution de l'émission du HI (moment 0, image 1) montre un disque de gaz incliné avec un important bras spiral vers l'Est. Le champ de vitesses (moment 1, image 2) est celui d'un disque en rotation avec des perturbations dans le Nord-Est et le Nord. La distribution de dispersion de la vitesse (moment 2, image 3) est presque uniforme avec un maximum au centre de la galaxie. L'application de la technique de réduction par le mélange de gaussiennes décrite à

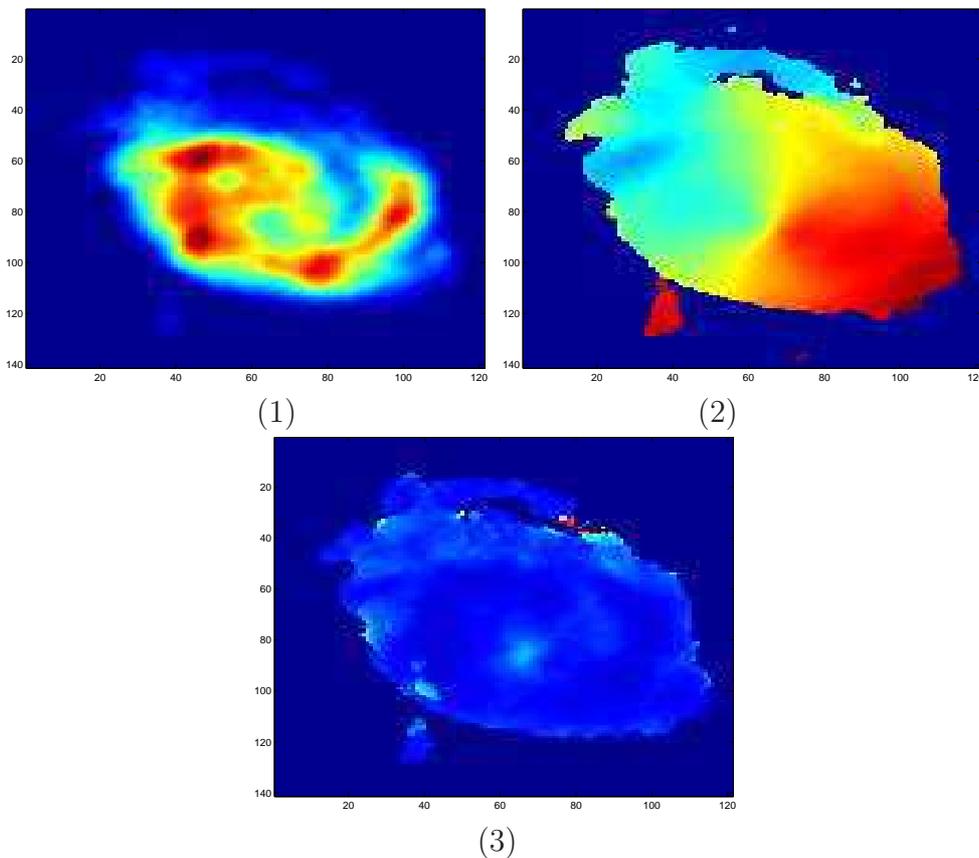


FIG. 4.6 – NGC4254 ([113]) : (1) la distribution de l'émission de la raie HI (moment 0). L'axe x correspond à l'ascension droite, l'axe y à la déclinaison. (2) le champ de vitesses (moment 1). (3) la dispersion des vitesses.

la Sect. 2.4.2, p. 44, avec un nombre de composantes égal à 6, nous a fourni les images paramètres de la Fig. 4.7 qui correspondent aux pondérations des gaussiennes choisies (Tab. 4.4).

La carte de segmentation à 7 classes, obtenue par arbre de Markov avec une attache

μ	12.3	16.0	20.3	25.0	29.2	33.0
σ	1.5	1.5	1.6	1.5	1.4	1.2

TAB. 4.4 – Base des six gaussiennes sélectionnées pour la réduction du cube HI de la galaxie spirale NGC4254

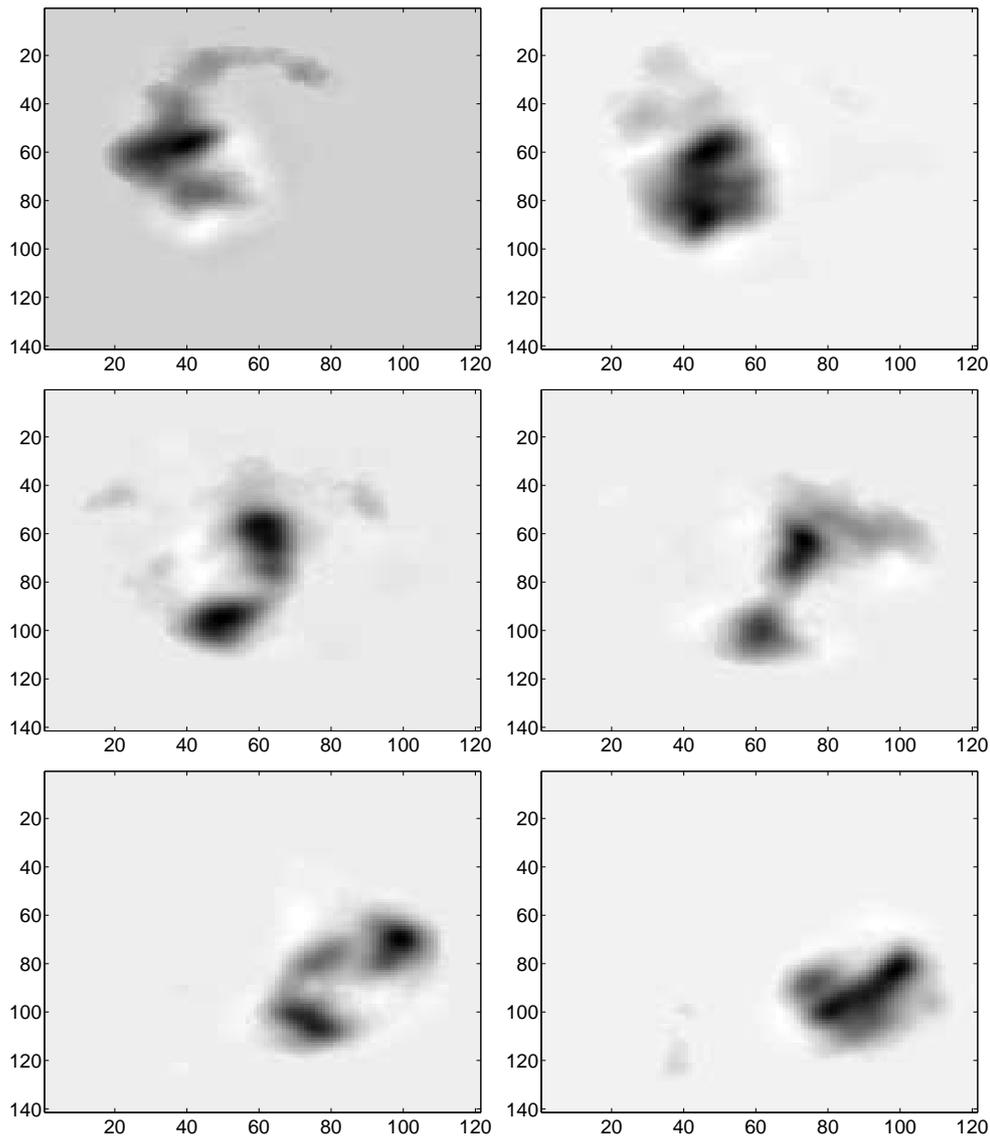


FIG. 4.7 – Les images paramètres correspondant aux poids issus de la modélisation du cube NGC4254 par un modèle de mélange des six gaussiennes de la Tab. 4.4. Ceci correspond à la réduction du cube original à 42 bandes en 6.

aux données gaussienne, ainsi que les spectres moyens par classe, réels et modélisés, sont présentés à la Fig. 4.8. Généralement les spectres réels sont bien approximés par les spectres modélisés, et les profils spectraux bien distincts pour toutes les classes. La comparaison entre la carte de segmentation et le champ de vitesses (Fig. 4.6-2) montre que la segmentation est faite généralement suivant la vitesse, *i.e.*, la position de la raie d'émission dans le spectre 1D de chaque pixel. La seule exception est la raie du profil spectral de la classe 4 située en dessous du profil moyen de la classe 5 qui est positionné au voisinage de la vitesse centrale du cube. Dans ce cas, la segmentation est également basée sur le maximum du profil spectral qui est, pour le cas de la classe 4, beaucoup plus petit que celui de la classe 5. Ainsi pour un nombre de classes suffisamment grand, la segmentation est également basée sur l'intensité maximale des profils spectraux.

Deux perturbations de la zone de vitesse dans la partie nord de la galaxie peuvent être identifiées dans la carte de segmentation de la Fig. 4.8 : (1) un écart au gradient symétrique de vitesse près du pixel (70,50) où la région correspondant aux classes 6 et 4 avance vers la région affectée à la classe 3, et (2) la région associée à la classe 1 dans le nord de la galaxie. Les deux perturbations sont identifiées en tant que classes distinctes (1 et 4) dans la Fig. 4.8. Nous concluons que la méthode fonctionne avec succès pour ce genre de cube. Pour ce cube relativement simple en données qui a servi comme validation, la carte de segmentation (Fig. 4.8) ne contient pas d'information supplémentaire à la carte des vitesses (Fig. 4.6-2). Cependant, la carte de segmentation pourrait être employée pour produire des masques isolant les régions cinématiquement intéressantes dans le nord de la galaxie (classe 1 et 4). Phookun et *al.* [113] ont dérivé une courbe de rotation à partir des données du cube, élaboré les diagrammes position-vitesse de la région Nord, et soustrait l'émission issue à partir du disque de gaz tournant pour isoler des régions perturbées cinématiquement. Notre méthode n'est pas destinée à remplacer ou concurrencer le travail d'analyse des astronomes mais à le faciliter [54].

4.3.3 Réduction/Segmentation jointe par la MACPP régularisée

La technique de MACPP régularisée par l'arbre de Markov caché, détaillée par l'Algo. 3.7, p. 78, est testée sur des images synthétiques à la Sect. 3.7, p. 78. Les résultats obtenus sont assez satisfaisants. Dans ce qui suit, nous allons appliquer cet algorithme sur deux images réelles, la première de télédétection, et la seconde astronomique.

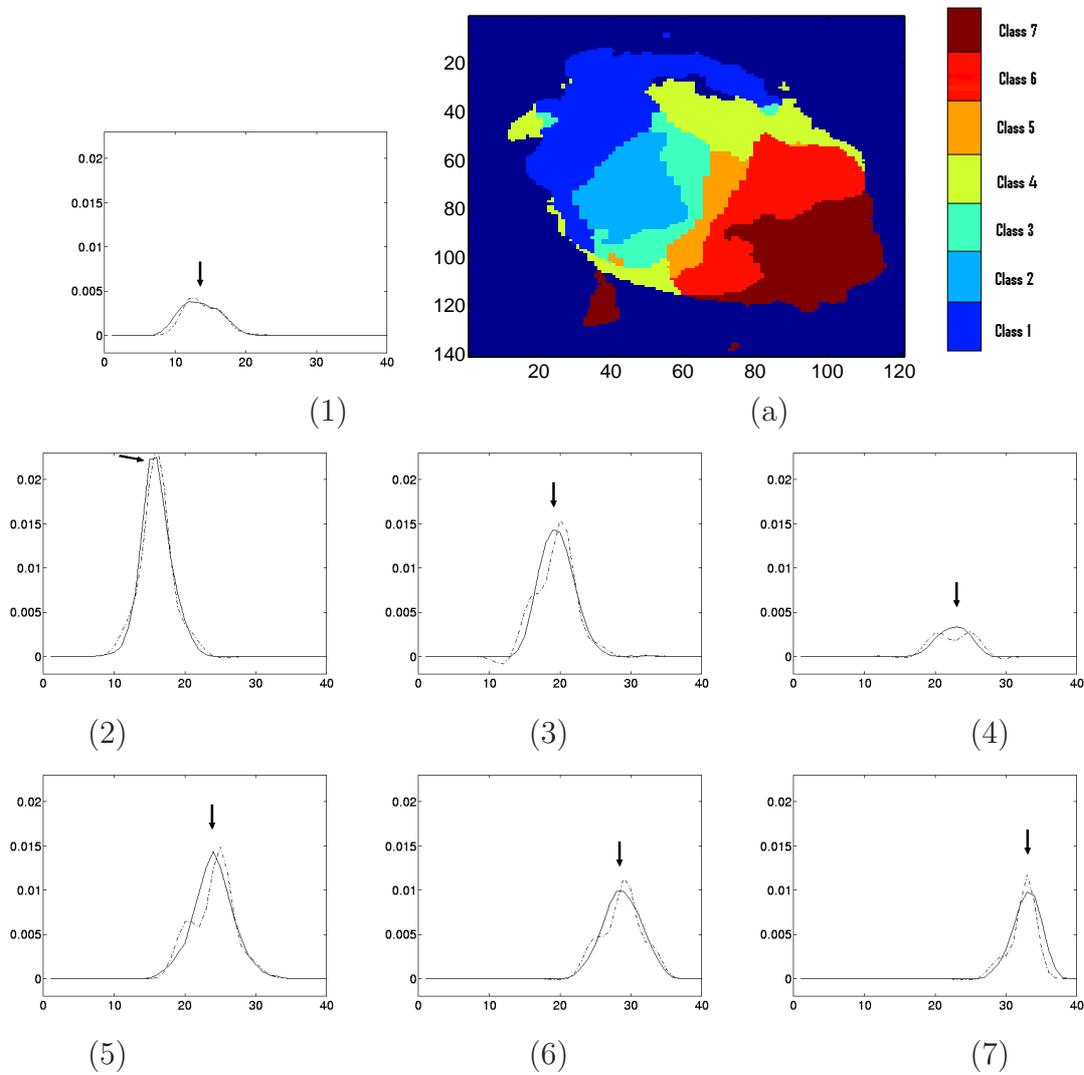


FIG. 4.8 – Résultats de segmentation en 7 classes des images réduites de la Fig. 4.7 : (a) carte de segmentation. (1-7) spectre moyen en chaque classe (trait plein : spectre observé, trait interrompu : spectre modèle). L'axe x représente la vitesse, et l'axe y l'intensité. Les pics des ces profils types sont indiqués par des flèches.

Image satellite 6-bandes de la région de Strasbourg

Il s'agit d'une image 6-bandes de télédétection de la région de Strasbourg (Fig. 4.9). Les longueurs d'ondes correspondantes sont respectivement : 1011, 1250, 1802, 2254, 1802, 2254 nm. Nous avons réalisé trois segmentations avec la MACPP régularisée avec un nombre croissant de classes. Les résultats sont présentés à la Fig. 4.10. La dimension réduite au sein de chaque ACPP est fixée à 2. On arrive à distinguer clairement dans la carte de segmentation à 4 classes les différentes régions présentes dans la scène : l'eau, végétation (séparée en deux classes) et les chemins entre les champs. Lorsqu'on augmente le nombre de classes (5 ou 6), la végétation est divisée en plusieurs zones en accord avec leurs comportements dans les différentes longueurs d'ondes. Les résultats nous semblent satisfaisants mais peuvent néanmoins être améliorés en utilisant une régularisation plus adaptée aux images de télédétection, présentant des formes géométriques aux contours précis, comme les champs de Markov.

Cube à 32 bandes autour de la raie OIII de l'étoile Eta Carinae

Il s'agit de 32 bandes dans le domaine du visible autour de la raie de d'oxygène OIII (489nm à 506nm) de l'étoile Eta Carinae (Fig. 4.11). On remarque principalement, la présence d'étoiles (sources ponctuelles) dans les bandes du début et de la fin du cube (Fig. 4.11 bandes 3 et 26) et d'un nuage (émission diffuse) dû à une nébuleuse dans les bandes centrales du cube (Fig. 4.11 bandes 12,14,16 et 18). Les moments d'ordres inférieurs à trois, du cube clipé⁵ à 3σ , sont donnés à la Fig. 4.12. On remarque la présence des étoiles dans le moment 0. Le moment 1 est peu informatif, tandis que le moment 2 présente des variations notables spécialement au centre de l'image. Nous avons opéré une segmentation à 6 et 8 classes avec la MACPP régularisée où la dimension de l'ACPP locale de chaque classe est fixée à 2. Les résultats de segmentation à 6 et à 8 classes sont donnés à la Fig. 4.13. Nous retrouvons dans les deux cartes les deux structures qui sont l'ensemble d'étoiles et le gaz de la nébuleuse, avec une meilleure séparation dans le cas de la segmentation 8 classes. Une plus profonde interprétation d'un point de vue astrophysique nécessite l'intervention d'un astronome.

⁵ce seuillage a été effectué pour rendre plus claires les cartes des moments. La segmentation cependant est faite sur le cube original

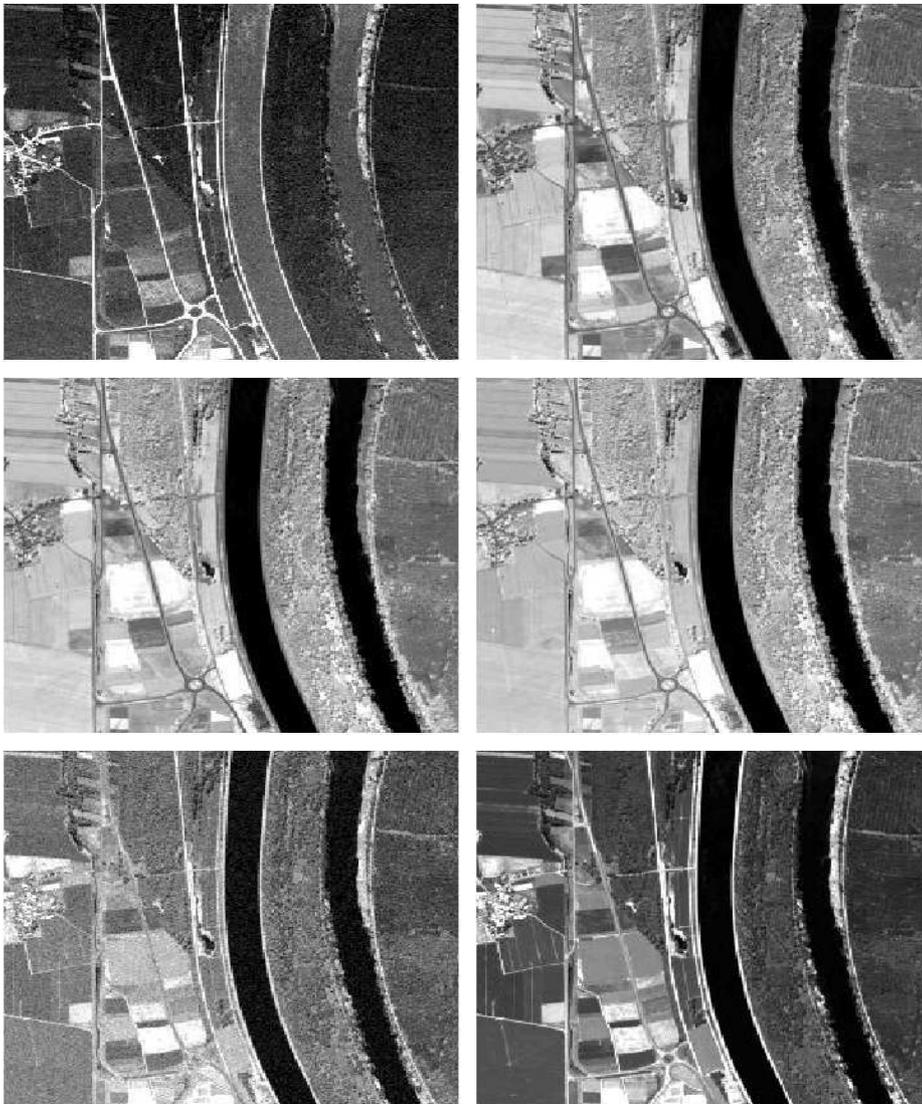


FIG. 4.9 – Images de télédétection de la région de Strasbourg. Dans l'ordre lexicographique, les longueurs d'ondes sont respectivement : 1011, 1250, 1802, 2254, 1802, 2254 nm

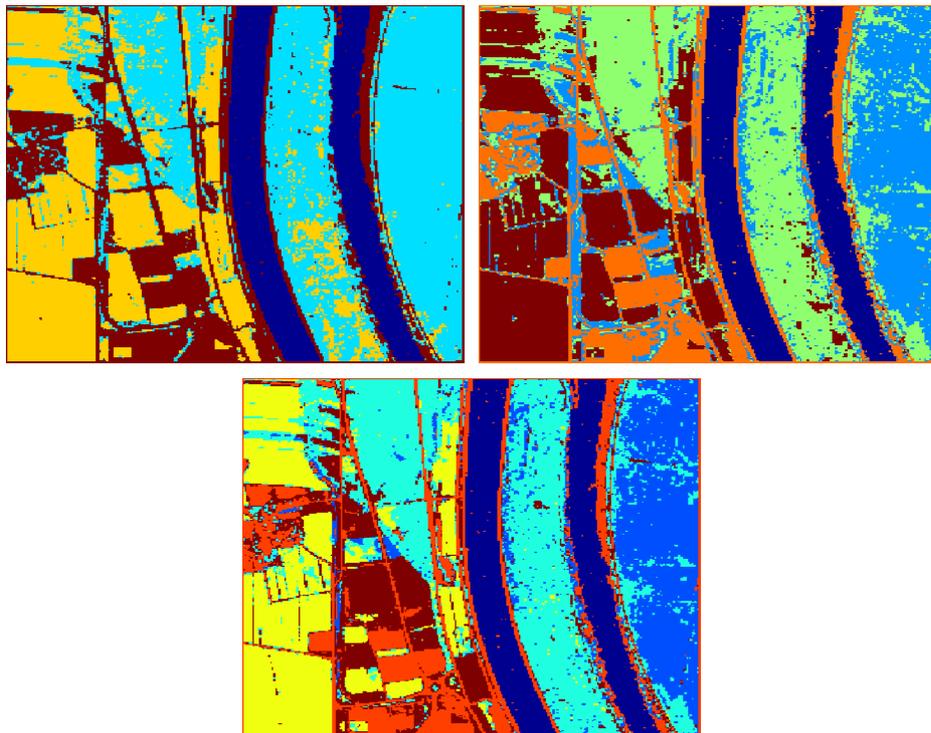


FIG. 4.10 – Segmentation des images de télédétection de la Fig. 4.9. Dans l'ordre lexicographique, carte de segmentation avec : 4, 5 et 6 classes.

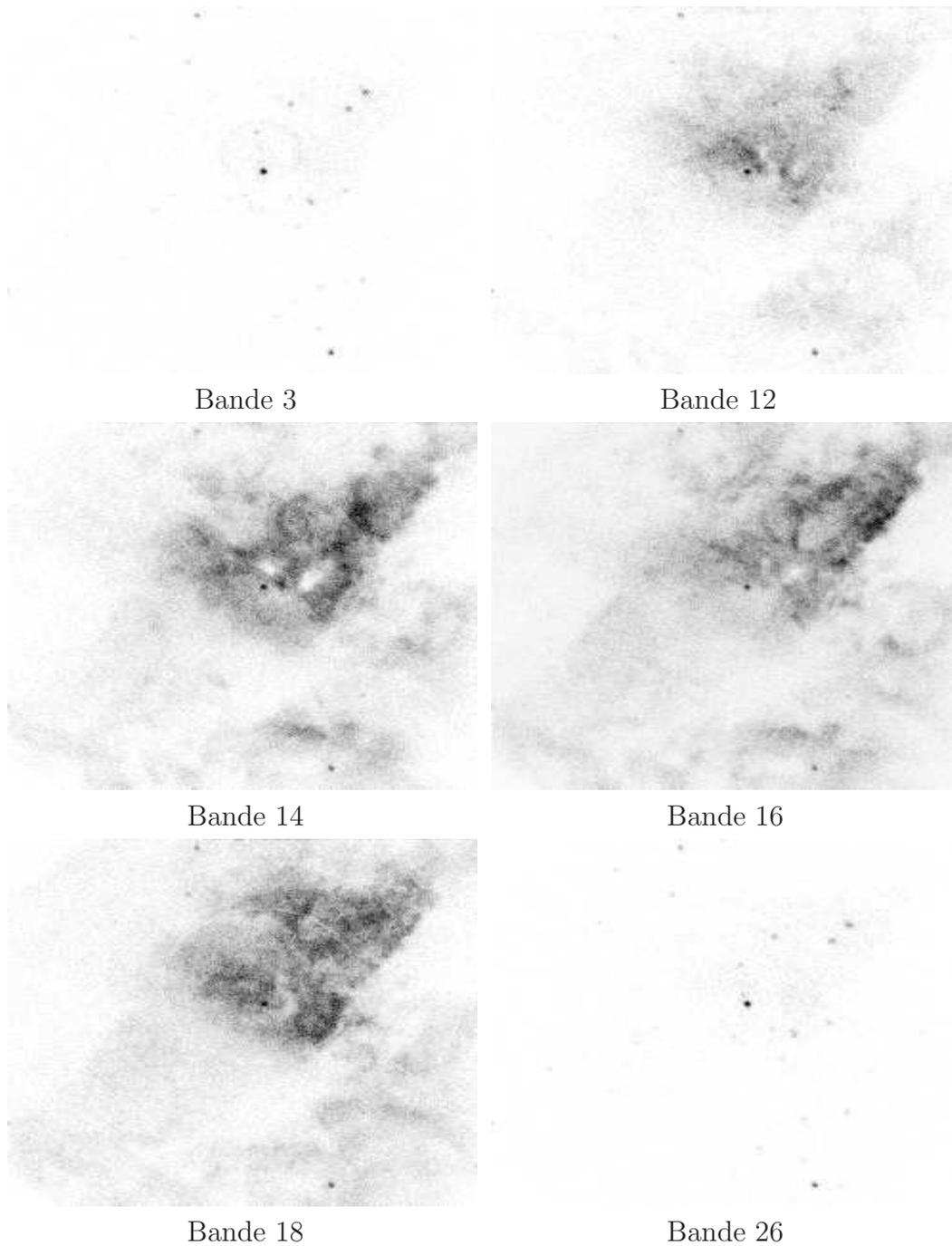


FIG. 4.11 – Étoile Eta Carinae : 6 bandes (en inverse vidéo) sélectionnées dans le cube à 32 bandes résultat de l'observation autour la raie de l'oxygène OIII (489nm à 506nm)

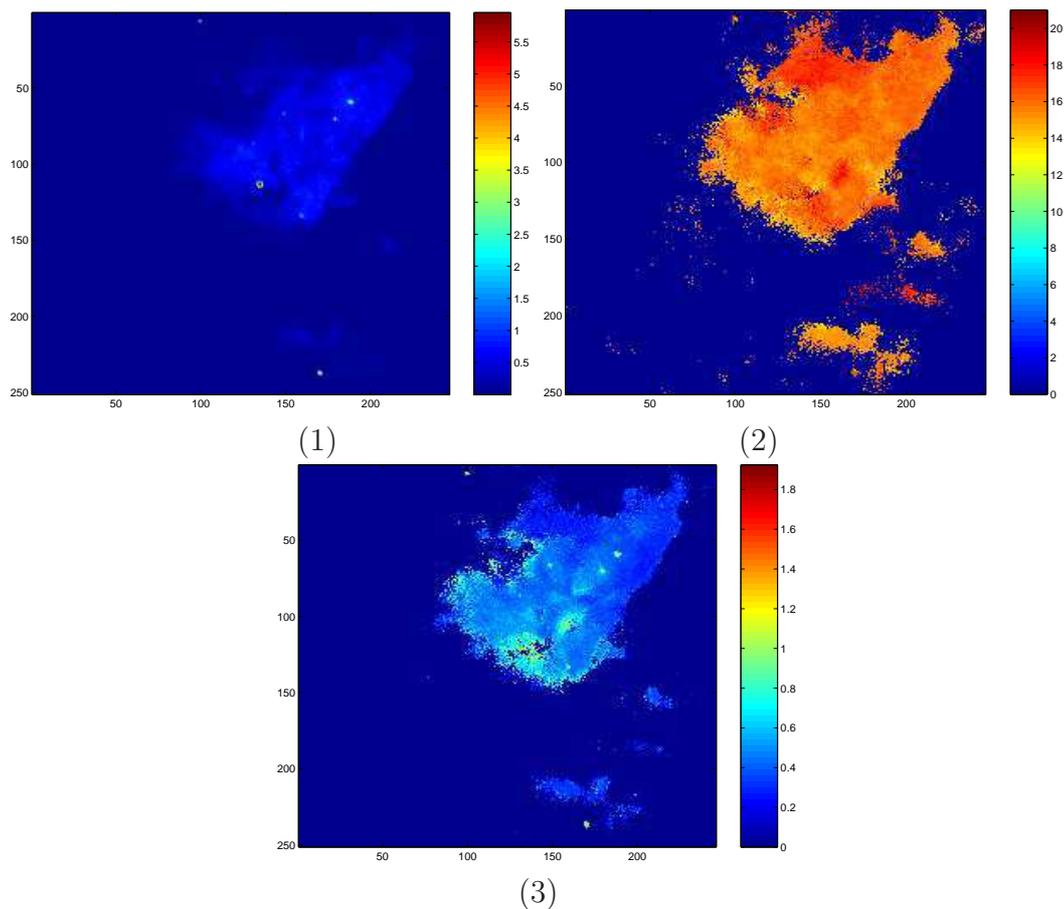


FIG. 4.12 – Eta Carinae : (1) la distribution de l'émission de la raie OII (moment 0), (2) le champ de vitesse (moment 1), (3) la dispersion des vitesses

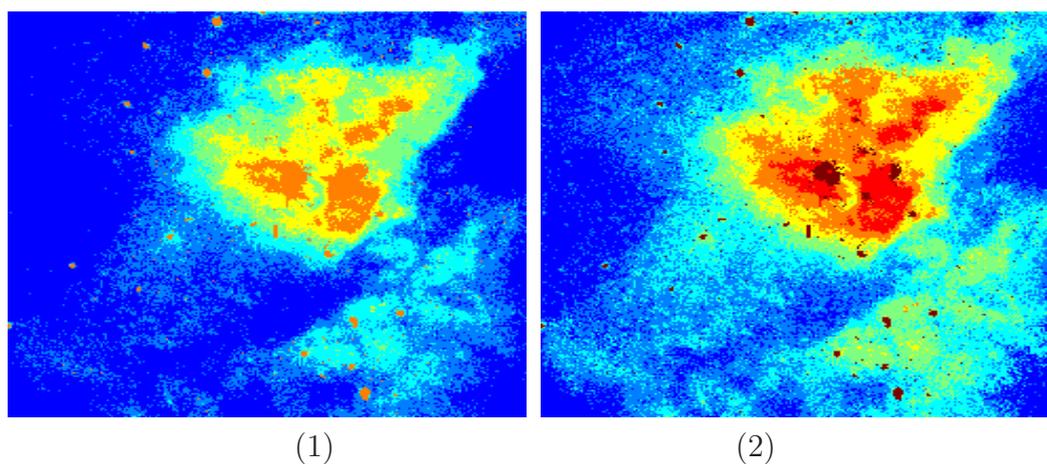


FIG. 4.13 – Résultats de segmentation du cube autour de la raie OIII de l'étoile Eta Carinae avec la MACPP régularisée : (1) Carte de segmentation à 6 classes, (2) Carte de segmentation à 8 classes.

4.4 Arbres de Markov et analyse multirésolution

Les arbres de Markov, par leur structure, se montrent des modèles très efficaces pour les données multirésolutions [143]. L'association de l'arbre de Markov caché à l'analyse multirésolution avec la transformée en ondelettes orthogonales [98], a été introduite par Crouse et *al.* [38]. Cette transformation présente des propriétés intéressantes telles que la localité, la compacité, la non gaussianité et la persistance à travers les échelles [38, 130]. Elle est néanmoins non invariante par translation et opère plutôt comme un détecteur de contours suivant des orientations données. Cependant, en astronomie il n'y a généralement pas de contours francs et les objets sont relativement diffus. C'est la raison pour laquelle on lui préfère une analyse isotropique de l'image comme la transformée en ondelettes à trous [66, 133, 135] ou l'analyse en pyramide avec une seule ondelette [135] obtenue par adaptation de la pyramide laplacienne de Burt et Adelson [23]. Cette dernière ayant la même structure en pyramide que le quadarbre sera donc utilisée. Dans ce qui suit on rappellera cette analyse brièvement et on suivra par sa modélisation par arbres de Markov.

4.4.1 Analyse en pyramide avec une seule ondelette

Soit $f(k)$ un signal discret 1D. Des approximations de plus en plus grossières de ce signal sont obtenues par l'application répétée d'un filtre passe bas suivi d'une décimation d'un facteur deux. Ainsi l'approximation $c_{j+1}(k)$ à l'échelle $j + 1$ est donnée par :

$$c_{j+1}(k) = \sum_l h(l - 2k)c_j(l), \quad (4.1)$$

où $c_j(k)$ est l'approximation à l'échelle j et h un filtre passe bas normalisé, symétrique et vérifiant la contrainte de contribution égale⁶[23]. Cette procédure (Fig. 4.14) est initialisée en posant $c_0(k) = f(k)$. Nous avons choisi une ondelette de fonction échelle B_3 -spline, assez semblable à la gaussienne, isotrope et compacte [107, 13]. Ceci donne un filtre h de coefficients $[\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}]$.

L'information détail $w_j(k)$ perdue lors du passage de la résolution j à la résolution $j + 1$ est calculée comme la différence entre l'approximation à l'échelle j et l'approximation non décimée à l'échelle $j + 1$ comme suit :

$$w_{j+1}(k) = c_j(k) - \tilde{c}_{j+1}(k) \quad (4.2)$$

où $\tilde{c}_{j+1}(k) = \sum_l h(l - k)c_j(l)$.

La procédure de reconstruction est la même que pour la pyramide laplacienne. Etant donné les détails et l'approximation à l'échelle $j + 1$, l'approximation à l'échelle j est calculée par :

$$c_j(k) = w_{j+1}(k) + \sum_l h(k - 2l)c_{j+1}(l). \quad (4.3)$$

Cette reconstruction n'est pas exacte [135]. Cependant, cette dernière peut être approchée itérativement par l'algorithme de Van Citter [33].

Pour une image F , on opère séparément avec le même filtre h sur les lignes puis sur les colonnes [135]. Ainsi on obtient une pyramide W où chaque plan W_j correspond à une résolution donnée.

⁶Burt et Adelson ont utilisé un filtre à 5 coefficients. Si $h(0) = a$, $h(-1) = h(1) = b$, and $h(-2) = h(2) = c$ alors la contrainte de contribution égale exige $a + 2c = 2b$.

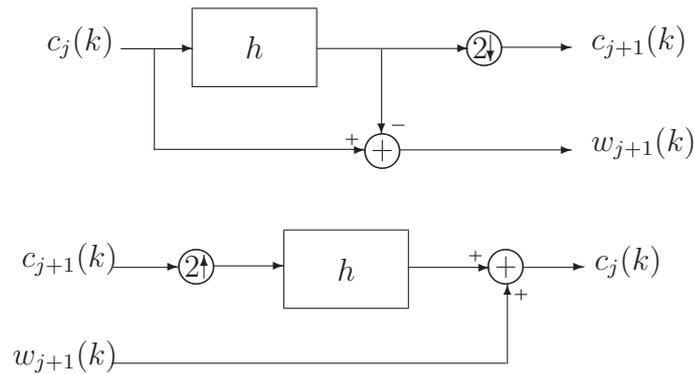


FIG. 4.14 – Analyse en pyramide avec une seule ondelette : étage d’analyse en haut, et étage de synthèse en bas

4.4.2 Modélisation par arbres de Markov

La modélisation par arbre de Markov caché introduite par Crouse et *al.* [38], permet de tenir compte des propriétés principales des coefficients issus d’une transformée en ondelettes. La densité non gaussienne piquée autour de zéro et à queue lourde, conséquence de la propriété de la compacité⁷, est prise en compte par le choix de deux classes (états cachés) gaussiennes. L’une de moyenne nulle et de faible variance associée aux coefficients d’amplitude faible, et l’autre toujours de moyenne nulle mais de forte variance associée aux coefficients d’amplitude forte. La propriété de persistance⁸ à travers les échelles est capturée par les liens interéchelles entre l’état caché d’un site et celui de son parent. De plus, le fait que chaque site dispose de quatre enfants voisins à l’échelle en dessous permet d’appréhender la propriété de localité⁹ en tenant compte du voisinage d’un coefficient lors de sa classification.

Pour les images astronomiques, nous avons utilisé le même modèle sur la décomposition en pyramide avec une seule ondelette. De plus nous avons choisi une loi gaussienne généralisée pour chacune des deux classes permettant une souplesse supplémentaire de modélisation des densités de chacune d’elles.

Par ailleurs, l’arbre de Markov couple étant une généralisation de l’arbre de Markov caché, il constitue lui aussi un modèle très approprié pour de telles données. Il prend en compte toutes les propriétés principales à savoir la non gaussianité, la persistance et la localité. Toutefois la modélisation est différente pour les deux dernières. Les coefficients ne sont pas liés uniquement à leurs états cachés mais aussi aux états cachés des enfants, du parents ainsi qu’aux coefficients des sites enfants et du parent. Cela permet de renforcer la modélisation de la persistance. De plus, la possibilité offerte d’utiliser des lois gaussiennes généralisées pour les deux classes permet une grande souplesse du modèle.

⁷Il existe un petit nombre de coefficients de forte amplitude et un grand nombre de coefficients de faible amplitude

⁸Les coefficients de fortes (resp. faibles) amplitudes ont tendance à se propager à travers les échelles

⁹Chaque coefficient constitue un contenu local, en espace et en échelle, du signal

4.5 Débruitage des images avec arbres de Markov

4.5.1 Débruitage par sélection des coefficients pertinents

La segmentation en deux classes des coefficients d'une décomposition en pyramide réalise une sélection automatique des coefficients pertinents. Ce choix ne tient pas compte seulement de l'amplitude des coefficients mais aussi de leur voisinage intra et inter-échelles ce qui constitue une amélioration par rapport aux seuillages classiques. En choisissant l'étiquette 1 pour la classe des coefficients à forte amplitude et 0 les autres, la carte de segmentation $X = \{X_s, s \in S\}$ représente le support du signal utile dans la pyramide des détails W [131, 12, 135]. Cependant la pyramide tronquée $W^t = X \odot W$ (\odot étant la multiplication élément par élément dans la pyramide) n'est pas une analyse pyramidale d'une image lisse puisque tous les coefficients non significatifs sont mis à zéro. On cherche alors une solution lisse \hat{F} qui minimise le critère J :

$$J = \| (W^t - X \odot O(\hat{F})) \| \quad (4.4)$$

où O est l'opérateur de l'analyse pyramidale. En pratique on utilise l'algorithme de Van Cittert [33, 135] (Algo : 4.1, p. 104).

Algo. 4.1 Algorithme de Van Cittert pour la reconstruction itérative d'images astronomiques

Entrée : La transformée pyramidale W et la carte de segmentation multirésolution X

- $W^t = X \odot W$
- $p = 0, \hat{F}^{[p]} = O^{-1}(W^t)$
où O^{-1} est l'opérateur de reconstruction Eq. 4.3

Répéter

- $W_r^{[p]} = (W^t - O(\hat{F}^{[p]})) \odot X$
- $F_r^{[p]} = O^{-1}(W_r^{[p]})$
- $\hat{F}^{[p+1]} = \hat{F}^{[p]} + F_r^{[p]}$
- $p = p + 1$

Jusqu'à $\| W_r^{[p]} \|$ faible

Sortie : $\hat{F} = \hat{F}^{[p+1]}$

Résultats sur une image HDF

Nous avons appliqué cette technique sur l'image monobande des galaxies high-z de la zone "Hubble Deep Field" observée avec le télescope spatial Hubble sur la longueur d'onde 814 nm. Cette image a été bruitée (Fig.4.15) avec un bruit gaussien et un rapport signal à bruit de 3db. Nous avons utilisé les quadarbres de Markov caché et couple avec des attaches aux données gaussiennes et gaussiennes généralisées avec une décomposition pyramidale avec l'ondelette de fonction échelle B-3 spline décrite plus haut. Les modèles dans le cas gaussien généralisé donnent de meilleurs résultats et donc on peut dire que notre utilisation de ce modèle est justifiée. De plus dans tous les cas, on remarque une meilleure performance de l'arbre couple. Ceci est sans doute du au renforcement des liens entre sites parent et enfants.

4.5.2 Débruitage par rétrécissement des coefficients de détails

Il s'agit d'opérer une pondération de chaque coefficient détail par un facteur entre 0 et 1 [44]. Chaque coefficient est supposé résulter de la somme de deux composantes l'une due au bruit et l'autre au signal utile. L'atténuation est alors proportionnelle à la contribution de la composante de bruit dans la somme. Disposant des modélisations par arbre de Markov, un choix naturel du facteur de pondération serait la probabilité *a posteriori* de la classe des coefficients à forte amplitude. En effet, les coefficients de faibles amplitude correspondant vraisemblablement au bruit auront une faible probabilité d'appartenir à cette classe et se verront fortement atténués. Suivant ce principe, la pyramide de détails W est rétrécie en \tilde{W} comme suit :

$$\forall s \in S^n, n = 0, \dots, R : \tilde{W}_s = \xi_s(1) W_s \quad (4.5)$$

où W_s et \tilde{W}_s sont les coefficients, respectivement originaux et rétrécis associés au site s , et $\xi_s(1) = p(x_s = 1/W)$, la probabilité *a posteriori* d'avoir la classe des coefficients (étiquetée 1) forts au site s . L'image débruitée est reconstruite à partir de \tilde{W} en utilisant l'Eq. 4.3.

Résultats sur une image HDF

Nous avons appliqué cette technique sur la même image astronomique utilisée dans le cas de la sélection des coefficients que nous avons bruitée de la même manière (Fig. 4.16). Nous avons utilisé les mêmes modèles markoviens et la même analyse multirésolution. Les mêmes remarques peuvent être émises avec une diminution claire des écarts de performances. De plus, le rétrécissement semble réduire les effets du bruit existant dans les coefficients à forte amplitude et qui n'est donc pas modifié dans le cas de la sélection.

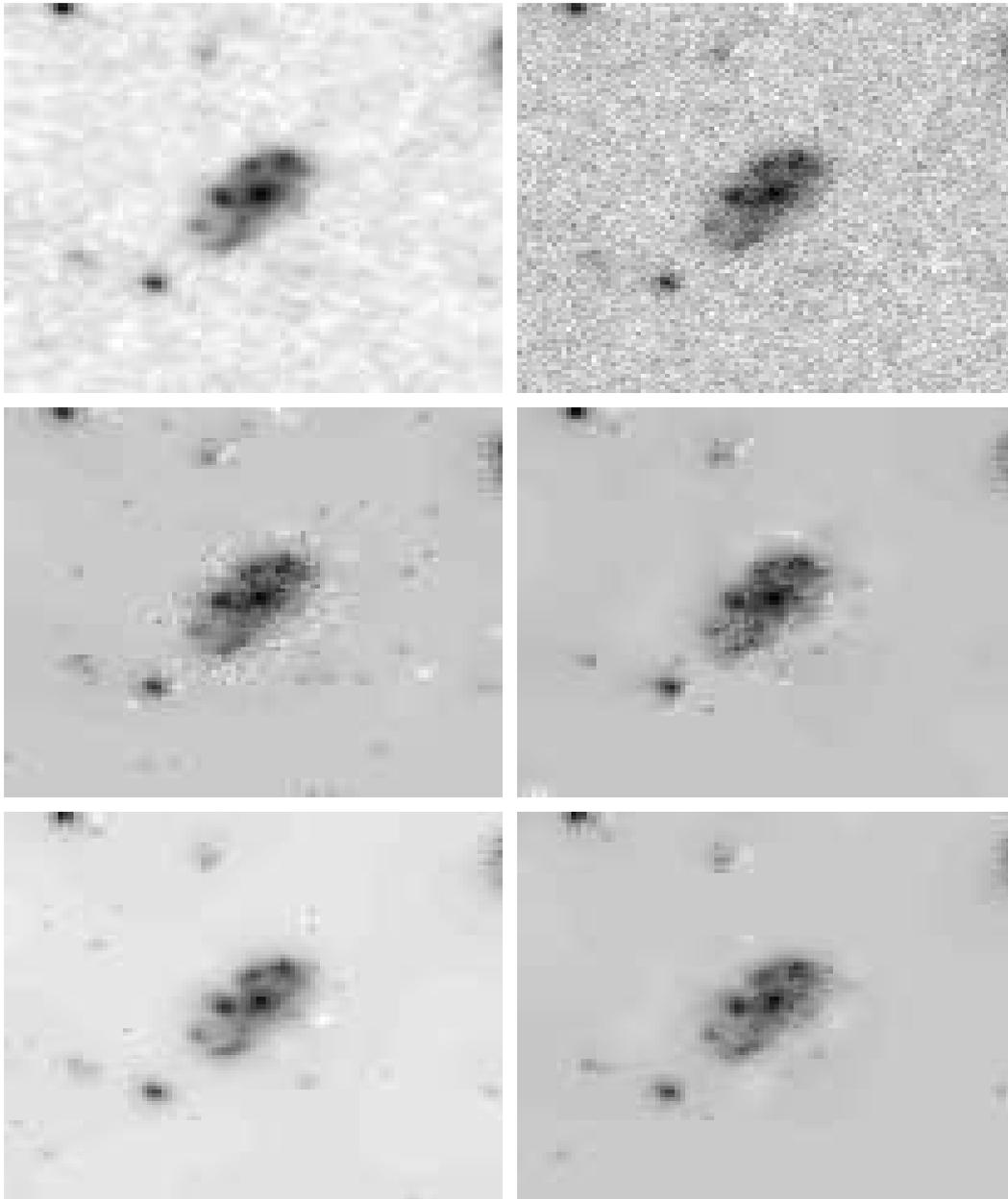


FIG. 4.15 – Débruitage par sélection des coefficients pertinents. L'image originale (hdf4-378 de longueur d'onde 814nm) en haut à gauche (en inverse vidéo), bruitée telle que le rapport signal à bruit est de 3db (haut à droite), est débruitée avec une segmentation deux classes avec respectivement, dans l'ordre lexicographique, l'arbre de Markov caché, et l'arbre de Markov couple, tous deux dans le cas gaussien, l'arbre de Markov caché, et l'arbre de Markov couple, tous deux dans le cas gaussien généralisé.

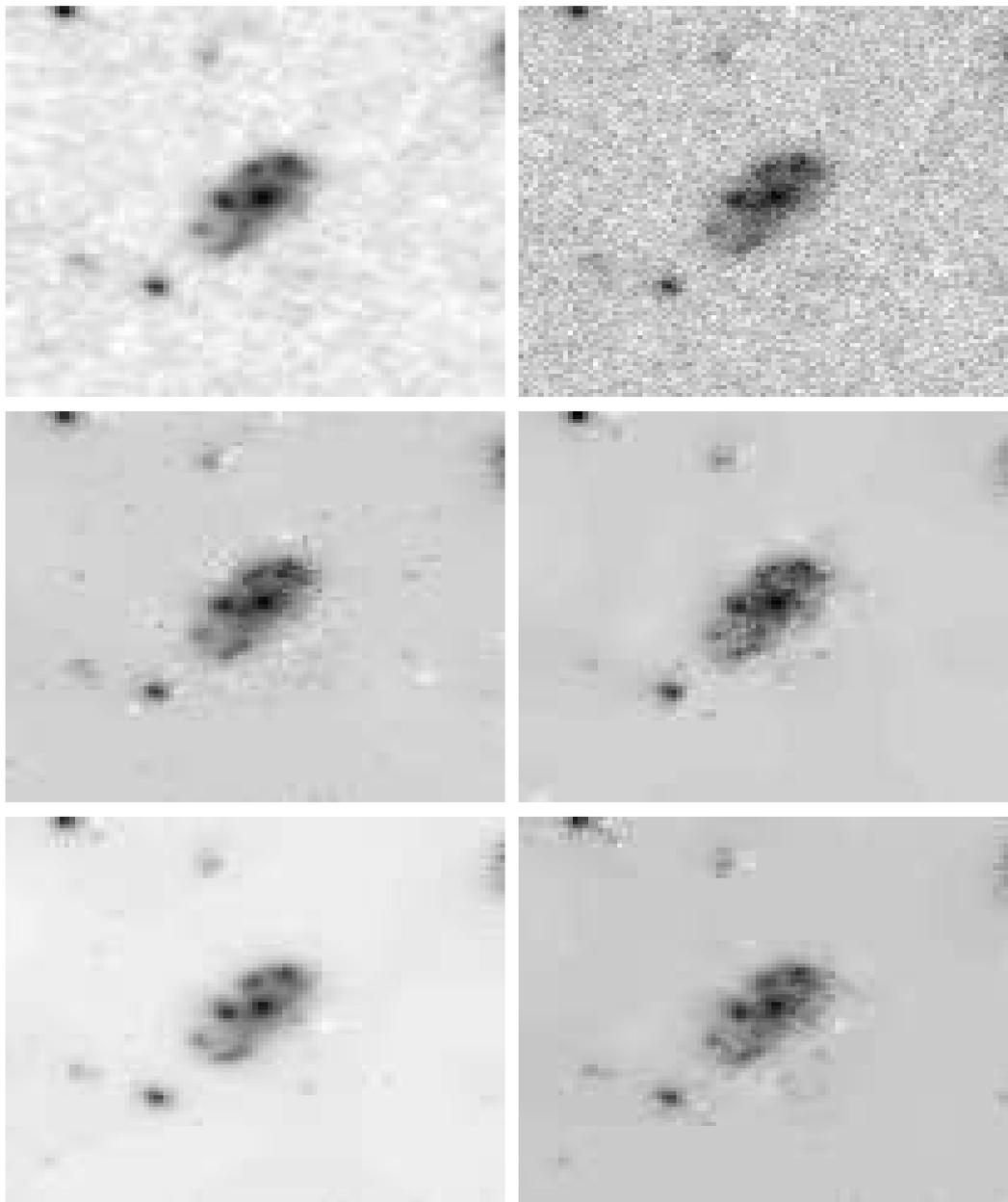


FIG. 4.16 – Débruitage par rétrécissement des coefficients de détails. L'image originale (hdf4-378 de longueur d'onde 814nm) en haut à gauche (en inverse vidéo), bruitée telle que le rapport signal à bruit est de 3db (en haut à droite), est débruitée avec une segmentation deux classes avec respectivement, dans l'ordre lexicographique, l'arbre de Markov caché, et l'arbre de Markov couple, tous deux dans le cas gaussien, l'arbre de Markov caché, et l'arbre de Markov couple, tous deux dans le cas gaussien généralisé.

4.6 Fusion d'images sur l'arbre de Markov caché

La fusion d'images [16] multibandes est un outil très utile en astronomie et dans d'autres disciplines telles que la télédétection et la médecine. Elle permet de synthétiser l'information présente dans les différentes bandes dans une image unique. Cette opération devient plus difficile en présence du bruit. Cependant, l'analyse multirésolution s'avère être un choix intéressant pour la fusion d'images [145] et le débruitage [44].

Pour une image \mathcal{F} à N bandes, on analyse chaque bande séparément ce qui donne N pyramides multirésolutions W^b , $b = 1, \dots, N$. Ces N pyramides sont combinées dans une unique Pyramide Multirésolution Multiband (PMM) \mathbf{W} en considérant les coefficients détails $W_j^1(k), \dots, W_j^N(k)$, de la position spatiale k à l'échelle j comme composantes d'un vecteur unique $\mathbf{W}_j(k)$. Il faut alors considérer trois types de voisinages à savoir le voisinage spatial intra-échelles (propriété de localité), le voisinage inter-échelles (propriété de persistance) et le voisinage inter-bandes (corrélacion inter-bande). L'arbre de Markov caché tel qu'il a été introduit dans [38] ne prend en compte que les deux premiers. Dans le chapitre 2, nous avons présenté l'arbre de Markov caché avec une modélisation du terme d'attache aux données dans le cas multidimensionnel qui nous permet de traiter des observations multibandes et multirésolutions [51]. L'association de ce modèle à la PMM \mathbf{W} prend compte des trois types de voisinages et réalise une réelle fusion multibande et multirésolution lors de la sélection des coefficients.

Une fois la carte de segmentation en deux classes obtenue, les coefficients sélectionnés sont fusionnés en utilisant une des deux règles suivantes :

1. $\forall s \in S^k : W_s^{fused} = \frac{\sum_{i=1}^N x_s W_s^i}{N}$
2. $\forall s \in S^k : W_s^{fused} = \frac{\sum_{i=1}^N \sigma_i^k x_s W_s^i}{\sum_{i=1}^N \sigma_i^k x_s}$
 σ_i^k étant l'écart type de la $i^{ième}$ marginale de la densité multidimensionnelle associée à la classe 1 à l'échelle k .

Pour les approximations jugées assez fiables (peu affectées par le bruit), une simple moyenne de toutes les bandes est effectuée.

La reconstruction est faite par l'algorithme de Van Cittert [33]. On opère de la même manière que la procédure Algo. 4.1, en introduisant la pyramide résultat de fusion W^{fused} , la moyenne des approximations et la carte de segmentation. La technique de fusion est résumée dans la Fig. 4.17.

Résultats sur les images HDF

La technique de fusion décrite à la Fig. 4.17 a été testée sur une image astronomique multibande réelle des galaxies high-z de la zone "Hubble Deep Field" observée avec le télescope spatial Hubble sur six longueurs d'ondes, de l'ultraviolet à l'infrarouge (Fig.4.18). Pour augmenter la difficulté, nous avons rajouté un bruit gaussien à avec un rapport signal sur bruit de 3db en chaque bande. Nous avons opéré une analyse multirésolution sur 4 échelles pour chaque bande avec l'analyse pyramidale avec une seule ondelette décrite plus haut. Les résultats de la fusion avec les deux règles évoquées ci-avant

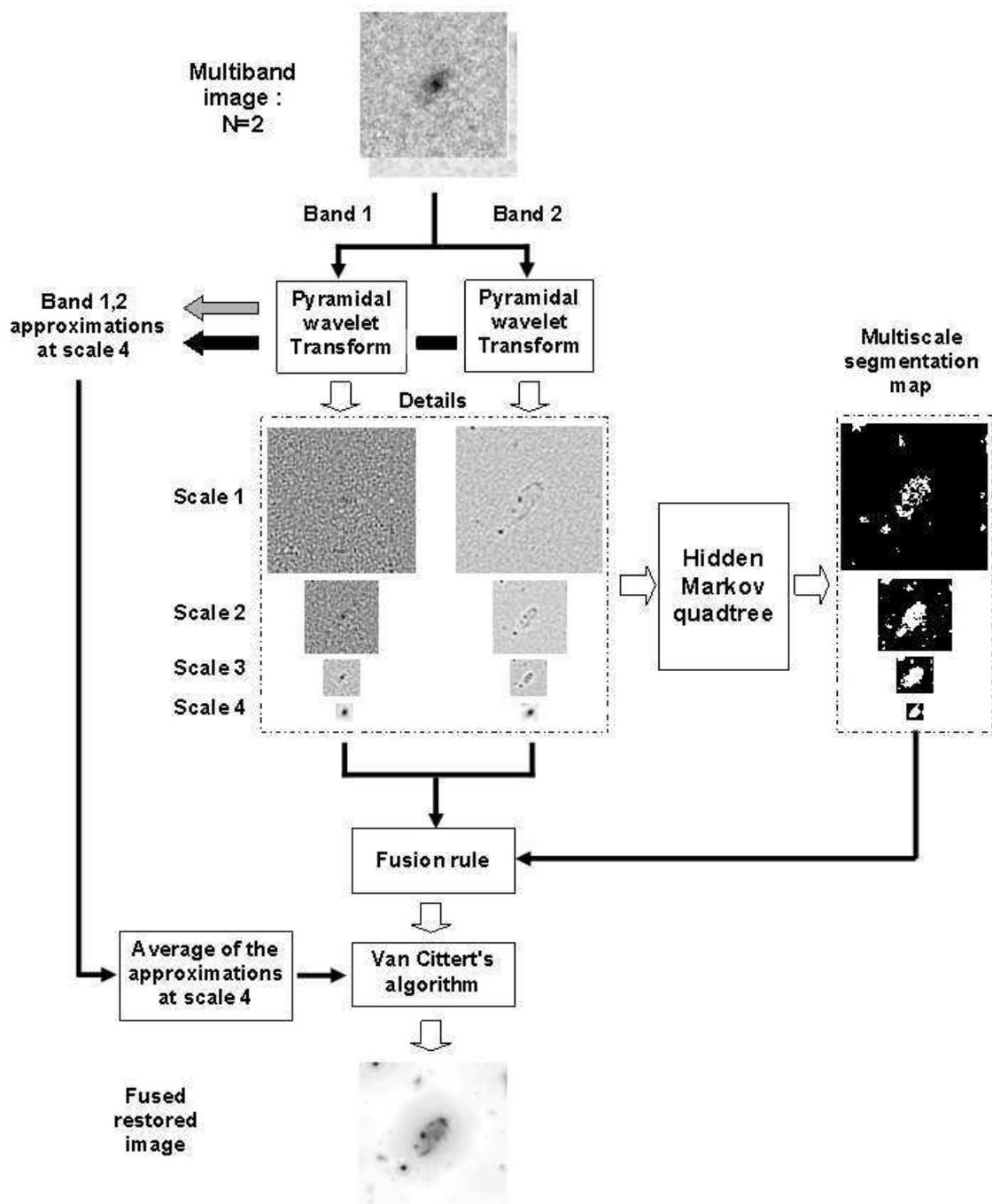


FIG. 4.17 – Procédure de fusion des images astronomiques bruitées

sont comparés avec le résultat obtenu avec le simple moyennage des bandes, respectivement pour le cas gaussien Fig.4.19 et le cas gaussien généralisé Fig.4.20.

Notre technique donne des résultats plus lisses et de loin meilleurs que la simple moyenne. L'image fusionnée résume les propriétés principales de toutes les bandes dans une unique image débruitée et met en valeur la structure globale de la galaxie. Ceci est dû au fait que la fusion est réalisée sur les coefficients significatifs uniquement. On voit ainsi l'efficacité de la modélisation markovienne pour la sélection des structures réelles de l'image. Il faut noter aussi que la différence entre les deux règles est à peine visible. De plus on ne distingue pas d'améliorations assez nette pour le modèle gaussien généralisé.

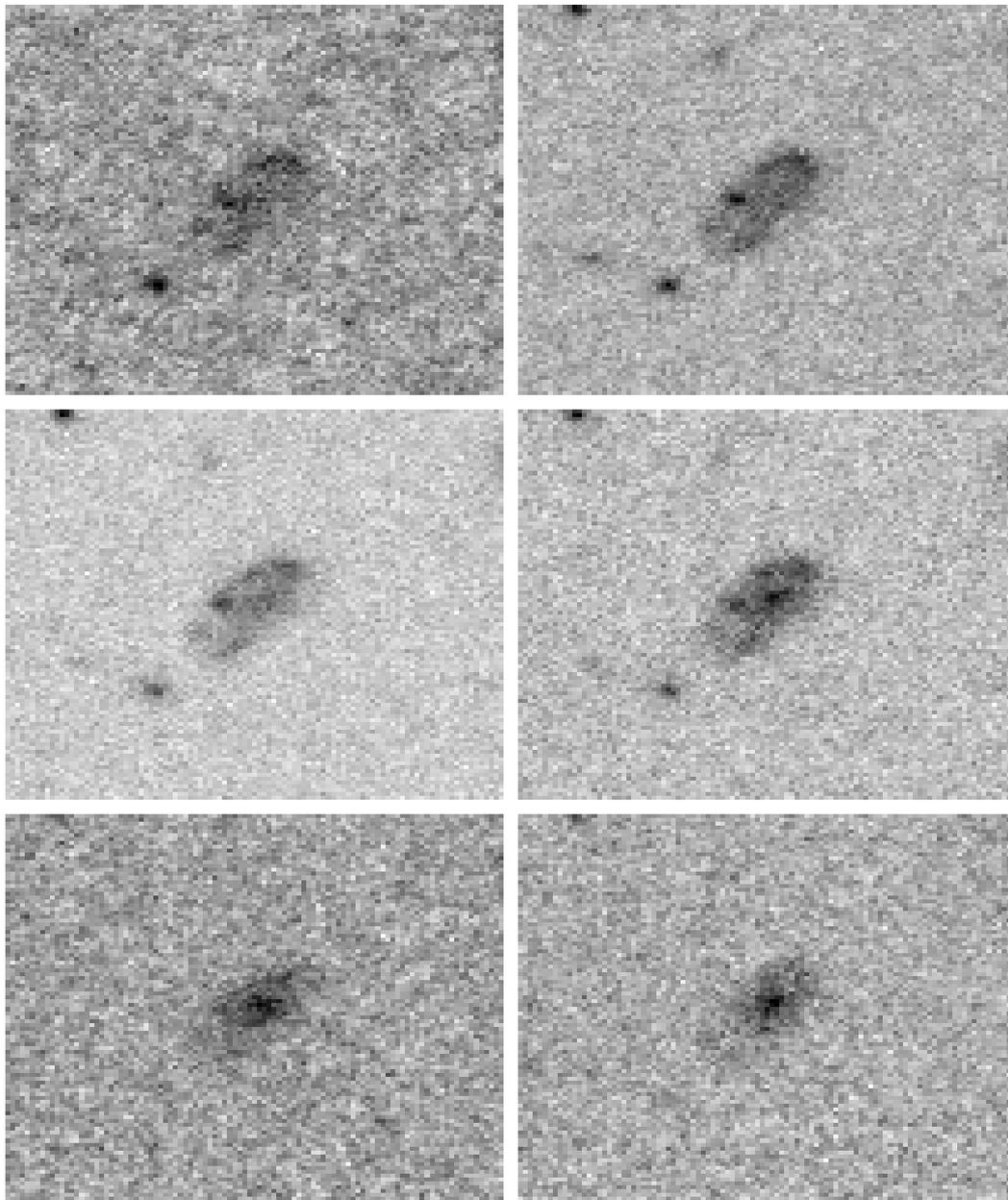


FIG. 4.18 – L'image multibande originale (en inverse vidéo) : hdf4-378. 6 bandes correspondant aux longueurs d'ondes (en nm) : 300, 450, 606, 814, 1100 et 1600.

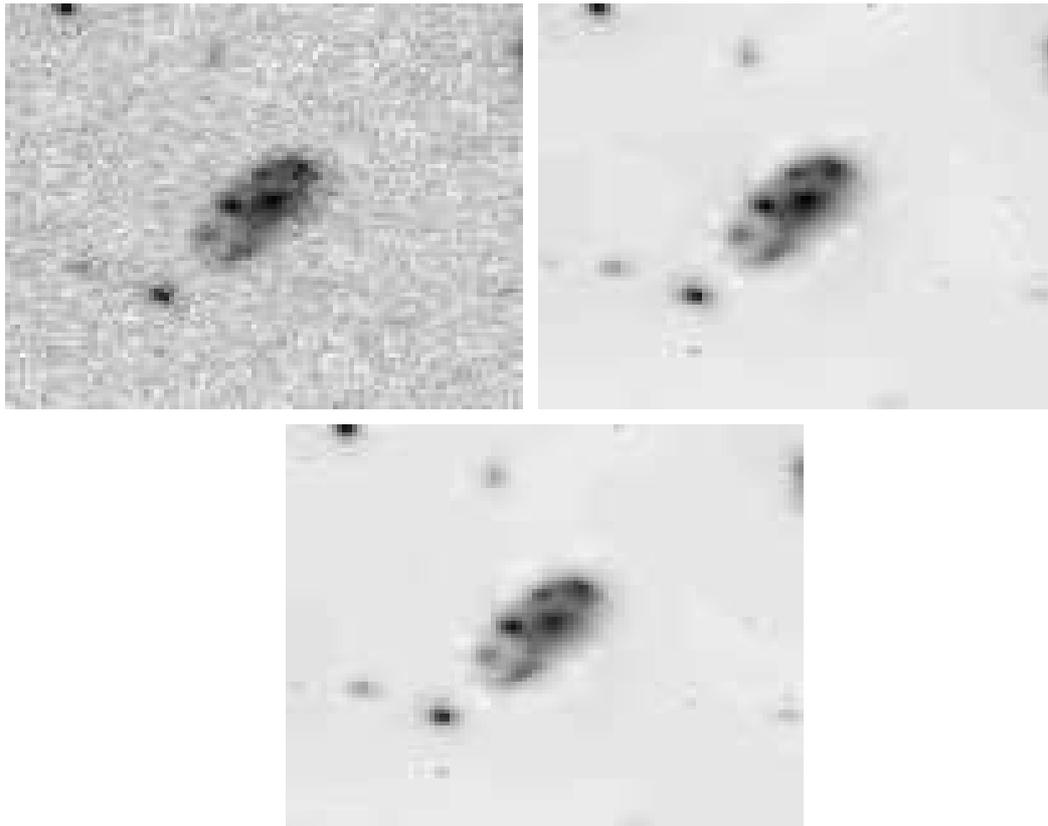


FIG. 4.19 – Comparaison des résultats de la fusion de l'image multibande de la Fig.4.18 (en inverse vidéo) avec une attache aux données gaussienne. En haut à gauche : la moyenne des bandes. En haut à droite : la technique proposée avec la règle 1. En bas : la technique proposée avec la règle 2 .

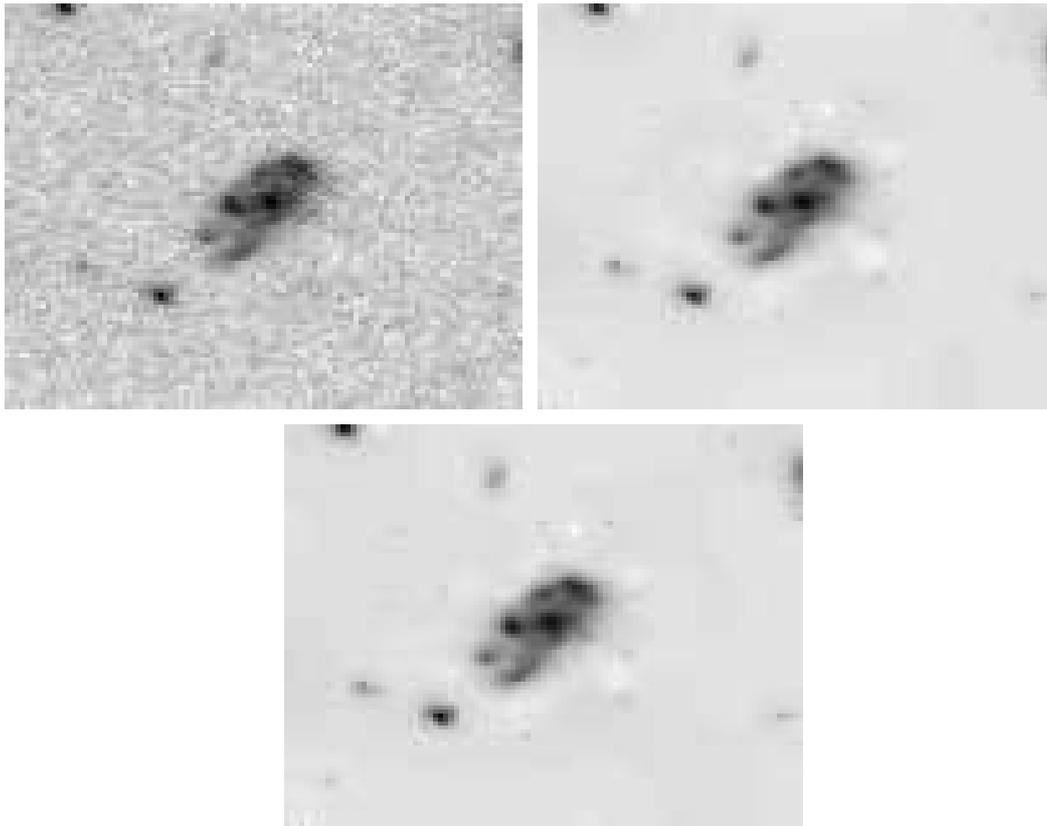


FIG. 4.20 – Comparaison des résultats de la fusion de l'image multibande de la Fig.4.18 (en inverse vidéo) avec une attache aux données gaussienne généralisée. En haut à gauche : la moyenne des bandes. En haut à droite : la technique proposée avec la règle 1. En bas : la technique proposée avec la règle 2 .

4.7 Conclusion

Dans ce chapitre, nous avons montré l'intérêt des techniques développées dans les chapitres précédents pour l'imagerie astronomique. Cependant le travail est loin d'être achevé notamment la validation de la régularisation de la MACPP sur des images astronomiques, une meilleure exploitation de la segmentation en deux classes des pyramides de détails pour le débruitage et la fusion, et l'extension de l'arbre couple pour le cas multi-longueur d'ondes.

Conclusion générale

Cette thèse avait pour but principal la segmentation d'images multispectrales astronomiques. Un problème majeur auquel se heurtent les algorithmes de segmentation lorsque le nombre de bandes est important est le phénomène Hughes. En effet, malgré l'information supplémentaire apportée par l'ajout d'une bande spectrale, les performances de la classification diminuent à cause de la perte de précision de l'estimation des paramètres du modèle. Il est donc impératif de réduire le nombre de bandes en écartant l'information redondante tout en gardant celle qui se révèle pertinente. Dans ce cadre nous avons proposé trois schémas de traitement pour les images astronomiques : 1) le regroupement de bandes suivis de projections locales, 2) la réduction des cubes radio par mélange de gaussiennes et 3) la régularisation du mélange d'analyseurs en composantes principales probabilistes.

La première méthode consiste à regrouper, par un algorithme de coalescence hiérarchique, les bandes spectrales suivant un critère de similarité multiéchelle, basé sur une mesure statistique (l'histogramme normalisé) et une mesure géométrique (le barycentre). Les groupes ainsi établis contiennent des images assez ressemblantes ce qui justifie l'utilisation, au sein de chaque groupe, d'une projection locale linéaire (ACP ou ACI) pour trouver une ou plusieurs images représentatives. Cette technique intéressante pour un nombre de bandes ne dépassant pas la vingtaine, dépend de deux paramètres (le nombre de groupes et la dimension réduite par groupe) dont la détermination automatique pourrait être envisagée. Le choix du nombre de groupes est difficile et peut être guidé par le nombre de classes ou une limite maximale de la distance utilisée pour le regroupement. La dimension réduite par groupe peut être plus facilement choisie en se basant sur les valeurs propres associées à l'ACP.

Le deuxième schéma de traitement proposé est réservé aux cubes radio-astronomiques. Elle consiste à utiliser un modèle de mélange de plusieurs gaussiennes prédéterminées pour approximer le spectre en chaque pixel, de dimension égale au nombre de bandes, de manière à ce que seules les amplitudes suffisent à le coder. Ceci revient à la projection des spectres sur une base de fonctions gaussiennes. Chaque pixel est ainsi représenté par quelques paramètres donnant lieu à une image multi-composantes de paramètres de dimension beaucoup plus faible que l'originale. L'extension de cette technique à d'autres types de cubes d'images peut se faire par le choix d'autres types de fonctions telles que les splines par exemple.

La troisième approche développée est basée sur le modèle de mélange d'ACP probabilistes. Ce modèle bien adapté pour la modélisation de données multidimensionnelles, ne

prend pas en compte l'existence d'éventuels liens entre les échantillons, comme c'est le cas pour les pixels d'une image où la corrélation entre voisins ne peut être ignorée. Nous avons donc proposé une régularisation markovienne à la solution fournie par le mélange d'ACPP permettant de prendre en compte le contexte spatial d'un pixel lors de sa classification. Chaque classe est décrite par une ACPP locale caractérisée par un nombre de vecteurs propres égal à la dimension réduite locale choisie pour cette ACPP et qui constitue une bonne signature de la classe. Ceci est très utile dans le cas où les classes correspondent à des réalités physiques telles que des objets astronomiques, des types de sol ou des étendues d'eau. Cette technique peut être étendue dans quatre directions. La première consiste à utiliser d'autres mélanges de projections locales comme le mélange d'analyseurs factoriaux ou le mélange d'analyseurs en composantes indépendantes. Cette extension, possible dès que l'on dispose d'une formulation statistique d'une projection locale, permet de mieux s'adapter avec les structures locales des pixels que le mélange basé sur l'ACPP. La deuxième extension possible est la segmentation d'images multispectrales multirésolutions. Ceci peut être fait très facilement en traitant les données des échelles en dessus de la base du quadarbre de la même façon que le sont les données associés aux feuilles dans la version actuelle de la méthode. La troisième possibilité est de choisir une régularisation basée sur un autre a priori markovien que l'arbre caché, approche qui a été développée dans cette thèse. Le choix du modèle Markovien dépend essentiellement de la nature des images de l'application. La quatrième possibilité consisterait à exploiter ce modèle pour obtenir une fusion des images du cube. En effet, en choisissant le premier axe de l'ACPP de chaque classe, nous pouvons obtenir une fusion de tout le cube en une unique image par une somme pondérée de toutes les images représentatives des classes obtenues par projection sur les premiers axes des ACPP de toutes les classes. La pondération en chaque pixel est fournie naturellement par la probabilité a posteriori des classes en chaque position étant données les observations. Sur le même principe, nous pouvons imaginer de restreindre la dimension réduite à trois pour obtenir un résultat de fusion coloré.

Que les tâches de réduction et de segmentation soient menées séparément, comme c'est le cas des deux premières techniques, ou conjointement, comme c'est le cas de la technique de la régularisation du mélange d'ACPP, nous avons toujours utilisé une modélisation markovienne hiérarchique par quadarbre de Markov caché pour tenir compte de la corrélation spatiale entre pixels voisins, tout en introduisant de nouvelles modélisations du terme d'attache aux données. En effet, en plus du couplage avec le mélange d'ACPP, nous avons présenté une autre modélisation en nous basant sur la théorie des copules. Cette modélisation permet le calcul de la loi jointe de variables aléatoires corrélées dont les lois marginales sont quelconques. Nous avons détaillé le cas des marginales gaussiennes généralisées qui nous ont permis de couvrir une large classe de lois symétriques.

Le quadarbre de Markov avec le terme d'attache aux données gaussien généralisé multidimensionnel ainsi défini est bien adapté aux données multibandes et multirésolutions. Il tient spécialement compte des trois principales propriétés des coefficients de détails de telles analyses, qui sont la localité, la non-gaussiannité et la persistance à travers les échelles. Nous l'avons ainsi appliqué avec succès pour la modélisation statistique de données issues d'une analyse multirésolution par décomposition pyramidale avec une seule

ondelette d'images astronomiques. Cela nous a permis de développer des procédures de débruitage d'images monobandes et de débruitage/fusion d'images multibandes basées sur la sélection de la classe des coefficients pertinents ou sur la rétrécissement par un facteur égal à la probabilité a posteriori de cette classe. Néanmoins notre exploitation de la modélisation par quadarbre de ces données multirésolution reste fruste et nécessite une réflexion plus approfondie qu'il convient de poursuivre notamment pour les images colorées.

Dans le cadre de l'exploration de modèles de segmentation pour les données multirésolutions, nous avons étudié le quadarbre de Markov couple. Nous avons étendu ce modèle pour le cas d'une attache aux données non gaussienne. Nous avons détaillé le cas gaussien généralisé permettant de couvrir une large catégorie de lois symétriques. Appliqué à l'analyse multirésolution par la pyramide avec une seule ondelette, ce modèle s'est révélé plus pertinent que le modèle du quadarbre caché surtout pour la modélisation de la propriété de persistance à travers les échelles. Cela nous a permis de débruiter des images monobandes avec sélection et rétrécissement des coefficients pertinents. La même remarque que pour le quadarbre caché concernant une meilleure exploitation du modèle dans le domaine ondelettes peut être émise. De plus, l'extension de ce modèle au cas multibande semble théoriquement assez facile. Cependant, cette extension risque de poser un problème de dépassement inférieur de capacité avec la modélisation actuelle de la passe montante sur cet arbre. Il serait peut être préférable de reformuler cette étape pour calculer des probabilités $p(z_s/y_{>s})$ au lieu de $p(y_{>s}/z_s)$.

Enfin, il est très important d'insister sur la validation sur des données réelles, qui valorise le travail effectué, et procure un retour de l'expert nous permettant d'ajuster nos traitements et les améliorer. Dans ce domaine, il nous semble important de sensibiliser et gagner la confiance des utilisateurs potentiels, les astronomes, à ce genre de travaux pour qu'ils s'impliquent d'avantage et amènent leur savoir-faire, ce qui ne peut être que bénéfique pour les deux communautés.

Publications de l'auteur

Journaux

F. Flitti, Ch. Collet, B. Vollmer and F. Bonnarel : Multiband segmentation of a spectroscopic line data cube : application to the HI data cube of the spiral galaxy NGC 4254 *EURASIP journal on Applied Signal Processing, special issue on Applications of Signal Processing in Astrophysics and Cosmology, Vol. 2005, N°15, Aug. 2005.*

M. Petremand, M. Louys, Ch. Collet and **F. Flitti** : Représentation couleur d'images multispectrales astronomiques - Color display for multiwavelength astronomical images. *Traitement du signal, special issue, Vol.21, N°6, Dec. 2004.*

Conferences

F. Flitti, Ch. Collet and Eric Slezak : Astronomical multiband image fusion and restoration using pyramidal analysis and markovian segmentation. *The Eighth International Conference on Information Fusion, ISIF'05, July 2005, Philadelphia, PA, USA.*

F. Flitti, Ch. Collet and A. Joannic-Chardin : Unsupervised multiband image segmentation using hidden Markov quadtree and copulas. *The International Conference on Image Processing, ICIP'05, September 2005, Genova, Italy.*

F. Flitti, Ch. Collet and Eric Slezak : Wavelet Domain astronomical multiband image fusion and restoration using Markov quadtree and copulas. *The 13th European Signal Processing Conference, EUSIPCO'05, September 2005, Antalya, Turkey.*

Ch. Collet and **F. Flitti** : Variations on Markovian Quadtree Model for Multiband Astronomical Image Analysis. *International symposium on Applied Stochastic Models and Data Analysis, ASMDA'05, May 2005, Brest, France.*

F. Flitti, Ch. Collet, B. Wollmer and F. Bonnarel : Data reduction of hyperspectral radio-astronomical images for galaxy cluster segmentation. *The Fourth International Conference on Physics in Signal and Image Processing, PSIP'05, 31 Jan.-2 Feb. 2005, Toulouse, France.*

F. Flitti, Ch. Collet and F. Bonnarel : Reduction methods for multispectral image analysis . *iAstro Workshop, COST action 283, Computational and Information infrastructure in the Astronomical DataGrid, October 2003, Observatoire de Nice, France.*

F. Flitti, Ch. Collet and F. Bonnarel : ACP et ACI pour la réduction de données en imagerie astronomique multispectrale . *19ème colloque sur le traitement du signal et des images, GRESTI, September 2003, Paris, France.*

Bibliographie

- [1] R. Azencott and C. Graffigne. Non supervised segmentation using multi-level Markov random fields. In *Proc. 11th Intern. Conf. on Pattern Recognition*, pages 201–204, The Hague, The Netherlands, September 1992.
- [2] S. De Backer. *Unsupervised Pattern Recognition Dimensionality reduction and Classification*. PhD thesis, University of Antwerp, 2002.
- [3] M. Barlaud, L. Blanc-Ferraud, and P. Charbonnier. Image coding using multiresolution markov random fields. In *SPIE, Image Processing Algorithms and Techniques 3*, San Jose, CA, 1992.
- [4] S.T. Barnard. Stochastic stereo matching over scale. In *Intern. J. Comput. Vis.*, volume 3, pages 17–32, 1989.
- [5] M. Basseville, A. Benveniste, K.C. Chou, S.A. Golden, R. Nikoukhah, and A.S. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Trans. Information Theory*, 2(38) :766–784, March 1992.
- [6] L.E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic function of markov processes. In *Inequalities*, volume 3, pages 1–8, 1972.
- [7] B. Benmiloud. *Chaînes de Markov cachées et cegmentation statistique non supervisée de séquence d'images*. Thèse de doctorat (PhD thesis), Université de Paris VII, décembre 1994.
- [8] B. Benmiloud and W. Pieczynski. Estimation des paramètres dans les chaînes de markov cachées et segmentation d'images. *Traitement du Signal*, 12(5) :433–454, 1995.
- [9] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, 36 :192–236, 1974.
- [10] J. Besag. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*, pages 616–618, 1977.
- [11] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48 :259–302, 1986.
- [12] A. Bijaoui and F. Rué. A multiscale vision model adapted to the astronomical images. *Signal Processing*, 46 :345–362, 1995.
- [13] A. Bijaoui, J.-L. Starck, and F. Murtagh. Restauration des images multi-échelles par l'algorithme à trous. *Traitement du Signal*, 3 :11, 1994.

- [14] J. Bilmes. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report ICSI-TR-97-021, University of Berkeley, <http://ssli.ee.washington.edu/people/bilmes/mypapers/em.pdf>, 1997.
- [15] L. Blanc-feraud and M. Barlaud. Restauration d'image bruitée par analyse multirésolution et champs de Markov. In *13^{ème} Colloque sur le Traitement du Signal et des Images (GRETSI)*, pages 829–832, Juan-Les-Pins, septembre 1991.
- [16] I. Bloch and H. Maître. Fusions de données en traitement d'images : modèles d'information et décisions. *Traitement du signal*, 11(6) :435–446, 1994.
- [17] C. Bouman and B. Liu. Segmentation of textured images using a multiple resolution approach. In *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pages 1124–1127, New York, 1988.
- [18] C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(2) :99–113, February 1991.
- [19] C. Bouman and M. Shapiro. Multispectral image segmentation using a multiscale model. In *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, volume 3, pages 565–568, San Francisco, 1992.
- [20] C.A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2) :162–177, 1994.
- [21] N. Brunel and W. Pieczynski. Unsupervised signal restoration using copulas and pairwise Markov chains. Saint-Louis Missouri, 2003. IEEE Workshop on Statistical Signal Processing.
- [22] N. Brunel, W. Pieczynski, and S. Derrode. Copulas in vectorial hidden markov chains for multicomponent image segmentation. In *Proc. ICASSP*, Philadelphia, USA, March 2005.
- [23] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Trans. Communications*, 31(4) :532–540, April 1983.
- [24] G. Celeux, D. Chauveau, and J. Diebolt. On stochastic versions of the EM algorithm. Report de recherche 2514, INRIA, March 1995.
- [25] G. Celeux and J. Diebolt. L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités. *Revue de Statistique Appliquée*, 34(2), 1986.
- [26] G. Celeux and J. Diebolt. A stochastic approximation type EM algorithm for the mixture problem. In *Stochastics and Stochastics Reports*, volume 41, pages 119–134, 1992.
- [27] B. Chalmond. Image restoration using an estimated Markov model. *Signal Processing*, 15(2) :115–129, September 1988.
- [28] B. Chalmond. An iterative Gibbsian technique for reconstruction of m-ary images. *Pattern Recognition*, 22(6) :747–761, 1989.
- [29] P. Charbonnier, L. Blanc-Feraud, and M. Barlaud. Noisy image restoration using multiresolution Markov random fields. *Journal of Visual Communication and Image Representation*, 3(4) :338–346, December 1992.

- [30] A. Chardin and P. Pérez. Semi-iterative inferences with hierarchical energy-based models for image analysis. *Proc. International Workshop EMMCVPR'99 : Energy Minimisation Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science*, 1654, 1999.
- [31] J. Y. Chen. *Image Database Management Using Similarity Pyramids*. PhD thesis, Perdue University, May 1999.
- [32] K.C. Chou, S.A. Golden, and A.S. Willsky. Multiresolution stochastic models, data fusion and wavelet transforms. *Signal Processing*, 34(3) :257–282, 1993.
- [33] P. H. Van Cittert. Zum einfluss der spaltbreite auf die intensitätsverteilung in spektrallinien ii. *Z. Physik*, 69 :298, 1931.
- [34] C. Collet, P. Thourel, M. Mignotte, P. Pérez, and P. Bouthemy. Une nouvelle approche en traitement d'images sonar haute résolution : la segmentation markovienne hiérarchique multimodèle. *Traitement du signal*, 15(3) :231–250, 1998.
- [35] Ch. Collet and F. Flitti. Variations on Markovian quadtree model for multiband astronomical image analysis. In *Proc. ASMDA*, Brest, France, May 2005.
- [36] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36 :287–314, 1994.
- [37] F. De Coulon. *Théorie et traitement des signaux*, volume tome VI of *Traité d'électricité*. Presses polytechniques romandes, Lausanne, 1984.
- [38] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. on Image Processing*, 46(4), April 1998.
- [39] J.P. Delmas. Relations entre les algorithmes d'estimation itérative EM et ICE avec exemples d'applications. In 15th colloque GRETSI, pages 185–188, Antibes, France, 1995.
- [40] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Royal Statistical Society, serie B*(39) :1–38, 1977.
- [41] S. Derrode, G. Mercier, and W. Pieczynski. Unsupervised multicomponent image segmentation combining a vectorial hmc model and ICA. *ICIP Int. conf. on Image Proc.*, pages CD–Rom, September 2003.
- [42] X. Descombes, M Sigelle, and F. Prêteux. Estimating gaussian markov random field parameters in a nonstationary framework : Application to remote sensing imaging. *IEEE Trans. Image Processing*, 8(4) :490–503, 1999.
- [43] P.A. Devijver. Baum's forward-backward algorithm revisited. *Pattern Recognition Letters*, 39(1) :369–373, 1985.
- [44] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81 :425–455, September 1994.
- [45] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2001.
- [46] D. Egret. Et le ciel devint virtuel. *La Recherche*, (369) :Novembre, 2003.
- [47] K. Falconer. *Fractal Geometry*. John Wiley and Son Ltd.

- [48] P.W. Fieguth. *Application of Multiscale Estimation to Large Scale Multidimensional Imaging and Remote Sensing Problems*. Phd thesis, MIT, June 1995.
- [49] F. Flitti, Ch. Collet, and F. Bonnarel. Acp et Aci pour la réduction de données en imagerie astronomique multispectrale. *Congrès GRETSI*, Septembre 2003.
- [50] F. Flitti, Ch. Collet, and F. Bonnarel. Reduction methods for multispectral image analysis. In *iAstro Workshop, COST action 283, Computational and Information infrastructure in the Astronomical DataGrid*, Nice, France, October 2003.
- [51] F. Flitti, Ch. Collet, and A. Joannic-Chardin. Unsupervised multiband image segmentation using hidden Markov quadtree and copulas. In *Proc. ICIP*, Genova, Italy, september 2005.
- [52] F. Flitti, Ch. Collet, and Eric Slezak. Astronomical multiband image fusion and restoration using pyramidal analysis and Markovian segmentation. In *Proc. ISIF*, Philadelphia, USA, July 2005.
- [53] F. Flitti, Ch. Collet, and Eric Slezak. Wavelet domain astronomical multiband image fusion and restoration using Markov quadtree and copulas. In *Proc. EUSIPCO*, Antalya, Turkey, september 2005.
- [54] F. Flitti, Ch. Collet, B. Vollmer, and F. Bonnarel. Multiband segmentation of a spectroscopic line data cube : application to the HI data cube of the spiral galaxy NGC 4254. *EURASIP EURopian journal on Applied Signal Processing, special issue on Applications of Signal Processing in Astrophysics and Cosmology*, 15 :2546–2558, 2005.
- [55] F. Flitti, Ch. Collet, B. Wollmer, and F. Bonnarel. Data reduction of hyperspectral radio-astronomical images for galaxy cluster segmentation. In *Proc. PSIP*, Toulouse, France, February 2005.
- [56] G.D. Forney. The Viterbi Algorithm. *Proceedings of the IEEE*, 61(3) :268–278, March 1973.
- [57] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A 9*, pages 53–77, 1951.
- [58] Z. Ghahramani and G.E. Hinton. The EM algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1, University of Toronto, <http://www.gatsby.ucl.ac.uk/~zoubin/papers.html>, 1996.
- [59] B. Gidas. A renormalization group approach to image processing problem. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(2) :164–180, 1989.
- [60] N. Giordana. *Segmentation non supervisée d'images multispectrales par chaînes de Markov cachées*. Thèse de doctorat, Institut National des Télécommunications (INT), décembre 1996.
- [61] N. Giordana and W. Pieczynski. Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5) :465–475, 1997.
- [62] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2001.
- [63] W. Härdle and L. Simar. *Applied Multivariate Statistical Analysis*. Method and Data technologies, <http://www.mdtech.de/>, October 2003.

- [64] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12) :1217–1232, December 1993.
- [65] F. Heitz, P. Pérez, and P. Bouthemy. Multiscale minimisation of global energy functions in some visual recovery problems. In *Computer Vision Graph. and Image Proces. : Image Understanding*, volume 59, pages 125–134, January 1994.
- [66] M. Holschneider, R. Kronland-Martinet, J. Morlet, and A. Grossmann. A real-time algorithm for signal analysis with the help of the wavelet transform. In J. M. Combes, A. Grossmann, and Ph. Tchamitchian, editors, *Wavelets, Time-Frequency Methods and Phase Space*, pages 286–297. Springer-Verlag, 1989.
- [67] P. J. Huber. Projection pursuit with discussion. *The Annals of Statistics*, 13(2) :435–525, 1985.
- [68] G.F. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Information Theory*, 14(1) :55–63, 1968.
- [69] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans.on Neural Networks*, 10(3) :626–634, May 1999.
- [70] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley and Sons, 2001.
- [71] A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing Systems 10 (NIPS*97)*, pages 273–279. MIT Press, 1998.
- [72] A. Hyvärinen and E. Oja. Independent Component Analysis : Algorithms and Applications. *Neural Networks*, 13(4-5) :411–430, 2000.
- [73] M. Jamshidian and R. Jennrich. Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association*, 88 :222–228, 1993.
- [74] L. Jimenez and D. Landgrebe. High dimensional feature reduction via projection pursuit. *School of Electrical and Computer Engineering, Purdue University, West Lafayette in 47907-1285*, April 1995.
- [75] I. T. Jolliffe. *Principal Component Analysis*. New York : Springer-Verlag, 1986.
- [76] M. Jones and R. Sibson. What is projection pursuit? *J. of the Royal Statistical Society*, Ser. A(150) :1–36, 1987.
- [77] B. H. Juang and L. R. Rabiner. The segmental k-means algorithm for estimating parameters of hidden Markov models. *IEEE Trans. On ASSP*, 38(9) :1639–1641, September 1990.
- [78] N. Kambhatla and T. K. Leen. Fast non-linear dimension reduction. *Advances in Neural Information Processing Systems*, June 1994.
- [79] Z. Kato. *Modélisations markoviennes multirésolutions en vision par ordinateur. Application á la segmentation d’images SPOT*. Thèse de doctorat (phd thesis), Université de Nice, Inria Sophia-Antipolis, déc. 1994.
- [80] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, (220) :671–680, 1983.

- [81] W. J. Krzanowski and F. H. C. Marriott. *Multivariate Analysis Part 2 : Classification, Covariance Structures and Repeated Measurements*. London : Edward Arnold, 1994.
- [82] J.M. Laferté. *Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques. Application à la fusion de données multirésolution*. Thèse de doctorat, Université de Rennes 1, IRISA, Octobre 1996.
- [83] J.-M. Laferté, P. Pérez, and F. Heitz. Discrete markov image modeling and inference on the quad-tree. *IEEE Trans. Image Process.*, 9(3) :390–404, March 2000.
- [84] G. N. Lance and W. T. Williams. A general theory of classificatory sorting strategies 1. hierarchical systems. *Computer Journal*, 9(4) :373–380, 1966.
- [85] P. Lanchantin and W. Pieczynski. Unsupervised restoration of hidden non stationary markov chain using evidential priors. *IEEE Transactions on Signal Processing*, 53(8) :3091–3098, 2005.
- [86] D. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley and Sons, 2003.
- [87] K. Lange. A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society*, B 57 :425–437, 1995.
- [88] K. Lange. A quasi-Newton acceleration of the EM algorithm. *Statistica sinica*, 5 :1–18, 1995.
- [89] S. L. Lauritzen. *Complex Stochastic Systems*, chapter Causal Inference From Graphical Models. CHAPMAN and HALL/CRC, 2001.
- [90] T.W. Lee, M.S. Lewicki, and T.J. Sejnowski. ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10) :1078–1089, October 2000.
- [91] M. Lennon, G. Mercier, M.C. Mouchot, and L. Hubert-Moy. Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images. In *IGARSS 2003 : International Conference on Geoscience and Remote Sensing, Sydney, Australia*, July 2001.
- [92] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, 2nd edition, 2001.
- [93] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. J. Wiley and Sons, New York, 1987.
- [94] T. A. Louis. Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society*, B 44 :226–233, 1982.
- [95] M. Luetttgen. *Image Processing with Multiscale Stochastic Models*. Phd thesis, MIT Laboratory of Information and Decision Systems, May 1993.
- [96] M. R. Luetttgen, W. C. Karl, A. S. Willsky, and R. R. Tenney. Multiscale representations of Markov random fields. *IEEE Transaction on Signal Processing*, 41(12) :3377–3396, December 1993.
- [97] M.R. Luetttgen, W.C. Karl, A.S. Willsky, and R. Tenney. Multiscale representation of markov random fields. *IEEE Trans. Image Process.*, 41(12) :3377–3395, December 1993.

- [98] S. Mallat. *Une Exploration des Signaux en Ondelettes*. École polytechniques edition, 2000.
- [99] G. F. McLachlan and T. Krishnan. *The EM algorithm and Extensions*. John Wiley and Sons, 1997.
- [100] X.L. Meng and D.B. Rubin. Maximum likelihood estimation via the ecm algorithm : A general framework. *Biometrika*, 80(2) :267–278, 1993.
- [101] Jordan M.I. and Weiss Y. *Handbook of Neural Networks and Brain Theory*, chapter Probabilistic Inference in Graphical Models. MIT Press, 2002.
- [102] M. Mignotte. *Segmentation d’images sonar par approche markovienne hiérarchique non supervisée et classification d’ombres portées par modèles statistiques*. Thèse de doctorat (PhD thesis), Université de Bretagne Occidentale - Ecole Navale (Laboratoire GTS), juillet 1998.
- [103] A. Mohammadpour, O. Féron, and A. Mohammad-Djafari. Bayesian segmentation of hyperspectral images. In *Proceedings of MaxEnt04*, pages 25–30, Max-Planck Institute, Garching, Munich, Germany, July 2004.
- [104] E. Monfrini, J. Lecomte, F. Desbouvries, and W. Pieczynski. Image and signal restoration using pairwise markov trees. Saint-Louis Missouri, 2003. IEEE Workshop on Statistical Signal Processing.
- [105] E. Monfrini, T. Ledru, and W. Pieczynski. Segmentation non supervisée d’images par arbres de Markov cachés. In *17ème colloque GRETSI*, volume CD-Rom, pages 43–46, Vannes, France, September 1999.
- [106] K. Murphy. An introduction to graphical models. Technical report, Intel Research, citeseer.ist.psu.edu/article/murphy01introduction.html, 2001.
- [107] F. Murtagh, J.-L. Starck, and A. Bijaoui. Multiresolution in Astronomical Image Processing : A General Framework. *The International Journal of Image Systems and Technology*, 6 :332–338, 1995.
- [108] R. B. Nelsen. *An introduction to copulas*. Lecture Notes in Statistics. Springer, New York, 1998.
- [109] D. Nuzillard and A. Bijaoui. Blind source separation and analysis of multispectral astronomical images. *Astronomy and Astrophysics*, 174(1) :129–138, November 1993.
- [110] M.T. Orchard. Predictive motion field segmentation for image sequence coding. *IEEE Transaction on Circuits and Systems for Video Technology*, 3(1) :54–70, 1993.
- [111] A. Papoulis. *Probability random variables and stochastic processes*. Mac Graw Hill, 3rd edition, 1991.
- [112] P. Pérez. Modèles et algorithmes pour l’analyse probabiliste des images. Université de Rennes 1, December 2003. Habilitation à diriger des recherches.
- [113] B. Phookun, S.N. Vogel, and L.G. Mundy. Ngc 4254 : a spiral galaxy with an m=1 mode and infalling gas. *Astrophysical Journal*, 418 :113–122, November 1993.
- [114] K. K. Phoon. Application of fractile correlations and copulas to non-gaussian random vectors. In *Proceedings, Second International ASRANet(Network for Integrating Structural Safety, Risk, and Reliability) Colloquium*, Barcelona, Spain, July 2004.

- [115] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Traitement du signal*, 11(2) :141–153, 1994.
- [116] W. Pieczynski. Arbre de Markov Couple - pairwise Markov Tree. *Comptes rendus de l'Académie des Sciences - Mathématiques*, Ser. I 335(1) :79–82, 2002.
- [117] W. Pieczynski. Arbre de Markov Triplet et fusion de Dempster-Shafer - Triplet Markov Tree and Dempster-Shafer fusion. *Comptes rendus de l'Académie des Sciences - Mathématiques*, Ser. I 336(10) :869–872, 2003.
- [118] W. Pieczynski, J. Bouvrais, and C. Michel. Unsupervised bayesian fusion of correlated sensors. In *First International Conference on Multisource-Multisensor Information Fusion*, Las Vegas, Nevada, USA, June 6-9 1998.
- [119] P. Pérez, F. Heitz, and P. Bouthemy. *Maximum Entropy and Bayesian Methods*, chapter Global bayesian estimation, constrained multiscale Markov random fields and the analysis of visual motion, pages 383–388. Kluwer Academic, 1993.
- [120] J. N. Provost. *Classification bathymétrique en imagerie multispectrale SPOT*. Thèse de doctorat, Université de Bretagne Occidentale - Ecole Navale (Laboratoire GTS), Juin 2001.
- [121] J. N. Provost, C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy. Hierarchical markovian segmentation of multispectral images for the reconstruction of water depth maps. *Computer Vision and Image Understanding*, 93(2) :155–174, February 2004.
- [122] L. R. Rabiner, B. H. Juang, S. E. Levinson, and M. M. Sondhi. Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities. *AT-T Tech. Journal*, 64(6) :1211–1234, July-August 1985.
- [123] L. R. Rabiner, J. G. Wilpon, and B. H. Juang. A segmental k-means training procedure for connected word recognition. *AT-T Tech. Journal*, 64(3) :21–40, May 1986.
- [124] C. Regazzoni, F. Arduini, and G. Vernazza. A multilevel gmrf-based approach to image segmentation and restoration. *Signal Processing* 34, pages 43–67, 1993.
- [125] H. Rehrauer, K. Seidel, and M. Datcu. Multiscale image segmentation with a dynamic label tree. In T. I. Stein, editor, *Proc. of the IEEE International Geoscience and Remote Sensing Symposium*, pages 1772–1774, 1998.
- [126] G. Rellier. *Analyse de textures dans l'espace hyperspectral par des méthodes probabilistes*. Thèse de doctorat, Univ. Sophia Antipolis - INRIA, novembre 2002.
- [127] G. Rellier, X. Descombes, F. Falzon, and J. Zérubia. La poursuite de projection pour la classification d'image hyperspectrales texturée. *Rapport INRIA*, (4152), mars 2001.
- [128] Ch. Robert. *The Bayesian Choice : from Decision-Theoretic Motivations to Computational Implementation*. Springer-Verlag, New York, second edition, 2001.
- [129] S. Roberts and W. Penny. Mixtures of independent component analysers. In *Proceedings of ICANN*, Vienna, August 2001.
- [130] J. K. Romberg, H. Choi, and R. G. Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden markov models. *IEEE Trans. Image Processing*, 10(7) :1056–1068, July 2001.

- [131] F. Rué and A. Bijaoui. A multiscale vision model to analyse field astronomical images. *Experimental Astronomy*, 7 :129–160, 1997.
- [132] J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman and Hall/CRC, 1997.
- [133] M. Shensa. The discrete wavelet transform : Wedding the À trous and mallat algorithms. *IEEE Transactions on Signal Processing*, Volume 40(10) :2464–2482, October 1992.
- [134] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, pages 229–231, 1959.
- [135] J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis : The Multiscale Approach*. Cambridge University Press, 1998.
- [136] M. A. Tanner. *Tools for statistical inference*. Springer-Verlag, 1996.
- [137] M. E. Tipping and C. Bishop. Mixtures of Probabilistic Principal Component Analysers. *Neural Computation*, (11) :443–482, 1999.
- [138] M. E. Tipping and C. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistic Society, Series B*, 61 :611–622, Part 3.
- [139] M.K. Varanasi and B. Aazhang. Parametric generalized gaussian density estimation. *J. Acoust. Soc. Amer*, 86 :1404–1415, 1989.
- [140] E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*. Communications and Control Engineering Series. Springer, London, 1997.
- [141] G.C. Wei and M.A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85 :699–704, 1990.
- [142] C.K.L. Williams and A.J. Adams. *Advances in Neural Information Processing Systems*, chapter DTs : Dynamic trees. MIT Press, 1999.
- [143] A. S. Willsky. Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE*, 90(8) :1396–1458, August 2002.
- [144] J. Zhang, J. Modestino, and D.A. Langan. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Trans. on Image Processing*, 3(4) :404–420, 1994.
- [145] Z. Zhang and R. S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*, 87(8) :1315 – 1326, August 1999.

Résumé :

Le développement de nouveaux capteurs multispectraux en imagerie astronomique permet l'acquisition de données d'une grande richesse. Néanmoins, la classification d'images multidimensionnelles se heurte souvent au phénomène de Hughes : l'augmentation de la dimensionnalité s'accompagne d'un accroissement du nombre de paramètres du modèle et donc inévitablement une baisse de précision de leur estimation entraînant une dégradation de la qualité de la segmentation. Il est donc impératif d'écartier l'information redondante afin de réaliser des opérations de segmentation ou de classification robustes. Dans le cadre de cette thèse, nous avons proposé deux méthodes de réduction de la dimensionnalité pour des images multispectrales : 1) le regroupement de bandes suivis de projections locales ; 2) la réduction des cubes radio par un modèle de mélange de gaussiennes. Nous avons également proposé un schéma de réduction/segmentation jointe basé sur la régularisation du mélange d'analyseurs en composantes principales probabilistes (MACPP). En ce qui concerne la tâche de segmentation, nous avons choisie une approche bayésienne s'appuyant sur des modèles hiérarchiques récents à base d'arbres de Markov caché et couple. Ces modèles permettent en effet un calcul rapide et exact des probabilités *a posteriori*. Pour le terme d'attache aux données, nous avons utilisée la loi gaussienne multidimensionnelles classique, la loi gaussienne généralisée multidimensionnelles formulée grâce à la théorie des copules et la vraisemblance par rapport au modèle de l'ACP probabiliste (dans le cadre de la MACPP régularisée). L'apport majeur de ce travail consiste donc à proposer différents modèles markoviens hiérarchiques de segmentation adaptés aux données multidimensionnelles multirésolutions. Leur exploitation pour des données issues d'une analyse par ondelettes adaptée au contexte astronomique nous a permis de développer des techniques de débruitage et de fusion d'images astronomiques multispectrales nouvelles. Tous les algorithmes sont non supervisés et ont été validés sur des images synthétiques et réelles.

Mots-clés : Quadarbre markovien, copules, images multispectrales, réduction de dimensionnalité, segmentation, classification, multiresolution, fusion, astronomie.

Abstract :

The development of astronomical multispectral sensors allows data of a great richness. Nevertheless, the classification of multidimensional images is often limited by Hughes phenomenon : when dimensionality increases the number of parameters of the model grows and the precision of their estimates falls inevitably, therefore the quality of the segmentation dramatically decreases. It is thus imperative to discard redundant information in order to carry out robust segmentation or classification. In this thesis, we have proposed two methods for multispectral image dimensionality reduction : 1) bands regrouping followed by local projections ; 2) radio cubes reduction by a mixture of Gaussians model. We have also proposed joint reduction/segmentation scheme based on the regularization of the mixture of probabilistic principal components analyzers (MPPCA). For the segmentation task, we have used a Bayesian approach based on hierarchical Markov models namely the hidden Markov tree and the pairwise Markov tree. These models allow fast and exact computation of the *a posteriori* probabilities. For the data driven term, we have used three formulations : 1) the classical multidimensional Gaussian distribution 2) the multidimensional generalized Gaussian distribution formulated using copulas theory 3) the likelihood of the probabilistic PCA model (within the framework of the regularized MPPCA). The major contribution of this work consists in introducing various hierarchical Markov models for multidimensional and multiresolution data segmentation. Their exploitation for data issued from wavelets analysis, adapted to the astronomical context, enabled us to develop new denoising and fusion techniques of multispectral astronomical images. All our algorithms are unsupervised and were validated on synthetic and real images.

Key words : Markov quadtree, copulas, multispectral images, dimensionality reduction, segmentation, classification, multiresolution, fusion, astronomy.