



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne
pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention : Traitement du signal et télécommunications
Ecole doctorale Matisse

présentée par

Pierre Formont

préparée à SONDRA-SUPELEC — EA4454
SUPELEC ONERA NUS DSO Research Alliance
Institut d'Électronique et de Télécommunications de Rennes (I.E.T.R.)

Outils statistiques et géométriques pour la classification des images SAR polarimétriques hautement texturées

Thèse soutenue à SUPÉLEC
le 10 décembre 2013

devant le jury composé de :

Florence TUPIN

Professeur, Institut Telecom, Paris / *rapporteur*

Carlos LÓPEZ-MARTÍNEZ

Professeur Associé, Universitat Politècnica de Catalunya, Barcelone, Espagne / *rapporteur*

Cédric RICHARD

Professeur, Université de Nice / *examineur*

Frédéric PASCAL

Professeur Assistant, SONDRA-Supélec, Gif-sur-Yvette
examineur

Jean-Philippe OVARLEZ

Maître de recherche, ONERA, Palaiseau / *examineur*

Torbjørn ELTOFT

Professeur, University of Tromsø, Norvège / *examineur*

Gabriel VASILE

Chargé de recherche, GIPSA-lab, Grenoble / *invité*

Laurent FERRO-FAMIL

Professeur, Université de Rennes 1, France / *directeur de thèse*

Remerciements

Je tiens tout d'abord à remercier les différentes équipes qui m'ont accueillies tout au long de cette thèse, notamment mon équipe officielle de l'ONERA qui, malgré mes apparitions sporadiques, s'est toujours montrée chaleureuse. Un grand merci notamment à Nicolas et Elise pour les discussions scientifiques, toujours intéressantes ; à Olivier pour les échanges de bouquins et m'avoir fait découvrir Alain Damasio ; et à Eric pour les nombreuses discussions et les conseils que tu m'as apporté, ainsi que ta sélection de Blu-Ray (Space Battleship Yamato, gros film). Je tiens également à remercier l'ONERA d'une manière générale ainsi que la DGA pour avoir financé ces travaux de thèse.

J'adresse de même un énorme merci à toute l'équipe de SONDRRA, où j'ai passé le plus clair de mon temps pendant ces quatre années de thèse. J'ai réellement adoré le temps passé en votre compagnie et l'ambiance du labo va me manquer dans les prochaines années. Merci à Régis pour ton aide lors de la préparation de la soutenance et à Laetitia pour nos discussions matinales les jours où j'arrivais à me motiver pour arriver tôt. Un grand merci aux autres thésards, sans qui le labo ne serait pas aussi accueillant et agréable : les anciens, notamment Chin, Fredou et Karim pour votre bonne humeur et les soirées à Singapour (un p'tit PES ?) mais aussi les plus récents, Jérôme pour nos virées en Azeroth et sur le terrain de foot, Alice et Arnaud pour les coinches du midi et Azza pour m'accompagner au bout du monde en conférence. Je ne peux évidemment pas ne pas mentionner mes co-bureaux de choc : Mélanie, repartie dans ses vertes contrées à l'autre bout de la planète, ta joie de vivre n'a pas d'égale, et Joana pour ton entrain, les pauses clopes et la *sobrasada*, amuse-toi bien à Singapour ! Mes remerciements s'étendent aussi naturellement à Marc Lesturgie, pour toutes les opportunités que SONDRRA m'a apportées et la confiance que tu m'as accordée. Enfin, je me dois de remercier Anne-Hélène, ainsi qu'Elisabeth de l'ONERA, pour avoir géré toutes les formalités administratives et pour leur patience extrême vis-à-vis de ma relation très nonchalante avec les deadlines.

Je voudrais adresser un remerciement particulier à mes différents encadrants pour leurs nombreuses qualités, aussi bien scientifiques qu'humaines, que j'ai pu apprécier pendant ces quelques années et leur investissement dans ce travail. Merci donc à Frédéric Pascal, grand maître des statistiques et du foot en salle et à Jean-Philippe Ovarlez, gourou du radar et du choix des restos pendant les confs. J'ai beaucoup apprécié travailler avec vous, tant pour vos compétences scientifiques que pour tous les à-côtés : les discussions, les belotes, les foots, les restos, les confs et j'en passe. Merci aussi à Gabriel Vasile pour m'avoir éclairé sur un certain nombre de choses et pour tes idées, toujours novatrices. Enfin, merci à mon directeur de thèse, Laurent Ferro-Famil, pour nos discussions, plus souvent internationales que françaises, mais toujours productives.

Je tiens à remercier également ma famille et mes amis qui ont toujours été là, dans les bons comme dans les mauvais moments. Merci à mes grands-parents, mes parents, mon frère et ma sœur, ne changez jamais. Merci aux amis de longue date : Jonathan et Pierre, anciens colocs et compagnons thésards rhônalpins ; Quentin et Anaëlle, pour les soirées autour d'un rougail saucisse et les midis autour de gyozas ; Mily, *êl síla erin lú e-govaned vîn* ; et l'Infamous Mumble Clan pour les innombrables soirées passées ensemble dans cette ambiance chaotique mais efficace. Merci à tous les gens que j'ai pu rencontrer à Supélec autour d'un ballon de foot ou au détour d'un couloir, ceux qui sont déjà partis comme ceux qui sont restés : Benji, Romaric, Nabil, Thang, Islam, Christophe, Marc et les autres. Et, évidemment, un incommensurable merci aux deux larrons avec qui j'ai partagé bien des joies et des déboires, Laurie et Alexandre. Laurie, ma camarade d'embourbement dans la rédaction, j'espère que tes projets actuels t'emmènent vers des cieux plus radieux. Alex, on va se le faire un jour ce roadtrip, et il y aura du Mike Brant dans la playlist.

Enfin, merci à toi, ma Lucile, sans qui j'aurai tout envoyé balader il y a bien longtemps. Merci, du fond du cœur.

Table des matières

Introduction	1
1 État de l'art	5
1.1 Radar SAR polarimétrique	5
1.1.1 Imagerie radar	5
1.1.2 Radar à Synthèse d'Ouverture	6
1.1.3 Polarimétrie	8
1.2 Traitements classiques en classification SAR polarimétrique	13
1.2.1 Hypothèse de bruit gaussien	13
1.2.2 Estimation de la matrice de covariance	13
1.2.3 Décompositions polarimétriques	14
1.2.4 Classifieur Wishart	18
1.3 Modèle de vecteurs aléatoires sphériquement invariants	24
1.3.1 Définition	24
1.3.2 L'Estimateur du Point Fixe	25
1.3.3 Propriétés statistiques du FPE	26
1.3.4 Estimation de la texture	27
1.3.5 Influence de l'estimateur de la matrice de covariance sur les résultats de classification	28
1.4 Conclusion	30
2 Classification non supervisée	33
2.1 Test de Box	33
2.1.1 Inconvénients du classifieur Wishart	33
2.1.2 Test d'égalité de matrices de covariance	34
2.1.3 M-Test de Box	45
2.2 Algorithmes et résultats	47
2.2.1 Algorithmes	47
2.2.2 Résultats	50
2.3 Implémentation du test de Box au sein d'autres algorithmes	54
2.3.1 Application avec une segmentation initiale	54
2.3.2 Application sans segmentation initiale	63
2.3.3 Application à la classification hiérarchique	65
2.4 Conclusion	70

Table des matières

3 Centres de classe et géométrie riemannienne	71
3.1 Calcul des centres de classe pour le FPE	71
3.2 Contribution de la géométrie riemannienne	74
3.2.1 Cas des matrices réelles définies positives	75
3.2.2 Extension aux matrices hermitiennes définies positives	80
3.3 Simulations	81
3.4 Données réelles	86
3.4.1 Utilisation de la distance de Wishart	86
3.4.2 Utilisation de la distance géométrique	88
3.4.3 Répartition des pixels dans l'espace $H-\alpha$	89
3.5 Conclusion	93
4 Extension à l'imagerie hyperspectrale	95
4.1 Principe de l'imagerie hyperspectrale	95
4.1.1 Principe de l'imagerie hyperspectrale	95
4.1.2 Classification pour l'imagerie hyperspectrale	96
4.2 Application : données réelles	98
4.2.1 Jeu de données	98
4.2.2 Algorithmes et résultats	99
4.3 Conclusion	106
Conclusion	107
Annexe A Statistical Classification for Heterogeneous Polarimetric SAR Images	111
Annexe B English summary	123
Annexe C Liste des publications	127
Bibliographie	129

Table des figures

1.1	Spectre électromagnétique	6
1.2	Première image SAR acquise, en 1957	6
1.3	Schéma d'un système SAR	7
1.4	Ellipse de polarisation	11
1.5	Voisinage pour l'estimation de la matrice de covariance	14
1.6	Baie de San Francisco dans la base de Pauli : $S_{HH} - S_{VV}$, $\sqrt{2}S_{HV}$, $S_{HH} + S_{VV}$	16
1.7	Baie de San Francisco, angle α	18
1.8	Baie de San Francisco, entropie H	19
1.9	Plan H- α	19
1.10	Décomposition de Cloude-Pottier sur l'image AIRSAR de San Francisco	20
1.11	Algorithme du classifieur Wishart	22
1.12	Jeu de données : Brétigny	23
1.13	Comparaison de la classification Wishart avec une classification sur la puissance	24
1.14	Répartition des valeurs propres suivant l'estimateur de la matrice de covariance	28
1.15	Comparaison des résultats de la décomposition de Cloude-Pottier entre la SCM et le FPE	29
1.16	Comparaison des résultats du classifieur Wishart entre la SCM et le FPE avec initialisation par la décomposition de Cloude-Pottier	30
2.1	Algorithme du classifieur Box	48
2.2	Classifieur Wishart avec le FPE	49
2.3	Résultats de classification avec la SCM et carte associée.	51
2.4	Résultats de classification avec le FPE et carte associée.	53
2.5	Algorithme de segmentation par croissance de régions.	55
2.6	Jeu de données : Toulouse	56
2.7	Segmentation sur l'image de Brétigny	56
2.8	Classification avec initialisation par croissance de régions : degrés de liberté théoriques	58
2.9	Classification avec initialisation par croissance de régions : comparaison entre la SCM et le FPE	59
2.10	Grand jeu de données de Brétigny : image originale et image segmentée.	61
2.11	Classification avec initialisation par croissance de régions sur un grand jeu de données	62
2.12	K-moyennes++ sans segmentation initiale : comparaison entre la SCM et le FPE.	63
2.13	K-moyennes++ sans segmentation initiale : répartition des classes dans le plan H- α	64
2.14	Exemple de classification hiérarchique	65
2.15	Classification hiérarchique utilisant différentes fonctions de lien avec un taux de fausse alarme réglé à 0,1%	68
2.16	Comparaison de la SCM et du FPE pour la classification hiérarchique avec un taux de fausse alarme réglé à 0,01%	69
2.17	Comparaison de la distribution des données avec une distribution du χ^2 en fonction de la fonction de lien utilisée.	70
3.1	Calcul des centres de classe	72
3.2	Comparaison des résultats du classifieur Wishart suivant la méthode de calculs des centres : traditionnelle ou réestimation de la matrice.	73
3.3	Zones d'extraction des matrices de covariance de référence pour la simulation.	82
3.4	Structure de l'image simulée	83

Table des figures

3.5	Simulation de données SAR polarimétriques k-distribuées.	83
3.6	Simulations avec initialisation aléatoire et distance de Wishart.	84
3.7	Simulations avec initialisation par décomposition de Cloude-Pottier et distance de Wishart.	84
3.8	Simulations avec initialisation aléatoire et distance géométrique.	85
3.9	Simulations avec initialisation par décomposition de Cloude-Pottier et distance géométrique.	85
3.10	Image réelle avec initialisation aléatoire et distance de Wishart.	86
3.11	Image réelle avec initialisation par décomposition de Cloude-Pottier et distance de Wishart.	87
3.12	Image réelle avec initialisation aléatoire et distance géométrique.	88
3.13	Image réelle avec initialisation par décomposition de Cloude-Pottier et distance géométrique.	89
3.14	Répartition par classe des pixels dans l'espace $H-\alpha$ - FPE, décomposition de Cloude-Pottier et distance de Wishart	90
3.15	Répartition totale des pixels dans l'espace $H-\alpha$ - FPE, décomposition de Cloude-Pottier et distance de Wishart	92
4.1	Différence entre distance de Mahalanobis et distance euclidienne	97
4.2	Jeu de données hyperspectral Hyperion.	98
4.3	Résultats de classification en utilisant la SCM, le SAM et la moyenne arithmétique	100
4.4	Résultats de classification en utilisant la SCM et la moyenne arithmétique	102
4.5	Résultats de classification en utilisant la SCM et la moyenne géométrique	103
4.6	Résultats de classification en utilisant le FPE et la moyenne arithmétique	104
4.7	Résultats de classification en utilisant le FPE et la moyenne géométrique	105

Symboles et notations

\mathbb{N}	Ensemble des nombres entiers
\mathbb{R}	Ensemble des nombres réels
\mathbb{C}	Ensemble des nombres complexes
$M_n(\mathbb{C})$	Ensemble des matrices de dimension $n \times n$ à coefficients complexes
\mathbf{k}	Vecteur de données polarimétriques
\mathbf{T}	Matrice de covariance polarimétrique
$\hat{\mathbf{T}}$	Estimateur de la matrice de covariance
$\hat{\mathbf{T}}_{SCM}$	Estimateur de la matrice de covariance dans le cas gaussien (Sample Covariance Matrix)
$\hat{\mathbf{T}}_{FPE}$	Estimateur de la matrice de covariance dans le cas non gaussien (Fixed Point Estimator)
T	Opérateur de transposition
H	Opérateur de transposition-conjugaison
$\text{Tr}(\mathbf{M})$	Trace de la matrice \mathbf{M}
$\det(\mathbf{M})$	Déterminant de la matrice \mathbf{M}
$\ \mathbf{M}\ $	Norme de la matrice \mathbf{M}
$p(x)$	Densité de probabilité de la variable aléatoire x
$p(x y)$	Densité de probabilité de la variable aléatoire x sachant y
$E[x]$	Espérance mathématique de la variable aléatoire x
$\xrightarrow[N \rightarrow +\infty]{Pr}$	Convergence en probabilité quand N tend vers l'infini
$\xrightarrow[N \rightarrow +\infty]{}$	Convergence classique (presque sûre) quand N tend vers l'infini
H_0	Hypothèse nulle du test d'égalité (matrices égales)
H_1	Hypothèse alternative du test d'égalité (matrices différentes)
$L(x; H, \theta)$	Vraisemblance de la variable aléatoire x , sous l'hypothèse H en fonction des paramètres θ
Λ	Rapport de vraisemblance
P_{FA}	Probabilité de fausse alarme
$\Gamma(x)$	Fonction Gamma de x
$K(n, m)$	Fonction de Bessel modifiée du second ordre de n et m
$\mathcal{N}(m, \sigma^2)$	Loi normale réelle univariée (moyenne m , variance σ^2)
$\mathcal{N}(\mathbf{m}, \mathbf{M})$	Loi normale réelle multivariée (moyenne \mathbf{m} , covariance \mathbf{M})
$\mathcal{CN}(\mathbf{m}, \mathbf{M})$	Loi normale complexe multivariée (moyenne \mathbf{m} , covariance \mathbf{M})
$\chi^2(n)$	Loi du χ^2 à n degrés de liberté
$\gamma(\nu, \theta)$	Loi Gamma de paramètre de forme ν et paramètre d'échelle θ
$\mathcal{W}(N, m, \mathbf{M})$	Loi de Wishart de paramètres N et m et de matrice \mathbf{M}

Introduction

La cartographie de la Terre à grande échelle peut se faire de deux manières principales : l'observation par des systèmes d'acquisition d'images optiques ou par des systèmes radars. L'utilisation de systèmes radars apporte plusieurs avantages par rapport aux systèmes optiques, notamment l'indépendance par rapport à l'éclairage de la zone à observer et la capacité à traverser la couche nuageuse. Ainsi, les données peuvent être acquises à toute heure du jour et de la nuit et par tous les temps. Ces propriétés ont largement contribué au déploiement très important des systèmes radars dans de nombreuses applications, que ce soit dans le domaine civil ou dans le domaine militaire. Ces applications incluent notamment la détection de bateaux, la détection d'icebergs, la classification de l'occupation des sols et bien d'autres.

Le Radar à Synthèse d'Ouverture, appelé *Synthetic Aperture Radar* (SAR) en anglais est un système d'imagerie radar qui permet d'atteindre en pratique des résolutions spatiales élevées. Ceci en fait la base de tous les systèmes d'acquisition d'images radar à l'heure actuelle. Un système SAR fonctionne en émettant une onde électromagnétique en direction du sol et en mesurant à la fois la phase et l'amplitude du signal de retour (appelé signal rétrodiffusé). Les données ainsi mesurées sont donc complexes. De plus, dû à la propriété de polarisation des ondes électromagnétiques, des systèmes SAR appelés PolSAR (pour *Polarimetric SAR*) sont capables d'émettre et de recevoir des ondes polarisées orthogonalement à la fois en émission et en réception. Pour chaque cellule de résolution du système radar, on dispose ainsi de plusieurs mesures grâce aux différentes combinaisons possibles de polarisation, contenues dans une matrice appelée matrice de diffusion.

En raison de la nature aléatoire de l'environnement observé par un radar, où des modifications dans l'espace ou dans le temps peuvent intervenir de manière très rapide, le signal rétrodiffusé est représenté par une variable aléatoire. Traditionnellement, pour des radars disposant d'une faible résolution, chaque cellule de résolution contient un très grand nombre de réflecteurs sur lesquels l'onde provenant du radar va venir se réfléchir. Le signal mesuré pour cette cellule de résolution est donc la somme des signaux élémentaires sur chacun des réflecteurs présents au sein de la cellule. Comme ce nombre est important, le Théorème Central Limite implique que le signal rétrodiffusé est distribué selon une loi gaussienne. Ainsi, les données SAR polarimétriques sont généralement modélisées par une loi gaussienne complexe multivariée.

Ces données polarimétriques contiennent les relations entre les différents canaux de polarisation qui peuvent être reliées à des informations physiques sur la scène imagée. Pouvoir analyser ces relations permet de remonter aux processus physiques se déroulant lorsqu'une onde électromagnétique rencontre un objet et ainsi de pouvoir caractériser cet objet. La nature aléatoire de ces interactions peut être étudiée en considérant les moments du second ordre des données, à savoir leur matrice de covariance.

Comme il a été mentionné précédemment, les images SAR polarimétriques peuvent être utilisées pour de nombreuses applications et en particulier, pour la classification de l'occupation des sols. Ainsi, la classification des images SAR polarimétriques est un domaine de recherche très actif. Deux approches principales sont apparues dans la littérature correspondante.

La première approche est de classifier les pixels grâce à leurs caractéristiques physiques. Plusieurs décompositions de matrices ont été proposées à ce sujet :

- des décompositions cohérentes, basées sur la matrice de diffusion : décomposition de Cameron [19] ou de Krogager [48], afin de caractériser les réflecteurs spéculaires. C'est une approche déterministe.

Introduction

- des décompositions non-cohérentes, basées sur la matrice de covariance : décomposition de Freeman [34] ou de Cloude-Pottier [24], qui permettent de caractériser les réflecteurs étendus. C'est une approche statistique, non-déterministe.

La deuxième approche est de classifier les images à partir de leurs propriétés statistiques. Kong a proposé une mesure de distance pour un type de données SAR polarimétriques [46], qui a été étendue par la suite par van Zyl [74], Yueh [82] et Lee [51]. La mesure de distance proposée par Lee est basée sur la distribution de Wishart associée à la matrice de covariance et résultant de l'hypothèse de la distribution gaussienne du signal rétrodiffusé.

Cependant, les systèmes d'acquisitions PolSAR récents sont capables d'obtenir des images en très haute résolution, de l'ordre du décimètre. Dans ce cas, le nombre de réflecteurs à l'intérieur de chaque cellule de résolution diminue de manière drastique et varie grandement d'une cellule à l'autre. Ceci entraîne une hétérogénéité dans la distribution du signal rétrodiffusé qui n'est pas prise en compte par la distribution gaussienne. Il est donc nécessaire de modéliser le signal de manière à pouvoir tenir compte de cette hétérogénéité. Pour cela, le modèle des vecteurs aléatoires sphériquement invariants (ou *Spherically Invariant Random Vectors*, soit SIRV), initialement proposé par Yao [81] pour la théorie de l'information, a été introduit dans la communauté de la détection radar, notamment par Farina [33]. Ce modèle théorique a fait l'objet de validations par un grand nombre de campagnes de mesures, notamment par le MIT (Massachusetts Institute of Technology) [12] ou l'ONERA (The French Aerospace Lab, anciennement Office National d'Études et Recherches Aérospatiales), ainsi que par les travaux, entre autres, de Trunk [71], Goldman [37], Ward [76] ou Watts [77].

Par ailleurs, une grande partie des algorithmes de classification pour les images SAR polarimétriques et en particulier le classifieur Wishart proposé par Lee dans [50] fait appel à la moyenne de matrices de covariance pour calculer la matrice correspondant au centre de chaque classe. Ce centre est calculé comme la moyenne arithmétique des matrices. Cependant, les travaux récents de Barbaresco [6–8], inspirés par les travaux de Moakher [55], ainsi que les travaux de Devlaminck [30] dans le cadre de l'imagerie optique, ont permis de mettre en avant le fait que l'espace des matrices de covariance, à savoir l'espace des matrices hermitiennes définies positives, n'est pas un espace euclidien mais est doté d'une structure de variété riemannienne. Il est alors nécessaire d'employer des outils prenant en compte cette structure non-euclidienne. C'est notamment le cas de la géométrie de l'information et des travaux de Rao [4, 17] qui permettent d'obtenir de manière relativement simple une expression de la moyenne de matrices hermitiennes définies positives.

Le but de ce travail de thèse est donc d'étudier l'impact de ces nouveaux outils sur la classification d'images SAR polarimétriques en termes de précision de classification et d'interprétation du sens physique des résultats de classification, au travers des décompositions polarimétriques usuelles.

Ce manuscrit est construit en quatre chapitres, qui abordent chacun un des points principaux présentés dans les paragraphes précédents. Le premier chapitre expose dans un premier temps le principe de l'imagerie SAR et de la polarimétrie, au travers de la caractérisation de la propagation des ondes électromagnétiques à partir des équations de Maxwell. Les outils statistiques classiques employés pour la classification polarimétrique sont ensuite rappelés, à savoir l'hypothèse d'une distribution gaussienne du signal rétrodiffusé et l'estimation de la matrice de covariance de ce signal, communément appelée la Sample Covariance Matrix (SCM). Par la suite, des décompositions polarimétriques, basées notamment sur la matrice de covariance et destinées à expliquer la nature des phénomènes physiques se déroulant dans une cellule de résolution, ainsi qu'un algorithme de classification classique (appelé classifieur Wishart) sont présentés. Enfin, le modèle de signal non-gaussien est introduit avec l'estimateur adapté à cette hypothèse ainsi que les propriétés de l'algorithme permettant d'obte-

nir cet estimateur et les propriétés statistiques de l'estimateur proprement dit.

Le deuxième chapitre est consacré à l'étude d'un test statistique d'égalité de matrices de covariance et à son application dans des algorithmes de classification polarimétrique. L'expression du test pour le cas gaussien est rappelée puis son extension au cas SIRV est présentée. Dans chaque cas, deux scénarios sont envisagés : un scénario où une des deux matrices de covariance est connue et un scénario où les deux matrices de covariance sont inconnues. Le test fournit une distance statistique entre deux matrices, qui est employée dans trois algorithmes de classifications différents : le classifieur Wishart, un algorithme développé pendant ce doctorat pour tirer parti de la région critique fournie par le test et un algorithme de segmentation hiérarchique. Les résultats sont présentés sur des données réelles, acquises par le système RAMSES de l'ONERA.

Dans le troisième chapitre, les principes de la géométrie de l'information sont rappelés, ainsi que la manière de calculer la moyenne géométrique de matrices de covariance et la distance géométrique entre deux matrices de covariance. Ces résultats sont d'abord présentés pour le cas de matrices réelles symétriques définies positives avant d'être étendus au cas des matrices hermitiennes définies positives. L'apport de cette technique est ensuite étudiée sur des images simulées puis sur des images réelles. Une interprétation polarimétrique des résultats est proposée.

Le quatrième et dernier chapitre présente des résultats préliminaires sur des travaux d'extension des méthodes développées pour le cas de données SAR polarimétriques et leur potentialité d'application sur des données hyperspectrales. L'imagerie hyperspectrale consiste à acquérir une image dans un grand nombre de bandes spectrales situées dans la région du spectre électromagnétique contenant la lumière visible. Malgré quelques différences, ces deux types de données sont suffisamment proches pour justifier de cette tentative d'extension de ces méthodes, à savoir un modèle de signal non-gaussien et l'emploi de la géométrie de l'information. Des premiers résultats sont présentés sur des données réelles de la NASA (National Aeronautics and Space Administration).

Enfin, nous concluons ce travail en soulignant les points importants qui s'en dégagent et en proposant des perspectives pour poursuivre ces travaux.

Sommaire

1.1	Radar SAR polarimétrique	5
1.1.1	Imagerie radar	5
1.1.2	Radar à Synthèse d'Ouverture	6
1.1.3	Polarimétrie	8
1.2	Traitements classiques en classification SAR polarimétrique	13
1.2.1	Hypothèse de bruit gaussien	13
1.2.2	Estimation de la matrice de covariance	13
1.2.3	Décompositions polarimétriques	14
1.2.4	Classifieur Wishart	18
1.3	Modèle de vecteurs aléatoires sphériquement invariants	24
1.3.1	Définition	24
1.3.2	L'Estimateur du Point Fixe	25
1.3.3	Propriétés statistiques du FPE	26
1.3.4	Estimation de la texture	27
1.3.5	Influence de l'estimateur de la matrice de covariance sur les résultats de classification	28
1.4	Conclusion	30

Ce chapitre présente la problématique étudiée dans cette thèse. Les principes du radar SAR polarimétrique ainsi que ceux de la classification sont rappelés. La modélisation utilisée pour le signal rétrodiffusé, appelée modèle SIRV, est ensuite introduite. La définition et les propriétés principales des SIRV seront présentées. Enfin, l'estimateur de la matrice de covariance associé au cas SIRV utilisé tout au long de cette thèse est introduit.

1.1 Radar SAR polarimétrique

1.1.1 Imagerie radar

Un radar est à l'origine un système permettant de déterminer la distance et la vitesse d'un objet en utilisant des ondes radio. L'origine du terme "radar" est d'ailleurs l'acronyme anglais **RA**dio **D**etection **A**nd **R**anging. Mais de nombreuses autres applications sont possibles dans un grand nombre de domaines. Une de ces applications est l'imagerie radar : au lieu d'utiliser la lumière visible pour imager une zone (comme dans l'imagerie optique classique), on utilise des ondes électromagnétiques de fréquences beaucoup plus courtes, les ondes radio (voir Fig. 1.1). Ce système actif fonctionne à n'importe quelle heure du jour et de la nuit puisque la source d'illumination est le radar lui-même. Certaines bandes de fréquence permettent aussi de traverser la couche nuageuse, permettant ainsi de faire des acquisitions par temps couvert.

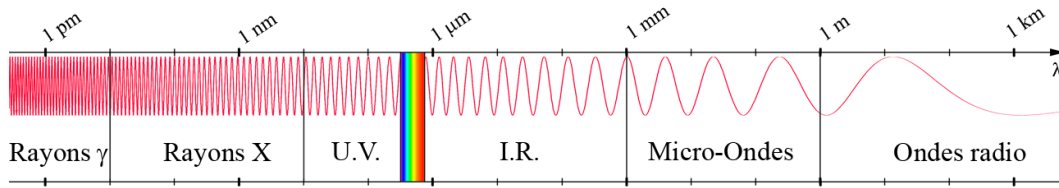


FIGURE 1.1 – Spectre électromagnétique

1.1.2 Radar à Synthèse d'Ouverture

Le Radar à Synthèse d'Ouverture (RSO), appelé Synthetic Aperture Radar (SAR) en anglais, est un système radar permettant d'améliorer grandement la résolution en azimut par rapport à un système radar imageur classique. Les premiers systèmes SAR ont été développés au début des années 50 et les premières images acquises à la fin de cette décennie. La Fig. 1.2 montre la première image SAR acquise par l'Université du Michigan en 1957, grâce à des lampes à mercures. Ce n'est qu'avec l'avènement du laser et des avancées en informatique et en électronique que la qualité des images s'est améliorée. Depuis le lancement de SEASAR, le premier SAR numérique embarqué sur un satellite, en 1978, les systèmes SAR ont permis d'apporter une quantité importante d'information pour l'étude de la Terre.

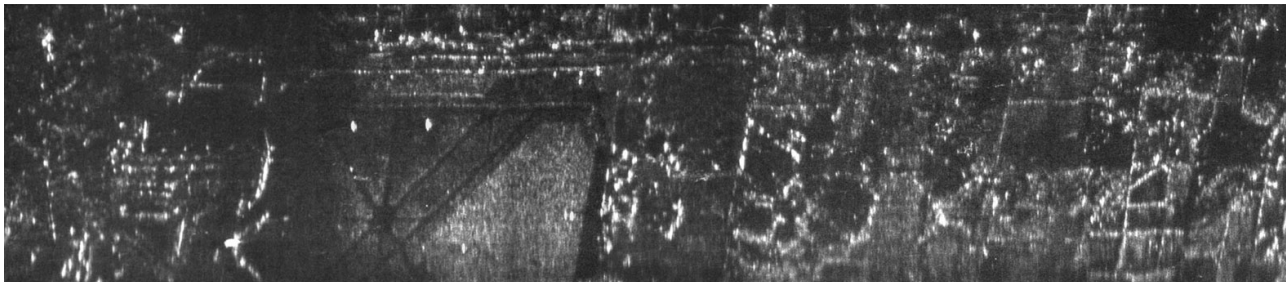


FIGURE 1.2 – Première image SAR acquise, en 1957

Un radar imageur fonctionne en émettant des ondes électromagnétiques dans une direction donnée, généralement vers le sol. La surface illuminée diffuse l'onde incidente dans toutes les directions, dont la direction d'émission. La partie qui est réémise vers le radar est enregistrée et traitée afin de former l'image radar. On appelle cette partie le signal rétrodiffusé. La Fig. 1.3 présente, de manière schématisée, le fonctionnement d'un système SAR. Un porteur — avion ou satellite — se déplace selon une trajectoire donnée et illumine une zone du sol.

La résolution spatiale est un point très important de tout système imageur. C'est la capacité d'un radar à séparer deux objets voisins selon une certaine distance. Si la résolution est suffisamment grande, les deux objets voisins seront placés dans deux cases de résolutions séparées. Sinon, ils seront situés dans la même case par combinaison de l'énergie réfléchie par les deux objets.

Cependant, la résolution est directement liée à la durée des impulsions comprimées émises par le radar : pour augmenter la résolution, il faut réduire la durée des impulsions ou augmenter la bande émise du signal si les impulsions sont codées. Il est aussi nécessaire que ces impulsions aient une énergie élevée afin de minimiser les effets de bruit induits par les rayonnements naturels mais surtout les bruits de mesure inhérents au système radar et de pouvoir détecter les signaux émis : on cherche à maximiser le rapport signal sur bruit. La génération

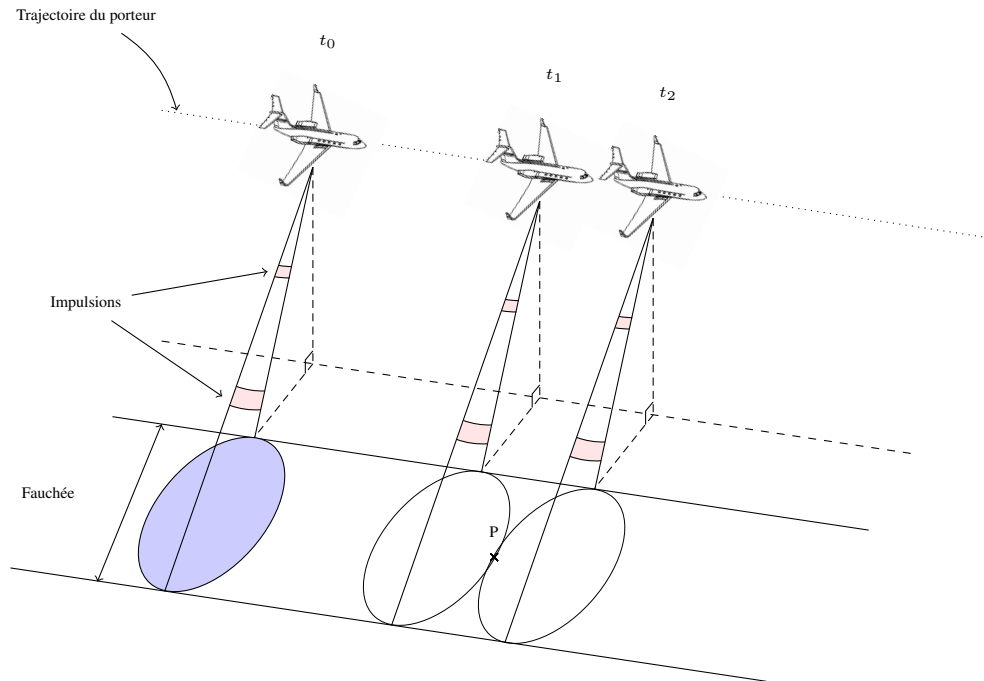


FIGURE 1.3 – Schéma d'un système SAR

d'impulsions courtes à grande énergie étant trop complexe à réaliser en pratique, les impulsions utilisées sont plus longues et codées mais ont une énergie suffisamment grande.

La résolution en distance δ_r est donnée par l'équation suivante [29] :

$$\delta_r = \frac{c}{2\beta}, \quad (1.1)$$

avec c la célérité de la lumière et β la bande de fréquences émises.

La résolution dans la direction perpendiculaire à la distance, aussi appelée azimut, est donnée, de manière approchée, par l'équation suivante [29] :

$$\delta_a = \theta R = R \frac{\lambda}{L}, \quad (1.2)$$

avec R la distance entre le SAR et le sol, θ l'ouverture angulaire du SAR, λ la longueur d'onde du signal et L la taille de l'antenne en azimut.

Pour améliorer la résolution en azimut, la manière la plus efficace serait d'augmenter la taille de l'antenne. Or, pour atteindre des bonnes résolutions, la taille de l'antenne nécessaire serait trop élevée pour être réalisable en pratique. C'est ici qu'intervient le principe du système SAR : la synthèse d'ouverture. Dû à l'ouverture du faisceau incident, un même point est illuminé sur une certaine durée lors du déplacement du radar (cf. point P sur la Fig. 1.3, illuminé sur l'intervalle $[t_1, t_2]$). En intégrant les signaux rétrodiffusés pour chaque point le long de la trajectoire du système radar, on dispose alors d'une antenne synthétique virtuelle dont la taille est beaucoup plus élevée que celle de l'antenne réelle.

On peut alors considérer l'ouverture angulaire et la longueur de l'antenne synthétique [29] :

$$\theta_{SA} = \frac{\lambda}{2L_{SA}} \quad \text{et} \quad L_{SA} \leq \theta R = R \frac{\lambda}{L}. \quad (1.3)$$

On a alors une résolution en azimuth donnée par [29] :

$$\delta_a \geq R\theta_{SA} = \frac{L}{2}. \quad (1.4)$$

1.1.3 Polarimétrie

La polarimétrie est l'étude de la polarisation des ondes électromagnétiques. Cette section est consacrée à la présentation des différents concepts utilisés pour les radars SAR polarimétriques. En premier lieu, les mécanismes de propagation des ondes électromagnétiques seront rappelés. Ensuite, la matrice de diffusion sera introduite. Enfin, la matrice de covariance polarimétrique sera présentée.

1.1.3.1 Propagation des ondes électromagnétiques

Une onde électromagnétique est la représentation ondulatoire d'un rayonnement électromagnétique, phénomène qui se manifeste par la présence d'un champ électrique couplé à un champ magnétique. La propagation des ondes électromagnétiques est régie par les équations de Maxwell, qui décrivent le comportement des champs électrique et magnétique dans l'espace et au cours du temps. Ces équations sont les suivantes :

$$\vec{\nabla} \wedge \vec{E}(\vec{r}, t) = -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t}, \quad (1.5)$$

$$\vec{\nabla} \wedge \vec{H}(\vec{r}, t) = \vec{J}_T(\vec{r}, t) + \frac{\partial \vec{D}(\vec{r}, t)}{\partial t}, \quad (1.6)$$

$$\vec{\nabla} \cdot \vec{D}(\vec{r}, t) = \rho(\vec{r}, t), \quad (1.7)$$

$$\vec{\nabla} \cdot \vec{B}(\vec{r}, t) = 0, \quad (1.8)$$

avec \vec{r} le vecteur position (le point de l'espace où sont évaluées les différentes quantités), t l'instant où sont évaluées les différentes quantités, $\vec{E}(\vec{r}, t)$ le champ électrique, $\vec{H}(\vec{r}, t)$ le champ magnétique, $\vec{D}(\vec{r}, t)$ l'induction électrique, $\vec{B}(\vec{r}, t)$ l'induction magnétique, $\vec{J}_T(\vec{r}, t)$ la densité de courant et $\rho(\vec{r}, t)$ la densité de charge libre [69]. $\vec{\nabla}$ est l'opérateur nabla, $\vec{\nabla} \wedge \vec{x}$ l'opérateur rotationnel et $\vec{\nabla} \cdot \vec{x}$ l'opérateur divergence [67].

La densité de courant $\vec{J}_T(\vec{r}, t) = \vec{J}_s(\vec{r}, t) + \vec{J}_c(\vec{r}, t)$ est composée de deux termes :

- $\vec{J}_s(\vec{r}, t)$ est un terme induit par la présence de sources dans le milieu,
- $\vec{J}_c(\vec{r}, t) = \sigma \vec{E}(\vec{r}, t)$ dépend de la conductivité σ du milieu.

Les relations entre champs et inductions sont les suivantes :

$$\begin{aligned} \vec{D}(\vec{r}, t) &= \varepsilon \vec{E}(\vec{r}, t) + \vec{P}(\vec{r}, t), \\ \vec{B}(\vec{r}, t) &= \mu \left(\vec{H}(\vec{r}, t) + \vec{M}(\vec{r}, t) \right), \end{aligned} \quad (1.9)$$

avec $\vec{P}(\vec{r}, t)$ le vecteur polarisation, $\vec{M}(\vec{r}, t)$ le vecteur aimantation, ε la permittivité du milieu et μ la perméabilité du milieu.

Si le milieu est linéaire, homogène et isotrope, alors

$$\vec{\mathbf{P}}(\vec{\mathbf{r}}, t) = \vec{\mathbf{M}}(\vec{\mathbf{r}}, t) = \vec{\mathbf{J}}_s(\vec{\mathbf{r}}, t) = \vec{\mathbf{0}}. \quad (1.10)$$

Les Eqs. (1.7), (1.9) et (1.10) impliquent la relation suivante :

$$\vec{\nabla} \cdot \vec{\mathbf{E}}(\vec{\mathbf{r}}, t) = \frac{\rho(\vec{\mathbf{r}}, t)}{\varepsilon}. \quad (1.11)$$

On dispose de la relation de calcul vectoriel suivante [67] :

$$\vec{\nabla} \wedge [\vec{\nabla} \wedge \vec{\mathbf{x}}] = \vec{\nabla} [\vec{\nabla} \cdot \vec{\mathbf{x}}] - \Delta \vec{\mathbf{x}}, \quad (1.12)$$

avec $\vec{\mathbf{x}}$ un vecteur quelconque et Δ l'opérateur Laplacien.

En appliquant la relation de l'Eq. (1.12) au vecteur champ électrique $\vec{\mathbf{E}}(\vec{\mathbf{r}}, t)$ et d'après les Eqs. (1.5) et (1.11), on obtient l'équation de propagation suivante :

$$\Delta \vec{\mathbf{E}}(\vec{\mathbf{r}}, t) - \mu\varepsilon \frac{\partial^2 \vec{\mathbf{E}}(\vec{\mathbf{r}}, t)}{\partial t^2} - \mu\sigma \frac{\partial \vec{\mathbf{E}}(\vec{\mathbf{r}}, t)}{\partial t} = \frac{\vec{\nabla} \rho(\vec{\mathbf{r}}, t)}{\varepsilon}, \quad (1.13)$$

avec $\vec{\nabla} \cdot \vec{\mathbf{x}}$ l'opérateur gradient [67].

On suppose de plus que le milieu est libre de charges, *i.e.* $\frac{\vec{\nabla} \rho(\vec{\mathbf{r}}, t)}{\varepsilon} = \vec{\mathbf{0}}$. L'Eq. (1.13) devient alors :

$$\Delta \vec{\mathbf{E}}(\vec{\mathbf{r}}, t) - \mu\varepsilon \frac{\partial^2 \vec{\mathbf{E}}(\vec{\mathbf{r}}, t)}{\partial t^2} - \mu\sigma \frac{\partial \vec{\mathbf{E}}(\vec{\mathbf{r}}, t)}{\partial t} = 0. \quad (1.14)$$

Une solution de l'Eq. (1.13) particulièrement adaptée à l'étude de la polarisation est l'onde plane monochromatique [69]. On peut alors simplifier grandement l'étude de l'équation de propagation en considérant le vecteur champ électrique $\vec{\mathbf{E}}(\vec{\mathbf{r}}, t)$ comme la partie réelle du vecteur complexe $\underline{\vec{\mathbf{E}}}(\vec{\mathbf{r}}, t)$ défini de la manière suivante :

$$\underline{\vec{\mathbf{E}}}(\vec{\mathbf{r}}, t) = \underline{\vec{\mathbf{E}}_0} \exp\left(j\left(\omega t - \underline{\mathbf{k}} \cdot \vec{\mathbf{r}}\right)\right) = \underline{\vec{\mathbf{E}}}(\vec{\mathbf{r}}) \exp(j\omega t), \quad (1.15)$$

avec $\underline{\vec{\mathbf{E}}_0}$ l'amplitude complexe du vecteur $\underline{\vec{\mathbf{E}}}(\vec{\mathbf{r}}, t)$, $\underline{\mathbf{k}}$ le vecteur d'onde et ω la pulsation de l'onde.

L'Eq. (1.14) peut donc s'écrire de la manière suivante :

$$\Delta \underline{\vec{\mathbf{E}}}(\vec{\mathbf{r}}) + \omega^2 \mu\varepsilon \left(1 - j \frac{\sigma}{\varepsilon\omega}\right) \underline{\vec{\mathbf{E}}}(\vec{\mathbf{r}}) = 0. \quad (1.16)$$

On définit l'amplitude complexe du vecteur d'onde $\underline{\mathbf{k}}$, notée \underline{k} , par :

$$\underline{k} = \sqrt{\omega^2 \mu\varepsilon \left(1 - j \frac{\sigma}{\varepsilon\omega}\right)} = \beta - j\alpha. \quad (1.17)$$

On peut représenter le champ électrique dans une base orthogonale $(\vec{\mathbf{x}}, \vec{\mathbf{y}}, \vec{\mathbf{z}})$ telle que $\vec{\mathbf{k}} = \vec{\mathbf{z}}$. Ceci permet

d'exprimer le vecteur $\vec{E}(\vec{r})$ de la manière suivante :

$$\begin{aligned} \vec{E}(\vec{r}) &= \underline{E}_0 \exp\left(-j\vec{k}\vec{r}\right) \\ \Rightarrow \underline{E}(z) &= \underline{E}_0 \exp(-\alpha z - j\beta z) \quad \text{avec } E_{0z} = 0, \end{aligned} \quad (1.18)$$

car le vecteur d'onde est orthogonal au vecteur champ électrique.

Dans le domaine temporel, l'Eq. (1.18) devient :

$$\underline{E}(z, t) = \begin{bmatrix} E_{0x} e^{-\alpha z} \cos(\omega t - kz + \delta_x) \\ E_{0y} e^{-\alpha z} \cos(\omega t - kz + \delta_y) \\ 0 \end{bmatrix}. \quad (1.19)$$

Le terme en α est un terme d'atténuation commun à toutes les composantes du vecteur, il n'est donc pas relié à la polarisation de l'onde. On peut donc considérer $\alpha = 0$. L'évolution spatiale d'une onde plane monochromatique au cours du temps est donc donnée par :

$$\underline{E}(z, t) = \begin{bmatrix} E_{0x} \cos(\omega t - kz + \delta_x) \\ E_{0y} \cos(\omega t - kz + \delta_y) \\ 0 \end{bmatrix}. \quad (1.20)$$

1.1.3.2 Ellipse de polarisation

L'Eq. (1.20) donne l'expression de l'évolution spatiale d'une onde plane monochromatique au cours du temps. Pour représenter plus simplement cette évolution, on se place à une position donnée dans l'espace et on observe l'évolution au cours du temps [47, 56]. La relation entre les 2 composantes du vecteur $\underline{E}(z, t)$ est donnée par :

$$\left(\frac{E_x(z_0, t)}{E_{0x}}\right)^2 + \left(\frac{E_y(z_0, t)}{E_{0y}}\right)^2 - 2\frac{E_x(z_0, t)E_y(z_0, t)}{E_{0x}E_{0y}} \cos(\delta_y - \delta_x) = \sin^2(\delta_y - \delta_x). \quad (1.21)$$

L'Eq. (1.21) est l'équation d'une ellipse, appelée ellipse de polarisation qui est représentée Fig. 1.4 et peut être caractérisée par 3 paramètres :

— son amplitude A , déterminée par :

$$A = \sqrt{E_{0x}^2 + E_{0y}^2}, \quad (1.22)$$

— son orientation ϕ , déterminée par :

$$\tan 2\phi = 2\frac{E_{0x}E_{0y}}{E_{0x}^2 - E_{0y}^2} \cos \delta \quad \text{avec } \delta = \delta_y - \delta_x, \quad (1.23)$$

— son ellipticité τ , déterminée par :

$$|\sin 2\tau| = 2\frac{E_{0x}E_{0y}}{E_{0x}^2 + E_{0y}^2} |\sin \delta|. \quad (1.24)$$

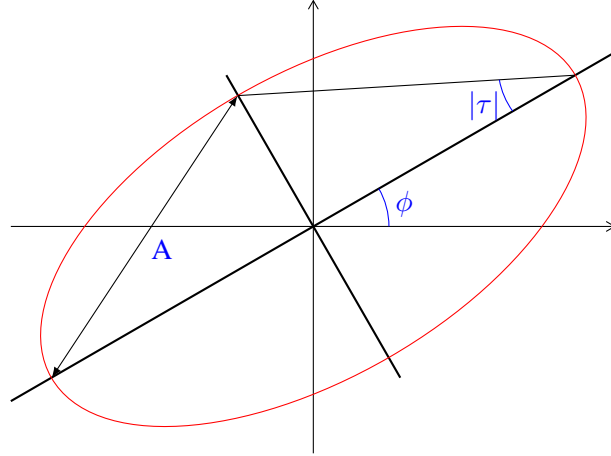


FIGURE 1.4 – Ellipse de polarisation

1.1.3.3 Vecteur de Jones

Le vecteur de Jones permet de représenter la polarisation d'une onde en utilisant le minimum d'information possible [5, 11]. Le champ électrique décrit par l'Eq. (1.20) s'écrit :

$$\begin{aligned} \vec{\underline{E}}(z, t) &= \begin{bmatrix} E_{0x} \cos(\omega t - kz + \delta_x) \\ E_{0y} \cos(\omega t - kz + \delta_y) \\ 0 \end{bmatrix} = \text{Re} \left\{ \begin{bmatrix} E_{0x} e^{j\delta_x} \\ E_{0y} e^{j\delta_y} \\ 0 \end{bmatrix} e^{-jkz} e^{j\omega t} \right\}, \\ &= \text{Re} \left\{ \vec{\underline{E}}(z) e^{j\omega t} \right\}. \end{aligned} \quad (1.25)$$

Le vecteur de Jones $\underline{\mathbf{E}}$ est donc défini par :

$$\underline{\mathbf{E}} = \vec{\underline{E}}(z) \Big|_{z=0} = \begin{bmatrix} E_{0x} e^{j\delta_x} \\ E_{0y} e^{j\delta_y} \end{bmatrix}, \quad (1.26)$$

puisque le terme en z est nul.

On peut aussi l'écrire en fonction des caractéristiques de l'ellipse de polarisation :

$$\begin{aligned} \underline{\mathbf{E}} &= A e^{j\alpha} \begin{bmatrix} \cos \phi \cos \tau - j \sin \phi \sin \tau \\ \sin \phi \cos \tau + j \cos \phi \sin \tau \end{bmatrix}, \\ &= A e^{j\alpha} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} \cos \tau \\ j \sin \tau \end{bmatrix}. \end{aligned} \quad (1.27)$$

1.1.3.4 Matrice de diffusion

Lorsqu'une onde électromagnétique se propage, elle peut rencontrer un objet et interagir avec celui-ci. Dans ce cas, une partie de l'énergie est absorbée par l'objet et le reste est réfléchi ou diffusé. Les propriétés de polarisation de l'onde réfléchie peuvent être différentes de celles de l'onde incidente : si l'on peut mesurer ces changements, il est alors raisonnable d'imaginer pouvoir caractériser l'objet rencontré par l'onde incidente.

Le vecteur de Jones présenté dans la section précédente permet de caractériser l'état de polarisation d'une onde. On peut alors relier le vecteur de Jones de l'onde incidente $\underline{\mathbf{E}}_I$ avec celui de l'onde réfléchie $\underline{\mathbf{E}}_R$ grâce à

la matrice de diffusion, aussi appelée matrice de Sinclair [47] :

$$\underline{\mathbf{E}}_R = \frac{e^{-jkr}}{r} \mathbf{S} \underline{\mathbf{E}}_I, \quad (1.28)$$

avec $\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$ la matrice de diffusion et les S_{ij} les coefficients de diffusion complexes [53]. Les éléments diagonaux de \mathbf{S} sont appelés les coefficients co-polarisés (co-pol) car ils ont la même polarisation en émission et en réception. Les éléments non-diagonaux sont appelés cross-polarisés (cross-pol).

Dans cette thèse, la base de polarisation employée est la base de polarisation cartésienne horizontale-verticale (H,V) car les images employées sont décrites dans cette base. On a alors :

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}. \quad (1.29)$$

On se place aussi dans la configuration dite monostatique, où l'antenne d'émission sert aussi d'antenne de réception. La matrice de Sinclair est alors symétrique et les coefficients cross-pol égaux : $S_{HV} = S_{VH}$.

1.1.3.5 Matrices de cohérence et de covariance polarimétrique

Quand un radar de type SAR illumine une surface, le signal rétrodiffusé pour une cellule de résolution donnée est la somme des réflexions du signal incident sur tous les réflecteurs élémentaires présents au sein de la cellule. Ainsi, la distance entre chaque réflecteur élémentaire et le radar varie, ce qui rend les ondes réfléchies incohérentes en phase. Cette incohérence en phase, qu'on peut rapprocher des phénomènes de type marche aléatoire, se traduit sur les images par la présence d'un bruit, appelé *speckle* [38].

L'approche cohérente *via* l'ellipse de polarisation et les vecteurs de Jones n'est donc plus applicable dans ce cas. Il est alors plus aisé de traiter le signal comme une variable aléatoire et d'en étudier les moments du second ordre. On peut réécrire la matrice de Sinclair de l'Eq. (1.29) sous forme vectorielle en la projetant sur une base orthogonale. Deux bases couramment utilisées sont la base lexicographique et la base de Pauli. Le vecteur résultant d'une telle projection est appelé vecteur cible et usuellement noté \mathbf{k} :

$$\begin{aligned} \mathbf{k}_L &= [S_{HH} \quad \sqrt{2}S_{HV} \quad S_{VV}]^T, \\ \mathbf{k}_P &= \frac{1}{\sqrt{2}} [S_{HH} + S_{VV} \quad S_{HH} - S_{VV} \quad 2S_{HV}]^T, \end{aligned} \quad (1.30)$$

où T indique l'opérateur de transposition. Les termes en $\sqrt{2}$ sont introduits pour garder une puissance, aussi appelée *span*, constante :

$$\text{span}(\mathbf{S}) = \|\mathbf{k}_L\|^2 = \|\mathbf{k}_P\|^2 = |S_{HH}|^2 + |S_{VV}|^2 + 2|S_{HV}|^2. \quad (1.31)$$

Les moments du second ordre sont donnés par les matrices de covariance \mathbf{T} et de cohérence \mathbf{C} polarimétriques de la manière suivante :

$$\begin{aligned} \mathbf{T} &= \mathbb{E} [\mathbf{k}_L \mathbf{k}_L^H], \\ \mathbf{C} &= \mathbb{E} [\mathbf{k}_P \mathbf{k}_P^H], \end{aligned} \quad (1.32)$$

où $\mathbb{E}[\cdot]$ représente l'espérance mathématique et H l'opérateur de transposition conjuguaison.

1.2 Traitements classiques en classification SAR polarimétrique

1.2.1 Hypothèse de bruit gaussien

1.2.1.1 Une seule polarisation

Si l'on fait les hypothèses suivantes :

- il y a un grand nombre de réflecteurs dans chaque cellule de résolution,
- la distance au radar est beaucoup plus grande que les longueurs d'onde utilisées par ce radar,
- la surface illuminée est homogène et très rugueuse vis-à-vis de la longueur d'onde utilisée par le radar SAR.

Alors, on peut représenter la phase de la somme des réflexions élémentaires par une distribution uniforme entre $-\pi$ et π . D'après le Théorème Central Limite, les parties réelle et imaginaire de cette somme, x et y , étant indépendantes et identiquement distribuées (i.i.d.), cette somme converge vers une loi gaussienne de moyenne nulle et de variance $\frac{\sigma^2}{2}$.

1.2.1.2 Multi-polarisation

Dans le cas d'une acquisition avec les trois canaux de polarisation (HH , HV et VV), le speckle apparaît non seulement dans les images d'intensité mais aussi dans les relations entre polarisations. Donc les hypothèses mentionnées précédemment pour le cas d'une seule polarisation s'appliquent de la même manière. C'est-à-dire que si l'on considère l'un des vecteurs \mathbf{k} de l'Eq. (1.30), alors ses composantes sont des variables aléatoires suivant une loi gaussienne complexe, car leurs parties réelles et imaginaires sont i.i.d. gaussiennes. Donc, d'après [38], le vecteur \mathbf{k} suit une loi gaussienne complexe circulaire multivariée de moyenne nulle et de matrice de covariance \mathbf{T} dont la densité de probabilité (ddp) s'écrit :

$$p(\mathbf{k}) = \frac{1}{\pi^m |\mathbf{T}|} \exp(-\mathbf{k}^H \mathbf{T}^{-1} \mathbf{k}). \quad (1.33)$$

1.2.2 Estimation de la matrice de covariance

L'expression de la matrice de covariance ou de la matrice de cohérence est donnée par l'Eq. (1.32). Par souci de simplicité, dans le reste de ce document, on parlera indistinctement de matrice de covariance bien qu'on utilise la base de Pauli et qu'on devrait utiliser le terme matrice de cohérence. Cette matrice est en pratique inconnue, elle doit donc être estimée. Traditionnellement, on a besoin de plusieurs réalisations de la variable aléatoire représentant le vecteur cible afin de pouvoir l'estimer. Cependant, en pratique, lors de l'acquisition d'une image SAR, on ne dispose malheureusement que d'une seule réalisation temporelle du vecteur cible. Il est en effet peu, voire pas du tout, réalisable de faire plusieurs "passes" (ou acquisitions) rapprochées dans le temps avec un avion ou un satellite afin d'obtenir les différentes réalisations. Pour pallier ce problème pratique, on utilise alors un voisinage spatial dans l'image du pixel considéré \mathbf{k} composé d'un certain nombre de ses voisins en supposant que ce voisinage local est homogène au sens statistique, c'est-à-dire que les propriétés statistiques y varient peu. Ce voisinage est aussi appelé données secondaires. La Fig. 1.5 illustre ce concept.

On considère donc le N -échantillon suivant :

$$(\mathbf{k}_1, \dots, \mathbf{k}_N) \quad \text{avec les } \mathbf{k}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}) \text{ i.i.d.} \quad (1.34)$$

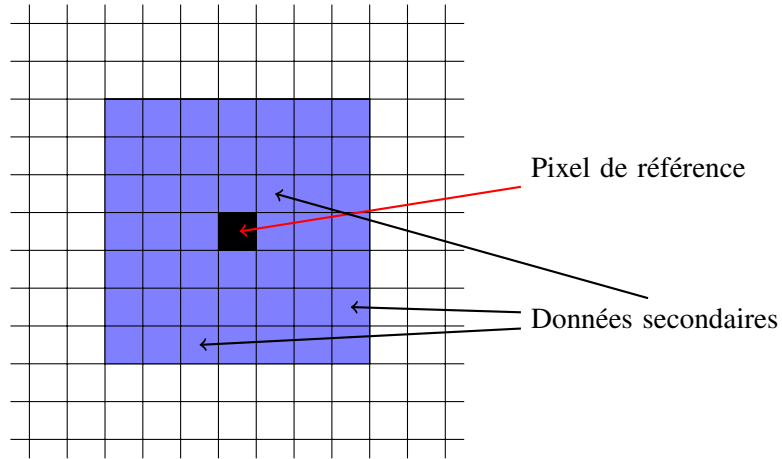


FIGURE 1.5 – Voisinage pour l'estimation de la matrice de covariance

Dans ce cas, l'Estimateur du Maximum de Vraisemblance (EMV), noté $\hat{\mathbf{T}}_{SCM}$, de la matrice de covariance \mathbf{T} de \mathbf{k} est donné par la Sample Covariance Matrix (SCM) ou matrice de covariance empirique :

$$\hat{\mathbf{T}}_{SCM} = \frac{1}{N} \sum_{i=1}^N \mathbf{k}_i \mathbf{k}_i^H \quad (1.35)$$

La méthode d'estimation par maximum de vraisemblance sera développée plus en détail dans le chapitre 2.

En notant $\mathbf{A} = N\hat{\mathbf{T}}_{SCM}$, sous l'hypothèse gaussienne mentionnée précédemment pour les \mathbf{k}_i , \mathbf{A} suit une distribution de Wishart complexe à $2N$ degrés de liberté (ddl) et de matrice de covariance \mathbf{T} caractérisée par la ddp suivante [59] :

$$p(\mathbf{A}) = \frac{|\mathbf{A}|^{N-m} \exp(-\text{Tr}(\mathbf{T}^{-1}\mathbf{A}))}{K(N, m) |\mathbf{T}|^N}, \quad (1.36)$$

avec $\text{Tr}(\cdot)$ l'opérateur trace, m la dimension du vecteur \mathbf{k} et $K(\cdot; \cdot)$ la fonction de Bessel modifiée du second ordre :

$$K(N, m) = \pi^{\frac{1}{2}m(m-1)} \prod_{i=0}^{m-1} \Gamma(n-i), \quad (1.37)$$

avec $\Gamma(\cdot)$ la fonction Gamma d'Euler [1].

Comme nous nous plaçons dans le cas d'une configuration monostatique, le vecteur \mathbf{k} est constitué de trois composantes (HH , VV et HV) et on a donc $m = 3$.

On obtient directement la distribution de $\hat{\mathbf{T}}_{SCM}$ grâce à l'Eq. (1.36) :

$$p(\hat{\mathbf{T}}_{SCM}) = \frac{N^{m(N-m)} |\hat{\mathbf{T}}_{SCM}|^{N-m} \exp(-N\text{Tr}(\mathbf{T}^{-1}\hat{\mathbf{T}}_{SCM}))}{K(N, m) |\mathbf{T}|^N}. \quad (1.38)$$

1.2.3 Décompositions polarimétriques

Les décompositions polarimétriques sont des outils servant à décrire et interpréter les données polarimétriques. Pour les cibles distribuées, il est intéressant d'envisager le fait qu'il y ait un mécanisme polarimétrique

dominant ou moyen dans chaque cellule de résolution. Il est alors possible de caractériser ce mécanisme de plusieurs manières :

- en utilisant une analyse de la matrice de Sinclair \mathbf{S} ,
- en utilisant un modèle de décomposition de la matrice de covariance \mathbf{T} ,
- en exploitant la décomposition en valeurs propres et vecteurs propres de la matrice de covariance \mathbf{T} .

Comme il existe de nombreuses décompositions de chaque type, on ne s'intéressera ici qu'à deux décompositions, celles de Pauli [53] et de Cloude-Pottier [24], qui sont les plus répandues et les plus simples à mettre en œuvre.

1.2.3.1 Analyse de la matrice de Sinclair : décomposition de Pauli

La base de Pauli a déjà été présentée dans la Section 1.1.3.5. Il s'agit de représenter la matrice de Sinclair \mathbf{S} par la combinaison linéaire des matrices de Pauli qui caractérisent chacune un mécanisme physique de polarisation différent que l'on cherche à identifier :

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} = \frac{1}{\sqrt{2}} \left(a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix} \right), \quad (1.39)$$

avec

$$a = \frac{S_{HH} + S_{VV}}{\sqrt{2}}, \quad b = \frac{S_{HH} - S_{VV}}{\sqrt{2}}, \quad c = \frac{S_{HV} + S_{VH}}{\sqrt{2}} \quad \text{et} \quad d = j \frac{S_{HV} - S_{VH}}{\sqrt{2}}. \quad (1.40)$$

En configuration monostatique, le terme d est nul car $S_{HV} = S_{VH}$. D'après [25], les coefficients a , b et c représentent respectivement un mécanisme de :

- simple rebond : c'est-à-dire de diffusion sur une surface,
- double-rebond : typiquement de réflexion sur un bâtiment,
- diffusion de volume : souvent présent dans les zones de végétation.

Il est important de noter que lorsqu'un coefficient ne domine pas par rapport aux autres, l'interprétation devient très difficile.

La Fig. 1.6 présente le résultat d'une décomposition de Pauli sur une image de référence, acquise par le système AIRSAR de la NASA en 1988 au dessus de la baie de San Francisco en Californie. La représentation colorée est $S_{HH} + S_{VV}$ en bleu, $S_{HH} - S_{VV}$ en rouge et $\sqrt{2}S_{HV}$ en vert.

La surface de l'océan est représentée presque entièrement en bleu, ce qui correspond au terme $S_{HH} + S_{VV}$. Ce terme est caractéristique d'une diffusion sur une surface, ce qui est cohérent dans le cas de l'océan. De même, les zones de végétation ressortent en vert, ce qui correspond au terme $\sqrt{2}S_{HV}$ et donc bien à une diffusion de volume. Finalement, les bâtiments ressortent principalement en rouge, c'est-à-dire que le terme en $S_{HH} - S_{VV}$ est plus important, ce qui indique bien un mécanisme de type double rebond.

1.2.3.2 Décomposition de Cloude-Pottier

La décomposition de Pauli permet d'avoir un premier aperçu des différents mécanismes de rétrodiffusion dans une image mais n'en permet pas une investigation plus poussée. C'est pourquoi Cloude et Pottier ont introduit la décomposition suivante dans [24].



FIGURE 1.6 – Baie de San Francisco dans la base de Pauli : $S_{HH} - S_{VV}$, $\sqrt{2}S_{HV}$, $S_{HH} + S_{VV}$

1.2.3.2.1 Paramètres de la décomposition La décomposition de Cloude-Pottier se base sur une décomposition en valeurs propres et vecteurs propres de la matrice de covariance polarimétrique \mathbf{T} . L'estimateur $\hat{\mathbf{T}}_{SCM}$ est une matrice hermitienne définie positive, ses valeurs propres λ_i sont donc réelles et positives. Par ailleurs, d'après [24] ses vecteurs propres \mathbf{u}_i peuvent se mettre sous la forme :

$$\mathbf{u}_i = \begin{bmatrix} \cos \alpha_i \exp(j(\delta_i + \phi_i)) \\ \sin \alpha_i \cos \beta_i \exp(j(\delta_i + \phi_i)) \\ \sin \alpha_i \sin \beta_i \exp(j(\gamma_i + \phi_i)) \end{bmatrix}, \quad (1.41)$$

avec β_i un angle d'orientation préféré par le réflecteur engendrant le mécanisme de rétrodiffusion présent principalement dans la cellule étudiée et $i = 1, \dots, m$. α_i , γ_i , δ_i et ϕ_i sont des paramètres du modèle n'ayant pas de signification physique simple.

La décomposition de Cloude-Pottier est basée sur l'idée qu'au sein de chaque pixel, il existe un mécanisme polarimétrique associé à chacun des vecteurs propres et dont l'intensité dépend de la valeur propre associée.

1.2. Traitements classiques en classification SAR polarimétrique

On peut définir des pseudo-probabilités, notées P_i , qui représentent le poids de chaque valeur propre et qui sont définies de la manière suivante :

$$P_i = \frac{\lambda_i}{\sum_{k=1}^m \lambda_k} \quad (1.42)$$

Il est alors possible d'extraire trois paramètres des valeurs propres et des vecteurs propres :

- l'entropie polarimétrique H , qui représente le degré de chaos à l'intérieur de la cellule de résolution. Si H est faible, alors un mécanisme de rétrodiffusion dominant est présent et identifiable facilement dans la cellule. Au contraire, si l'entropie est élevée, il est difficile d'extraire un mécanisme principal. Elle est définie par :

$$H = - \sum_{k=1}^m P_k \log_m P_k. \quad (1.43)$$

- l'angle α , qui représente le type de mécanisme dominant. Cet angle est directement relié aux phénomènes physiques intervenant dans le processus de rétrodiffusion. Il est invariant par rotation autour de l'axe de visée du radar. L'angle α varie entre 0° où il représente une diffusion sur une surface et 90° où il représente la diffusion par un réflecteur appelé dièdre. D'autres mécanismes sont représentés par des valeurs intermédiaires de l'angle α : par exemple, des nuages de particules anisotropes, des mécanismes de double rebond, etc. Il est défini par :

$$\alpha = \sum_{k=1}^m P_k \alpha_k, \quad (1.44)$$

où les α_k ont été définis à l'Eq. (1.41).

- l'anisotropie A , qui représente l'importance relative des valeurs propres secondaires, donc des mécanismes de rétrodiffusion secondaires. C'est un paramètre complémentaire de l'entropie H lorsque celle-ci atteint des valeurs intermédiaires. L'apport de l'anisotropie est de mieux caractériser les mécanismes dans ce scénario. Elle est définie dans le cas monostatique par :

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \quad (1.45)$$

Les Fig. 1.7 et 1.8 représentent respectivement les paramètres α et H de l'image de la baie de San Francisco présentée précédemment, la couleur bleue représentant la valeur la plus faible et la couleur rouge la plus élevée.

Pour la Fig. 1.7, on retrouve l'océan en bleu, donc avec des valeurs de α proche de 0 , ce qui est cohérent avec l'interprétation physique de l'angle α qui indique une diffusion de surface pour des valeurs faibles. Pour la Fig. 1.8, l'entropie prend des valeurs très faibles sur la surface de l'océan, ce qui indique qu'un seul mécanisme est présent, le mécanisme de rétrodiffusion par une surface indiqué par l'angle α . Certains pixels de l'océan prennent des valeurs élevées, indiquant la présence de plusieurs mécanismes. Ces points représentent des bateaux qui entraînent l'apparition de mécanismes supplémentaires. Sur tout le reste de l'image, l'entropie prend des valeurs élevées, ce qui indique la présence de plusieurs mécanismes de polarisation différents au sein de chaque pixel. Ceci est dû à la résolution faible de l'image qui se caractérise par la présence de plusieurs objets différents au sein de chaque cellule de résolution qui vont réfléchir le signal de manière différente et créer des mécanismes de rétrodiffusion différents, entraînant une entropie élevée.

1.2.3.2.2 Plan H- α Une fois que les paramètres ont été calculés pour une matrice de covariance, il est possible de représenter plus finement le mécanisme sous-jacent grâce au couple H- α . Pour cela, on représente le plan des couples H- α réalisables comme sur la Fig. 1.9.

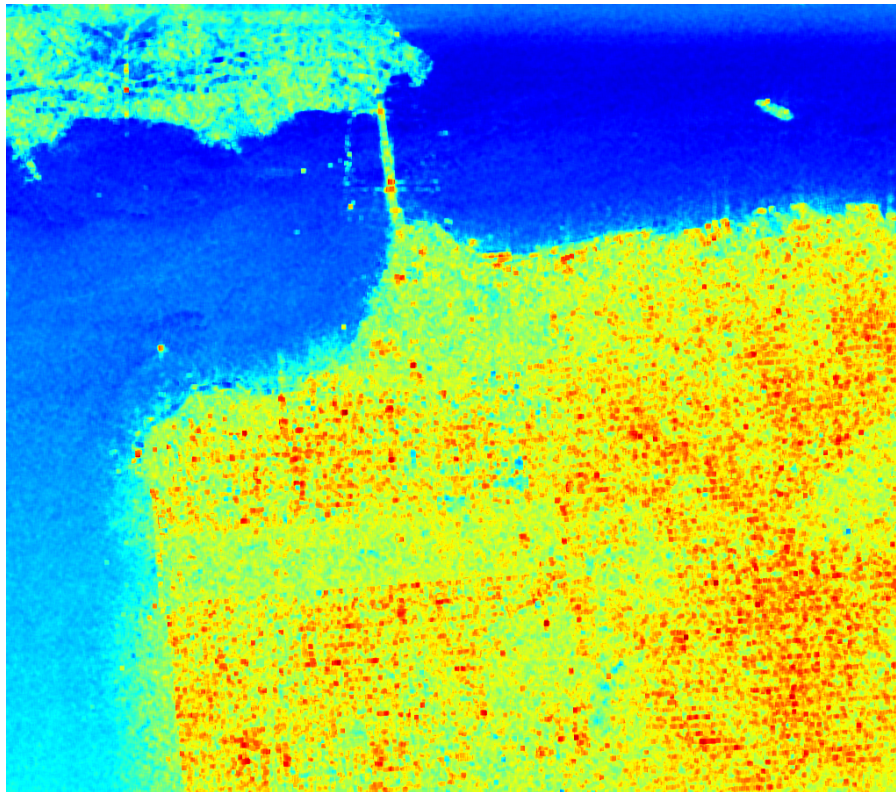


FIGURE 1.7 – Baie de San Francisco, angle α

La courbe rouge délimite deux parties du plan : à l'intérieur de la surface délimitée par la courbe rouge, l'ensemble des couples physiquement réalisables. Les courbes pointillées bleues délimitent les zones correspondant à différents mécanismes de polarisation. Ces zones ont été définies de manière relativement arbitraire dans [24] et sont à adapter en fonction des données. Elles fournissent néanmoins une idée approximative des différents mécanismes de polarisation présents dans une image. Le lecteur est invité à se référer à [24] pour plus d'informations sur les différentes zones.

Il est alors possible de représenter la répartition des pixels dans les différentes classes données par la décomposition de Cloude-Pottier, comme présenté sur la Fig. 1.10(a), ainsi que leur répartition dans le plan $H-\alpha$, comme présenté sur la Fig. 1.10(b). Il est notamment possible d'interpréter la classe représentée en bleu, qui correspond à la surface de l'océan et possède à la fois une entropie et un angle α faible. Ceci représente un mécanisme de diffusion par une surface appelé diffusion de Bragg [24]. Ce type de mécanisme est typique de la réflexion d'ondes électromagnétiques sur l'océan [62], ce qui est en concordance avec la décomposition de Cloude-Pottier.

1.2.4 Classifieur Wishart

Dans [46], une distance adaptée aux données SAR polarimétriques *single-look* a été proposée pour réaliser une classification supervisée d'images SAR à partir des vecteurs de données \mathbf{k} . Elle a été étendue dans [51] au cas, beaucoup plus répandu, des données SAR *multi-look* en utilisant la matrice de covariance des données et en tirant parti de la distribution de cette matrice de covariance. Une image *multi-look* est réalisée en divisant l'antenne virtuelle en plusieurs segments. Sur chacun de ces segments, une image SAR est formée ; elles sont ensuite sommées pour réaliser l'image *multi-look*. Ceci a pour effet de réduire le bruit de speckle, naturellement

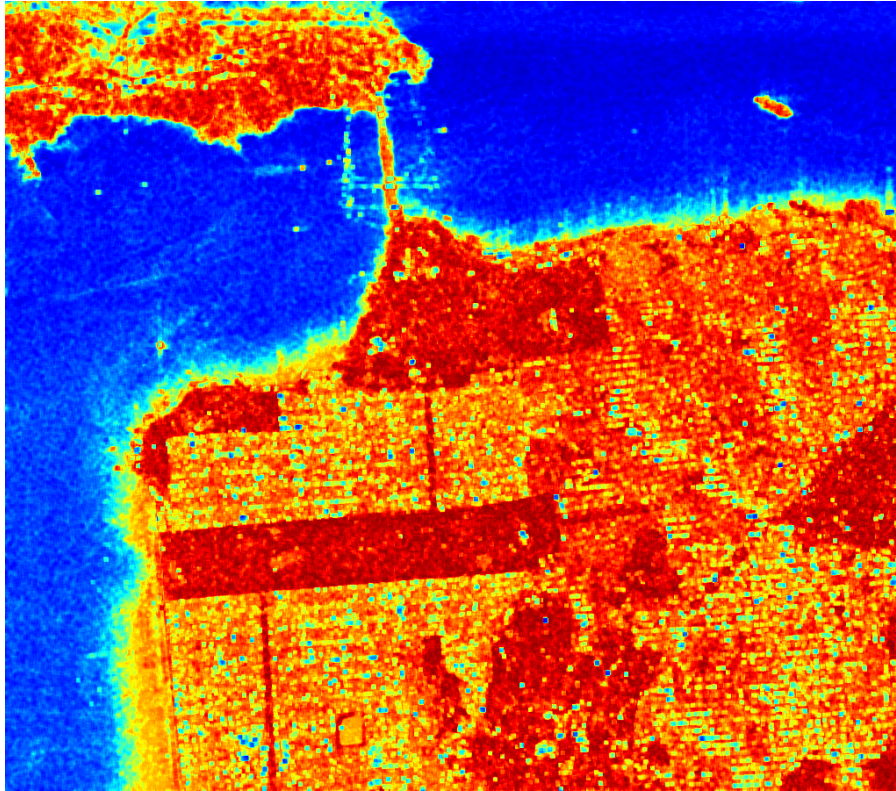


FIGURE 1.8 – Baie de San Francisco, entropie H

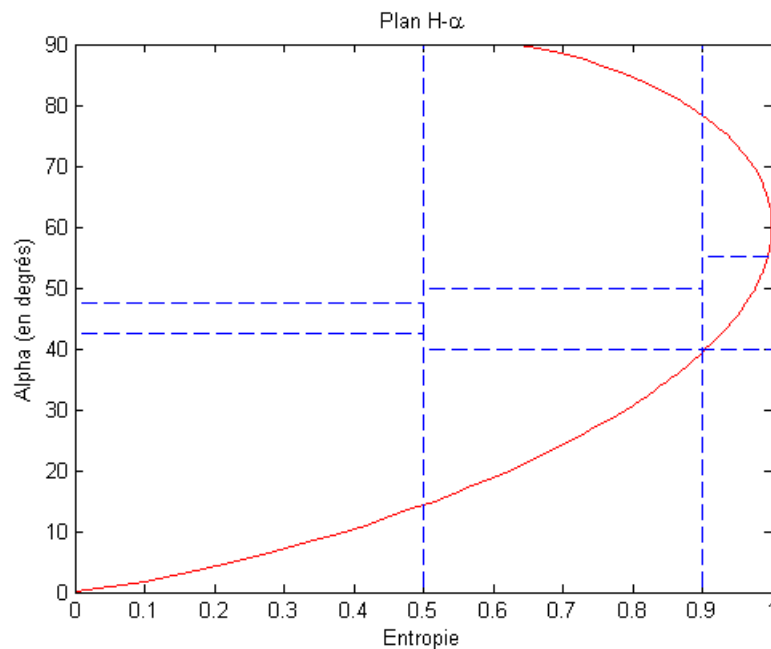


FIGURE 1.9 – Plan $H-\alpha$

présent dans les images SAR polarimétriques. Les données SAR polarimétriques étant distribuées selon une loi gaussienne complexe multivariée, leur matrice de covariance suit une distribution de Wishart, qui a été présentée à l'Eq. (1.36). Cette distance, appelée distance de Wishart, a ensuite été utilisée dans [50] au sein

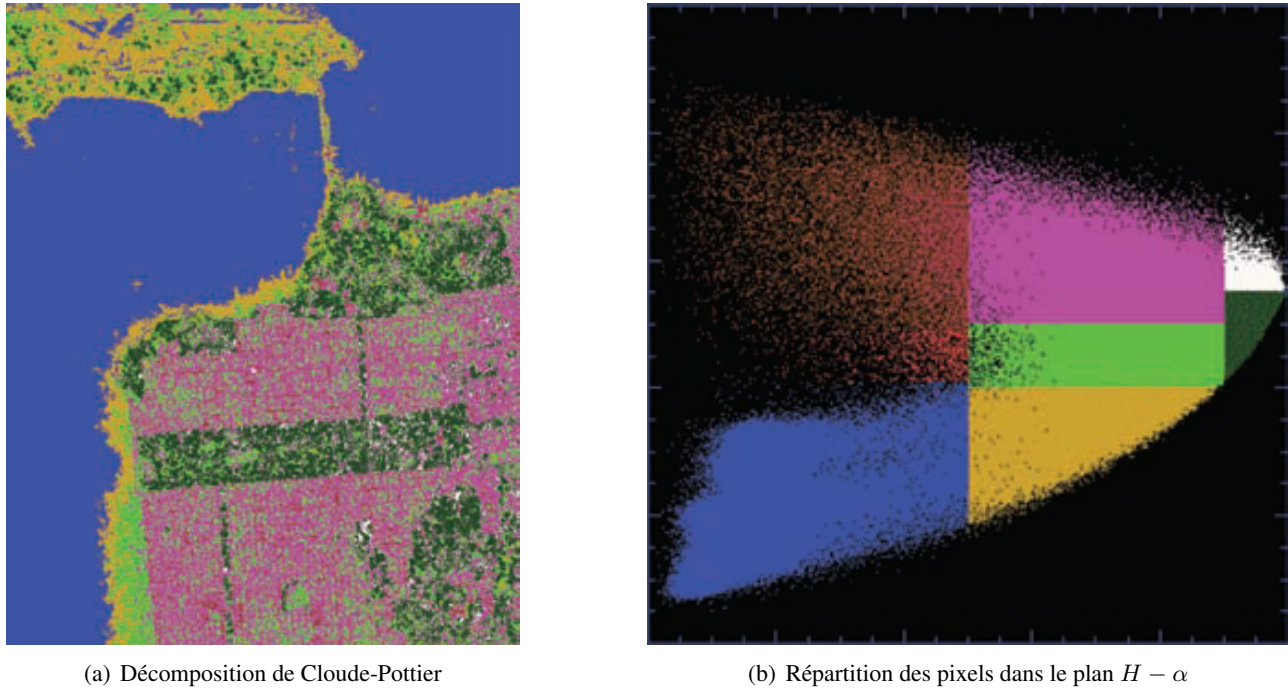


FIGURE 1.10 – Décomposition de Cloude-Pottier sur l'image AIRSAR de San Francisco

d'un algorithme de classification non supervisée, appelé classifieur de Wishart, qui est identique à celui proposé dans [51], excepté la phase d'initialisation qui est réalisée en utilisant la décomposition de Cloude-Pottier présentée précédemment. La distance et le classifieur de Wishart sont ici présentés.

1.2.4.1 Distance de Wishart

D'après l'Eq. (1.36), la matrice $\mathbf{A} = N\hat{\mathbf{T}}_{SCM}$ suit une distribution de Wishart à $2N$ degrés de liberté et de matrice de covariance \mathbf{T} dont la ddp est :

$$p(\mathbf{A}) = \frac{|\mathbf{A}|^{N-m} \exp(-\text{Tr}(\mathbf{T}^{-1}\mathbf{A}))}{K(N, m) |\mathbf{T}|^N}. \quad (1.46)$$

Dans [51], Lee *et al.* ont proposé une distance, appelée *distance de Wishart*, basée sur cette distribution en utilisant une approche bayésienne. On rappelle ici la manière d'obtenir cette distance. Une matrice \mathbf{A} est assignée dans la classe ω_i si elle vérifie la propriété suivante :

$$P(\omega_i|\mathbf{A}) > P(\omega_j|\mathbf{A}) \quad \forall j \neq i. \quad (1.47)$$

La règle de Bayes donne :

$$P(\omega_i|\mathbf{A}) = \frac{p(\mathbf{A}|\omega_i) P(\omega_i)}{p(\mathbf{A})}. \quad (1.48)$$

où $P(\omega_i)$ la probabilité *a priori* de la classe ω_i .

La quantité $p(\mathbf{A})$ est indépendante du choix de la classe, on peut donc l'ignorer. \mathbf{A} est donc assignée à ω_i si :

$$p(\mathbf{A}|\omega_i) P(\omega_i) > p(\mathbf{A}|\omega_j) P(\omega_j) \quad \forall j \neq i, \quad (1.49)$$

1.2. Traitements classiques en classification SAR polarimétrique

avec $p(\mathbf{A}|\omega_i)$ la densité de probabilité suivante :

$$p(\mathbf{A}|\omega_i) = \frac{|\mathbf{A}|^{N-m} \exp(-\text{Tr}(\mathbf{T}_i^{-1}\mathbf{A}))}{K(N, m) |\mathbf{T}_i|^N}, \quad (1.50)$$

avec $\mathbf{T}_i = \text{E}[\mathbf{k}\mathbf{k}^H|\omega_i]$.

En prenant le logarithme de l'Eq. (1.46), on obtient la distance entre la matrice \mathbf{A} et le centre \mathbf{T}_i de la classe ω_i :

$$d(\mathbf{A}, \omega_i) = N \ln |\mathbf{T}_i| + N \text{Tr}(\mathbf{T}_i^{-1}\mathbf{A}) - \ln(P(\omega_i)) - (N - m) \ln |\mathbf{A}| + \ln(K(N, m)). \quad (1.51)$$

Les deux derniers termes ne dépendent pas de ω_i , ils peuvent donc être ignorés. De même, on suppose que la probabilité de chaque classe $P(\omega_i)$ est *a priori* équivalente, on peut donc également ignorer ce terme. La distance de Wishart est alors donnée par l'équation suivante :

$$d_W(\mathbf{A}, \omega_i) = \ln |\mathbf{T}_i| + \text{Tr}(\mathbf{T}_i^{-1}\mathbf{A}). \quad (1.52)$$

À noter que l'emploi du terme *distance* est un abus de langage car la fonction obtenue dans l'Eq. (1.52) ne possède pas toutes les caractéristiques d'une distance au sens propre. Notamment, elle n'est pas symétrique et ne vérifie pas l'inégalité triangulaire. Le terme de mesure serait plus approprié ; elle est d'ailleurs appelée *Wishart distance measure* dans la littérature.

On montrera dans le chapitre 2 que l'on peut retrouver l'expression de cette distance par un test statistique d'égalité de matrices de covariance et dont l'Eq. (1.52) n'est qu'un cas particulier.

1.2.4.2 Algorithme du classifieur

L'algorithme du classifieur Wishart est un simple algorithme de type K-moyennes. On peut le décomposer en trois étapes :

1. Initialisation : les pixels sont initialement répartis dans un nombre de classes fixé. Traditionnellement, la décomposition de Cloude-Pottier de la Section 1.2.3.2 est utilisée pour répartir les pixels en huit classes correspondant aux huit zones du plan H- α (délimitées par les pointillés bleus dans la Fig. 1.9).
2. Calcul des centres de classes : on calcule la matrice de centre de classe \mathbf{T}_i en prenant la moyenne des matrices de covariances de chacun des pixels appartenant à la classe i :

$$\mathbf{T}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \hat{\mathbf{T}}_{SCM}^k \quad \text{avec } \hat{\mathbf{T}}_{SCM}^k \in \omega_i. \quad (1.53)$$

3. Réassignation des pixels : les pixels sont ensuite de nouveau répartis dans les différentes classes grâce à la distance de Wishart de l'Eq. (1.52).
4. L'algorithme se poursuit jusqu'à un critère d'arrêt : nombre d'itérations, nombre de pixels changeant de classe dans une itération, etc.

Son fonctionnement est décrit sous forme de diagramme dans la Fig. 1.11.

1.2.4.3 Application du classifieur sur des données réelles

Intéressons nous maintenant à un jeu de données qui sera fortement utilisé dans la suite du manuscrit, acquis en bande X dans la région de Brétigny par le système RAMSES de l'ONERA avec une résolution de l'ordre de

1.3m à la fois en distance et en azimut. Il est représenté dans la base de Pauli sur la Fig. 1.12, conjointement avec l’image optique de la même zone acquise plusieurs années après, ce qui explique la présence de bâtiments supplémentaires. Ce jeu de données présente une grande diversité d’un point de vue polarimétrique, à cause de la présence de nombreux champs, de plusieurs bâtiments, d’un parking et de zones arborées.

L’application du classifieur Wishart à ce jeu de données entraîne les résultats de classification présentés sur la Fig. 1.13(a). On observe principalement une bonne classification des champs mais sur les bâtiments, la classification n’est pas très bonne puisqu’ils sont représentés individuellement par plusieurs classes. De même sur la zone urbaine du coin supérieur gauche et sur le bâtiment 3, on remarque que les points les plus brillants ne font pas partie de la même classe (ou des mêmes classes) que le reste du bâtiment. Ce phénomène laisse imaginer que la classification Wishart présente des limitations sur les zones non-naturelles de l’image.

Pour mettre ceci en évidence, une classification de type K-moyennes est appliquée sur ce même jeu de données, mais cette fois-ci sur l’image d’intensité (ou *span*), donnée par l’Eq. (1.31). On suppose ici que l’intensité est modélisée par une loi appelée K-distribution. En effet, dans le cas où les données suivent une distribution gaussienne, leur amplitude (la racine carrée de l’intensité) est distribuée selon une loi de Rayleigh [73], qui modélise de manière satisfaisante la distribution de données acquises pour des résolutions faibles. Cependant, lorsque la résolution diminue — et devient donc meilleure — le modèle de Rayleigh ne permet souvent pas de modéliser correctement la distribution de certaines de ces données. On fait alors appel à d’autres distributions comme la K-distribution [44] ou la loi de Weibull [78]. La K-distribution présente un intérêt particulier pour le cas du radar SAR polarimétrique puisqu’elle est dérivée d’un processus de diffusion physique [44], qu’elle se comporte comme une distribution de Rayleigh dans le cas de données homogènes et qu’elle

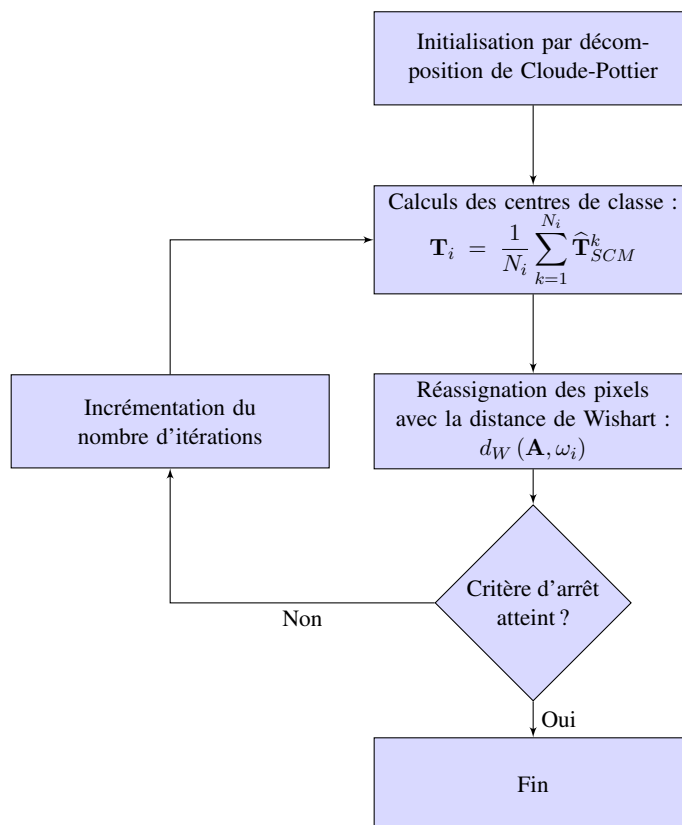


FIGURE 1.11 – Algorithme du classifieur Wishart



(a) Brétigny dans la base de Pauli : $S_{HH} - S_{VV}$, S_{HV} , $S_{HH} + S_{VV}$

(b) Image Google Earth

FIGURE 1.12 – Jeu de données : Brétigny

a été appliquée avec succès pour modéliser certains types de signaux radar [83]. La K-distribution peut être caractérisée par le produit d’une amplitude distribuée selon une loi de Rayleigh avec une variable aléatoire suivant une loi Gamma [72]. La loi Gamma est caractérisée par deux paramètres, un paramètre de forme, noté ν , et un paramètre d’échelle, noté θ . La ddp d’une variable aléatoire, notée τ , suivant une loi Gamma de paramètres $(\nu, 2/b^2)$ est donnée par l’équation suivante :

$$p(\tau) = \left(\frac{b^2}{2}\right)^\nu \frac{\tau^{\nu-1}}{\Gamma(\nu)} \exp\left(-\frac{b^2\tau}{2}\right), \quad (1.54)$$

avec Γ la fonction Gamma d’Euler [1].

Le résultat de la classification par un algorithme K-moyennes appliqué sur l’image d’intensité est présenté sur la Fig. 1.13(b). En comparant la Fig. 1.13(a) à la Fig. 1.13(b), il apparaît que les résultats de classification selon les deux approches sont très similaires. Cependant, l’approche de la Fig. 1.13(b) ne fait appel qu’à l’intensité de chaque pixel, c’est-à-dire la trace de la matrice de covariance, et ne prend absolument pas en compte les éléments non-diagonaux de cette matrice de covariance qui contiennent une partie de l’information polarimétrique puisqu’ils représentent les relations entre les différents canaux de polarisation. L’approche de la Fig. 1.13(a) utilise, elle, la matrice de covariance dans son ensemble. Il convient alors de s’interroger sur la manière dont cette information polarimétrique est extraite si une classification utilisant toute la matrice de covariance polarimétrique des pixels avec une focalisation sur les mécanismes de rétrodiffusion et leur interprétation donne des résultats extrêmement similaires à une classification uniquement sur la puissance des pixels.

Plus précisément, il est important de considérer l’hypothèse de la distribution gaussienne du signal. Comme il a été indiqué précédemment, certains scénarios demandent l’utilisation de distributions plus complexes dans le cas où la distribution gaussienne ne permet pas de modéliser correctement le fouillis. Il est possible qu’une mauvaise adéquation du modèle et de l’estimateur de la matrice de covariance, la SCM, soit à l’origine de cette “pollution” de la classification polarimétrique par l’information de puissance. En effet, les systèmes d’acquisition d’images SAR récents sont capables d’obtenir des images avec des résolutions de l’ordre du mètre, voire

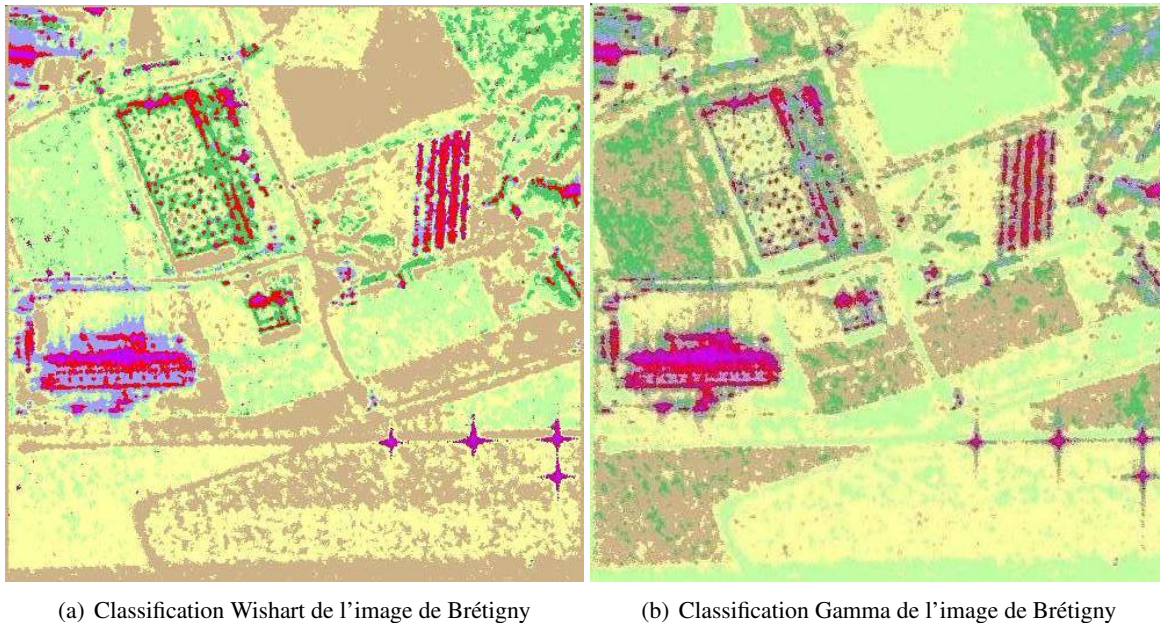


FIGURE 1.13 – Comparaison de la classification Wishart avec une classification sur la puissance

de la dizaine de centimètres. À ce niveau de résolution, il apparaît que les méthodes de classification basées sur la matrice de covariance apportent peu d'informations supplémentaires par rapport à des classifications basées sur l'intensité de chacun des pixels, cf. Fig. 1.13.

En effet, lorsqu'on considère des images SAR haute résolution, une des hypothèses fondamentales menant à la représentation gaussienne du vecteur cible \mathbf{k} , à savoir l'hypothèse qu'un grand nombre de réflecteurs est présent dans chaque cellule de résolution n'est plus vérifiée [66]. En réalité, le nombre de réflecteurs peut grandement varier d'une cellule de résolution à l'autre, ce qui entraîne une hétérogénéité du signal, c'est-à-dire que le signal est localement Gaussien avec une puissance aléatoire. Il est alors nécessaire de prendre en compte ces variations de puissance locales afin d'avoir une meilleure modélisation du signal et ainsi d'améliorer les performances des classifieurs. Pour cela, nous allons considérer un modèle de signal non gaussien appelé le modèle SIRV.

1.3 Modèle de vecteurs aléatoires sphériquement invariants

Dans cette section, le modèle non gaussien de description du signal SAR polarimétrique qui sera employé tout au long du manuscrit est introduit. Dans un premier temps, le modèle en lui-même est présenté puis un estimateur de matrice de covariance appliqué à ce modèle, l'estimateur du Point Fixe, est défini. Cet estimateur est obtenu par un algorithme itératif dont on rappellera les propriétés. Enfin, les propriétés statistiques importantes de l'estimateur du Point Fixe seront présentées.

1.3.1 Définition

Les vecteurs aléatoires sphériquement invariants, ou Spherically Invariant Random Vectors (SIRV), ont été introduits par Yao en 1973 [81] dans la communauté du traitement de l'information. Ils ont été appliqués dans la communauté radar à plusieurs reprises, notamment pour la modélisation du fouillis de mer [2, 27, 35, 83]. Les SIRV permettent de décrire des processus aléatoires non gaussiens, comme la K-distribution, les lois de

1.3. Modèle de vecteurs aléatoires sphériquement invariants

Weibull ou de KummerU, etc. L'idée de ce modèle est de pouvoir améliorer la représentation du signal, supposé localement gaussien mais sujet à des variations spatiales de puissance qui le rendent hétérogène.

Un SIRV est donc un processus gaussien non homogène de puissance aléatoire. Au lieu de représenter le vecteur cible \mathbf{k} comme un vecteur aléatoire gaussien complexe, on le représente par le produit d'un vecteur aléatoire gaussien complexe, noté \mathbf{x} et appelé *speckle* de moyenne nulle et de matrice de covariance $\mathbf{M} = E[\mathbf{x}\mathbf{x}^H]$ et d'une variable aléatoire positive, notée τ et appelée *texture* :

$$\mathbf{k} = \sqrt{\tau}\mathbf{x}. \quad (1.55)$$

La ddp de \mathbf{k} est donnée par l'équation suivante :

$$p_{\mathbf{k}}(\mathbf{k}) = \int_0^{+\infty} \frac{1}{(\pi\tau)^m |\mathbf{M}|} \exp\left(-\frac{\mathbf{k}^H \mathbf{M}^{-1} \mathbf{k}}{\tau}\right) p_{\tau}(\tau) d\tau, \quad (1.56)$$

où $p_{\tau}(\cdot)$ est la ddp inconnue de la variable τ .

C'est le choix de la ddp de la texture τ qui va permettre de décrire un grand nombre de processus stochastiques différents. Par exemple, si τ suit une loi Gamma, le vecteur \mathbf{k} suivra une K-distribution, présentée précédemment. Si τ suit une loi de Fisher, \mathbf{k} suivra une loi de KummerU [14]. Le cas d'une distribution gaussienne pour \mathbf{k} est obtenu en choisissant un Dirac comme loi de texture. Dans le cadre de données SAR polarimétriques, la texture représente les variations locales de puissance, d'une cellule de résolution à l'autre. C'est donc cette variable qui modélise l'hétérogénéité spatiale du signal.

Le terme \mathbf{x} est appelé *speckle* et contient l'information polarimétrique effective, dénuée de la puissance contenue dans la texture, au travers de sa matrice de covariance \mathbf{M} .

La texture modélise donc l'hétérogénéité du signal, c'est-à-dire les variations locales de puissance d'une cellule de résolution à l'autre. La distribution SIRV est donc gaussienne, conditionnellement à une cellule de résolution.

1.3.2 L'Estimateur du Point Fixe

Dans le cas d'un signal modélisé par un SIRV, la SCM de l'Eq. (1.35) n'est plus adaptée car ce n'est plus l'estimateur optimal au sens du maximum de vraisemblance. Les travaux de Gini et Greco [36, 39] ont permis d'établir un estimateur du maximum de vraisemblance exact de la matrice de covariance d'un SIRV, toujours en utilisant un N -échantillon de données secondaires $(\mathbf{k}_1, \dots, \mathbf{k}_N)$ — comme dans le cas de l'estimation par la SCM pour le cas gaussien — lorsque la texture τ est considérée comme déterministe et inconnue. Dans le cas plus général où τ est considérée comme aléatoire, ce même estimateur est un estimateur du maximum de vraisemblance approché. Contrairement à la SCM où une expression analytique en fonction des données secondaires \mathbf{k}_i existe, cet estimateur est la solution d'une équation implicite. Plus particulièrement, il est donné par la solution de l'équation de point fixe suivante :

$$\hat{\mathbf{T}} = \frac{m}{N} \sum_{i=1}^N \frac{\mathbf{k}_i \mathbf{k}_i^H}{\mathbf{k}_i^H \hat{\mathbf{T}}^{-1} \mathbf{k}_i}. \quad (1.57)$$

Cet estimateur est donc appelé estimateur du point fixe, ou *Fixed Point Estimate* (FPE). Les travaux de Pascal [60] ont permis d'établir que la solution de cette équation existe et est unique à un facteur près. En effet, si $\hat{\mathbf{T}}_{FPE}$ est une solution de l'Eq. (1.57), alors pour tout $a \in \mathbb{C}$, $a\hat{\mathbf{T}}_{FPE}$ est aussi solution de l'Eq. (1.57). Il

est donc aisé de voir qu'il y a unicité de la structure de la matrice de covariance, indépendamment du facteur a . Pour assurer une unicité complète, la matrice de covariance est normalisée selon la \mathbf{M} -normalisation [28], c'est-à-dire de telle manière que $\text{Tr}(\widehat{\mathbf{T}}) = m$.

De manière pratique, il n'existe pas de solution analytiquement calculable de l'Eq. (1.57) mais l'algorithme itératif convergent suivant permet d'obtenir une solution :

$$\begin{cases} \widehat{\mathbf{T}}(0) = \mathbf{M}_0, \\ \widehat{\mathbf{T}}(k+1) = \frac{m}{N} \sum_{i=1}^N \frac{\mathbf{k}_i \mathbf{k}_i^H}{\mathbf{k}_i^H (\widehat{\mathbf{T}}(k))^{-1} \mathbf{k}_i}. \end{cases} \quad (1.58)$$

avec $\widehat{\mathbf{T}}(k)$ la valeur de l'estimée à l'itération k et \mathbf{M}_0 une matrice d'initialisation. D'après [60], cet algorithme converge quelque soit le point de départ vers une solution unique à un facteur d'échelle près. La matrice identité \mathbf{I} est généralement utilisée comme initialisation et la \mathbf{M} -normalisation employée pour garantir une solution unique.

Il est intéressant de noter que les données secondaires \mathbf{k}_i sont aussi modélisées par des SIRV. On peut alors les remplacer dans l'Eq. (1.57) par leur expression donnée par l'Eq. (1.55). On a alors :

$$\widehat{\mathbf{T}}_{FPE} = \frac{m}{N} \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \widehat{\mathbf{T}}_{FPE}^{-1} \mathbf{x}_i}, \quad (1.59)$$

et le FPE ne dépend que des variables de speckle \mathbf{x}_i et pas des textures τ_i .

1.3.3 Propriétés statistiques du FPE

L'algorithme itératif présenté précédemment fournit donc un estimateur de la matrice de covariance. Les travaux de Pascal [61] ont permis de démontrer trois propriétés statistiques importantes de cet estimateur qui justifient son utilisation dans un environnement SAR non gaussien.

Premièrement, le FPE est un estimateur sans biais de la matrice de covariance des \mathbf{x}_i , c'est-à-dire :

$$\mathbb{E} \left[\widehat{\mathbf{T}}_{FPE} \right] - \mathbf{M} = \mathbf{0}. \quad (1.60)$$

Son espérance statistique est donc égale à la "vraie" matrice de covariance des \mathbf{x}_i .

De plus, c'est un estimateur consistant de \mathbf{M} , ce qui s'exprime suivant l'équation suivante :

$$\widehat{\mathbf{T}}_{FPE} \xrightarrow[N \rightarrow +\infty]{Pr} \mathbf{M}, \quad (1.61)$$

avec $\xrightarrow[N \rightarrow +\infty]{Pr}$ la convergence en probabilité. Ce résultat implique que, plus on utilise de données secondaires pour calculer le FPE, plus on s'approche de la véritable valeur de \mathbf{M} , la matrice de covariance des \mathbf{x}_i .

Enfin, on dispose du résultat suivant sur la distribution asymptotique du FPE. On notera ici en exposant le nombre de données secondaires utilisées pour réaliser l'estimation de la matrice de covariance. Pour un

1.3. Modèle de vecteurs aléatoires sphériquement invariants

N -échantillon de données secondaires \mathbf{k}_i , on a :

$$\sqrt{N} \left(\text{vec} \left(\widehat{\mathbf{T}}_{SCM}^N \right) - \text{vec} (\mathbf{M}) \right) \xrightarrow[N \rightarrow +\infty]{} \mathcal{CN} (\mathbf{0}, \boldsymbol{\Sigma}), \quad (1.62)$$

$$\sqrt{N} \left(\text{vec} \left(\widehat{\mathbf{T}}_{FPE}^N \right) - \text{vec} (\mathbf{M}) \right) \xrightarrow[N \rightarrow +\infty]{} \mathcal{CN} \left(\mathbf{0}, \frac{m+1}{m} \boldsymbol{\Sigma} \right), \quad (1.63)$$

$$\Leftrightarrow \sqrt{N \frac{m}{m+1}} \left(\text{vec} \left(\widehat{\mathbf{T}}_{FPE}^N \right) - \text{vec} (\mathbf{M}) \right) \xrightarrow[N \rightarrow +\infty]{} \mathcal{CN} (\mathbf{0}, \boldsymbol{\Sigma}), \quad (1.64)$$

avec $\text{vec} (\mathbf{M})$ l'opérateur transformant la matrice \mathbf{M} de taille $n \times p$ en un vecteur colonne de taille np et $\boldsymbol{\Sigma}$ une matrice dépendant des données [61].

Il apparaît alors que le FPE estimé avec N données secondaires converge vers la même loi que la SCM estimée avec le même nombre de données mais avec un terme sous la racine différent. En notant ce terme $N_1 = N \frac{m}{m+1}$, on peut réécrire l'Eq. (1.64) de la manière suivante :

$$\sqrt{N_1} \left(\text{vec} \left(\widehat{\mathbf{T}}_{FPE}^{\frac{m+1}{m} N_1} \right) - \text{vec} (\mathbf{M}) \right) \xrightarrow[N \rightarrow +\infty]{} \mathcal{CN} (\mathbf{0}, \boldsymbol{\Sigma}) \quad (1.65)$$

On peut en conclure, d'après les Eqs. (1.62) et (1.65), que le FPE se comporte de manière asymptotique, c'est-à-dire quand N est suffisamment grand, comme une matrice de Wishart (puisque c'est la distribution que suit la SCM) avec un nombre de degrés de liberté légèrement plus élevé puisque multiplié par un facteur $\frac{m+1}{m}$. Ce résultat est extrêmement important puisqu'il permet d'étendre au FPE des résultats nécessitant une hypothèse gaussienne sur la distribution des données, que nous verrons au chapitre 2. Il faut toutefois noter qu'il s'agit d'un résultat asymptotique, ce qui implique que l'on ne dispose pas forcément d'information sur la distribution du FPE pour un nombre fini d'échantillons et que l'approximation par une distribution de Wishart n'est effectivement qu'une approximation. Cependant, en situation réelle, l'expérience montre que si N n'est pas "trop petit", l'approximation est vérifiée.

1.3.4 Estimation de la texture

Un estimateur de la texture est parfois nécessaire dans certaines applications. Il est donné par l'expression suivante [75] :

$$\widehat{\tau}_i = \frac{\mathbf{x}_i^H \mathbf{M}^{-1} \mathbf{x}_i}{m}. \quad (1.66)$$

A partir de cette expression, Vasile dérive dans [75] l'expression de l'estimateur du *span* d'un pixel, donnée par l'équation suivante :

$$\widehat{\mathbf{s}}_{FP} = \mathbf{k}_i^H \widehat{\mathbf{T}}_{FPE}^{-1} \mathbf{k}_i, \quad (1.67)$$

Cette expression est identique à celle du Polarimetric Whitening Filter (PWF) développé par Novak dans [58] afin de réduire l'effet de chatoiement observable sur les images SAR dû au bruit de *speckle* et dont l'expression est rappelée ici :

$$\widehat{\mathbf{s}}_{SCM} = \mathbf{k}_i^H \widehat{\mathbf{T}}_{SCM}^{-1} \mathbf{k}_i, \quad (1.68)$$

avec pour seule différence, l'utilisation du FPE au lieu de la SCM.

On peut aussi rapprocher cette expression de la distance de Mahalanobis [54] dont on parlera plus en détail au chapitre quatre.

1.3.5 Influence de l'estimateur de la matrice de covariance sur les résultats de classification

1.3.5.1 Décomposition de Cloude-Pottier

On s'intéresse ici à l'influence de l'estimateur de la matrice de covariance sur la décomposition de Cloude-Pottier. Comme cette décomposition est basée sur l'importance relative des valeurs propres de la matrice de covariance, il est raisonnable de penser que les différences entre la SCM et le FPE devraient être minimales. En effet, si le FPE ne prend en compte que l'information polarimétrique et la SCM l'information polarimétrique multipliée par un certain coefficient, les valeurs propres devraient être sensiblement identiques à un facteur près, qui disparaît dans le calcul de l'entropie. Si l'on s'intéresse aux valeurs propres calculées pour tous les pixels de la Fig. 1.12(a) présentées sur les Fig. 1.14(a) et 1.14(b), il apparaît que les courbes de répartition sont très différentes. En effet, la M-normalisation utilisée dans le calcul du FPE met une contrainte sur la valeur des valeurs propres.

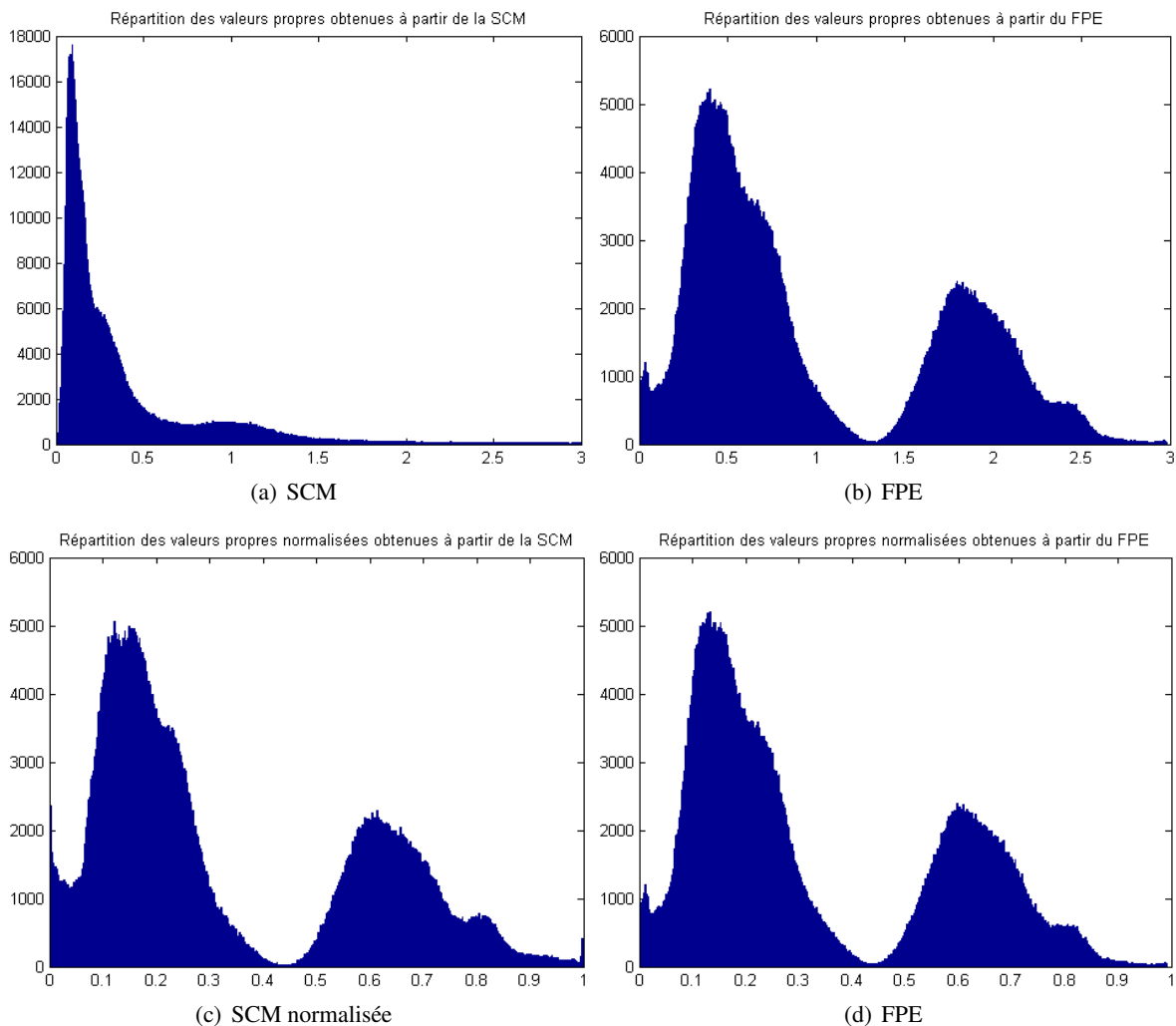


FIGURE 1.14 – Répartition des valeurs propres suivant l'estimateur de la matrice de covariance

Après application de la M-normalisation sur la SCM, on obtient effectivement des répartitions extrêmement similaires entre SCM et FPE, ce qui rejoint l'interprétation polarimétrique du FPE par rapport à la SCM, à savoir que les deux matrices contiennent la même information polarimétrique mais que la SCM est "polluée" par l'information de puissance. Ces résultats sont présentés sur les Fig. 1.14(c) et 1.14(d). La Fig. 1.14(d) est

bien évidemment identique à la Fig. 1.14(b) puisque la M-normalisation a déjà été appliqué dans le calcul du FPE.

Par comparaison avec l'Eq. (1.59), on peut écrire la SCM sous hypothèse SIRV :

$$\hat{\mathbf{T}}_{SCM} = \frac{1}{N} \sum_{i=1}^N \tau_i \mathbf{x}_i \mathbf{x}_i^H. \quad (1.69)$$

La variable de texture est toujours présente pour la SCM, ce qui explique les résultats présentés dans les Fig. 1.13(a) et 1.13(b), qui étaient toutes deux dominées par l'information de puissance. Le FPE ne dépendant pas de cette information, il est raisonnable de penser que les résultats de classification par le classifieur de Wishart seront différents de ceux obtenus avec l'intensité et auront une interprétation polarimétrique plus cohérente.

Les résultats de la décomposition de Cloude-Pottier sur l'image de Brétigny sont présentés sur la Fig. 1.15. On remarque en effet que les résultats sont quasiment identiques quel que soit le choix de l'estimateur de la matrice de covariance.

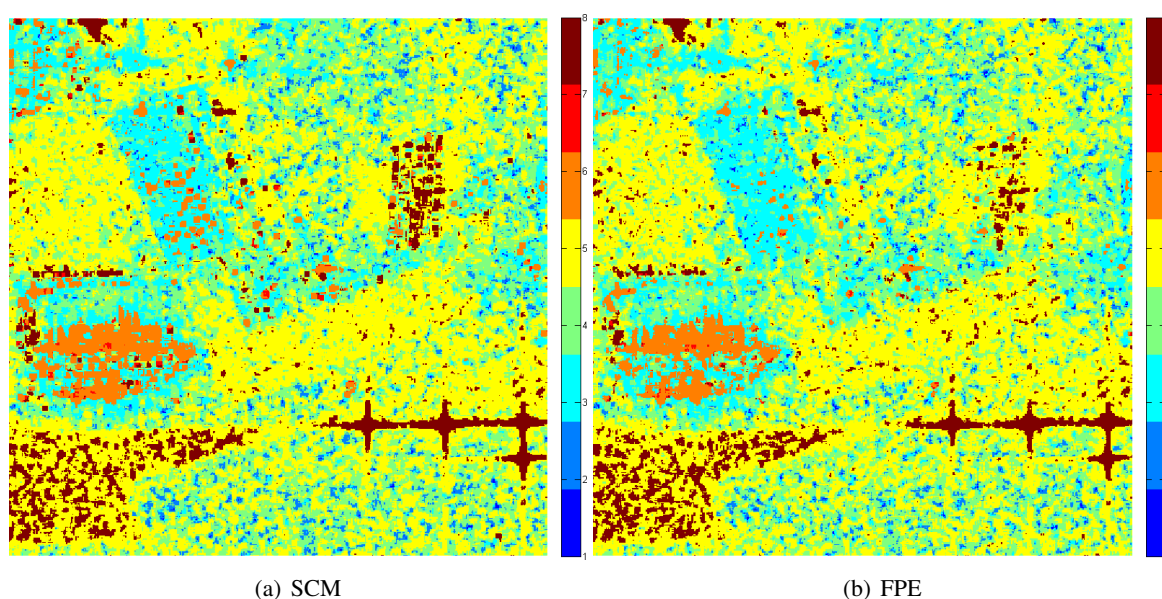


FIGURE 1.15 – Comparaison des résultats de la décomposition de Cloude-Pottier entre la SCM et le FPE

Ceci renforce une fois de plus l'idée que la même information polarimétrique est contenue dans la SCM et le FPE.

1.3.5.2 Classifieur Wishart

Si l'on remplace la SCM par le FPE dans le classifieur Wishart, on obtient des résultats de classification très différents. La Fig. 1.16 présente la différence entre les résultats du classifieur Wishart avec la SCM et avec le FPE à partir d'une initialisation commune. A noter qu'il n'y a aucun choix réfléchi sur l'attribution des couleurs aux classes.

Les différences entre les résultats issus des deux méthodes sont très nettement marquées. L'influence de la puissance dans la SCM est clairement visible avec les zones homogènes en puissance, comme les champs,

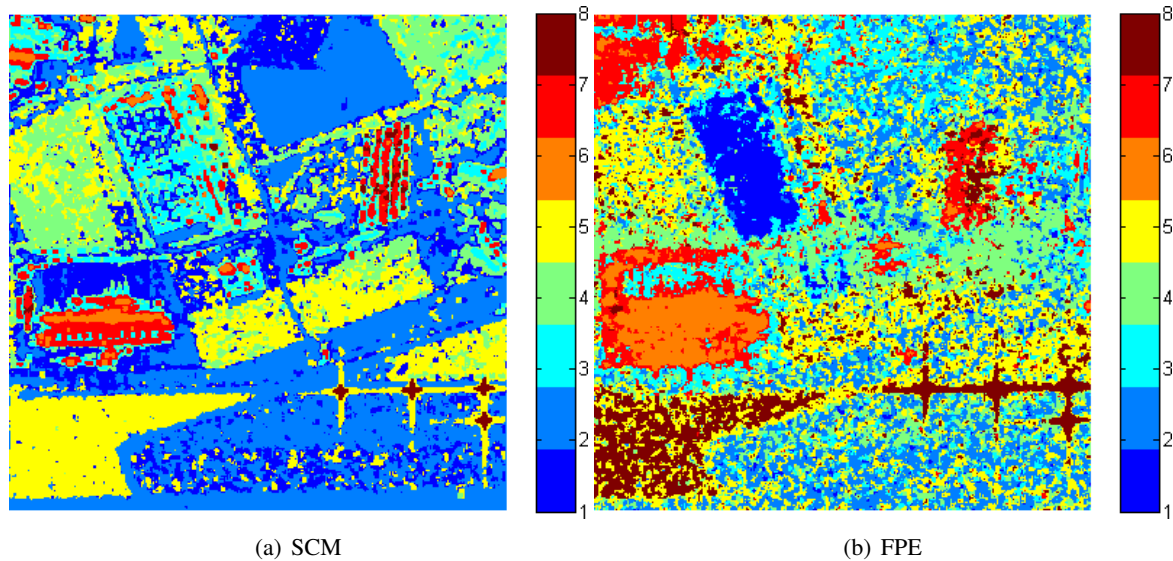


FIGURE 1.16 – Comparaison des résultats du classifieur Wishart entre la SCM et le FPE avec initialisation par la décomposition de Cloude-Pottier

qui ressortent avec des contours très nets. Dans le coin inférieur droit, les coins réflecteurs qui ont servi à la calibration du radar ont leur centre situé dans la classe marron, au même titre que certains pixels situés sur les différents bâtiments et qui correspondent à des points extrêmement brillants. Sur les zones hétérogènes en puissance, comme les bâtiments, on reconnaît très facilement les détails des structures que l’on peut observer sur la Fig. 1.12 et qui polluent la classification. En utilisant le FPE, au contraire, ces zones hétérogènes ont des contours beaucoup plus nets qui traduisent *a priori* une homogénéité en polarimétrie, non perturbée par la puissance. La classification sur les zones gaussiennes apparaît cependant moins performante visuellement qu’en utilisant la SCM ce qui peut s’expliquer pour plusieurs raisons.

Premièrement, la SCM est l’estimateur optimal dans le cas gaussien, contrairement au FPE, ce qui explique ses meilleures performances. Deuxièmement, les résultats de la SCM apparaissent meilleurs sur les zones gaussiennes mais on a pu observer sur les Fig. 1.13(a) et 1.13(b) que l’information de puissance était prédominante en utilisant la SCM. Cette information de puissance est d’ailleurs aisée à reconnaître, même sur les images satellitaires ou dans la base de Pauli (cf Fig. 1.12(a) et 1.12(b)) où l’on peut facilement reconnaître les contours des différentes zones. *A contrario*, l’information polarimétrique est une information “cachée” qui n’apparaît pas au premier abord. Si l’on observe la décomposition de Cloude-Pottier sur la Fig. 1.15, qui donne des informations sur les mécanismes polarimétriques de l’image, on remarque le même genre de “patches” qu’avec le classifieur Wishart-FPE, ce qui conforte l’idée que l’information apportée par ce classifieur est bien l’information polarimétrique. Enfin, la plupart des travaux menés sur la classification SAR polarimétrique dans la littérature est basée sur la SCM, donc plus sur la puissance que sur l’information polarimétrique sous-jacente. Tout ceci complique l’interprétation de ces résultats car il y a peu, voire pas, de travaux auxquels les comparer.

1.4 Conclusion

Ce chapitre a permis de rappeler les concepts de base qui servent de fondements pour ce travail de thèse. Le principe du radar à synthèse d’ouverture (SAR) est introduit, de même que la polarimétrie, qui sert à identifier les mécanismes physiques de rétrodiffusion. Les traitements classiques en classification SAR polarimétrique exploitent cette information polarimétrique principalement au travers de la matrice de covariance, estimée par

la méthode du maximum de vraisemblance en supposant une distribution gaussienne du signal. Cependant, cette hypothèse s'avère erronée lorsque la résolution du système radar augmente, c'est pourquoi un modèle de signal non gaussien, le modèle SIRV, est proposé. Ce modèle décompose le signal en un produit d'un vecteur gaussien, appelé speckle et qui modélise l'information polarimétrique, et d'une variable aléatoire, appelée texture et qui modélise les variations locales de puissance. L'estimateur du maximum de vraisemblance de la matrice de covariance sous hypothèse SIRV, appelé l'Estimateur du Point Fixe (FPE), possède des propriétés intéressantes, notamment d'être indépendant de la variable de texture. Les résultats de classification par un classifieur classique, le classifieur Wishart, mettent en évidence des différences importantes dans l'utilisation d'un estimateur de matrice de covariance sous hypothèse gaussienne ou sous hypothèse SIRV.

Classification non supervisée

Sommaire

2.1	Test de Box	33
2.1.1	Inconvénients du classifieur Wishart	33
2.1.2	Test d'égalité de matrices de covariance	34
2.1.3	M-Test de Box	45
2.2	Algorithmes et résultats	47
2.2.1	Algorithmes	47
2.2.2	Résultats	50
2.3	Implémentation du test de Box au sein d'autres algorithmes	54
2.3.1	Application avec une segmentation initiale	54
2.3.2	Application sans segmentation initiale	63
2.3.3	Application à la classification hiérarchique	65
2.4	Conclusion	70

Ce chapitre est consacré à l'introduction et l'étude d'un nouvel algorithme de classification polarimétrique, appelé classifieur de Box pour remplacer ou compléter le classifieur Wishart. En effet, il est intéressant de se pencher sur les limitations du classifieur Wishart. Nous verrons qu'il est possible de pallier certaines de ses limitations. Le classifieur de Box sera tout d'abord présenté au sens statistique du terme dans une première partie. Puis des algorithmes basés sur ce test seront proposés et appliqués sur des images réelles. Enfin, une interprétation polarimétrique de ces résultats sera donnée.

2.1 Test de Box

2.1.1 Inconvénients du classifieur Wishart

Le classifieur Wishart a été présenté dans la section 1.2.4. Pour rappel, il s'agit d'un algorithme de type K-moyennes classique utilisant une distance adaptée au cas des matrices de covariance polarimétriques. Cet algorithme permet d'obtenir des classifications de bonne qualité mais présente toutefois quelques inconvénients dont un particulièrement important.

Tout d'abord, le nombre de classes de l'algorithme est un paramètre choisi par l'utilisateur. Sans information *a priori* sur le jeu de données, il est très difficile de trouver un nombre de classes optimal. Une première étape d'initialisation est donc souvent nécessaire et parfois coûteuse en temps de calcul. La manière traditionnelle d'initialiser le classifieur Wishart est d'utiliser la décomposition de Cloude-Pottier, décrite à la section 1.2.3.2, et d'assigner à chaque pixel une classe initiale dépendant de cette décomposition, fixant ainsi le nombre

initial de classes à huit. Par ailleurs, l'utilisation de la distance de Wishart est basée sur une hypothèse de distribution gaussienne du signal. Or, comme nous l'avons constaté précédemment, l'hypothèse gaussienne n'est plus vérifiée dans les images SAR haute résolution. Il serait donc nécessaire d'étendre la distance utilisée sous hypothèse non-gaussienne ou en s'affranchissant entièrement de la distribution du signal. Finalement, l'inconvénient majeur du classifieur Wishart est qu'un pixel est forcément assigné, par construction, à une des classes préexistantes, même si ce pixel est très éloigné des différents centres de classe.

Des méthodes ont été proposées pour augmenter le nombre de classes et améliorer la représentation graphique des résultats [32] ou pour diviser et fusionner les classes existantes [52], afin d'avoir une plus grande diversité. Dans [43], Ince présente une revue des classifieurs proposés dans la littérature depuis la publication de l'algorithme du classifieur Wishart. Ces algorithmes présentent au moins un des trois inconvénients majeurs suivants : une incapacité à trouver un optimum global comme solution du problème de classification, une incapacité à trouver un nombre optimal de classes et une sensibilité à l'initialisation. L'approche employée au cours de ce doctorat a été principalement de prendre en compte le fait que certains pixels ne doivent pas être "rangés" dans des classes existantes à cause de leur trop grande différence par rapport aux pixels déjà classés. La structure de l'algorithme employé et qui sera présenté dans la suite de ce manuscrit fait qu'un des inconvénients relevés par Ince peut être retiré, à savoir l'obtention automatique d'un nombre optimal de classes. Les problèmes de dépendance à l'initialisation sont aussi résolus.

Pour prendre en compte la possibilité que des pixels soient trop éloignés des pixels déjà connus, une structure particulièrement appropriée est celle d'un test d'hypothèse. En effet, ce type de test permet de tester la validité d'une hypothèse et fournit un seuil au-delà duquel l'hypothèse est rejetée. On cherche ici à savoir si les matrices de covariance de deux pixels sont égales et le rejet de l'hypothèse d'égalité permettra alors de prendre en compte les pixels trop différents. On montrera aussi que la distance de Wishart est un cas particulier du test considéré.

2.1.2 Test d'égalité de matrices de covariance

On cherche donc à "comparer" deux matrices de covariances empiriques l'une avec l'autre, sans hypothèse sur leur distribution. Plus exactement, on cherche à savoir si deux échantillons $\mathbf{y}^{(1)} = (\mathbf{y}_1^{(1)}, \mathbf{y}_2^{(1)}, \dots, \mathbf{y}_{N_1}^{(1)})$ et $\mathbf{y}^{(2)} = (\mathbf{y}_1^{(2)}, \mathbf{y}_2^{(2)}, \dots, \mathbf{y}_{N_2}^{(2)})$ ont la même matrice de covariance. Ainsi, on pourra comparer à l'aide d'un test d'hypothèse les pixels par leurs matrices de covariance estimées avec le FPE, et de cette manière rassembler les pixels avec les mêmes propriétés polarimétriques au sein d'une même classe. Le test d'hypothèses est donné, de manière très générale par :

$$\begin{cases} H_0 : \mathbf{T}_1 = \mathbf{T}_2, \\ H_1 : \mathbf{T}_1 \neq \mathbf{T}_2, \end{cases} \quad (2.1)$$

avec \mathbf{T}_1 et \mathbf{T}_2 les matrices de covariance de $\mathbf{y}^{(1)}$ et $\mathbf{y}^{(2)}$ respectivement.

Une manière très conventionnelle d'envisager ce test est d'utiliser la méthode du rapport de vraisemblance [57]. On définit une fonction de vraisemblance pour chacune des deux hypothèses H_0 et H_1 . La fonction de vraisemblance est une fonction des paramètres du modèle qui donne, connaissant un ensemble d'observations, la probabilité de ces observations connaissant la valeur des paramètres du système, regroupés dans un vecteur noté θ . Dans le cas d'observations i.i.d., elle est alors donnée par le produit des probabilités des observations (lois marginales) connaissant la valeur des paramètres. L'expression de la vraisemblance est donc donnée, par exemple pour $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$, par l'équation suivante :

$$L(\mathbf{y}; H, \theta) = \prod_{i=1}^{N_1} f(\mathbf{y}_i | H, \theta), \quad (2.2)$$

avec H l'hypothèse H_0 ou l'hypothèse H_1 et $f(\mathbf{y}_i | H, \theta)$ la ddp de \mathbf{y}_i sous l'hypothèse H .

Le rapport de vraisemblance permet alors de déterminer quelle hypothèse est la plus vraisemblable selon les observations disponibles. Il s'exprime de la manière suivante :

$$\Lambda = \frac{L(\mathbf{y}; H_1, \theta)}{L(\mathbf{y}; H_0, \theta)} = \frac{L(\mathbf{y}; \theta_1)}{L(\mathbf{y}; \theta_0)}, \quad (2.3)$$

avec θ_0 et θ_1 la valeur des paramètres suivant l'hypothèse H_0 et H_1 respectivement.

Notons qu'il est également possible de l'exprimer sous la forme du rapport de log-vraisemblance qui est souvent plus simple à manipuler :

$$\Lambda = \ln \left(\frac{L(\mathbf{y}; H_1, \theta)}{L(\mathbf{y}; H_0, \theta)} \right). \quad (2.4)$$

Afin de quantifier les performances du test, deux types d'erreurs sont utilisées :

- erreur de première espèce : obtenue quand H_0 est vraie et H_1 décidée. Ce type d'erreur est communément appelé *probabilité de fausse alarme* (ou P_{FA}) dans la communauté radar. En effet, la présence de bruit seul est généralement associée à l'hypothèse H_0 et la présence d'une cible plus du bruit à l'hypothèse H_1 .
- erreur de deuxième espèce : obtenue quand H_1 est vraie et H_0 décidée. Ce type d'erreur est communément appelé *probabilité de non détection* dans la communauté radar car il représente le fait de décider de l'absence du signal alors qu'il est présent.

En réalité, il est très difficile de s'affranchir totalement de ces erreurs, étant donné la nature fortement aléatoire de l'environnement radar. Néanmoins, il est possible de les minimiser suivant un certain nombre de critères. L'un des plus largement répandus est le critère de Neyman-Pearson qui cherche à minimiser la probabilité de non détection pour une probabilité de fausse alarme fixée, notée α . Dans la pratique, ce taux de fausse alarme est réglé par l'utilisateur et est généralement très faible.

Si les hypothèses H_0 et H_1 sont des hypothèses simples, alors le test donné par :

$$\Lambda = \frac{L(\mathbf{y}; H_1, \theta)}{L(\mathbf{y}; H_0, \theta)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (2.5)$$

est le test Uniformément le Plus Puissant (UPP) au sens de Neyman-Pearson, i.e. le test qui, pour une probabilité de fausses alarmes α (ou erreur de première espèce) fixée, a la plus petite erreur de seconde espèce (i.e. probabilité de non-détection) parmi les tests de $P_{FA} = \alpha$. Cette approche permet donc de maximiser la probabilité de détection.

Le terme η de l'Eq. (2.5) est communément appelé *seuil de détection* et est obtenu pour un taux de fausse alarme α donné. Il se calcule grâce à l'équation suivante :

$$\alpha = P(\Lambda(\mathbf{y}; H_0, \theta) > \eta), \quad (2.6)$$

avec $P(x > a)$ la probabilité que x soit supérieur à a .

Chapitre 2. Classification non supervisée

Cette approche n'est malheureusement pas applicable directement dans notre cas. En effet, chacune des deux hypothèses prend en compte un certain nombre de paramètres inconnus. Le lemme de Neyman-Pearson ne peut être appliqué puisque nous sommes en présence d'hypothèses composites, étant donné que nous allons manipuler des systèmes contenant des paramètres inconnus. On considère alors le Rapport de Vraisemblance Généralisé (RVG, ou GLRT pour Generalized Likelihood Ratio Test) qui consiste à remplacer les paramètres inconnus du modèle par leurs estimées au sens du maximum de vraisemblance. Bien que n'étant pas un test UPP, le RVG donne très souvent de bons résultats et est très répandu dans la littérature.

Le RVG s'exprime alors de la manière suivante :

$$\Lambda = \frac{\sup_{\theta} L(\mathbf{y}; H_1, \theta)}{\sup_{\theta} L(\mathbf{y}; H_0, \theta)}. \quad (2.7)$$

Dans le cadre de notre test d'hypothèse de l'Eq. (2.1), il est important de s'intéresser aux différents cas qui peuvent se présenter. On peut en distinguer deux principaux :

— une des matrices de covariance peut être connue, par exemple $\mathbf{T}_2 = \mathbf{T}_0$ connue. Dans ce cas, le test d'hypothèses revient à :

$$\begin{cases} H_0 : \mathbf{T} = \mathbf{T}_0, \\ H_1 : \mathbf{T} \neq \mathbf{T}_0. \end{cases}$$

— les deux matrices de covariances sont inconnues. Le test correspond alors à :

$$\begin{cases} H_0 : \mathbf{T}_1 = \mathbf{T}_2 = \mathbf{T}, \\ H_1 : \mathbf{T}_1 \neq \mathbf{T}_2. \end{cases}$$

Nous allons étudier ces deux cas pour les deux hypothèses sur la distribution des vecteurs $\mathbf{y}^{(1)}$ et $\mathbf{y}^{(2)}$ qui nous intéressent, à savoir la distribution gaussienne classique et la distribution plus générale issue du modèle SIRV.

2.1.2.0.1 Cas gaussien

On suppose que les vecteurs $\mathbf{y}^{(1)}$ et $\mathbf{y}^{(2)}$ sont des vecteurs d'observations gaussiens complexes circulaires de moyenne nulle. Pour un vecteur \mathbf{y} , avec $\mathbf{y} = \mathbf{y}^{(1)}$ ou $\mathbf{y}^{(2)}$, on note \mathbf{y}_i une observation. L'expression de la densité de probabilité de \mathbf{y}_i est bien connue et dépend de la matrice de covariance \mathbf{T} :

$$f(\mathbf{y}_i, \mathbf{T}) = \frac{1}{\pi^m |\mathbf{T}|} \exp(-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i) \quad (2.8)$$

avec m le nombre de composantes du vecteur $\mathbf{y}_i^{(1)}$.

En remplaçant l'Eq. (2.8) dans l'Eq. (2.2), on obtient l'expression de la vraisemblance de \mathbf{y} dans le cas gaussien :

$$L(\mathbf{y}; H, \mathbf{T}) = \prod_{i=1}^N \frac{1}{\pi^m |\mathbf{T}|} \exp(-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i), \quad (2.9)$$

avec \mathbf{T} la matrice de covariance, à savoir le paramètre inconnu du modèle.

L'EMV de \mathbf{T} dans le cas gaussien est la SCM de l'Eq. (1.35) dont nous rappelons ici l'expression :

$$\hat{\mathbf{T}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^H. \quad (2.10)$$

Nous allons nous intéresser aux deux cas mentionnés précédemment, à savoir :

- en considérant un paramètre connu, $\mathbf{T} = \mathbf{T}_0$,
- avec les deux matrices inconnues.

Cas $\mathbf{T} = \mathbf{T}_0$ connu :

On suppose ici que \mathbf{T}_2 est connue et on cherche à savoir si l'échantillon $\mathbf{y}^{(1)}$ suit la distribution caractérisée par \mathbf{T}_0 . Comme \mathbf{T}_2 est connue, il n'est pas nécessaire de différencier $\mathbf{y}^{(1)}$ de $\mathbf{y}^{(2)}$, on utilisera donc la notation \mathbf{y} . Le test d'hypothèse s'exprime alors de la manière suivante :

$$\begin{cases} H_0 : \mathbf{T} = \mathbf{T}_0, \\ H_1 : \mathbf{T} \neq \mathbf{T}_0. \end{cases}$$

On a alors le RVG suivant :

$$\Lambda = \frac{\sup_{\theta} L(\mathbf{y}; H_1)}{\sup_{\theta} L(\mathbf{y}; H_0)}, \quad (2.11)$$

avec θ les paramètres du modèle. Ici, le seul paramètre, commun aux deux hypothèses, est la matrice de covariance \mathbf{T} . Sous l'hypothèse H_0 , \mathbf{T} est connue et est égale à \mathbf{T}_0 . En remplaçant la vraisemblance par son expression donnée par l'Eq. (2.9) dans l'Eq. (2.11), le RVG devient :

$$\Lambda = \frac{\sup_{\mathbf{T}} \prod_{i=1}^{N_1} \frac{1}{\pi^m |\mathbf{T}|} \exp(-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i)}{\prod_{i=1}^{N_1} \frac{1}{\pi^m |\mathbf{T}_0|} \exp(-\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i)}. \quad (2.12)$$

Pour calculer le terme au numérateur, il faut dériver l'expression contenue dans le sup (la vraisemblance sous l'hypothèse H_1) par rapport au paramètre inconnu, ici \mathbf{T} et annuler cette dérivée pour obtenir ainsi le maximum de la vraisemblance par rapport aux paramètres du système. Pour cela, on va prendre le logarithme de cette vraisemblance et transformer le terme dans l'exponentielle afin de faciliter le calcul. On cherche donc le sup de :

$$-mN_1 \ln \pi - N_1 \ln |\mathbf{T}| - \sum_{i=1}^{N_1} \mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i \quad (2.13)$$

Le terme dans la somme est un scalaire (forme quadratique), on peut donc appliquer l'opérateur Trace pour simplifier l'expression de la manière suivante :

$$\begin{aligned}
 \sum_{i=1}^{N_1} \mathbf{y}_i^H \mathbf{T} \mathbf{y}_i &= \text{Tr} \left(\sum_{i=1}^{N_1} \mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i \right), \\
 &= \text{Tr} \left(\sum_{i=1}^{N_1} \mathbf{T}^{-1} \mathbf{y}_i \mathbf{y}_i^H \right), \\
 &= \text{Tr} \left(\mathbf{T}^{-1} \sum_{i=1}^{N_1} \mathbf{y}_i \mathbf{y}_i^H \right), \\
 &= \text{Tr} \left(\mathbf{T}^{-1} N_1 \hat{\mathbf{T}}_1 \right), \\
 &= N_1 \text{Tr} \left(\mathbf{T}^{-1} \hat{\mathbf{T}}_1 \right),
 \end{aligned} \tag{2.14}$$

avec $\hat{\mathbf{T}}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{y}_i^{(1)} \mathbf{y}_i^{(1)H}$ la SCM obtenue à partir des $\mathbf{y}_i^{(1)}$.

On peut alors remplacer l'Eq. (2.14) dans l'Eq. (2.13). On va de plus dériver par rapport à \mathbf{T} pour trouver le maximum de la fonction. En réalité, on va dériver par rapport à \mathbf{T}^{-1} pour simplifier le calcul. On a alors :

$$\begin{aligned}
 \frac{d}{d\mathbf{T}^{-1}} \left(-mN_1 \ln \pi + N_1 \ln |\mathbf{T}^{-1}| - N_1 \text{Tr} \left(\mathbf{T}^{-1} \hat{\mathbf{T}}_1 \right) \right) &= N_1 \mathbf{T}^T - N_1 \hat{\mathbf{T}}_1 \\
 &= 0 \\
 \Leftrightarrow \hat{\mathbf{T}} &= \hat{\mathbf{T}}_1
 \end{aligned} \tag{2.15}$$

L'estimateur du maximum de vraisemblance de \mathbf{T} sous hypothèse H_1 est donc évidemment la SCM de l'Eq. (1.35). Ceci donne alors l'expression suivante pour le RVG :

$$\Lambda = \frac{\prod_{i=1}^{N_1} \frac{1}{\pi^m |\hat{\mathbf{T}}_1|} \exp \left(-\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i \right)}{\prod_{i=1}^{N_1} \frac{1}{\pi^m |\mathbf{T}_0|} \exp \left(-\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i \right)}, \tag{2.16}$$

$$= \left| \frac{\mathbf{T}_0}{\hat{\mathbf{T}}_1} \right|^{N_1} \exp \left(\sum_{i=1}^{N_1} \mathbf{y}_i^H \left(\mathbf{T}_0^{-1} - \hat{\mathbf{T}}_1^{-1} \right) \mathbf{y}_i \right). \tag{2.17}$$

Par les mêmes considérations que dans l'Eq. (2.14), on applique l'opérateur Trace au terme situé dans l'exponentielle :

$$\begin{aligned}
 \sum_{i=1}^{N_1} \mathbf{y}_i^H \left(\mathbf{T}_0^{-1} - \hat{\mathbf{T}}_1^{-1} \right) \mathbf{y}_i &= \text{Tr} \left(\sum_{i=1}^{N_1} \mathbf{y}_i^H \left(\mathbf{T}_0^{-1} - \hat{\mathbf{T}}_1^{-1} \right) \mathbf{y}_i \right) \\
 &= \text{Tr} \left(\left(\mathbf{T}_0^{-1} - \hat{\mathbf{T}}_1^{-1} \right) \sum_{i=1}^{N_1} \mathbf{y}_i \mathbf{y}_i^H \right) \\
 &= \text{Tr} \left(\left(\mathbf{T}_0^{-1} - \hat{\mathbf{T}}_1^{-1} \right) N_1 \hat{\mathbf{T}}_1 \right) \\
 &= N_1 \left(\text{Tr} \left(\mathbf{T}_0^{-1} \hat{\mathbf{T}}_1 \right) - m \right)
 \end{aligned} \tag{2.18}$$

D'où l'expression du RVG final :

$$\Lambda = \left| \frac{\mathbf{T}_0}{\widehat{\mathbf{T}}_1} \right|^{N_1} \exp \left(N_1 \text{Tr} \left(\mathbf{T}_0^{-1} \widehat{\mathbf{T}}_1 \right) - N_1 m \right) \quad (2.19)$$

En prenant le logarithme de l'Eq. (2.14), on obtient :

$$\begin{aligned} \Lambda' &= \ln \left| \frac{\mathbf{T}_0}{\widehat{\mathbf{T}}_1} \right|^{N_1} \exp \left(N_1 \text{Tr} \left(\mathbf{T}_0^{-1} \widehat{\mathbf{T}}_1 \right) - N_1 m \right) \\ &= N_1 \left(\ln |\mathbf{T}_0| - \ln |\widehat{\mathbf{T}}_1| + \text{Tr} \left(\mathbf{T}_0^{-1} \widehat{\mathbf{T}}_1 \right) - m \right). \end{aligned} \quad (2.20)$$

Cette expression est proche de l'expression de l'Eq. (1.51). En effet, si l'on considère l'approche du classifieur Wishart, elle consiste à comparer la matrice de covariance d'un pixel avec les matrices de centre de classe et de choisir la plus petite distance entre ces matrices. On peut considérer alors que c'est la matrice de centre de classe est connue et notée \mathbf{T}_0 et qu'on cherche à savoir si un pixel de matrice de covariance \mathbf{T} appartient à cette classe. On peut alors, comme lors de la dérivation de la distance de Wishart [51], à nouveau éliminer les termes qui ne contribuent pas à la classification (les facteurs N_1 , $\ln |\widehat{\mathbf{T}}_1|$ et m) et on retrouve l'expression de la distance de Wishart de l'Eq. (1.52) :

$$\Lambda' = \ln |\mathbf{T}_0| + \text{Tr} \left(\mathbf{T}_0^{-1} \widehat{\mathbf{T}}_1 \right). \quad (2.21)$$

Cas de deux matrices inconnues :

Le test d'hypothèse s'exprime alors de la manière suivante :

$$\begin{cases} H_0 : \mathbf{T}_1 = \mathbf{T}_2 = \mathbf{T}, \\ H_1 : \mathbf{T}_1 \neq \mathbf{T}_2. \end{cases}$$

Le rapport de vraisemblance devient donc :

$$\Lambda = \frac{\sup_{\mathbf{T}_1, \mathbf{T}_2} \left(\prod_{i=1}^{N_1} \frac{1}{\pi^m |\mathbf{T}_1|} \exp \left(-\mathbf{y}_i^{(1)H} \mathbf{T}_1^{-1} \mathbf{y}_i^{(1)} \right) \prod_{i=1}^{N_2} \frac{1}{\pi^m |\mathbf{T}_2|} \exp \left(-\mathbf{y}_i^{(2)H} \mathbf{T}_2^{-1} \mathbf{y}_i^{(2)} \right) \right)}{\sup_{\mathbf{T}} \left(\prod_{i=1}^{N_1+N_2} \frac{1}{\pi^m |\mathbf{T}|} \exp \left(-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i \right) \right)}, \quad (2.22)$$

avec $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{N_1}, \mathbf{y}_{N_1+1}, \dots, \mathbf{y}_{N_1+N_2}) = (\mathbf{y}_1^{(1)}, \dots, \mathbf{y}_{N_1}^{(1)}, \mathbf{y}_1^{(2)}, \dots, \mathbf{y}_{N_2}^{(2)})$.

Pour le numérateur, par les mêmes considérations que pour l'Eq. (2.15), on peut montrer que les EMV de \mathbf{T}_1 et \mathbf{T}_2 sont respectivement $\widehat{\mathbf{T}}_1$ et $\widehat{\mathbf{T}}_2$ en dérivant la vraisemblance par rapport à \mathbf{T}_1^{-1} et \mathbf{T}_2^{-1} respectivement. Pour le dénominateur, on obtient l'EMV de \mathbf{T} , $\widehat{\mathbf{T}}$, en dérivant encore une fois le logarithme de la vraisemblance

par rapport au paramètre inconnu \mathbf{T} :

$$\begin{aligned}
 \sup_{\mathbf{T}} \left(\prod_{i=1}^{N_1+N_2} \frac{1}{\pi^m |\mathbf{T}|} \exp -\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i \right) &\stackrel{\ln}{\leadsto} -(N_1 + N_2)m \ln \pi - (N_1 + N_2) \ln(|\mathbf{T}|) - \sum_{i=1}^{N_1+N_2} \mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i, \\
 &= (N_1 + N_2) \ln(|\mathbf{T}^{-1}|) - (N_1 + N_2)m \ln \pi - \text{Tr} \left(\mathbf{T}^{-1} (N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2) \right), \\
 \Rightarrow \frac{d}{d\mathbf{T}^{-1}} \left((N_1 + N_2) \ln(|\mathbf{T}^{-1}|) - (N_1 + N_2)m \ln \pi - \text{Tr} \left(\mathbf{T}^{-1} (N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2) \right) \right) &= 0, \\
 \Leftrightarrow (N_1 + N_2)\mathbf{T} - (N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2) &= 0, \\
 \Leftrightarrow \hat{\mathbf{T}} &= \frac{N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2}{N_1 + N_2}. \tag{2.23}
 \end{aligned}$$

Ce résultat n'est pas surprenant puisque $\hat{\mathbf{T}}$ est l'estimateur au sens du maximum de vraisemblance du vecteur \mathbf{y} composé de la concaténation des vecteurs \mathbf{y}_1 et \mathbf{y}_2 , de telle manière que $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2]$. En utilisant directement la formule de la SCM, on obtient l'expression suivante :

$$\begin{aligned}
 \hat{\mathbf{T}} &= \frac{1}{N_1 + N_2} \sum_{i=1}^{N_1+N_2} (N_1 + N_2) \mathbf{y}_i \mathbf{y}_i^H, \\
 &= \frac{1}{N_1 + N_2} \left(N_1 \sum_{i=1}^{N_1} \mathbf{y}_i \mathbf{y}_i^H + N_2 \sum_{i=N_1+1}^{N_1+N_2} \mathbf{y}_i \mathbf{y}_i^H \right), \\
 &= \frac{1}{N_1 + N_2} \left(N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2 \right), \\
 &= \frac{N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2}{N_1 + N_2}. \tag{2.24}
 \end{aligned}$$

On peut alors remplacer dans l'Eq. (2.22) \mathbf{T}_1 , \mathbf{T}_2 et \mathbf{T} par leurs EMV $\hat{\mathbf{T}}_1$, $\hat{\mathbf{T}}_2$ et $\hat{\mathbf{T}}$ pour obtenir le RVG suivant :

$$\Lambda = \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \prod_{i=1}^{N_1} \exp \left(\mathbf{y}_i^{(1)H} \left(\hat{\mathbf{T}}^{-1} - \hat{\mathbf{T}}_1^{-1} \right) \mathbf{y}_i^{(1)} \right) \prod_{i=1}^{N_2} \exp \left(\mathbf{y}_i^{(2)H} \left(\hat{\mathbf{T}}^{-1} - \hat{\mathbf{T}}_2^{-1} \right) \mathbf{y}_i^{(2)} \right). \tag{2.25}$$

En appliquant encore une fois l'opérateur Trace sur les termes situés dans les exponentielles, on obtient :

$$\begin{aligned}
 \Lambda &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp \left(\text{Tr} \left(\left(\hat{\mathbf{T}}^{-1} - \hat{\mathbf{T}}_1^{-1} \right) N_1 \hat{\mathbf{T}}_1 \right) \right) \exp \left(\text{Tr} \left(\left(\hat{\mathbf{T}}^{-1} - \hat{\mathbf{T}}_2^{-1} \right) N_2 \hat{\mathbf{T}}_2 \right) \right), \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp \left(\text{Tr} \left(N_1 \left(\hat{\mathbf{T}}^{-1} \hat{\mathbf{T}}_1 - \mathbf{I} \right) \right) + \text{Tr} \left(N_2 \left(\hat{\mathbf{T}}^{-1} \hat{\mathbf{T}}_2 - \mathbf{I} \right) \right) \right), \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp \left(\text{Tr} \left(\hat{\mathbf{T}}^{-1} \left(N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2 \right) - (N_1 + N_2)m \right) \right), \tag{2.26}
 \end{aligned}$$

avec \mathbf{I} la matrice identité.

Ceci entraîne l'expression finale du RVG, donné par l'équation suivante :

$$\frac{|\widehat{\mathbf{T}}|^{N_1+N_2}}{|\widehat{\mathbf{T}}_1|^{N_1} |\widehat{\mathbf{T}}_2|^{N_2}}. \quad (2.27)$$

Cette expression est l'inverse d'une distance communément appelée distance de Bartlett, en référence au mathématicien qui a introduit le rapport des moyennes géométriques et arithmétiques des variances, dans le cas scalaire, comme un test robuste d'égalité entre distributions [41]. Elle a été employée à plusieurs reprises avec succès dans la communauté SAR avec comme application la détection de changement [26] ou la segmentation d'images [10, 31].

2.1.2.0.2 Cas SIRV Nous supposons ici que les données contenues dans les vecteurs $\mathbf{y}^{(1)}$ et $\mathbf{y}^{(2)}$ ne suivent plus des lois gaussiennes mais sont décrits par les processus SIRV décrits à la Section 1.3. Rappelons la densité de probabilité d'un tel processus :

$$f_{\mathbf{y}}(\mathbf{y}; \tau, \mathbf{T}) = \int_0^{+\infty} \frac{1}{(\pi\tau)^m |\mathbf{T}|} \exp\left(-\frac{\mathbf{y}^H \mathbf{T}^{-1} \mathbf{y}}{\tau}\right) p_{\tau}(\tau) d\tau, \quad (2.28)$$

avec $\mathbf{y} = \sqrt{\tau} \mathbf{x}$, $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{T})$ et $\tau \sim p_{\tau}(\tau)$ inconnue.

Puisque les données sont encore une fois i.i.d., la vraisemblance est donnée par l'équation suivante :

$$L(\mathbf{y}; H, \theta) = \prod_{i=1}^{N_k} f_{\mathbf{y}_i}(\mathbf{y}_i; H, \theta), \quad (2.29)$$

avec H l'hypothèse H_0 ou H_1 et θ les paramètres du modèle, à savoir τ et \mathbf{T} .

La présence de l'intégrale dans l'Eq. (2.28) rend difficile le calcul de cette vraisemblance. On se place alors dans le cadre d'une texture τ déterministe mais inconnue. On a alors la vraisemblance suivante :

$$\begin{aligned} L(\mathbf{y}; H, \tau, \mathbf{T}) &= \prod_{i=1}^N \frac{1}{\pi^m \tau_i^m |\mathbf{T}|} \exp\left(-\frac{\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i}\right), \\ &= \frac{1}{\pi^{mN} |\mathbf{T}|^N} \prod_{i=1}^N \frac{1}{\tau_i^m} \exp\left(-\frac{\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i}\right). \end{aligned} \quad (2.30)$$

De la même manière que dans le cas gaussien, nous allons maintenant considérer les deux possibilités d'envisager le test : un paramètre connu et deux paramètres inconnus.

Cas d'un paramètre connu :

Le test d'hypothèses s'exprime alors de la manière suivante :

$$\begin{cases} H_0 : \mathbf{T} = \mathbf{T}_0, \\ H_1 : \mathbf{T} \neq \mathbf{T}_0. \end{cases}$$

Le rapport de vraisemblance s'écrit donc :

$$\begin{aligned} \Lambda &= \frac{\sup_{\theta} L(\mathbf{y}; H_1, \theta)}{\sup_{\theta} L(\mathbf{y}; H_0, \theta)}, \\ &= \frac{\sup_{\theta, H_1} \frac{1}{\pi^{mN_1} |\mathbf{T}|^{N_1}} \prod_{i=1}^{N_1} \frac{1}{\tau_i^m} \exp\left(\frac{-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i}\right)}{\sup_{\theta, H_0} \frac{1}{\pi^{mN_1} |\mathbf{T}|^{N_1}} \prod_{i=1}^{N_1} \frac{1}{\tau_i^m} \exp\left(\frac{-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i}\right)}. \end{aligned} \quad (2.31)$$

La paramètre θ regroupe les paramètres du système, à savoir la matrice de covariance \mathbf{T} . En effet, on considère que la texture τ n'est pas un paramètre à part entière du système. Cependant, cette texture n'est pas connue de manière générale, elle sera donc estimée directement à partir des données contenues dans le vecteur \mathbf{y} .

On peut alors écrire le rapport de vraisemblance de la manière suivante :

$$\Lambda = \frac{\sup_{\mathbf{T}, H_1} \frac{1}{\pi^{mN_1} |\mathbf{T}|^{N_1}} \prod_{i=1}^{N_1} \frac{1}{\tau_i^m} \exp\left(\frac{-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i}\right)}{\sup_{\mathbf{T}, H_0} \frac{1}{\pi^{mN_1} |\mathbf{T}|^{N_1}} \prod_{i=1}^{N_1} \frac{1}{\tau_i^m} \exp\left(\frac{-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i}\right)}. \quad (2.32)$$

Sous l'hypothèse H_0 , on sait que $\mathbf{T} = \mathbf{T}_0$ et sous l'hypothèse H_1 , l'EMV de \mathbf{T} est le FPE de l'Eq. (1.57) dont on rappelle ici l'expression :

$$\hat{\mathbf{T}} = \frac{m}{N_1} \sum_{i=1}^{N_1} \frac{\mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \hat{\mathbf{T}}^{-1} \mathbf{y}_i}.$$

De plus, on connaît l'EMV de la texture τ , donné par l'Eq. (1.66) :

$$\hat{\tau}_i = \frac{\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{m},$$

mais \mathbf{T} n'étant pas connu, on le remplace par son EMV. On a donc :

$$\hat{\tau}_i = \frac{\mathbf{y}_i^H \hat{\mathbf{T}}^{-1} \mathbf{y}_i}{m}. \quad (2.33)$$

Le RVG devient donc :

$$\begin{aligned}
 \Lambda &= \frac{\frac{1}{\pi^{mN_1}} \left| \widehat{\mathbf{T}} \right|^{N_1} \prod_{i=1}^{N_1} \frac{1}{\widehat{\tau}_i^m} \exp \left(\frac{-\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i}{\widehat{\tau}_i} \right)}{\frac{1}{\pi^{mN_1}} \left| \mathbf{T}_0 \right|^{N_1} \prod_{i=1}^{N_1} \frac{1}{\widehat{\tau}_i^m} \exp \left(\frac{-\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\widehat{\tau}_i} \right)}, \\
 &= \frac{\left| \mathbf{T}_0 \right|^{N_1} \prod_{i=1}^{N_1} \exp \left(-m \frac{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right)}{\prod_{i=1}^{N_1} \exp \left(-m \frac{\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right)}, \\
 &= \frac{\left| \mathbf{T}_0 \right|^{N_1}}{\left| \widehat{\mathbf{T}} \right|^{N_1}} \exp \left(m \sum_{i=1}^{N_1} \frac{\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} - mN_1 \right). \tag{2.34}
 \end{aligned}$$

On peut encore une fois appliquer l'opérateur Trace sur l'expression $\sum_{i=1}^{N_1} \frac{\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i}$ qui devient :

$$\begin{aligned}
 \sum_{i=1}^{N_1} \frac{\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} &= \text{Tr} \left(\sum_{i=1}^{N_1} \frac{\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right), \\
 &= \sum_{i=1}^{N_1} \text{Tr} \left(\frac{\mathbf{y}_i^H \mathbf{T}_0^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right), \\
 &= \sum_{i=1}^{N_1} \text{Tr} \left(\frac{\mathbf{T}_0^{-1} \mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right), \\
 &= \text{Tr} \left(\sum_{i=1}^{N_1} \frac{\mathbf{T}_0^{-1} \mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right), \\
 &= \text{Tr} \left(\mathbf{T}_0^{-1} \sum_{i=1}^{N_1} \frac{\mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \widehat{\mathbf{T}}^{-1} \mathbf{y}_i} \right), \\
 &= \text{Tr} \left(\mathbf{T}_0^{-1} \widehat{\mathbf{T}} \right), \tag{2.35}
 \end{aligned}$$

d'après l'Eq. (1.57) qui donne l'expression de $\widehat{\mathbf{T}}$ selon le FPE.

En injectant l'Eq. (2.35) dans l'Eq. (2.34) et en appliquant le logarithme, on obtient alors l'expression finale du logarithme du rapport de vraisemblance :

$$\Lambda' = \ln \Lambda = N_1 \left(\ln \left| \mathbf{T}_0 \right| - \ln \left| \widehat{\mathbf{T}} \right| + \text{Tr} \left(\mathbf{T}_0^{-1} \widehat{\mathbf{T}} \right) - m \right). \tag{2.36}$$

Remarquons que le cas d'un SIRV traité ici est en réalité un cas gaussien à variance inconnue et différente sur tous les \mathbf{y}_i . Rappelons aussi que $\widehat{\mathbf{T}}$ n'est pas obtenu à partir de la SCM, l'estimateur traditionnel, mais bien du FPE adapté au cas SIRV, donné par l'Eq. (1.57). Comme pour le cas gaussien, si l'on considère le cas où l'on veut placer un pixel dans une classe parmi d'autres, on peut utiliser ce test avec $\widehat{\mathbf{T}}$ l'estimée de la matrice de covariance du pixel et \mathbf{T}_0 le centre de la classe. Comme on va comparer les résultats de l'Eq. (2.36) entre

plusieurs classes (donc plusieurs \mathbf{T}_0), afin de placer le pixel dans la classe la plus proche, les termes $N_1 \ln |\hat{\mathbf{T}}|$ et $N_1 m$ sont constants pour un $\hat{\mathbf{T}}$ donné. On obtient alors l'expression de la distance de Wishart (cf. Eq. (1.52)) avec pour seule différence l'estimateur de la matrice de covariance \mathbf{T} , $\hat{\mathbf{T}}$ qui est le FPE et plus la SCM..

Cas de deux matrices inconnues :

Comme dans le cas gaussien, le test d'hypothèse s'exprime de la manière suivante :

$$\begin{cases} H_0 : \mathbf{T}_1 = \mathbf{T}_2 = \mathbf{T}, \\ H_1 : \mathbf{T}_1 \neq \mathbf{T}_2. \end{cases}$$

L'expression du rapport de vraisemblance est donc :

$$\Lambda = \frac{\sup_{\mathbf{T}_1, \mathbf{T}_2} \left(\frac{1}{\pi^{mN_1} |\mathbf{T}_1|^{N_1}} \prod_{i=1}^{N_1} \frac{1}{\tau_{1i}^m} \exp \left(\frac{-\mathbf{y}_i^{(1)H} \mathbf{T}_1^{-1} \mathbf{y}_i^{(1)}}{\tau_{1i}} \right) \frac{1}{\pi^{mN_2} |\mathbf{T}_2|^{N_2}} \prod_{i=1}^{N_2} \frac{1}{\tau_{2i}^m} \exp \left(\frac{-\mathbf{y}_i^{(2)H} \mathbf{T}_2^{-1} \mathbf{y}_i^{(2)}}{\tau_{2i}} \right) \right)}{\sup_{\mathbf{T}} \frac{1}{\pi^{m(N_1+N_2)} |\mathbf{T}|^{N_1+N_2} \left(\prod_{i=1}^{N_1+N_2} \frac{1}{\tau_i^m} \exp \left(\frac{-\mathbf{y}_i^H \mathbf{T}^{-1} \mathbf{y}_i}{\tau_i} \right) \right)}, \quad (2.37)$$

avec $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{N_1}, \mathbf{y}_{N_1+1}, \dots, \mathbf{y}_{N_1+N_2}) = (\mathbf{y}_1^{(1)}, \dots, \mathbf{y}_{N_1}^{(1)}, \mathbf{y}_1^{(2)}, \dots, \mathbf{y}_{N_2}^{(2)})$. Comme on a les relations suivantes pour tout $i \in [1, \dots, N_1 + N_2]$:

- $\mathbf{y}_i \sim \sqrt{\tau_i} \mathbf{x}_i$, avec $\mathbf{x}_i \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{M})$,
- $\mathbf{y}_i^{(1)} \sim \sqrt{\tau_{1i}} \mathbf{x}_i$, avec $\mathbf{x}_i \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{M}_1)$,
- $\mathbf{y}_i^{(2)} \sim \sqrt{\tau_{2i}} \mathbf{x}_i$, avec $\mathbf{x}_i \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{M}_2)$,

alors on a $\tau = (\tau_1, \dots, \tau_{N_1}, \tau_{N_1+1}, \dots, \tau_{N_1+N_2}) = (\tau_{11}, \dots, \tau_{1N_1}, \tau_{21}, \dots, \tau_{2N_2})$.

Comme pour les cas précédents, on va remplacer \mathbf{T}_1 , \mathbf{T}_2 , τ_i , τ_{1i} et τ_{2i} par leurs EMV respectifs. On arrive

alors au rapport de vraisemblance généralisé suivant :

$$\begin{aligned}
 \Lambda &= \frac{\frac{1}{\pi^{mN_1}} |\hat{\mathbf{T}}_1|^{N_1} \prod_{i=1}^{N_1} \frac{1}{\hat{\tau}_{1i}^m} \exp\left(\frac{-\mathbf{y}_i^{(1)H} \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i^{(1)}}{\hat{\tau}_{1i}}\right) \frac{1}{\pi^{mN_2}} |\hat{\mathbf{T}}_2|^{N_2} \prod_{i=1}^{N_2} \frac{1}{\hat{\tau}_{2i}^m} \exp\left(\frac{-\mathbf{y}_i^{(2)H} \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i^{(2)}}{\hat{\tau}_{2i}}\right)}{1}, \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2} \prod_{i=1}^{N_1} \exp\left(-m \frac{\mathbf{y}_i^{(1)H} \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i^{(1)}}{\mathbf{y}_i^{(1)H} \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i^{(1)}}\right) \prod_{i=1}^{N_2} \exp\left(-m \frac{\mathbf{y}_i^{(2)H} \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i^{(2)}}{\mathbf{y}_i^{(2)H} \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i^{(2)}}\right)}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2} \prod_{i=1}^{N_1} \exp\left(-m \frac{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i}\right) \prod_{i=N_1+1}^{N_2} \exp\left(-m \frac{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i}\right)}, \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp\left(\sum_{i=1}^{N_1} \frac{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i} + \sum_{i=N_1+1}^{N_2} \frac{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i} - (N_1 + N_2)m\right), \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp\left(\sum_{i=1}^{N_1} \text{Tr}\left(\frac{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i}\right) + \sum_{i=N_1+1}^{N_2} \text{Tr}\left(\frac{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i}{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i}\right) - (N_1 + N_2)m\right), \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp\left(\sum_{i=1}^{N_1} \text{Tr}\left(\frac{\hat{\mathbf{T}}_1^{-1} \mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \hat{\mathbf{T}}_1^{-1} \mathbf{y}_i}\right) + \sum_{i=N_1+1}^{N_2} \text{Tr}\left(\frac{\hat{\mathbf{T}}_2^{-1} \mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \hat{\mathbf{T}}_2^{-1} \mathbf{y}_i}\right) - (N_1 + N_2)m\right), \\
 &= \frac{|\hat{\mathbf{T}}|^{N_1+N_2}}{|\hat{\mathbf{T}}_1|^{N_1} |\hat{\mathbf{T}}_2|^{N_2}} \exp\left(\text{Tr}\left(\hat{\mathbf{T}}^{-1} [N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2]\right) - (N_1 + N_2)m\right). \tag{2.38}
 \end{aligned}$$

On obtient une expression identique au cas gaussien (cf. Eq. (2.26)). Cependant, dans le cas gaussien, on peut exprimer simplement $\hat{\mathbf{T}}$ en fonction de $\hat{\mathbf{T}}_1$ et de $\hat{\mathbf{T}}_2$ comme une combinaison linéaire, qui permet de simplifier l'Eq. (2.26) en l'Eq. (2.27). Dans le cas SIRV, il n'existe pas d'expression analytique pour donner $\hat{\mathbf{T}}$ en fonction de $\hat{\mathbf{T}}_1$ et $\hat{\mathbf{T}}_2$ à cause de la structure de l'estimateur du point fixe (cf Eq. (1.57)). Il n'est donc pas possible de simplifier plus avant cette expression.

2.1.3 M-Test de Box

2.1.3.1 Présentation du test

Le M-test de Box [15] [16] est une modification des rapports de vraisemblance présentés précédemment pour le cas d'un paramètre commun [65]. Il a été initialement développé pour le cas gaussien et se base sur les travaux de Bartlett [9] qui a proposé des exposants légèrement modifiés pour le cas scalaire afin que la distribution de la statistique du test (qui est donc un GLRT) soit plus proche d'une distribution du χ^2 que la distribution de la statistique du test classique. Ces travaux ont par la suite été étendus au cas multivarié et la statistique proposée est :

$$t = \frac{|\hat{\mathbf{T}}_1|^{\frac{\nu_1}{2}} |\hat{\mathbf{T}}_2|^{\frac{\nu_2}{2}}}{|\hat{\mathbf{T}}|^{\frac{\nu}{2}}} \tag{2.39}$$

Chapitre 2. Classification non supervisée

où les $\nu_i = N_i$ sont les ddl de \mathbf{T}_i et $\nu = N_1 + N_2$, le ddl de $\widehat{\mathbf{T}}$, c'est-à-dire le nombre d'échantillons utilisés pour réaliser l'estimation de la matrice de covariance. On remarque aisément que cette statistique est l'inverse de l'Eq. (2.21).

Cette statistique est comprise entre 0 et 1, les valeurs proches de 0 rejetant l'hypothèse H_0 et les valeurs proches de 1 l'acceptant. Pour illustrer ce principe, on peut réécrire l'Eq. (2.39) comme :

$$t = \frac{\left(\frac{|\widehat{\mathbf{T}}_1|}{|\widehat{\mathbf{T}}|} \right)^{\frac{\nu_1}{2}} \left(\frac{|\widehat{\mathbf{T}}_2|}{|\widehat{\mathbf{T}}|} \right)^{\frac{\nu_2}{2}}}{1}$$

Quand $\widehat{\mathbf{T}}_1 = \widehat{\mathbf{T}}_2 = \widehat{\mathbf{T}}$, on a $t = 1$. Quand $\widehat{\mathbf{T}}_1$ et $\widehat{\mathbf{T}}_2$ ne sont pas égales, la valeur de t diminue d'autant plus que les matrices sont "différentes".

Box [15] [16] a introduit des modifications de cette statistique afin d'obtenir une approximation χ^2 pour la distribution de t . On obtient l'approximation du χ^2 en modifiant la statistique t par :

$$u = -2(1 - c_1) \ln(t) \sim \chi^2\left(\frac{1}{2}m(m+1)\right), \quad (2.40)$$

avec

$$c_1 = \left(\frac{\sum_{i=1}^2 \frac{1}{\nu_i}}{\sum_{i=1}^2 \nu_i} \right) \left(\frac{2m^2 + 3m - 1}{6(m+1)} \right),$$

avec $\chi^2(a)$ la distribution du χ^2 avec a ddl. On peut alors, dans notre cas, remplacer t par la statistique de l'Eq. (2.21) et appeler $u = 2(1 - c_1)t$.

La région critique du test est donnée par l'équation suivante :

$$R_c = \left\{ u, u > \chi_{P_{FA}}^2\left(\frac{1}{2}m(m+1)\right) \right\} \quad (2.41)$$

avec P_{FA} la probabilité de fausse alarme et $\chi_{P_{FA}}^2(n)$ le quantile d'ordre $1 - P_{FA}$ de la loi du χ^2 à n ddl. Ce taux de fausse alarme est généralement réglé par l'utilisateur. Il est choisi de manière à être très faible et correspond au fait de rejeter l'égalité des matrices de covariance alors qu'elles sont en réalité égales.

On dispose ainsi d'un test statistique dérivé rigoureusement à partir du rapport de vraisemblance et modifié afin de suivre plus précisément une distribution largement connue, la distribution du χ^2 . On est donc désormais en mesure d'utiliser ce test comme une distance et rejeter les pixels (caractérisés par leur matrice de covariance) qui seraient trop éloignés des pixels déjà connus grâce à l'Eq. (2.41).

2.1.3.2 Extension au cas SIRV

L'extension au cas SIRV du test de Box est compliquée par l'expression du rapport de vraisemblance pour le cas où les deux matrices sont inconnues dans le cas SIRV, donnée par l'Eq. (2.38). En effet, le terme dans l'exponentielle ne se simplifie pas car il n'existe pas, contrairement au cas gaussien, d'expression analytique en fonction de $\widehat{\mathbf{T}}_1$ et $\widehat{\mathbf{T}}_2$ pour l'EMV de la matrice de covariance \mathbf{T} obtenue à partir des deux échantillons $\mathbf{y}^{(1)}$ et $\mathbf{y}^{(2)}$ regroupés en un seul échantillon \mathbf{y} . En effet, les estimations sont réalisées avec le FPE. Il est donc

très difficile, voire impossible, de déterminer la statistique de ce rapport de vraisemblance et d'en donner une approximation dans le cas SIRV.

Cependant, une des propriétés du FPE, comme indiqué dans la Section 1.3.3 est qu'il se comporte asymptotiquement comme une matrice de Wishart estimée avec un nombre de ddl légèrement supérieur. Il est alors envisageable d'étendre les résultats obtenus dans le cas gaussien directement sur des matrices estimées par le FPE en ajustant le nombre de ddl. Il faut toutefois noter que cette méthode n'est qu'une approximation car elle utilise des résultats asymptotiques alors que le nombre d'échantillons utilisés est fini. On remarque cependant qu'en pratique, en utilisant une fenêtre de taille 7×7 ou 9×9 pour l'estimation de la matrice, le FPE a effectivement un comportement proche de celui d'une matrice distribuée selon une loi de Wishart.

L'extension au cas SIRV donne donc la statistique non modifiée suivante :

$$t = \frac{|\widehat{\mathbf{T}}_1|^{\frac{\nu_1}{2}} |\widehat{\mathbf{T}}_2|^{\frac{\nu_2}{2}}}{|\widehat{\mathbf{T}}|^{\frac{\nu}{2}}}, \quad (2.42)$$

où les $\nu_i = \frac{m}{m+1} N_i$ sont les ddl de $\widehat{\mathbf{T}}_i$ et $\nu = \frac{m}{m+1} (N_1 + N_2)$, le ddl de $\widehat{\mathbf{T}}$.

La statistique finale est donnée, comme dans le cas gaussien, par l'Eq. (2.40).

2.2 Algorithmes et résultats

2.2.1 Algorithmes

Comme indiqué précédemment, l'un des inconvénients du classifieur Wishart est le nombre de classes fixé à l'origine à huit à partir de la décomposition de Cloude-Pottier. L'ajout d'un autre paramètre physique extrait de la décomposition en valeurs propres de la matrice de covariance, appelé l'anisotropie, permet de passer ce nombre de classes à 16 [63]. Il est aussi possible d'inclure le span, c'est-à-dire la puissance (voir Eq. (1.31), afin d'assurer une meilleure convergence [45]. Le nombre de classes peut également être modifié *a posteriori* en comparant les dispersions intra-classes et inter-classes [50] et en utilisant des décisions de type *split-and-merge* afin d'obtenir un nombre optimal de classes. Cependant, toutes ces techniques ont pour point de départ la décomposition de Cloude-Pottier à huit classes (seize si l'on inclue l'anisotropie) ou la décomposition de Freeman-Durden dans le cas de [52]. Il est donc intéressant de pouvoir utiliser un algorithme qui n'ait pas ce nombre de classes prédéterminé.

Par ailleurs, le test de Box fournit un moyen de rejeter, par construction, les pixels trop différents des classes déjà connues. D'où l'idée d'employer un algorithme qui augmente le nombre de classes à chaque itération en utilisant les pixels rejetés comme nouveau point de départ. L'algorithme utilisé est présenté dans la Fig. 2.1.

La mise en œuvre du test de Box est très directe au sein de l'algorithme : pour chaque pixel de l'image, on calcule la distance par rapport à chacun des centres de classe en utilisant l'Eq. (2.40) et on prend le minimum de cette distance. La différence par rapport au classifieur Wishart se situe dans l'étape suivante, c'est-à-dire dans la prise en compte de la région critique. Si ce minimum ne remplit pas le critère de seuil donné par la distribution du χ^2 alors on le place au sein de la classe de rejet. S'il le remplit, on le place au sein de la classe correspondante. On incrémente alors le nombre de classes en utilisant les pixels de la classe de rejet pour calculer une matrice de covariance qui sera le centre de la classe nouvellement créée.

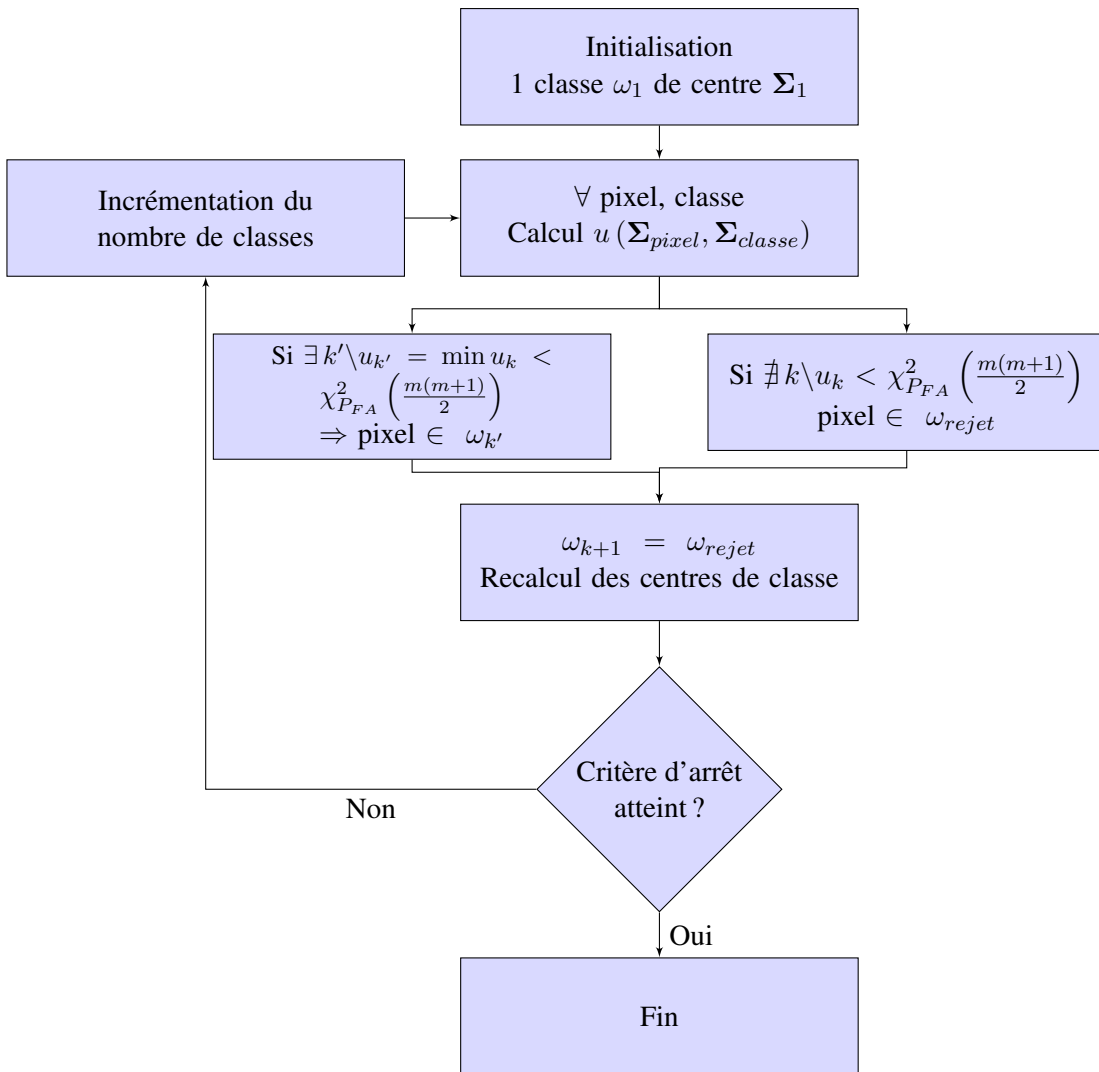
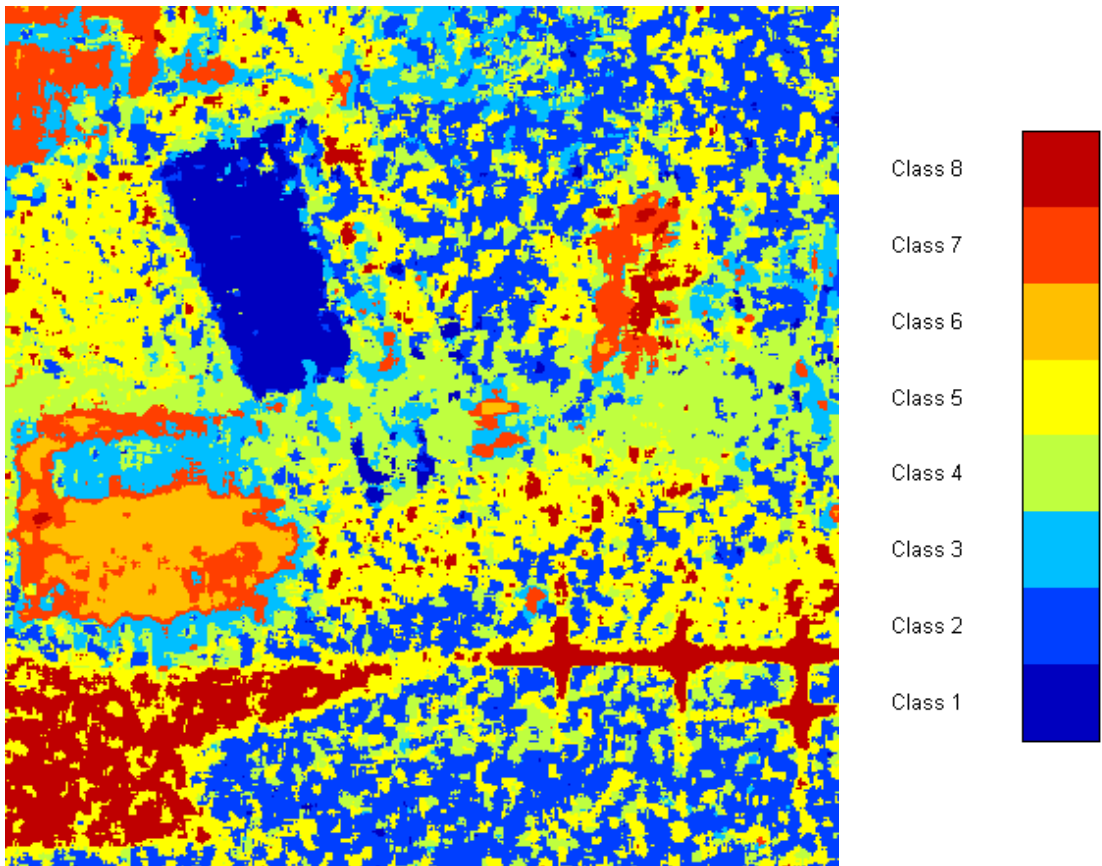


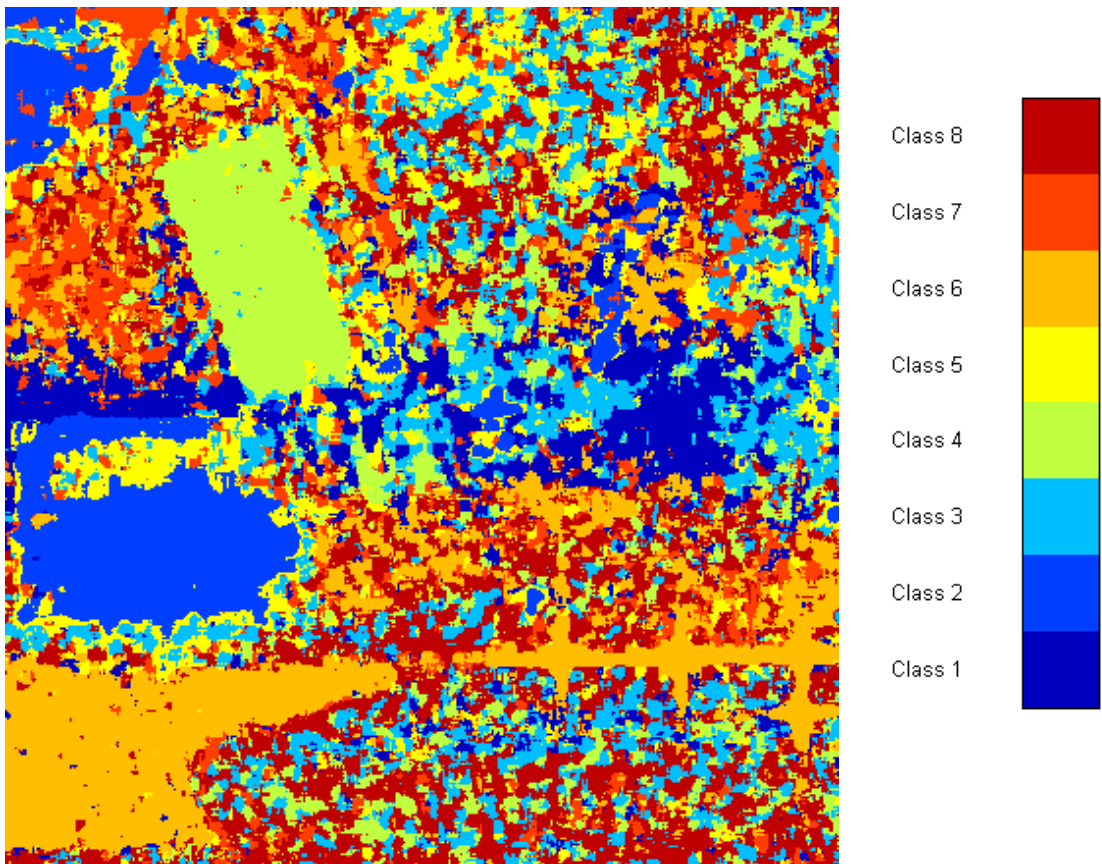
FIGURE 2.1 – Algorithme du classifieur Box

Le point le plus critique est l’initialisation puisque l’algorithme se construit à partir de la classe de rejet. Toute la construction risque donc d’être dépendante du choix de la première classe. Des essais préliminaires ont montré, comme on pouvait s’y attendre, qu’utiliser un unique pixel pour l’initialisation n’est pas acceptable car cela mène à de grandes disparités dans les résultats de classification finaux. Il est alors nécessaire de choisir une classe initiale qui permettra de construire la classification de manière à converger vers la même solution à chaque exécution de l’algorithme. Une manière de choisir cette classe est d’utiliser la décomposition de Cloude-Pottier de la section 1.2.3.2 afin d’avoir une classe initiale donnée par des considérations polarimétriques. Puisque la décomposition de Cloude-Pottier fournit huit classes, il suffit de choisir la classe la plus grande pour avoir une initialisation convenable.

Cependant, il s’est avéré inutile d’employer cette décomposition pour l’initialisation. En effet, même pour le classifieur Wishart, l’initialisation par la décomposition de Cloude-Pottier ne s’avère pas nécessaire. Une initialisation totalement aléatoire où chaque pixel est affecté à une classe parmi huit selon une loi uniforme donne quasiment les mêmes résultats de classification. La Fig. 2.2 présente les différences dans les résultats de classification d’après le classifieur Wishart utilisant le FPE selon l’initialisation. Il apparaît alors de manière évidente que les différences entre les deux classifications sont minimales, puisqu’aucun *a priori* n’est associé au



(a) Initialisation par la décomposition de Cloude-Pottier



(b) Initialisation aléatoire

FIGURE 2.2 – Classifieur Wishart avec le FPE

choix des couleurs, qui diffèrent entre les deux images.

A noter que les ddl des estimateurs de matrice de covariance dépendent du nombre de données utilisées pour l'estimation. Sauf mention contraire, les estimations sont faites dans la suite par une fenêtre glissante de taille 7×7 , de manière à avoir un nombre de données suffisant pour l'estimation et considérer une certaine homogénéité statistique au sein de la fenêtre.

2.2.2 Résultats

L'algorithme de la Section 2.2.1 a donc été appliqué sur l'image SAR de Brétigny de la Fig. 1.12(a). Pour rappel, cette image a été acquise en bande X avec une résolution en distance et en azimuth d'environ 1.3m. Ses dimensions sont de 501×501 pixels soit 251001 pixels. La mise en place de l'algorithme a nécessité quelques ajustements à cause du volume de données à considérer. En effet, pour chaque pixel et à chaque itération, il est nécessaire de calculer la matrice de covariance donc il est préférable de les calculer à l'avance et de les stocker pour éviter des calculs fastidieux à chaque opération. Pour le calcul du test proprement dit, des problèmes de stockage sont aussi apparus sur une machine standard (QuadCore, 8 Go de RAM). Le code a donc été vectorisé, ce qui a l'avantage de le rendre plus rapide et d'effectuer tous les calculs en une fois, ce qui élimine les variables de stockage intermédiaire, excepté les centres de classe qui sont recalculés à chaque itération.

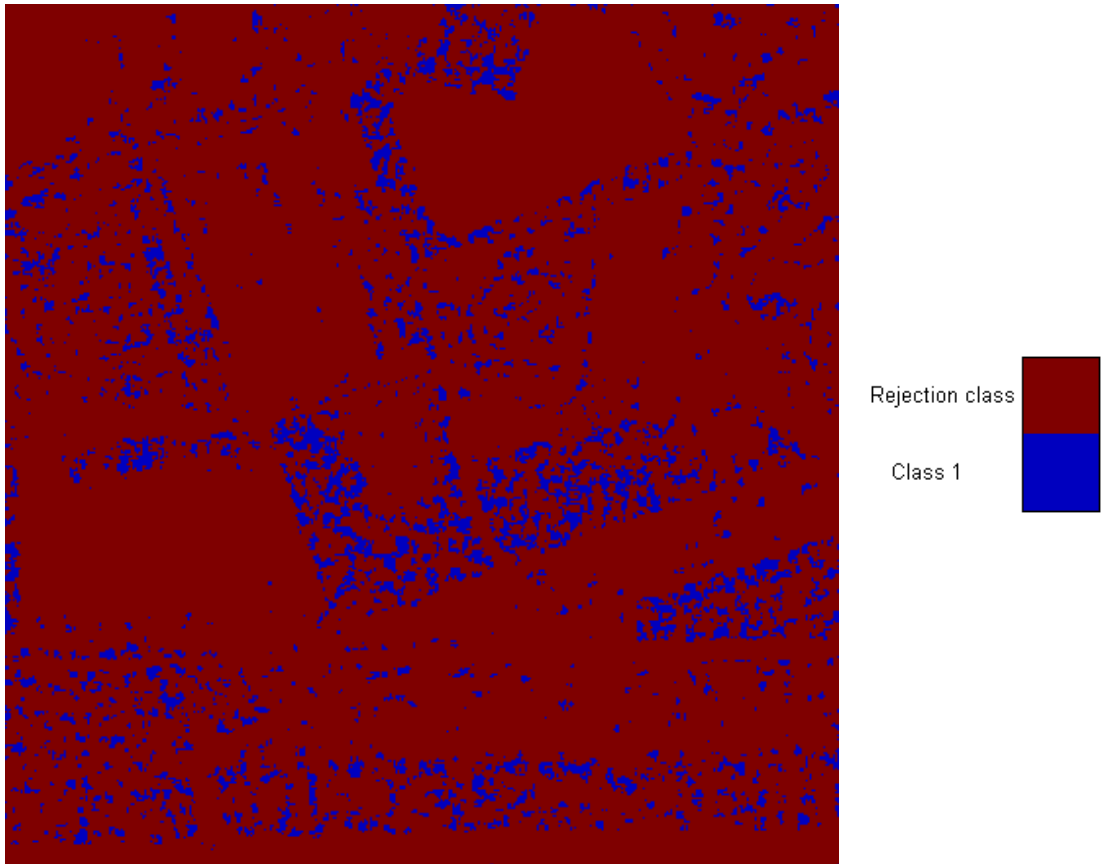
Les résultats de classification sont présentés après une itération et huit itérations, afin d'avoir un nombre de classes similaire au classifieur Wishart et de pouvoir faire des comparaisons. Les matrices de covariance sont estimées en utilisant la SCM ou le FPE avec les degrés de liberté correspondants.

2.2.2.1 Classification avec la SCM

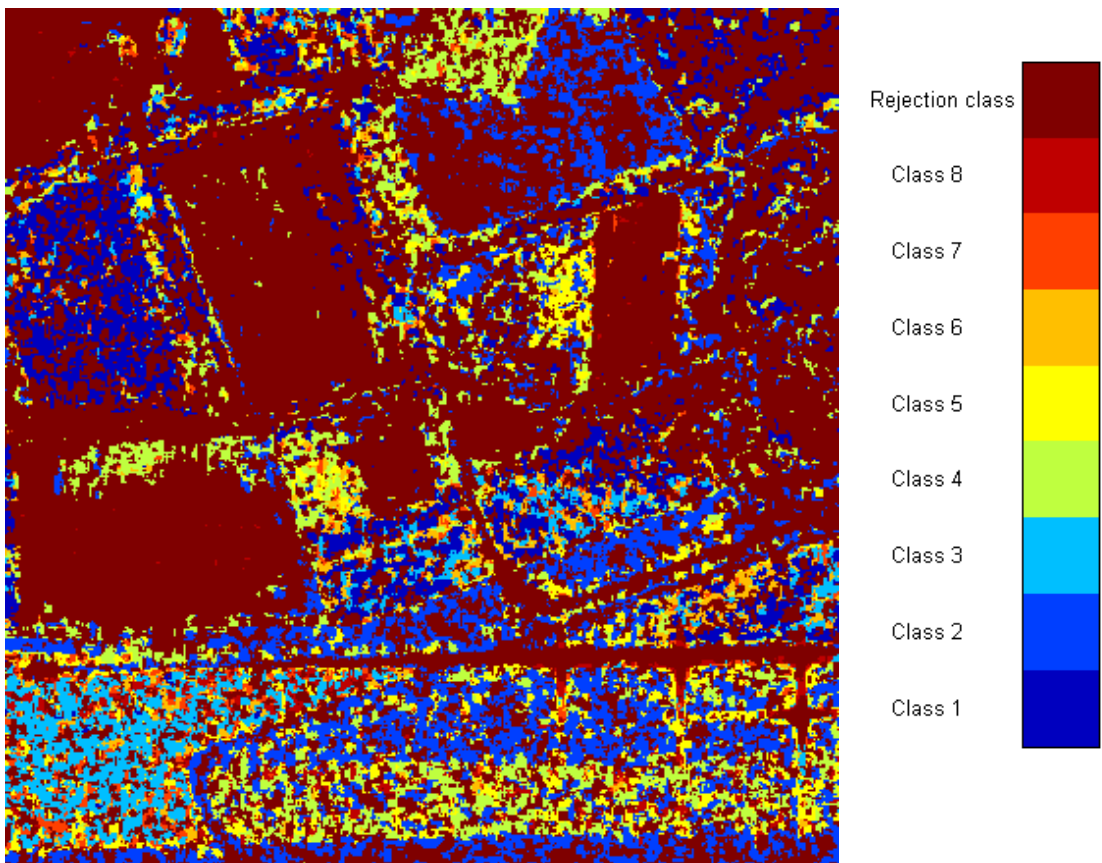
On se place ici sous hypothèse gaussienne. Les résultats au bout de une et huit itérations sont donnés sur la Fig. 2.3.

Au bout d'une itération, la plupart des pixels de l'image se retrouvent dans la classe de rejet et il est impossible de distinguer clairement un des éléments caractéristiques de l'image. Si l'on compare les résultats à ceux de la décomposition de Cloude-Pottier de la Fig. 1.15, les pixels de la première classe appartiennent à la zone jaune qui correspond à une surface aléatoire à entropie moyenne. Ces pixels appartenant en réalité à des champs, la classification semble être correcte.

Après huit itérations, il y a toujours de nombreux pixels rejetés dans les zones de champs, surtout dans le champ situé dans le haut de l'image. Cependant, la majeure partie des pixels appartenant à la classe de rejet appartiennent à deux types de structures : les zones urbaines (bâtiments, parking, coins réflecteurs, ...) et à la zone boisée sur la partie droite de l'image. Il est même possible d'identifier des petits bâtiments au centre de l'image, ce qui n'était pas possible avec le classifieur Wishart-FPE. Sur les zones gaussiennes, la classification par rapport au classifieur Wishart-FPE n'est pas réellement améliorée, on aperçoit toujours un mélange de classes, quoique moins poussé que sur la Fig. 1.16(b). On remarque toutefois deux classes qui semblent un peu plus cohérentes : la classe 3 en bleu ciel qui compose une grande partie du champ situé dans le coin inférieur gauche et la classe 5 en jaune qui est présente sur une large bande dans le champ situé dans le coin inférieur droit de l'image ainsi qu'entre les deux bâtiments au centre de l'image, à côté du parking et du champ situé dans la partie supérieure. Cette classe semble correspondre à un type de champ spécifique mais sans vérité terrain, il est difficile d'en tirer plus de conclusions.



(a) 1 itération



(b) 8 itérations

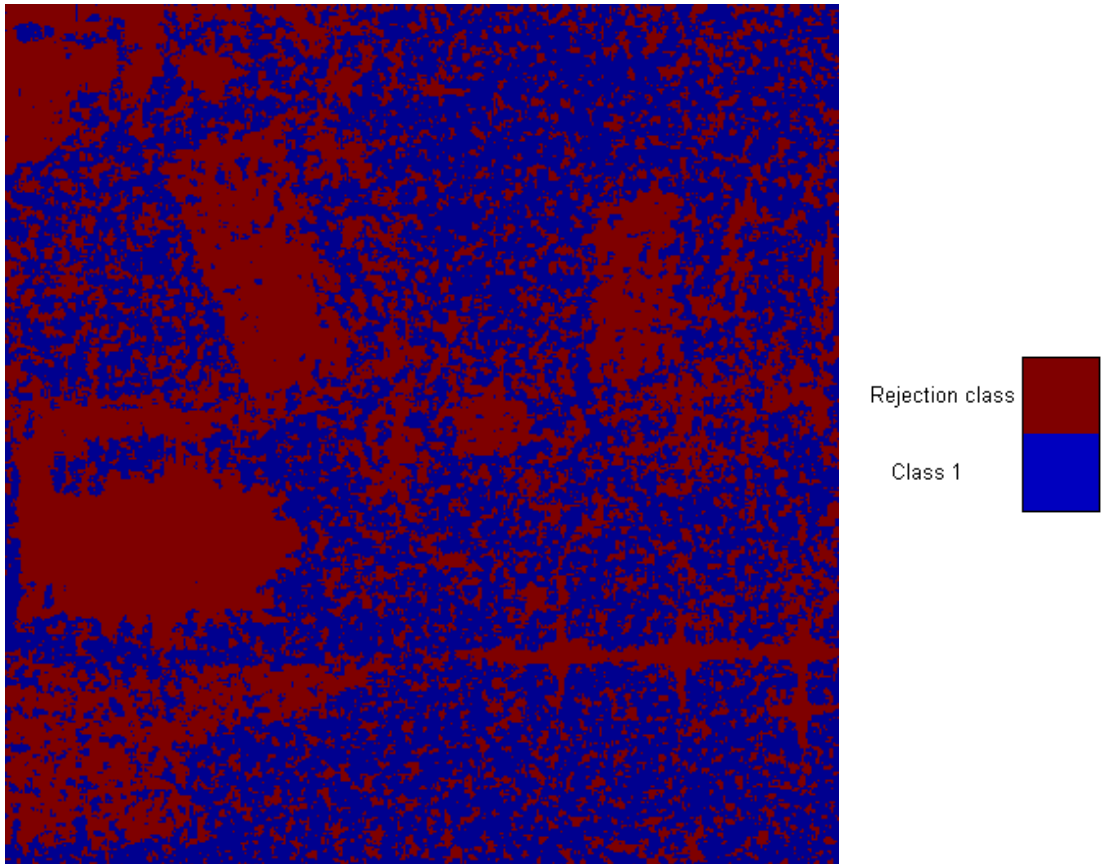
FIGURE 2.3 – Résultats de classification avec la SCM et carte associée.

2.2.2.2 Classification avec le FPE

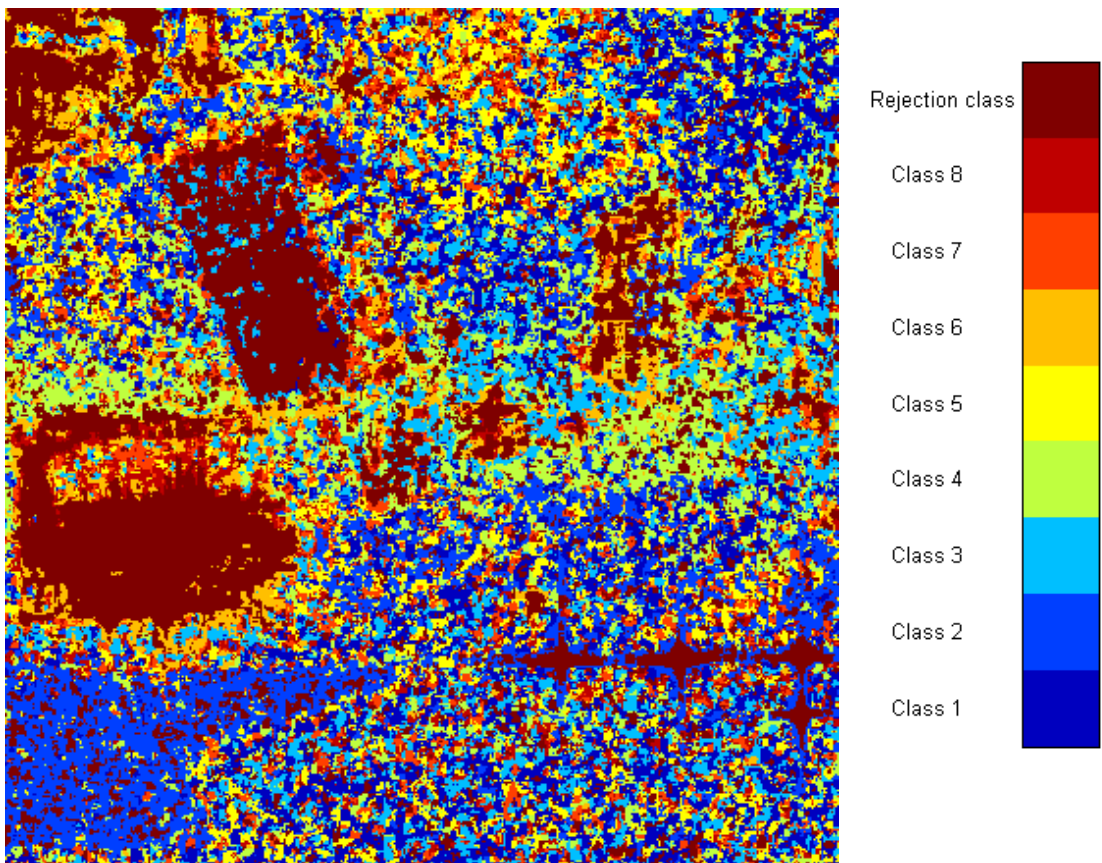
On se place ici sous hypothèse SIRV. Les résultats au bout de une et huit itérations sont donnés sur la Fig. 2.4.

Pour le FPE, les résultats de classification au bout d'une itération sont très encourageants. En effet, le nombre de pixels rejetés est très inférieur au cas de la SCM et les structures artificielles de la zone sont beaucoup plus visibles. Le nombre réduit de pixels rejetés est une indication d'une réduction de la fausse alarme (puisque le seuil de rejet est directement conditionné par le taux de fausse alarme réglé par l'utilisateur), ce qui montre que le FPE est mieux adapté que la SCM à un fouillis hétérogène. Au bout de huit itérations, les résultats sont assez proches de ceux obtenus avec la SCM en ce qui concerne les structures artificielles. Pour le reste de l'image, on retrouve un effet de *patching*, mais avec des patchs beaucoup plus petit qu'en utilisant la SCM. Ceci peut s'expliquer par le fait que la classification avec le FPE ne se fait pas sur la puissance : comme on estime la matrice de covariance avec les pixels voisins du pixel considéré, il se peut qu'il y ait des points très brillants qui vont contaminer les résultats de classification dans une zone assez large dans le cas de la SCM. Dans le cas du FPE, ces pixels n'ont que peu d'influence sur les résultats de classification, ce qui amène à des patchs plus petits.

Au final, l'augmentation du nombre de classes dans cet algorithme n'a que peu d'intérêt pour le cas du FPE puisqu'il n'y a quasiment pas de gain d'information en itérant de cette manière. Au bout d'une itération, on a déjà une bonne séparation entre zones "naturelles" et zones "artificielles" qui sont séparées entre classe 1 et classe de rejet.



(a) 1 itération



(b) 8 itérations

FIGURE 2.4 – Résultats de classification avec le FPE et carte associée.

2.3 Implémentation du test de Box au sein d'autres algorithmes

Étant donnés les résultats de l'algorithme du M-test de Box, à savoir bonne séparation entre zones naturelles et artificielles avec le FPE pour une classe mais peu d'améliorations en augmentant le nombre de classes, de nouvelles approches sont envisageables pour l'implémentation de ce test au sein d'algorithmes de classification. Deux voies sont alors possibles :

- améliorer l'initialisation : nous avons vu qu'une initialisation par décomposition de Cloude-Pottier ou une initialisation aléatoire donnaient des résultats sensiblement identiques. Dans [70], Trouvé a développé un algorithme de segmentation qui permet de présegmenter une image SAR polarimétrique selon un certain nombre de critères statistiques, graphiques et polarimétriques. Il est alors possible d'utiliser le résultat de cette segmentation pour initialiser l'algorithme du M-test de Box et utiliser les segments au lieu de simples pixels pour effectuer la classification.
- utiliser la statistique de test comme une distance entre matrices et l'employer dans un algorithme éprouvé : on peut remplacer la distance de Wishart dans le classifieur Wishart par la statistique du test de l'Eq. (2.40) en intégrant la classe de rejet donnée par la région critique du test (cf. Eq. (2.41)). On se retrouve alors avec un classifieur de type K-moyennes, qu'on appelle le classifieur Box qui permet de s'affranchir de l'attribution automatique des pixels à une des classes connues, grâce à la classe de rejet.

Ces deux approches ont été envisagées et sont présentées dans les sections suivantes. On verra que ces deux approches se rejoignent du fait de la réduction du nombre de données dans la première approche. La distinction principale se fera alors en fonction de l'application d'une segmentation initiale.

2.3.1 Application avec une segmentation initiale

2.3.1.1 Algorithme de segmentation

Pour pallier les problèmes de représentation des paramètres polarimétriques en milieu urbain, Nicolas Trouvé a eu l'occasion de développer un algorithme de croissance de segmentation pendant son doctorat à l'ONERA [70]. L'objectif initial de cet algorithme était simplement de faire une première segmentation pour améliorer l'estimation de paramètres polarimétriques dans les zones urbaines puis il a été grandement amélioré, au vu des premiers résultats encourageants. On rappelle ici rapidement les principaux éléments de cet algorithme de croissance de régions, résumé brièvement sur la Fig. 2.5.

L'image est initialement partitionnée en autant de segments qu'elle comporte de pixels. La distance entre chacun des segments est calculée suivant une distance statistique puis les deux régions les plus proches au sens de cette distance statistique sont fusionnées. De nombreux critères liés au traitement d'images sont pris en compte afin d'avoir une segmentation qui ait un aspect visuel cohérent avec l'image SAR de puissance. Ces critères graphiques fournissent un terme de pondération pour la distance statistique afin d'influencer la décision de fusionner ou non les deux régions. Le lecteur est fortement invité à consulter le chapitre 2 de [70] afin d'avoir de plus amples détails sur la mise en place de cet algorithme de segmentation. La Fig. 2.6(b) montre les résultats de la segmentation appliquée à l'image de la Fig. 2.6(a) acquise à Toulouse par le système SETHI de l'ONERA.

On peut remarquer une réduction très importante du speckle et la conservation des principales structures de l'image avec des frontières très nettes sur les bâtiments. Un autre avantage de cette segmentation est la réduction du nombre de données puisqu'on va considérer les segments de l'image au lieu des pixels.

2.3.1.2 Application

L'algorithme de croissance de régions a donc été appliqué sur l'image de Brétigny afin d'obtenir une segmentation appropriée pour servir de point de départ pour une classification. Plusieurs critères sont possibles

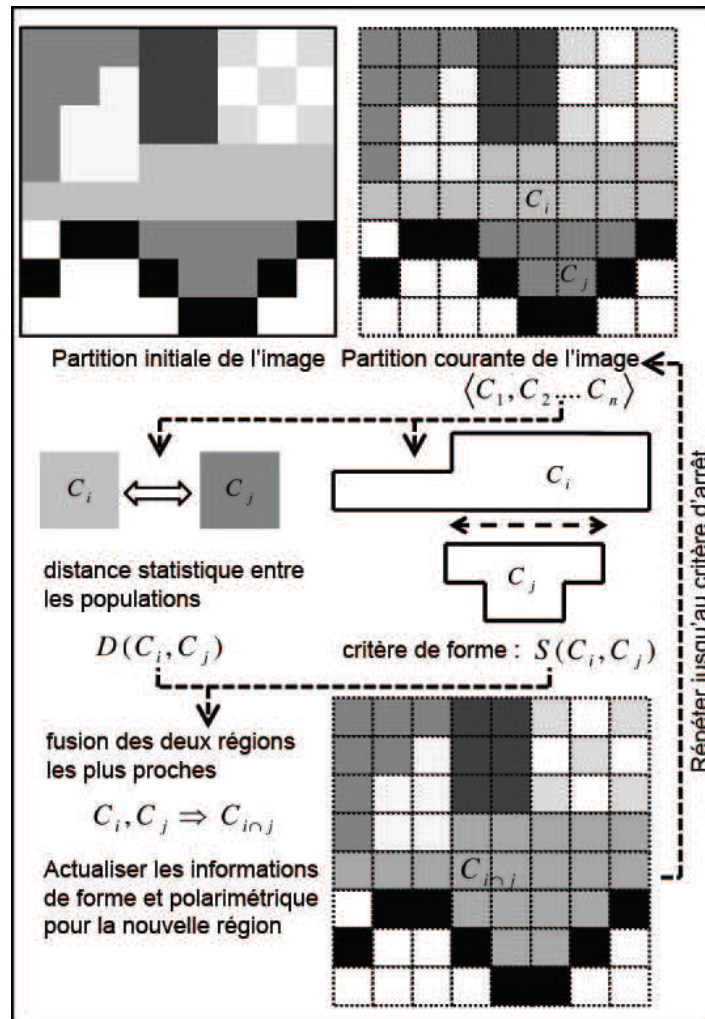


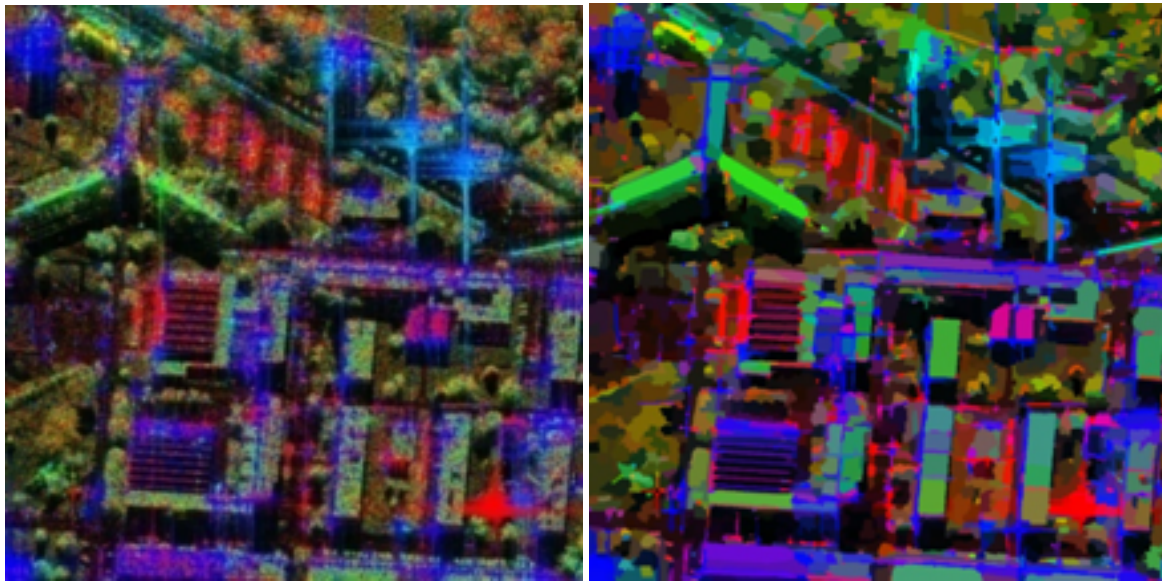
FIGURE 2.5 – Algorithme de segmentation par croissance de régions.

pour arrêter l'algorithme de croissance de régions, nous avons choisi d'avoir 1000 segments initiaux. L'image résultante est présentée sur la Fig. 2.7 dans la base de Pauli.

Comme indiqué dans la section précédente, le nombre de données est réduit en utilisant cette initialisation car on utilise les segments au lieu des pixels comme données. Ainsi, on passe d'une image composée de plus de 250000 pixels à une image composée de 1000 segments. Cependant, une application directe de l'algorithme de la Fig. 2.1 n'est pas optimale pour plusieurs raisons :

- étant donné le nombre réduit de données, il n'est pas aisé de choisir un grand nombre de ces données pour l'initialisation,
- choisir un nombre de données plus réduit pour l'initialisation peut mener à des divergences dans les résultats de classification,
- ces données sont de plus déjà segmentées : les choisir aléatoirement peut nuire à la classification si on regroupe initialement des régions avec des mécanismes polarimétriques différents.

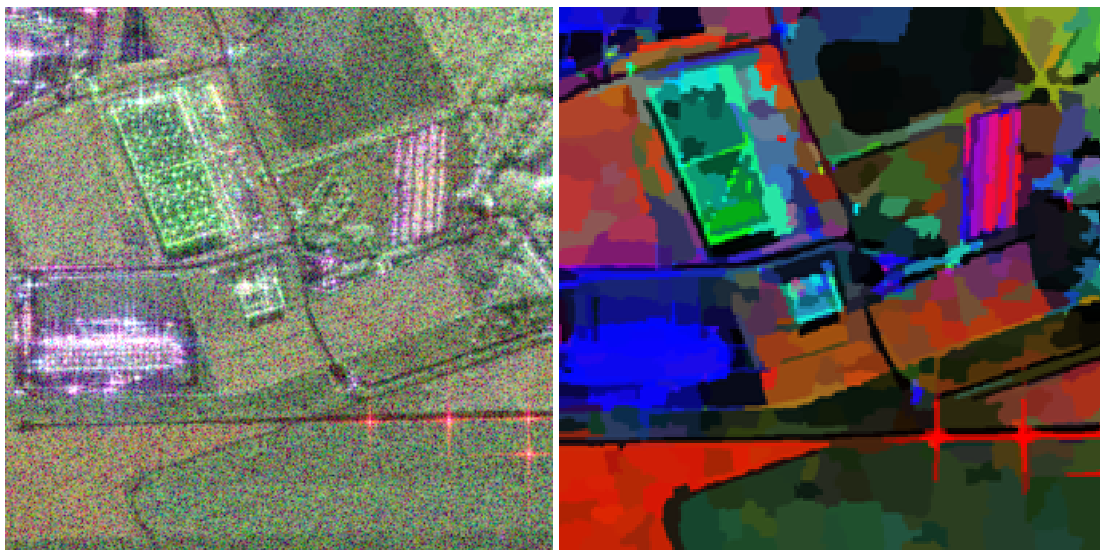
Pour toutes ces raisons, un autre algorithme est considéré, proche du K-moyennes : le K-moyennes++ [3]. On se rapproche alors de la deuxième solution envisagée, à savoir utiliser un algorithme éprouvé en utilisant le test de Box comme une distance. La différence entre les deux approches se situe donc au niveau de l'initiali-



(a) Image originale

(b) Image segmentée

FIGURE 2.6 – Jeu de données : Toulouse



(a) Image originale

(b) Image segmentée

FIGURE 2.7 – Segmentation sur l'image de Brétigny

sation : ici, on utilise une segmentation pour initialiser l'algorithme et on travaille sur les segments résultants de cette segmentation. Dans la section suivante, on n'utilisera pas de segmentation initiale et l'algorithme sera appliqué pixel par pixel.

Le K-moyennes++ diffère du K-moyennes uniquement par l'étape d'initialisation : dans le K-moyennes, les centres de classes sont choisis de manière aléatoire parmi les données. Pour le K-moyennes++, la procédure d'initialisation est la suivante :

1. le premier centre de classe est choisi au hasard parmi les données
2. la distance entre chacune des données et le centre le plus proche est calculée grâce à une certaine distance

D

3. le centre suivant est choisi parmi les données avec une probabilité uniforme, pondérée par le carré de D
4. les étapes 2 et 3 sont répétées jusqu'à avoir le nombre de classes désiré

L'avantage de cette procédure est d'avoir une initialisation qui améliore les résultats de classification par rapport à un K-moyennes classique en termes de vitesse de convergence et de réduction d'erreurs de classification. On utilise le test de Box de l'Eq. (2.40) comme une distance à la fois pour la distance D et la distance utilisée lors des itérations de l'algorithme (cf. étape 3 de la Fig. 1.11). On choisit ici aussi d'avoir huit classes, comme dans la section précédente.

Cependant, un autre phénomène est à prendre en compte avec cette segmentation initiale. Lors de la première implémentation du test de Box, les degrés de liberté dépendent du nombre de données utilisées pour réaliser l'estimation. Si ce nombre de données est faible, la valeur donnée par l'Eq. (2.40) est faible et donc souvent inférieure au seuil donné par l'Eq. (2.41), ce qui indique qu'on accepte l'hypothèse que les matrices de covariance sont égales. Au contraire, si le nombre de données est élevé, l'Eq. (2.40) donne des valeurs élevées et surpasse souvent le seuil de la région critique donc on rejette l'hypothèse d'égalité des matrices de covariance. Ce comportement est logique puisque si le nombre de données est faible, l'estimation obtenue n'est pas très bonne donc le test va être très tolérant sur des différences entre les deux matrices comparées. Mais si le nombre de données est élevé, cela veut dire qu'on a, *a priori*, une bonne estimation de la matrice de covariance donc de petites variations seront beaucoup plus significatives et pourront entraîner un rejet du test. Comme la taille des échantillons est relativement faible (carré de 7×7), le test est assez tolérant sur des différences entre les matrices. L'algorithme est ici stoppé quand la convergence est atteinte, c'est-à-dire quand les classes ne sont plus modifiées entre deux itérations.

Dans le cas présent, on ne compare plus des pixels mais des segments, qui ont des tailles parfois beaucoup plus importantes que les voisinages 7×7 des pixels. On a alors un test très intransigeant sur les différences entre matrices, puisqu'il va prendre des valeurs très élevées alors que le seuil n'est pas modifié, et qui rejette très souvent l'hypothèse nulle, à tel point qu'il est impossible d'en tirer des conclusions. La Fig. 2.8 montre des résultats de classification en utilisant le nombre de degrés de liberté théorique. Comme on peut le remarquer, la quasi-totalité des segments est rejetée, excepté la plupart des classes initiales et une partie des segments qui leur sont extrêmement similaires.

On peut alors considérer les ν_i comme des paramètres ajustables par l'utilisateur en fonction du jeu de données et du raffinement de la segmentation initiale afin de pallier ce problème de rejet trop important. On s'éloigne alors de la considération initiale d'un test statistique rigoureux puisque la valeur des exposants est modifiée, non pas pour coller plus précisément à une certaine distribution, mais afin d'avoir des résultats de classification exploitables. La Fig. 2.9 présente les résultats de classification obtenus pour une valeur de $\nu_i = 50$ avec la SCM et le FPE pour une même initialisation. On retrouve des comportements similaires à ceux de la Fig. 2.3 pour la SCM, à savoir une bonne séparation entre zones homogènes et zones hétérogènes, qui sont dans la classe de rejet. Cependant, la classification dans les zones homogènes est grandement améliorée par la segmentation initiale et on distingue clairement les contours de la plupart des champs. Pour le FPE, les résultats sont également assez similaires à ceux de la Fig. 2.4 avec beaucoup moins de rejets dans les zones hétérogènes. Un des grands bâtiments est clairement défini (classe 3) et on a de même une amélioration de la classification au sein des zones homogènes, au sens où les contours de certains champs sont beaucoup plus visibles. Néanmoins, on peut remarquer une bande au milieu de l'image qui est incorrectement classifiée au sein d'une seule classe (classe 5) et qu'on retrouve à chaque exécution de l'algorithme. La présence de cette bande est inexpliquée, les segments la composant ne partageant *a priori* pas de caractéristiques polarimétriques similaires.

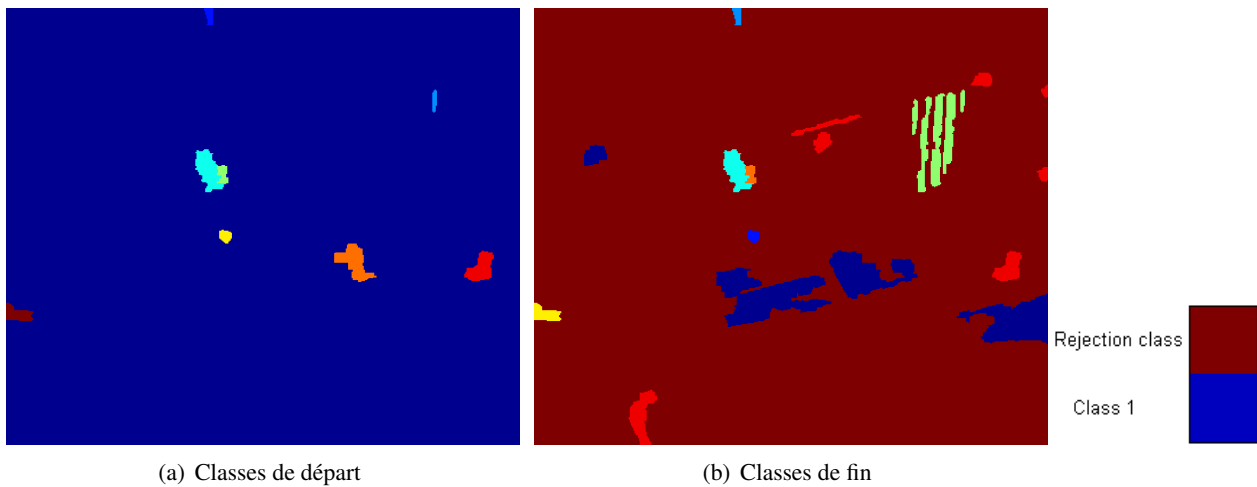


FIGURE 2.8 – Classification avec initialisation par croissance de régions : degrés de liberté théoriques

L'algorithme K-moyennes++ a aussi été appliqué sur un jeu de données beaucoup plus grand, qui correspond au jeu de données dont est issue l'image de la Fig. 1.12(a). L'image est de taille 3392 par 1533 pixels, ce qui rend indispensable l'application de la segmentation initiale pour pouvoir la traiter. En effet, les différents algorithmes utilisés font tous appel à la matrice de covariance du pixel. Pour avoir des traitements rapides, on calcule évidemment les matrices de covariance initialement puis on les stocke afin de ne pas avoir à les recalculer à chaque itération. Or, le nombre de pixels étant très élevé, on rencontre très rapidement des problèmes de mémoire sur des machines non-serveurs. Une autre solution est de recalculer les matrices dès que nécessaire mais le temps de calcul devient beaucoup trop grand. La segmentation initiale permet donc de passer d'une image composée de plus de cinq millions de pixels (Fig. 2.10(a)) à une image composée d'un nombre réduit de segments, par exemple 2500 sur la Fig. 2.10(b). Un rectangle blanc délimite le plus petit jeu de données utilisé précédemment. A noter que l'image de la Fig. 2.10(a) a été filtrée par filtrage de Lee [49] afin d'éliminer la majeure partie du speckle pour la visualisation. Ce filtrage n'a pas été appliqué sur les données pour la classification. En effet, l'estimation de la matrice de covariance effectuée déjà un filtrage par une fenêtre glissante. Il est vrai que ce filtrage est moins performant que le filtrage de Lee du point de vue du traitement d'images mais son apport pour la classification en utilisant le FPE ne s'est pas révélé probant sur des essais préliminaires.

Le nombre de 2500 segments a été choisi de manière relativement arbitraire : il permet d'avoir une segmentation relativement fine et d'avoir des traitements très rapides. Un nombre inférieur de segments occultait certains détails, principalement dans des zones urbaines et un nombre supérieur de segments ne permettait pas réellement une segmentation plus raffinée. Tout ceci est évidemment sujet à interprétation et dépend du jeu de données considéré. De même, le nombre de classes a été fixé arbitrairement à seize, avec les mêmes considérations que pour le nombre de segments : un trop petit nombre de classes rendait la classification trop grossière et un nombre trop élevé de classes rajoutait de la confusion sans réellement améliorer les résultats de classification. On peut néanmoins utiliser un nombre de classes compris entre 15 et 30 sur ce jeu de données sans réel argument pour un nombre spécifique par rapport aux autres.

Les résultats de classification au bout de dix itérations en utilisant le FPE sont présentés sur la Fig. 2.11. La Fig. 2.11(a) représente les résultats de classification bruts et la Fig. 2.11(b) les résultats de classification dans un jeu de couleurs créé de telle manière à se rapprocher d'une représentation dans la base de Pauli avec ici les surfaces en rouge, le double rebond en bleu et le volume toujours en vert. Les bandes supérieures et inférieures de l'image ne sont pas utiles pour la classification d'une manière générale car elles ne contiennent que du

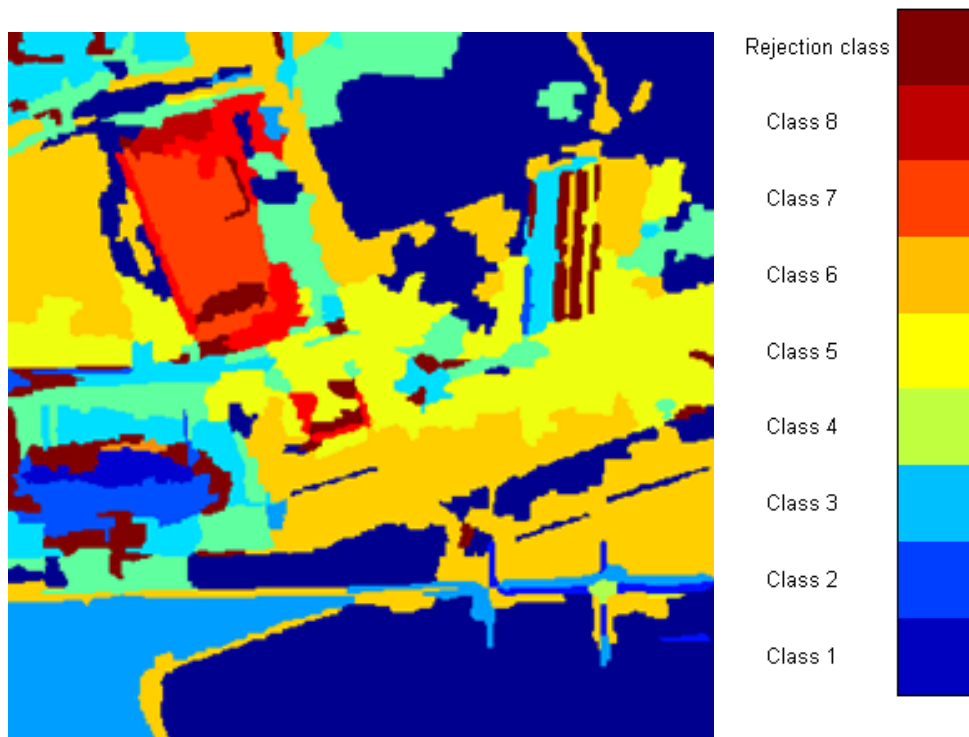
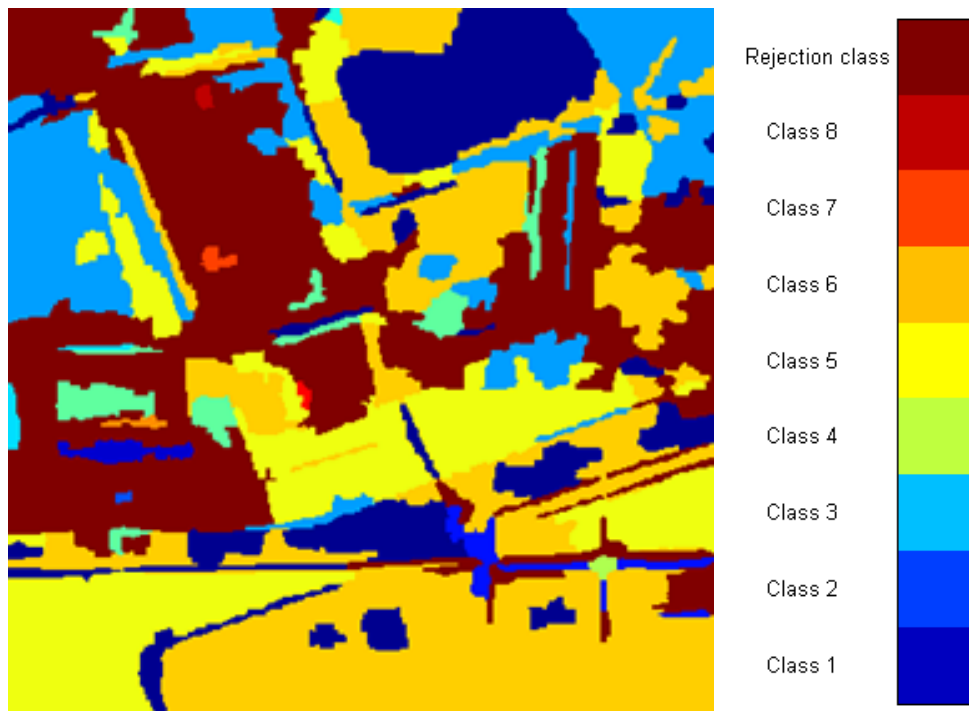
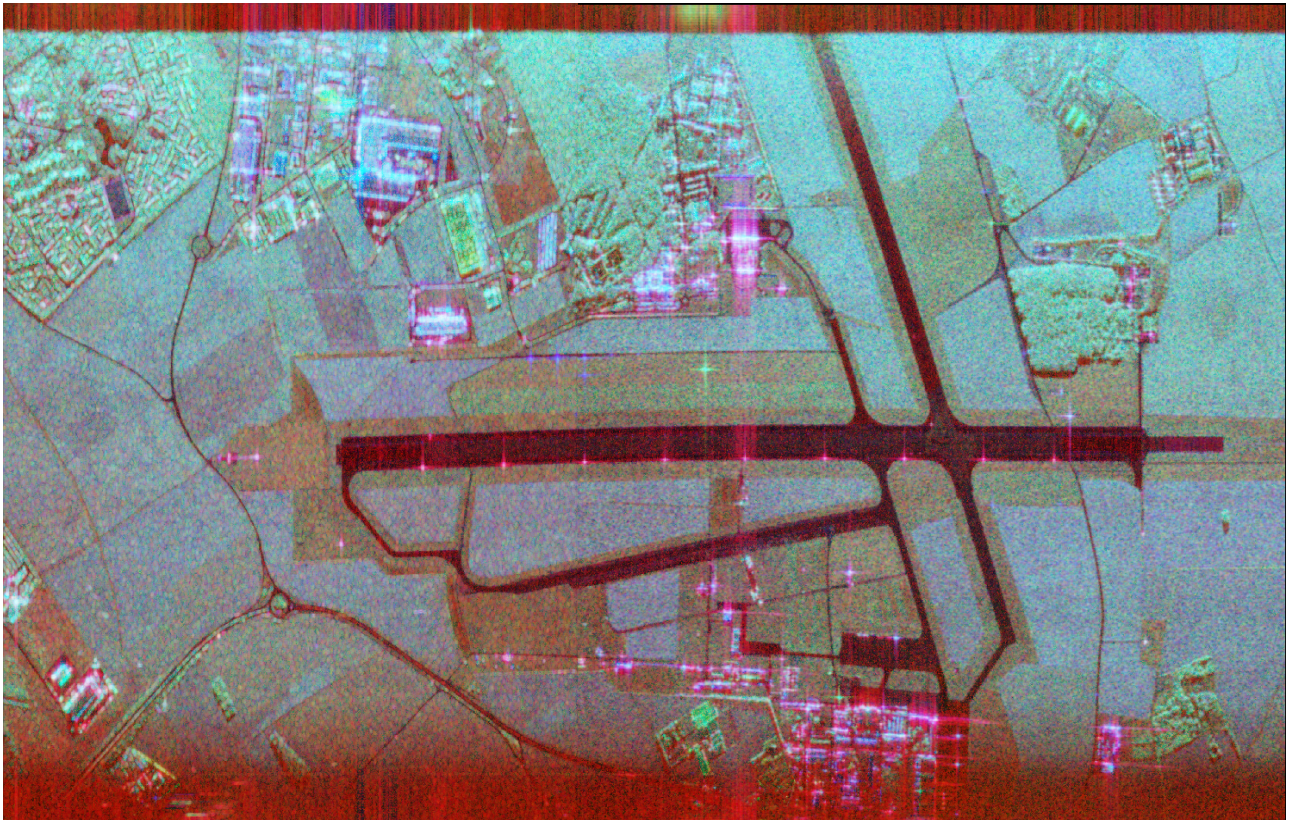


FIGURE 2.9 – Classification avec initialisation par croissance de régions : comparaison entre la SCM et le FPE

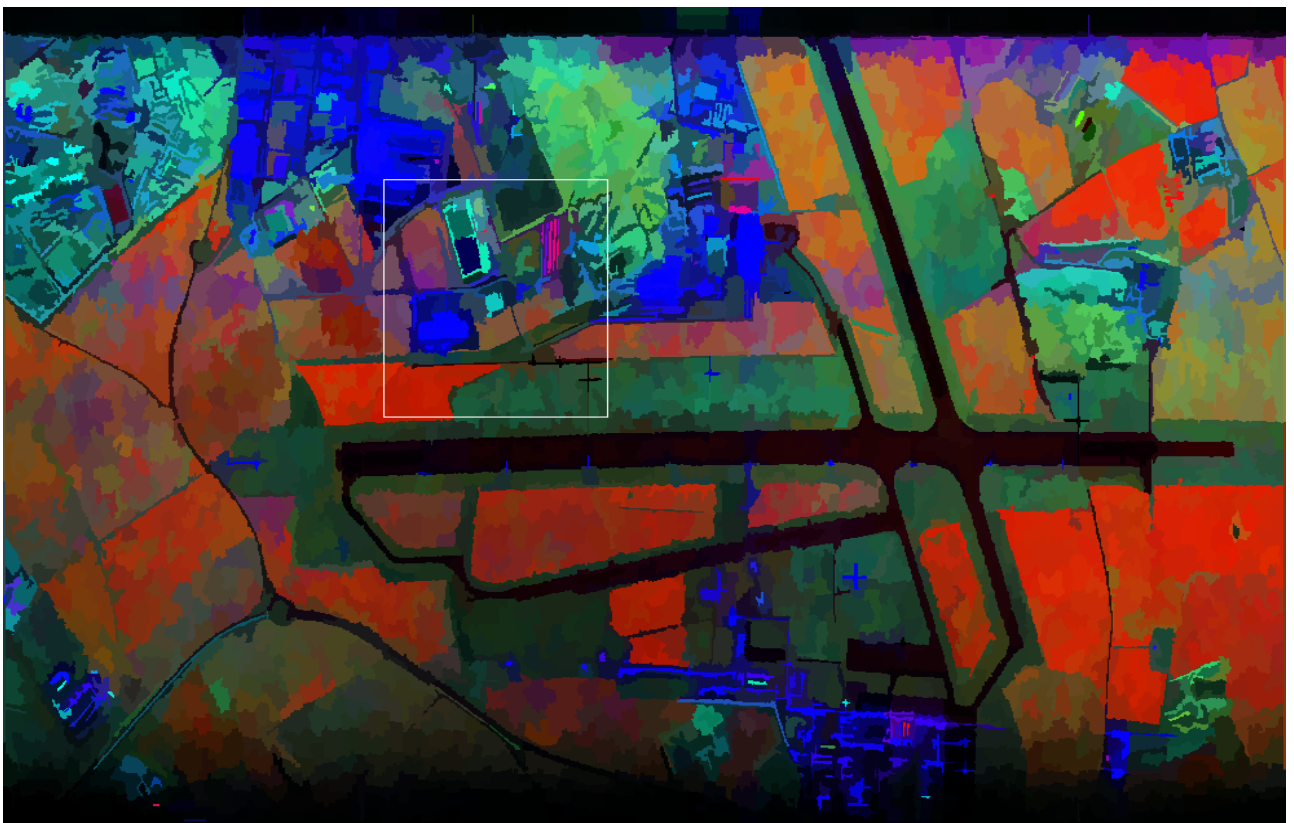
Chapitre 2. Classification non supervisée

bruit mais on peut voir qu'elles sont néanmoins relativement bien séparées du reste de l'image. Les champs de grande taille sont aussi correctement classifiés, bien que dans deux classes différentes : vert clair et orange. La classe bleue foncée représente principalement les zones boisées ainsi que certains champs, ce qui peut sembler erroné au premier abord. Cependant, si l'on observe l'image colorée suivant la décomposition de Pauli (Fig. 2.11(b) ou Fig. 2.10(a)), ces champs semblent avoir une composante de volume plus importante puisqu'ils ressortent principalement en vert. La classification est donc correcte pour cette classe. Dans les zones urbaines, on retrouve une multitude de classes auxquelles il est difficile d'attribuer une signification spécifique. La classe de rejet, en marron, est située principalement sur une bande horizontale, légèrement au dessus du milieu de l'image et regroupe différents types de structures de l'image : bâtiments, champs, forêts, ... On y retrouve notamment les deux grands bâtiments présents dans la sous-image étudiée précédemment. La séparation entre zones homogènes et zones hétérogènes par la classe de rejet qu'on avait pu observer dans les Fig. 2.3, 2.4 et 2.9 n'apparaît plus ici.

2.3. Implémentation du test de Box au sein d'autres algorithmes

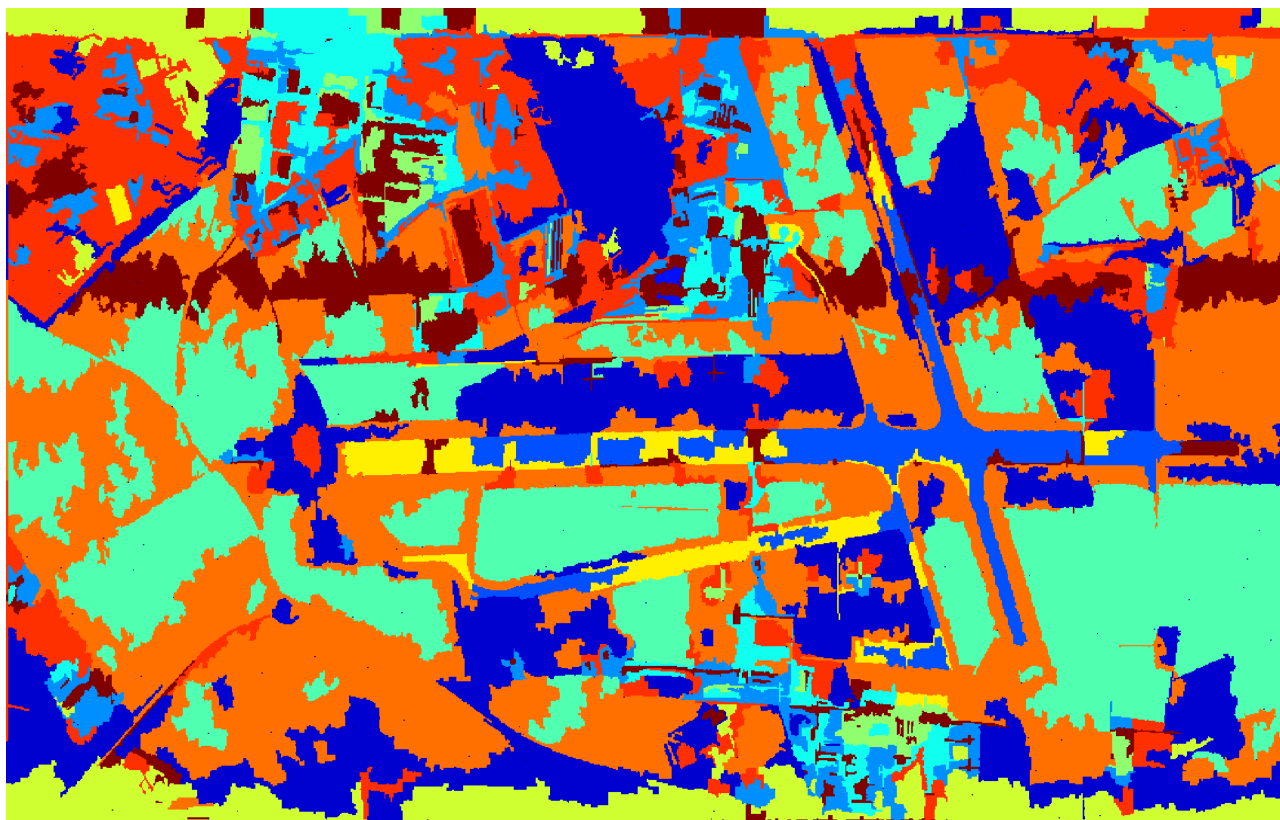


(a) Jeu de données dans la base de Pauli, filtre de Lee 7x7

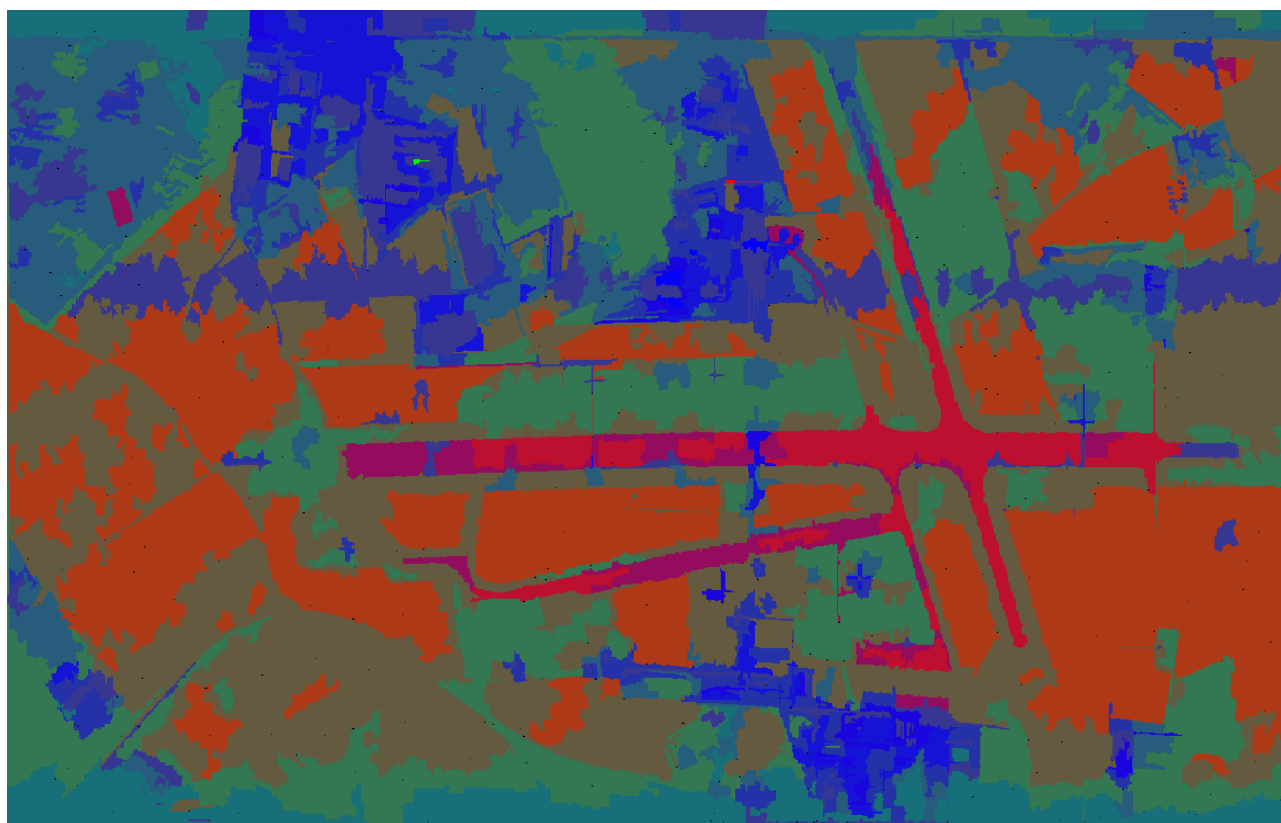


(b) Jeu de données segmenté

FIGURE 2.10 – Grand jeu de données de Brétigny : image originale et image segmentée.



(a) Classification finale



(b) Classification finale dans la base de Pauli

FIGURE 2.11 – Classification avec initialisation par croissance de régions sur un grand jeu de données

2.3.2 Application sans segmentation initiale

Comme on a pu l'observer précédemment, il est possible d'utiliser l'Eq. (2.40) comme une distance au sein d'un algorithme de classification. Les limitations introduites par la segmentation initiale et les résultats des Fig. 2.3 et 2.4 amènent donc à son utilisation au sein d'un algorithme K-moyennes (classifieur Wishart adapté à la distance de Box) ou K-moyennes++ sans segmentation initiale, directement sur les pixels de l'image. Le K-moyennes++ induit de légères variations dans les résultats de classification d'une exécution à l'autre mais pas à tel point que l'interprétation soit modifiée. Son utilisation à la place d'un classifieur Wishart adapté à la distance de Box et dont les classes sont initialisées par un grand nombre de pixels choisis aléatoirement dans l'image est ainsi possible mais sans réel gain ou perte de performances.

On applique ainsi un K-moyennes++ avec la distance de Box de l'Eq. (2.40), huit classes, un voisinage pour l'estimation de la matrice de covariance de taille 5×5 pixels. et un critère d'arrêt fixé quand moins de 5% des pixels changent de classe d'une itération à l'autre. La convergence dans ce cas est très rapide, de l'ordre de cinq ou six itérations, que ce soit avec la SCM ou le FPE. Les résultats sont présentés sur la Fig. 2.12. Ils sont très similaires à ceux des Fig. 2.3 et 2.4, c'est-à-dire que la séparation entre zones homogènes et zones hétérogènes est bien visible, avec les zones hétérogènes rejetées en immense majorité et placées dans la classe orange foncé, mais que la séparation au sein des deux types de zone l'est beaucoup moins. On peut néanmoins remarquer que certains champs, en bleu ciel principalement, ressortent plus visiblement avec la SCM qu'avec le FPE et de manière plus claire que sur la Fig. 2.3.

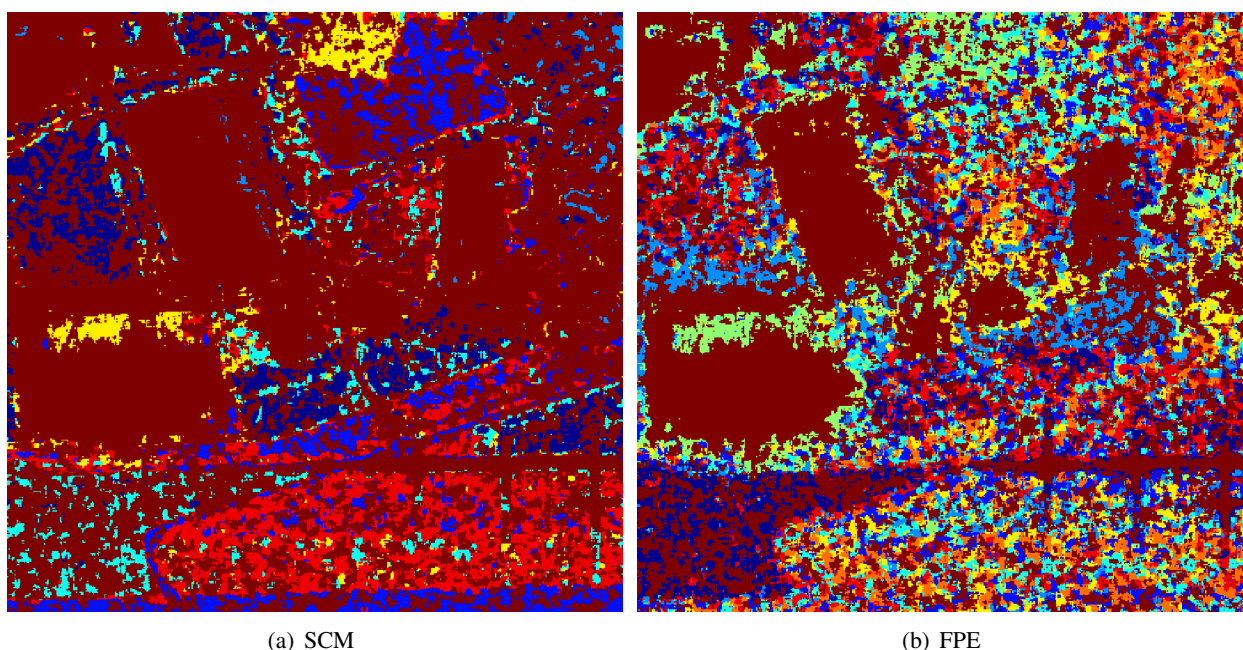
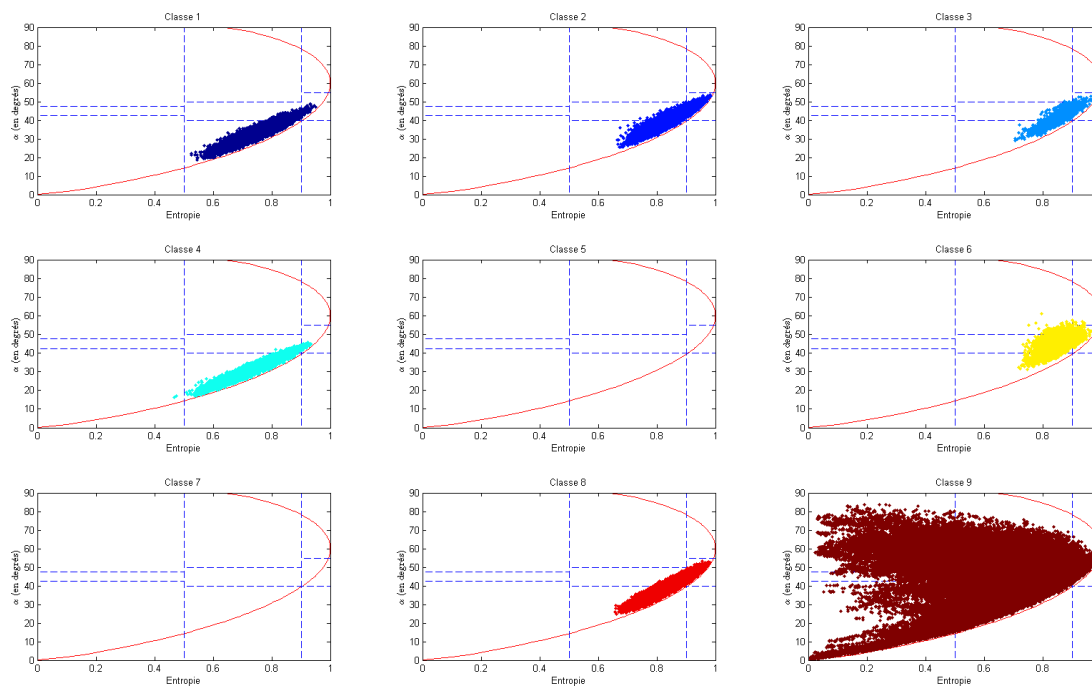


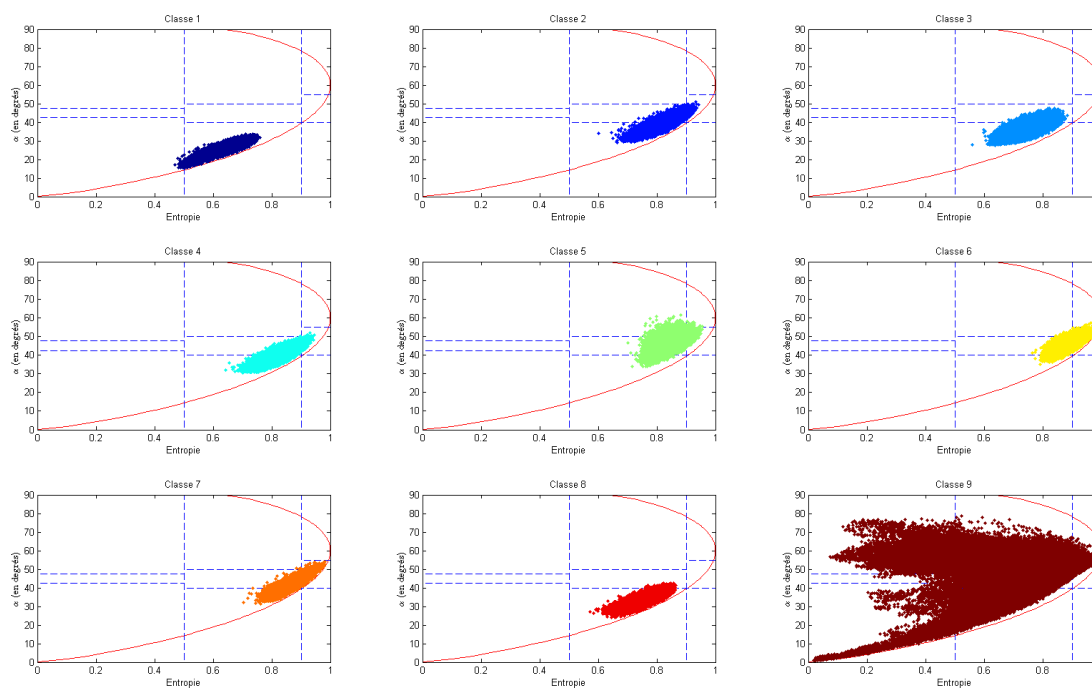
FIGURE 2.12 – K-moyennes++ sans segmentation initiale : comparaison entre la SCM et le FPE.

On peut également s'intéresser à la répartition des pixels de chaque classe dans le plan $H-\alpha$ de la Fig. 1.9, représentée dans la Fig. 2.13. La dispersion intra-classe est effectivement moindre dans le cas de la SCM mais on a dans les deux cas, une classe de rejet constituée d'un très grand nombre de pixels (environ 60% dans le cas de la SCM) et qui occupe tout l'espace $H-\alpha$. Il est encore une fois difficile de donner une réelle interprétation au sens polarimétrique de ces classifications si les pixels rejetés présentent tous les mécanismes polarimétriques présents dans l'image. La séparation entre les classes est aussi sujette à discussion, étant donné qu'elles se

situent toutes environ au même endroit du plan.



(a) SCM



(b) FPE

FIGURE 2.13 – K-moyennes++ sans segmentation initiale : répartition des classes dans le plan H- α .

2.3.3 Application à la classification hiérarchique

Des résultats de classification préliminaires ont été obtenus grâce à l'utilisation d'un algorithme de classification hiérarchique.

La classification hiérarchique est une méthode de classification dont le but est de créer une hiérarchie de classes. Cette hiérarchisation peut être réalisée de deux manières :

- de manière ascendante (ou *bottom-up*) : chaque élément destiné à être classifié est situé à l'origine dans sa propre classe et on fusionne les deux classes les plus proches à chaque itération,
- de manière descendante (ou *top-down*) : tous les éléments sont situés dans une seule classe à l'origine et cette classe est divisée à chaque itération.

L'approche employée ici est une approche ascendante avec comme éléments de classification les pixels de l'image SAR polarimétrique. Les résultats d'une classification hiérarchique sont souvent présentés sous la forme d'un dendrogramme, dont un exemple est donné sur la Fig. 2.14. Les nœuds situés en bas de l'arbre, appelés *feuilles*, représentent l'état initial de l'algorithme, à savoir chaque pixel dans une classe séparée. À chaque itération, les pixels les plus proches selon une certaine mesure de similarité sont fusionnés de manière à construire l'arbre et obtenir un nœud final contenant tous les pixels de l'image.

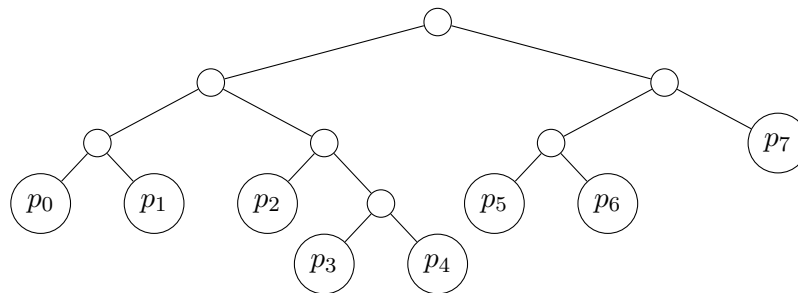


FIGURE 2.14 – Exemple de classification hiérarchique

Le point le plus crucial de la classification hiérarchique est de choisir à quelle hauteur sectionner l'arbre. En effet, sectionner l'arbre à une certaine hauteur permet d'obtenir les différentes classes. Par exemple, dans le cas de la Fig. 2.14, en coupant l'arbre à la hauteur 1, c'est-à-dire un niveau plus bas que le nœud racine, on obtient deux classes : une composée des pixels p_0 , p_1 , p_2 , p_3 et p_4 et une composée des pixels p_5 , p_6 et p_7 . En coupant l'arbre à la hauteur 2, on obtient quatre classes : une composée des pixels p_0 et p_1 , une de p_2 , p_3 et p_4 , une de p_5 et p_6 et la dernière du pixel p_7 . On procède de la même manière dans des arbres plus grands. Ce choix de la hauteur à laquelle couper l'arbre se fait généralement de manière empirique. Dans le cas de la classification des images SAR polarimétriques, en utilisant comme mesure de similarité la distance issue du M-test de Box, on dispose d'une valeur de hauteur obtenue directement à partir de la région critique. En effet, l'hypothèse H_0 (à savoir égalité des matrices de covariance des pixels considérés) est rejetée à partir d'un certain seuil, donné par la distribution χ^2 de la statistique du test (cf. Eq. (2.41)).

On implémente alors l'algorithme de classification hiérarchique de la manière suivante sur le jeu de données de Brétigny :

- calcul de la distance entre tous les pixels de l'image, deux-à-deux, avec l'Eq. (2.40). Etant donné le nombre de pixels dans l'image ($501 \times 501 = 251001$), il est pour l'instant nécessaire de sous-échantillonner l'image par un facteur trois en distance et en azimut afin de réduire le coût en mémoire.
- l'arbre est ensuite construit itérativement. On fusionne d'abord les deux pixels les plus proches puis on met à jour les distances entre ce groupe de pixels et les autres pixels. Pour cela, une fonction appelée

fonction de lien (*linkage function*) est nécessaire. En effet, la distance entre deux pixels est donnée par le M-test de Box. Mais la distance entre deux groupes de pixels nécessite d'être établie. Plusieurs fonctions de lien sont usuellement utilisées, comme par exemple :

- la distance minimale entre les éléments des deux groupes :

$$d(C_1, C_2) = \min (u(\mathbf{x}, \mathbf{y}); \mathbf{x} \in C_1, \mathbf{y} \in C_2),$$

- la distance maximale entre les éléments des deux groupes :

$$d(C_1, C_2) = \max (u(\mathbf{x}, \mathbf{y}); \mathbf{x} \in C_1, \mathbf{y} \in C_2),$$

- la distance moyenne entre les éléments des deux groupes :

$$d(C_1, C_2) = \frac{1}{N_1 N_2} \sum_{\mathbf{x} \in C_1} \sum_{\mathbf{y} \in C_2} u(\mathbf{x}, \mathbf{y}),$$

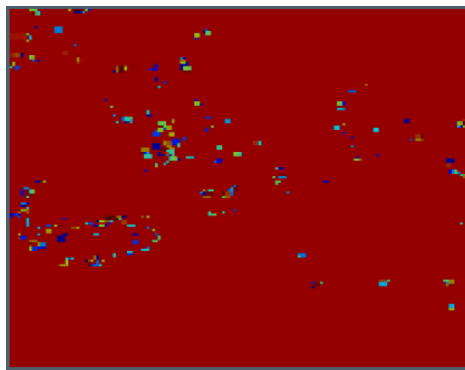
avec N_1 (resp. N_2) le cardinal de la classe C_1 (resp. C_2).

- la distance moyenne pondérée entre les éléments des deux groupes : si le groupe C_1 a été construit à partir des groupes C_a et C_b , alors la distance entre le groupe C_1 et un autre groupe C_2 est donnée de manière récursive par :

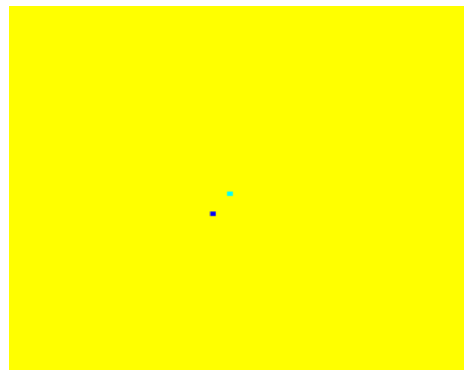
$$d(C_1, C_2) = \frac{d(C_a, C_2) + d(C_b, C_2)}{2}.$$

— une fois l'arbre construit, on coupe à la hauteur indiquée par la région critique et qui ne dépend que de la probabilité de fausse alarme. On obtient alors automatiquement le nombre de classes.

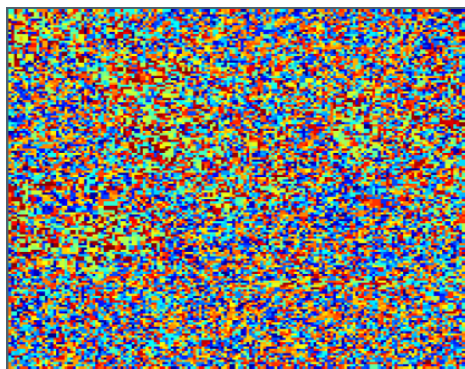
Le choix de la fonction de lien va influencer très fortement les résultats. En effet, utiliser par exemple la distance minimum entre les éléments des deux groupes est un critère très optimiste puisqu'il indique que la distance entre deux groupes de pixels est égal à la plus petite distance entre deux éléments de ces deux groupes et pourrait résulter en une sous-classification (pas assez de classes). Au contraire, la distance maximum est un critère très pessimiste et pourrait entraîner une sur-classification (trop de classes). Les résultats de la classification sont présentés sur la Fig. 2.15 pour l'utilisation de la SCM et du FPE avec un taux de fausse alarme réglé à 0,1%. On remarque en effet une sous-classification très importante avec l'utilisation de la distance minimum comme fonction de lien, ainsi qu'une forte sur-classification lors de l'utilisation de la distance maximum. Cette sur-classification apparaît aussi lors de l'utilisation de la distance moyenne pondérée (cf. Figs 2.15(g) et 2.15(h)). Au final, seuls les résultats de classification obtenus en utilisant la distance moyenne, observables sur les Figs 2.15(e) et 2.15(f), sont exploitables. La sur-classification est toujours présente dans ces deux images, surtout aux niveaux des zones urbaines mais de moindre manière, ce qui permet de reconnaître les éléments habituels de l'image : bâtiments et champs principalement. De manière à réduire cette sur-segmentation, on peut diminuer le taux de fausse alarme à 0,01%. Ce faisant, le seuil donné par la loi du χ^2 sera plus élevé. Ceci implique que l'arbre sera coupé à une hauteur plus élevée, plus proche du noeud racine et ainsi que le nombre de classes sera réduit. La Fig. 2.16 présente donc les résultats de la classification un taux de fausse alarme réglé à 0,01% et uniquement avec la distance moyenne comme fonction de lien, étant donné qu'il s'agit de la seule méthode fournissant des résultats exploitables. On peut voir que la sur-segmentation est effectivement réduite, notamment sur les zones naturelles pour la SCM où les champs ressortent de manière très forte. Pour le FPE, on a toujours une sur-segmentation sur les zones urbaines et bâtiments, mais moins marquée qu'avec un taux de fausse alarme réglé à 0,1% et on observe mieux leurs contours.



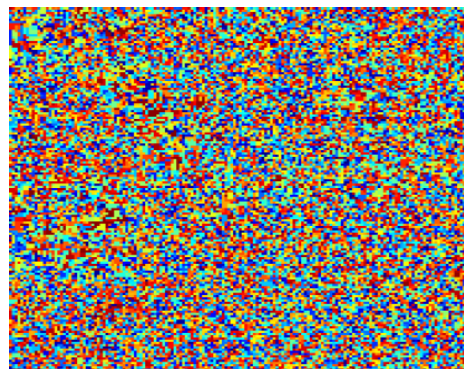
(a) SCM, distance minimum



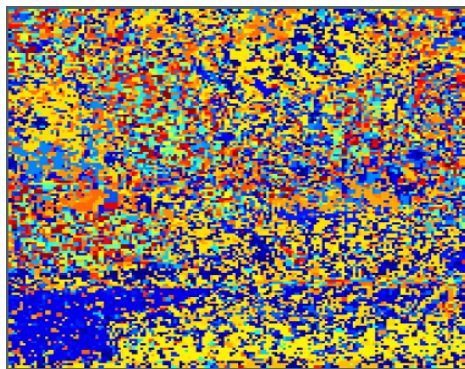
(b) FPE, distance minimum



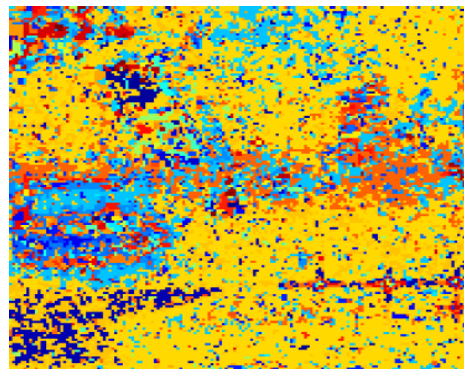
(c) SCM, distance maximum



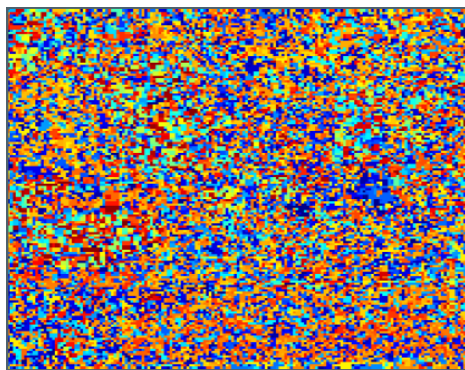
(d) FPE, distance maximum



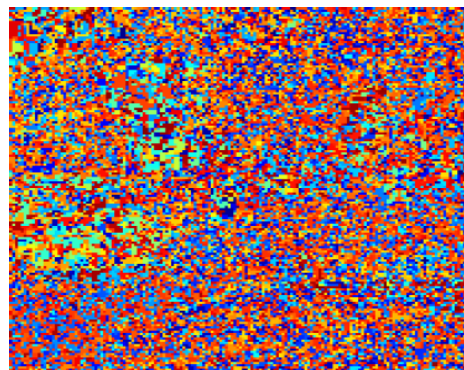
(e) SCM, distance moyenne



(f) FPE, distance moyenne



(g) SCM, distance moyenne pondérée



(h) FPE, distance moyenne pondérée

FIGURE 2.15 – Classification hiérarchique utilisant différentes fonctions de lien avec un taux de fausse alarme réglé à 0,1%

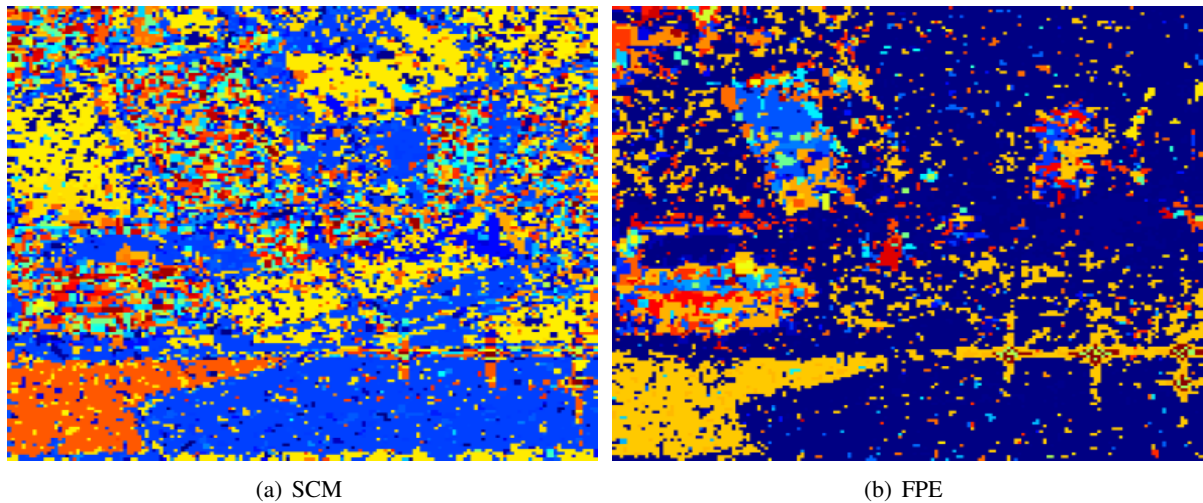


FIGURE 2.16 – Comparaison de la SCM et du FPE pour la classification hiérarchique avec un taux de fausse alarme réglé à 0,01%

De nombreux travaux restent encore à effectuer sur cette approche. En effet, il convient de s'assurer que la distribution des données s'approche bien d'une distribution du χ^2 . L'utilisation de la fonction de lien modifie en effet la distribution des données en calculant le minimum ou le maximum ou la moyenne de ces données. Des premiers résultats comparant la distribution des données en utilisant les différentes fonctions de lien avec une distribution du χ^2 sont présentés sur la Fig. 2.17. On s'aperçoit aisément qu'il existe un écart entre la loi théorique et la loi empirique, surtout pour le FPE. L'approximation χ^2 donnée par le test de Box est en effet proposée pour des données gaussiennes, c'est-à-dire des matrices suivant une loi de Wishart et le FPE ne se comporte comme une matrice de Wishart que de manière asymptotique. À distance finie, comme c'est le cas dans une application réelle, les écarts peuvent être importants. Ceci peut poser un problème puisque le seuil choisi pour couper l'arbre et obtenir le nombre de classes automatiquement dépend directement de cette approximation. On peut alors s'interroger sur la pertinence d'utiliser cette approche supposée rigoureuse en utilisant un seuil adapté à une distribution qui n'est pas, au final, réellement proche de la distribution empirique des données. Ce résultat est néanmoins à tempérer, étant donné que l'utilisation de la fonction de lien modifie la distribution des données.

Box [15] fournit aussi une approximation de la statistique du test par une loi de Fisher, il serait alors intéressant de considérer cette approximation et observer si le FPE est plus proche de cette distribution. On peut aussi approcher le problème de la distance de lien d'une autre manière. Plutôt que d'appliquer l'opérateur moyenne sur les distances entre pixels (donc entre matrices de covariance, cf. Eq. (2.40)), il est possible d'appliquer l'opérateur moyenne directement sur les matrices de covariance de chacun des deux groupes. Par exemple, si l'on veut calculer la distance entre deux noeuds A , composée de trente pixels et B , composé de vingt pixels, l'algorithme classique implique de calculer les distances deux à deux entre pixels du noeud A et pixels du noeud B (soit 30×20 distances) et d'en prendre la moyenne. Mais on connaît les matrices de covariance de chacun des pixels situés dans les noeuds A et B , il est alors possible de calculer la moyenne de ces matrices au sein de chaque noeud et de calculer simplement la distance entre ces deux matrices moyennes grâce à l'Eq. (2.40). On peut aussi considérer réestimer une matrice de covariance pour chacun des noeuds en utilisant la SCM ou le FPE à partir des pixels composant le noeud et encore une fois appliquer l'Eq. (2.40) pour obtenir la distance entre les deux matrices de covariance des noeuds. De cette manière, on peut aussi s'affranchir des considérations sur l'adéquation de la distribution du χ^2 avec la fonction de lien puisqu'on ne travaille plus sur les données transformées mais bien sur les données de base.

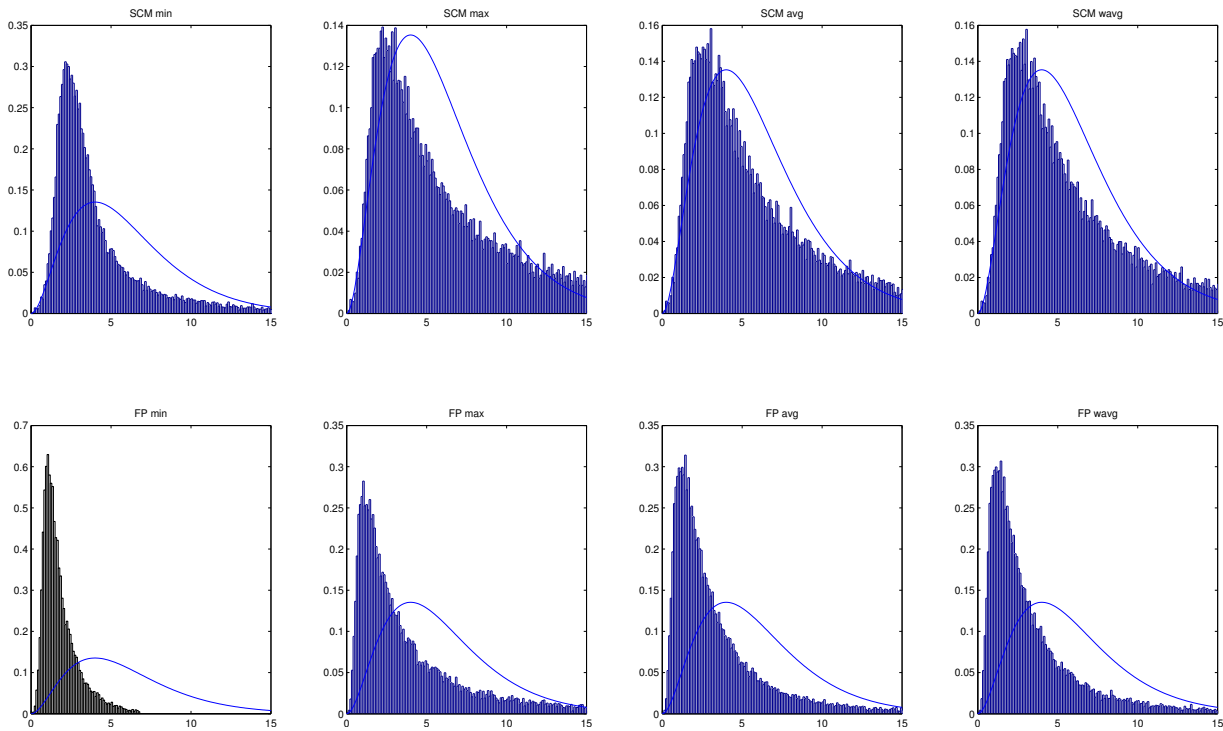


FIGURE 2.17 – Comparaison de la distribution des données avec une distribution du χ^2 en fonction de la fonction de lien utilisée.

2.4 Conclusion

Ce chapitre a permis d'introduire un test d'égalité de matrices de covariance pour la classification d'images SAR polarimétriques. Une définition rigoureuse du test a été donnée, avant d'établir la statistique de ce test selon plusieurs hypothèses : signal distribué selon une loi gaussienne ou selon un SIRV, matrices de covariance toutes deux inconnues ou une des deux matrices connue. Une approximation du χ^2 de la statistique du test pour le cas gaussien est rappelée puis étendue au cas SIRV. La statistique du test définit alors une distance, qui est implémentée dans plusieurs algorithmes de classification, et l'approximation du χ^2 fournit une région critique qui permet de rejeter les pixels trop différents des pixels déjà connus. Cette approche permet une bonne séparation entre zones naturelles et zones artificielles (bâtiments, parkings, coins réflecteurs, etc.) dans les images mais peu de séparation au sein de ces deux types de zone. L'approche par classification hiérarchique semble très prometteuse pour améliorer cette séparation.

Centres de classe et géométrie riemannienne

Sommaire

3.1	Calcul des centres de classe pour le FPE	71
3.2	Contribution de la géométrie riemannienne	74
3.2.1	Cas des matrices réelles définies positives	75
3.2.2	Extension aux matrices hermitiennes définies positives	80
3.3	Simulations	81
3.4	Données réelles	86
3.4.1	Utilisation de la distance de Wishart	86
3.4.2	Utilisation de la distance géométrique	88
3.4.3	Répartition des pixels dans l'espace $H-\alpha$	89
3.5	Conclusion	93

Ce chapitre est consacré à l'étude de l'impact du calcul des centres de classe lors de la classification. En premier lieu, l'influence du calcul par moyennage arithmétique traditionnel sera étudiée pour l'estimateur du point fixe puis l'utilisation de la géométrie riemannienne sera introduite. Nous verrons pourquoi le moyennage arithmétique (ou euclidien) n'est rigoureusement pas adapté au cas des matrices de covariance et l'intérêt d'employer une approche riemannienne.

3.1 Calcul des centres de classe pour le FPE

L'approche traditionnellement employée pour le calcul des centres de classes dans le classifieur Wishart est celle présentée dans l'Eq. (1.53), à savoir : le centre de classe est calculé comme la moyenne arithmétique des matrices de covariance des pixels appartenant à la classe considérée [50]. Elle découle directement de l'utilisation originelle de l'algorithme K-moyennes où les données sont représentées dans un espace euclidien, typiquement à deux ou trois dimensions. Dans ce cas, le centre de classe est le "barycentre" des données de la classe et est positionné spatialement au centre du cluster représenté dans l'espace correspondant aux données.

Dans le cas du radar, et plus précisément de la classification radar polarimétrique, les données sont des matrices hermitiennes définies positives. Il apparaît alors important de s'assurer que la méthode de calcul des centres de classe traditionnelle est bien adaptée aux données. Il s'avère d'autant plus que les données ne sont pas connues explicitement mais estimées à partir des données de base de l'image radar. Il y a alors deux approches possibles pour obtenir les centres de classe :

- approche 1, traditionnelle : utiliser l'approche traditionnelle de calculer la moyenne des matrices de la classe, ce qui entraîne une redondance des pixels (car chaque pixel est utilisé dans le calcul de plusieurs matrices de covariance donc il peut être utilisé plusieurs fois dans le calcul de la moyenne),

- approche 2, par estimation : revenir aux données pixel par pixel et estimer le centre de classe en utilisant tous les pixels de la classe.

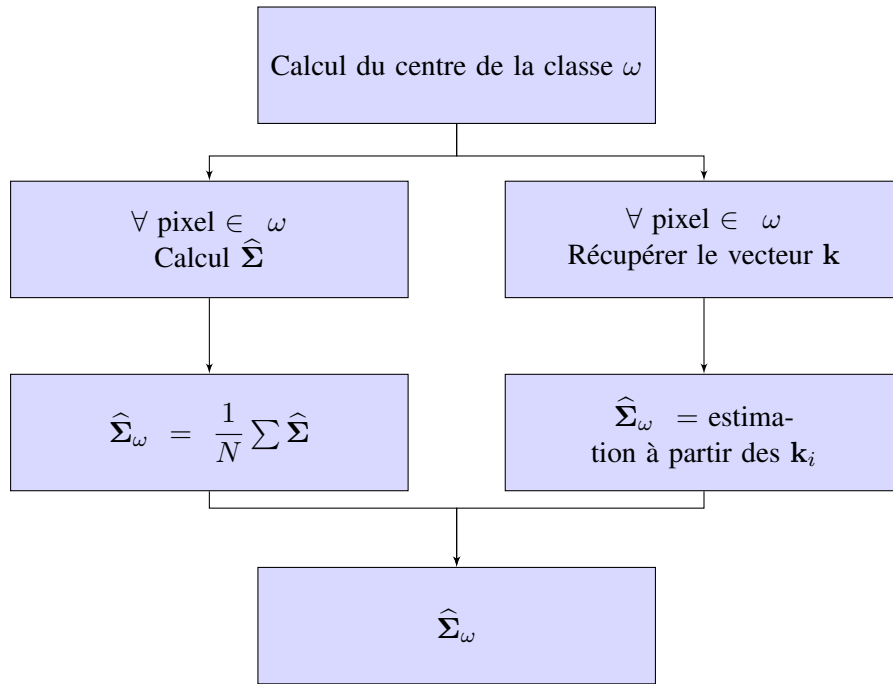


FIGURE 3.1 – Calcul des centres de classe

La Fig. 3.1 schématise ces deux approches. Dans le cas de la SCM, il est aisé de remarquer que les deux approches donnent un résultat relativement similaire, étant donné que la SCM est estimée, selon l’Eq. (1.35), comme la moyenne des produits des vecteurs \mathbf{k} situés dans le voisinage du pixel considéré avec leurs transposés conjugués. On a alors les deux situations suivantes :

- approche traditionnelle : $\hat{\Sigma}_\omega = \frac{1}{N} \sum \hat{\Sigma} = \frac{1}{N} \sum \sum_i \mathbf{k}_i \mathbf{k}_i^H$ avec les \mathbf{k}_i données secondaires utilisées pour l’estimation de la matrice de covariance de chaque pixel. Ici, \mathbf{k}_i peut être utilisé plusieurs fois,
- approche par estimation : $\hat{\Sigma}_\omega = \frac{1}{N} \sum \mathbf{k} \mathbf{k}^H$.

Pour l’approche traditionnelle, on a la même somme que dans l’approche par estimation, à laquelle est rajoutée la moyenne des données secondaires de chaque pixel. On a donc une somme beaucoup plus importante mais qui rajoute des termes *a priori* proches des termes déjà existants donc le moyennage devrait donner un résultat similaire selon les deux approches.

Dans le cas du FPE, l’estimation de la matrice est faite de manière non linéaire, selon l’Eq. (1.57). On peut donc s’attendre à avoir des résultats différents en suivant les deux approches. Il faut cependant prendre en compte le fait que les matrices estimées par le FPE sont normalisées pour avoir une trace égale à m , la taille du vecteur de données. On aura donc au final des centres de classe qui devraient être peu différents en norme suivant les deux approches. On discutera plus loin de l’influence de la distance entre matrices sur les comportements polarimétriques.

La Fig. 3.2 présente les résultats de classification par le classifieur Wishart avec initialisation par la décomposition de Cloude-Pottier pour la SCM et le FPE, selon les deux approches de calcul de centres de classe.

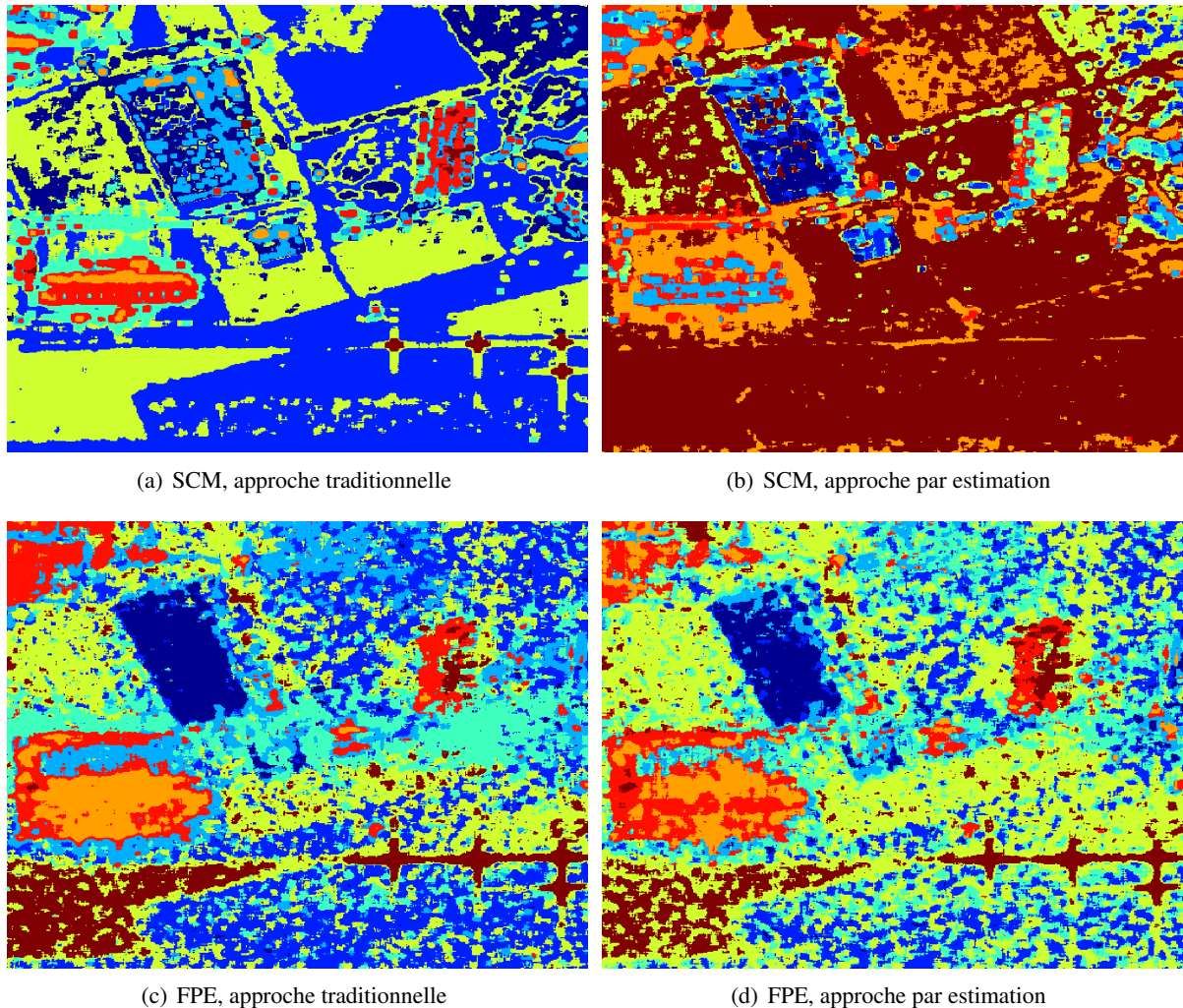


FIGURE 3.2 – Comparaison des résultats du classifieur Wishart suivant la méthode de calculs des centres : traditionnelle ou réestimation de la matrice.

Pour la SCM, les résultats sont en réalité très différents selon les deux approches. Avec l'approche traditionnelle, on observe encore une fois bien les contours des différents champs présents dans l'image, qui disparaissent complètement avec l'approche par estimation. Il est probable que la perte du peu d'information spatiale disponible pour la classification (à savoir l'information contenue dans l'estimation de la matrice de covariance par un voisinage spatial autour du pixel considéré) soit responsable de cette mauvaise classification. En effet, l'initialisation est effectuée grâce à la décomposition de Cloude-Pottier (cf. Fig. 1.15(a)) : si l'on considère par exemple les pixels situés dans la classe marron, ils composent principalement le champ situé dans le coin inférieur gauche, le parking et les coins réflecteurs. Les valeurs des matrices de covariance estimées par la SCM pour ces pixels varient énormément, de valeurs très faibles sur les champs à des valeurs très élevées sur les coins réflecteurs car la SCM n'est pas normalisée. Avec l'approche traditionnelle, ces fortes valeurs peuvent être compensées en considérant les pixels voisins. Avec l'approche par estimation, ces grandes variations sont probablement à l'origine de la moins bonne classification.

Pour le FPE, au contraire, les résultats sont très similaires, certainement dû au fait que les matrices de covariance estimées par le FPE sont normalisées, ce qui permet de comparer des matrices ayant le même ordre de grandeur. On peut néanmoins remarquer que les contours des bâtiments semblent moins bien dessinés avec l'approche par estimation, dû au plus faible nombre de données pour l'estimation.

Il apparaît donc que l'approche par moyennage des matrices semble plus appropriée pour la classification Wishart. Cependant, il est important de prendre en compte la structure des objets manipulés lors de ce moyennage, à savoir les matrices de covariance polarimétriques. Ces matrices sont des matrices hermitiennes définies positives et l'utilisation de la moyenne arithmétique suppose que ces matrices appartiennent à un espace euclidien. En réalité, l'espace des matrices hermitiennes définies positives est un espace riemannien (donc à structure courbe, contrairement aux espaces euclidiens à structure plane) et il est nécessaire de prendre en compte la structure de cette espace pour calculer la moyenne de matrices de covariance polarimétriques. Ceci peut être réalisé en considérant la géométrie de l'information, une branche de la géométrie riemannienne.

3.2 Contribution de la géométrie riemannienne

La géométrie riemannienne permet l'étude d'espaces courbes non-euclidiens en introduisant un objet mathématique appelé *variété*. Une variété est un espace topologique, au même sens qu'un espace vectoriel, qui est localement euclidien mais ne l'est généralement pas de manière globale. Un cercle est un exemple de variété dans le plan : on peut approcher localement la courbure du cercle par sa tangente mais le cercle dans son ensemble n'est pas euclidien. Dans un espace à trois dimensions, la surface d'une sphère est une variété : en cartographie, par exemple, il est possible de représenter une partie de la surface de la Terre sur une carte grâce à une projection mais on ne peut pas représenter la totalité des points sur une seule carte, quelle que soit la projection utilisée. Une variété est ainsi définie comme une collection de cartes, ou atlas. Des applications dites de changement de carte permettent de faire la transition entre les coordonnées d'une carte à l'autre. Si ces applications suivent certains critères de régularité, alors la variété est une variété différentielle, sur laquelle on peut employer le calcul différentiel et intégral.

En tout point d'une variété différentielle, on associe un espace vectoriel réel qui contient toutes les directions tangentes à la variété en ce point. Dans des cas simples, la notion d'espace tangent est triviale. Par exemple, en considérant le périmètre d'un cercle, l'espace tangent d'un point est la droite tangente en ce point. Pour la surface d'une sphère, c'est le plan tangent en ce point.

Et si une variété différentielle est munie d'une métrique riemannienne — c'est-à-dire pour chaque plan tangent à la surface, la donnée d'une forme bilinéaire symétrique définie positive différentiable — alors c'est une variété riemannienne. L'intérêt d'avoir affaire à une variété riemannienne, c'est qu'en plus de pouvoir appliquer des opérations de calcul différentiel et intégral, on peut définir la longueur entre deux points de la variété, ce qui permet la mise en place et l'étude d'angles et, ce qui nous intéresse, de distances.

Des travaux récents, notamment ceux de Barbaresco [6–8] ont permis de commencer à démocratiser l'utilisation de cette branche de la géométrie pour le radar. Ces travaux sont inspirés de ceux de Moakher [55], ainsi que ceux de Devlaminck [30] dans le cadre de l'imagerie optique. Ces derniers sont particulièrement intéressants puisqu'ils montrent une application directe et concrète de l'utilisation d'une mauvaise expression de la moyenne. En effet, considérons un rayon de lumière polarisé horizontalement passant à travers un polariseur linéaire parfait possédant un axe de polarisation rotatif. Si l'on s'intéresse au vecteur de Stokes — qui est une autre manière de représenter un état de polarisation [68], différente du vecteur de Jones vu au chapitre un — à

la sortie du système, on peut l'écrire de la manière suivante :

$$\left[\frac{S_0}{2}(1 + \cos 2\theta) \quad \frac{S_0}{2}(1 + \cos 2\theta) \cos 2\theta \quad \frac{S_0}{2}(1 + \cos 2\theta) \sin 2\theta \quad 0 \right]^T, \quad (3.1)$$

avec θ l'angle entre la polarisation horizontale et l'axe de rotation du polariseur et S_0 l'intensité de l'onde.

En considérant une distribution uniforme pour θ , entre $-\pi/2$ et $\pi/2$ par exemple, la moyenne euclidienne de ce vecteur de Stokes est un vecteur partiellement polarisé alors que le polariseur est linéaire et parfait. Ceci montre qu'une mauvaise expression de la moyenne peut amener à des conclusions erronées.

Dans le cas des matrices de covariance, on a affaire à des matrices hermitiennes définies positives. L'espace de ces matrices, noté $HPD(\mathbb{C})$, est une variété différentielle. Pour pouvoir étudier la distance entre deux matrices sur cette variété, il faut donc se munir d'une métrique riemannienne. Étudions d'abord le cas, plus simple, des matrices réelles symétriques définies positives et nous verrons par la suite comment généraliser les résultats obtenues aux matrices hermitiennes.

3.2.1 Cas des matrices réelles définies positives

Introduction :

Une des approches permettant de définir la métrique riemannienne sur l'espace de matrices réelles définies positives, noté $PD(\mathbb{R})$, a été introduite par Rao (voir [4] et [17]) et fait appel à la géométrie de l'information. Les travaux de Charon [22], que nous rappelons rapidement ici, donnent une version détaillée de cette méthode. On considère un espace E muni d'une mesure μ et une famille de fonctions paramétrées définies sur E , $\{p(x|\theta), \theta \in \Theta\}$ avec $\theta = (\theta_1, \dots, \theta_n)$ les paramètres et Θ l'espace des paramètres. Ces fonctions sont des densités de probabilités de la variable aléatoire X définie sur E .

On suppose de plus que cette famille de fonctions possède les propriétés suivantes :

$$\begin{aligned} \frac{\partial p(x|\theta)}{\partial \theta_i} \text{ existe, } \forall i \in \llbracket 1..n \rrbracket, \forall X \in E \text{ et } \forall \theta \in \Theta, \\ \int_E \frac{\partial p(x|\theta)}{\partial \theta_i} d\mu(x) = 0, \forall i \in \llbracket 1..n \rrbracket. \end{aligned} \quad (3.2)$$

On définit alors la matrice d'information de Fisher \mathbf{F} comme la matrice constituée des éléments f_{ij} suivants :

$$\begin{aligned} f_{ij} &= \mathbf{E} \left[\frac{\partial \ln p(X|\theta)}{\partial \theta_i} \frac{\partial \ln p(X|\theta)}{\partial \theta_j} \right], \\ &= \int_E \frac{\partial \ln p(x|\theta)}{\partial \theta_i} \frac{\partial \ln p(x|\theta)}{\partial \theta_j} p(x|\theta) d\mu(x). \end{aligned} \quad (3.3)$$

C'est une généralisation de l'information de Fisher, qui représente pour une variable aléatoire dépendant d'un seul paramètre l'information relative à ce paramètre.

Il importe maintenant de mesurer la distance entre deux densités de probabilités de paramètres voisins $p(x|\theta)$ et $p(x|\theta + d\theta)$. Pour cela, considérons le développement limité au premier ordre de leur différence

relative :

$$p(X|\theta + d\theta) - p(X|\theta) = \sum_{i=1}^n \frac{\partial p(X|\theta)}{\partial \theta_i} d\theta_i, \quad (3.4)$$

$$\frac{p(X|\theta + d\theta) - p(X|\theta)}{p(X|\theta)} = \sum_{i=1}^n \frac{\partial \ln p(X|\theta)}{\partial \theta_i} d\theta_i. \quad (3.5)$$

Rao a proposé de mesurer la longueur élémentaire ds en calculant la variance de l'Eq. (3.5). On a donc :

$$\text{Var} \left(\frac{p(X|\theta + d\theta) - p(X|\theta)}{p(X|\theta)} \right) = \mathbb{E} \left[\left(\sum_{i=1}^n \frac{\partial \ln p(X|\theta)}{\partial \theta_i} d\theta_i \right)^2 \right] - \mathbb{E} \left[\sum_{i=1}^n \frac{\partial \ln p(X|\theta)}{\partial \theta_i} d\theta_i \right]^2. \quad (3.6)$$

Or, on remarque que :

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ln p(X|\theta)}{\partial \theta_i} d\theta_i \right] &= \int_E \frac{\partial \ln p(x|\theta)}{\partial \theta_i} p(x|\theta) d\mu(x), \\ &= \int_E \frac{\partial p(x|\theta)}{\partial \theta_i} d\mu(x), \\ &= 0, \end{aligned} \quad (3.7)$$

d'après l'Eq. (3.2).

L'Eq. (3.6) devient donc :

$$\begin{aligned} \text{Var} \left(\frac{p(X|\theta + d\theta) - p(X|\theta)}{p(X|\theta)} \right) &= \mathbb{E} \left[\left(\sum_{i=1}^n \frac{\partial \ln p(X|\theta)}{\partial \theta_i} d\theta_i \right)^2 \right], \\ &= \sum_{i,j=1}^n f_{ij}(\theta) d\theta_i d\theta_j, \\ \text{d'où } ds^2 &= \sum_{i,j=1}^n f_{ij}(\theta) d\theta_i d\theta_j. \end{aligned} \quad (3.8)$$

On a donc défini une structure de variété riemannienne sur l'espace des paramètres Θ dont l'élément de longueur est donné par l'Eq. (3.8). Cette structure de variété riemannienne ainsi que l'expression de l'élément de longueur vont donc permettre de définir une distance entre deux éléments de Θ (qui sera réduit plus tard à $PD(\mathbb{R})$, l'espace des matrices symétriques définies positives) et par ce moyen, la moyenne au sens géométrique. Il est par ailleurs intéressant de noter qu'on peut retrouver ce résultat par une méthode qui peut sembler plus intuitive lorsqu'on manipule des distributions, à savoir en utilisant la divergence de Kullback-Leibler, qui est une mesure de similarité entre deux distributions.

Distance entre deux matrices de $PD(\mathbb{R})$:

Il s'agit maintenant de définir la distance entre deux points de Θ , notés θ^1 et θ^2 . Pour cela, on considère la courbe $\theta(t) = (\theta_1(t), \dots, \theta_n(t))$, de classe C^1 par morceaux dans \mathbb{R} , telle que $\theta(0) = \theta^1$ et $\theta(1) = \theta^2$ avec

$t \in [0, 1]$ et $\forall t \in [0, 1]$, $\theta(t) \in \Theta$. La longueur de cette courbe est donnée par :

$$\left| \int_0^1 \frac{ds}{dt} dt \right| = \left| \int_0^1 \left(\sum_{i,j=1}^n g_{ij}(\theta(t)) \dot{\theta}_i \dot{\theta}_j \right)^{1/2} dt \right|, \quad (3.9)$$

avec $\dot{\theta}_i = \frac{d\theta_i}{dt}$.

Il existe une infinité de courbes $\theta(t)$ reliant θ^1 à θ^2 . On s'intéresse à la courbe dont la longueur est minimale. Elle est appelée géodésique et c'est cette longueur minimale que l'on va choisir pour définir la distance entre les deux points θ^1 et θ^2 . L'expression de la distance est alors donnée par l'équation suivante :

$$d(\theta^1, \theta^2) = \inf_{t \rightarrow \theta(t)} \left| \int_0^1 \left(\sum_{i,j=1}^n g_{ij}(\theta(t)) \dot{\theta}_i \dot{\theta}_j \right)^{1/2} dt \right| \quad (3.10)$$

Malheureusement, il n'est pas possible d'expliciter les géodésiques plus amplement à partir de l'expression de l'Eq. (3.10) dans le cas général. On peut déterminer un système d'équations différentielles satisfait par ces géodésiques, en calculant la dérivée covariante — qui est un outil de la géométrie différentielle permettant de calculer la dérivée d'un champ de vecteurs sur une variété — du champ des vecteurs tangents à la géodésique. Cette dérivée étant nulle, on obtient un système de deux équations couplées qui ne sera pas explicité ici, puisque impossible à résoudre explicitement dans le cas général.

Revenons au cas particulier qui nous intéresse, à savoir le cas des images radar. Supposons pour l'instant que le vecteur de données \mathbf{k} suit une loi gaussienne multivariée de moyenne nulle. On réduit ainsi l'espace des paramètres Θ à l'espace des matrices symétriques définies positives $PD(\mathbb{R})$. Le paramètre θ est donc une matrice de $PD(\mathbb{R})$ notée \mathbf{M} . On connaît la loi de \mathbf{k} puisque c'est une loi gaussienne de ddp :

$$p(\mathbf{k}|\mathbf{M}) = \frac{1}{(2\pi)^{n/2} |\mathbf{M}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{k}^T \mathbf{M}^{-1} \mathbf{k}\right). \quad (3.11)$$

On peut ainsi calculer les éléments g_{ij} et l'Eq. (3.8) devient :

$$ds^2 = \frac{1}{2} \text{Tr} \left((\mathbf{M}^{-1} d\mathbf{M})^2 \right) \quad (3.12)$$

Le système d'équations différentielles du cas général est aussi simplifié et s'écrit sous la forme :

$$\begin{cases} \frac{d}{dt} \mathbf{M}^{-1} \dot{\mathbf{M}} = 0 \\ \text{Tr} \left((\mathbf{M}^{-1} \dot{\mathbf{M}})^2 \right) = 2 \end{cases} \quad (3.13)$$

Rappelons que $\mathbf{M}(t)$ est une courbe définie sur $[0, 1]$ telle que pour tout $t \in [0, 1]$, on a $\mathbf{M}(t) \in PD(\mathbb{R})$. La solution générale de la première relation de l'Eq. (3.13) est donnée par :

$$\mathbf{M}(t) = \mathbf{A}^T \exp(t\mathbf{B}) \mathbf{A} \quad (3.14)$$

avec \mathbf{A} et \mathbf{B} des matrices constantes inversibles et \mathbf{B} de déterminant "égal à 1. On obtient alors :

$$\mathbf{M}^{-1} \dot{\mathbf{M}} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A} \quad (3.15)$$

ce qui entraîne, d'après la deuxième relation de l'Eq. (3.13), que $\text{Tr}(\mathbf{B}^2) = 2$.

Les conditions de bord impliquent, d'après l'Eq. (3.14) et en appelant d_{12} la distance entre deux matrices \mathbf{M}_1 et \mathbf{M}_2 , les équations suivantes :

$$\begin{cases} \mathbf{A}^T \mathbf{A} = \mathbf{M}_1 \\ \mathbf{A}^T \exp(d_{12} \mathbf{B}) \mathbf{A} = \mathbf{M}_2 \end{cases} \quad (3.16)$$

avec $\mathbf{M}(0) = \mathbf{M}_1$ et $\mathbf{M}(d_{12}) = \mathbf{M}_2$.

Ceci entraîne alors que $d_{12} \mathbf{B} = \log \left((\mathbf{A}^{-1})^T \mathbf{M}_2 \mathbf{A}^{-1} \right)$. D'après cette expression, \mathbf{B} est donc une matrice symétrique puisque $(\mathbf{A}^{-1})^T \mathbf{M}_2 \mathbf{A}^{-1}$ est symétrique et que le logarithme conserve la symétrie. On a alors $\text{Tr}(\mathbf{B}^2) = \|\mathbf{B}\|_F^2$ avec $\|\dots\|_F$ la norme de Frobenius, et on sait que $\text{Tr}(\mathbf{B}^2) = 2$. On obtient donc l'expression de la géodésique entre deux matrices de $PD(\mathbb{R})$:

$$d(\mathbf{M}_1, \mathbf{M}_2) = \frac{1}{\sqrt{2}} \left\| \log \left((\mathbf{A}^{-1})^T \mathbf{M}_2 \mathbf{A}^{-1} \right) \right\|_F \quad (3.17)$$

On peut prouver [22] que le résultat de l'Eq. (3.17) ne dépend pas de la matrice \mathbf{A} choisie, tant qu'elle vérifie la propriété $\mathbf{A}^T \mathbf{A} = \mathbf{M}_1$. On peut donc réécrire l'Eq. (3.17) sous la forme suivante :

$$d(\mathbf{M}_1, \mathbf{M}_2) = \frac{1}{\sqrt{2}} \left\| \log \left((\mathbf{M}_1^{-1/2})^T \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right\|_F, \quad (3.18)$$

avec $\mathbf{M}_1^{1/2}$ n'importe quelle racine carrée de \mathbf{M}_1 .

Une autre formulation pour cette distance géométrique dépend des valeurs propres, notées λ_k de la matrice $(\mathbf{M}_1^{-1/2})^T \mathbf{M}_2 \mathbf{M}_1^{-1/2}$ et s'écrit de la manière suivante :

$$d(\mathbf{M}_1, \mathbf{M}_2) = \left[\frac{1}{2} \sum_{k=1}^n (\log \lambda_k)^2 \right]^{1/2}. \quad (3.19)$$

A noter que les λ_k sont aussi les valeurs propres généralisées de \mathbf{M}_1 et \mathbf{M}_2 , à savoir les racines de l'équation $|\lambda \mathbf{M}_2 - \mathbf{M}_1| = 0$. Et, plus important, on peut remplacer les expressions de \mathbf{A} et \mathbf{B} dans l'Eq. (3.14) afin d'obtenir une expression en tout point de la géodésique reliant \mathbf{M}_1 à \mathbf{M}_2 donnée par l'équation suivante :

$$\mathbf{M}(t) = (\mathbf{M}_1^{1/2})^T \exp \left(t \log \left((\mathbf{M}_1^{-1/2})^T \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right) \mathbf{M}_1^{1/2} \quad (3.20)$$

$$= (\mathbf{M}_1^{1/2})^T \left((\mathbf{M}_1^{-1/2})^T \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right)^t \mathbf{M}_1^{1/2}. \quad (3.21)$$

A partir de ce résultat, on peut aussi définir l'*application exponentielle* en tout point de la variété. L'expression de l'application exponentielle en un point \mathbf{M} de la variété pour un vecteur \mathbf{B} de l'espace tangent est donnée par :

$$\exp_{\mathbf{M}}(\mathbf{B}) = (\mathbf{M}^{1/2})^T \exp \left((\mathbf{M}^{-1/2})^T \mathbf{B} \mathbf{M}^{-1/2} \right) (\mathbf{M}^{1/2}). \quad (3.22)$$

Cette expression nous sera utile pour établir la moyenne de matrices de manière géométrique.

Moyenne de matrices dans $PD(\mathbb{R})$

Il existe plusieurs approches pour obtenir l'expression de la moyenne géométrique de matrices hermitiennes définies positives. Celle qui est présentée dans ce manuscrit est la plus directe et la plus intuitive si on la compare à l'approche traditionnelle pour le cas euclidien, à savoir l'approche par minimisation de la distance. En effet, considérons $(\mathbf{x}_i)_{i=1,\dots,N}$ N points de \mathbb{R}^M . Pour obtenir la moyenne au sens de la distance euclidienne de ces points (qu'on peut aussi appeler barycentre ou centre de masse), on va chercher la solution \mathbf{x} du problème de minimisation suivant :

$$\inf_{\mathbf{x} \in \mathbb{R}^M} \sum_{i=1}^N d(\mathbf{x}, \mathbf{x}_i)^2, \quad (3.23)$$

où $d(\cdot, \cdot)$ est la distance euclidienne.

On montre alors facilement que la solution \mathbf{x} de ce problème n'est autre que la moyenne arithmétique des \mathbf{x}_i et qu'il s'agit d'un minimum unique. Nous allons alors employer une approche similaire en nous plaçant dans $PD(\mathbb{R})$ et en résolvant le problème de minimisation suivant :

$$\inf_{\mathbf{M} \in PD(\mathbb{R})} \sum_{i=1}^N d(\mathbf{M}, \mathbf{M}_i)^2, \quad (3.24)$$

où $d(\cdot, \cdot)$ est la distance de l'Eq. (3.18), $(\mathbf{M}_i)_{i=1,\dots,N}$ l'ensemble des matrices dont on veut calculer la moyenne et \mathbf{M} la solution du problème.

En exprimant que le gradient de cette expression est nulle et par des considérations de calcul matriciel et différentiel, on arrive à la caractérisation suivante pour \mathbf{M} :

$$\sum_{i=1}^N \mathbf{M}_i \mathbf{M} = 0. \quad (3.25)$$

Cette expression indique que le gradient de l'Eq. (3.24), qu'on peut exprimer sous la forme $\sum_{i=1}^N \mathbf{M}_i \mathbf{M}$ est nul lorsque \mathbf{M} est la moyenne des matrices $(\mathbf{M}_i)_{i=1,\dots,N}$. Malheureusement, il n'existe pas de solution analytique à ce problème pour le cas $N > 2$. Un algorithme de descente de gradient utilisant l'application exponentielle de l'Eq. (3.22) permet d'approcher la solution de manière itérative. Il suffit de trouver une géodésique passant par un point quelconque de la variété et dont la tangente est dans la direction du gradient. On se déplace alors le long de cette géodésique sur une distance donnée par l'algorithme. Au nouveau point obtenu, on recalcule le gradient et on cherche de nouveau une géodésique dans sa direction, le long de laquelle on se déplace de la même distance. On continue de cette manière jusqu'à ce que la distance entre deux points au sens de la distance de l'Eq. (3.18) se trouve en dessous d'un certain seuil. La relation de récurrence qui sous-tend cet algorithme de descente de gradient est donnée par l'expression suivante :

$$\mathbf{P}_{n+1} = \left(\mathbf{P}_n^{1/2} \right)^T \exp \left(-\varepsilon \sum_{i=1}^N \log \left(\left(\mathbf{P}_n^{-1/2} \right)^T \mathbf{M}_i^{-1} \mathbf{P}_n^{-1/2} \right) \right) \mathbf{P}_n^{1/2}, \quad (3.26)$$

avec ε le pas du gradient et \mathbf{P}_i le résultat de l'algorithme à l'itération i .

3.2.2 Extension aux matrices hermitiennes définies positives

La partie précédente a permis d'obtenir une distance entre matrices symétriques définies positives basée sur la structure de l'espace $PD(\mathbb{R})$ les contenant. Il est maintenant temps d'étendre ces résultats aux matrices hermitiennes définies positives qui nous intéressent, car il s'agit des matrices de covariance que l'on cherche à étudier.

Pour cela, on étudie l'application Φ définie de la manière suivante :

$$\begin{aligned} \Phi : GL_n(\mathbb{C}) &\rightarrow GL_{2n}(\mathbb{R}) \\ \mathbf{A} + i\mathbf{B} &\rightarrow \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix}. \end{aligned} \quad (3.27)$$

On peut montrer que l'application Φ est injective et que l'image d'une matrice hermitienne définie positive est une matrice symétrique définie positive. Par ailleurs, l'application Φ possède les propriétés suivantes :

$$\Phi(\mathbf{M}_1)\Phi(\mathbf{M}_2) = \Phi(\mathbf{M}_1\mathbf{M}_2), \quad (3.28)$$

$$\Phi(\mathbf{M}^H) = \Phi(\mathbf{M})^T, \quad (3.29)$$

avec \mathbf{M}, \mathbf{M}_1 et $\mathbf{M}_2 \in GL_n(\mathbb{C})$.

On peut donc alors déterminer l'expression de la distance riemannienne pour des matrices hermitiennes définies positives (donc appartenant à $HPD(\mathbb{C})$) en exprimant la distance par l'Eq. (3.18) entre leurs images par Φ , qui sont des matrices symétriques définies positives (donc appartenant à $PD(\mathbb{R})$). On a alors l'expression suivante :

$$\begin{aligned} d_c(\mathbf{M}_1, \mathbf{M}_2)^2 &= d^2(\Phi(\mathbf{M}_1), \Phi(\mathbf{M}_2)) \\ &= \frac{1}{2} \left\| \log \left(\left(\Phi(\mathbf{M}_1)^{-1/2} \right)^T \Phi(\mathbf{M}_2) \Phi(\mathbf{M}_1)^{-1/2} \right) \right\|_F^2 \\ &= \frac{1}{2} \left\| \log \left(\Phi \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right) \right\|_F^2 \\ &= \frac{1}{2} \left\| \Phi \left(\log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right) \right\|_F^2 \\ &= \frac{1}{2} \text{Tr} \left(\Phi \left(\log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right) \Phi \left(\log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right)^T \right) \\ &= \frac{1}{2} \text{Tr} \left(\Phi \left(\log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right) \Phi \left(\log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right)^H \right) \\ &= \text{Tr} \left(\log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right)^H \right) \\ d_c(\mathbf{M}_1, \mathbf{M}_2)^2 &= \left\| \log \left(\left(\mathbf{M}_1^{-1/2} \right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right\|_F^2 \end{aligned} \quad (3.30)$$

On obtient donc un résultat identique au cas réel de l'Eq. (3.18), au facteur $\frac{1}{\sqrt{2}}$ près. Là encore, le terme $\mathbf{M}_1^{1/2}$ désigne n'importe quelle racine carrée de \mathbf{M}_1 , à savoir une matrice \mathbf{A} qui vérifie l'équation $\mathbf{A}\mathbf{A}^H = \mathbf{M}_1$. L'Eq. (3.19) est toujours valable dans ce cadre, avec les λ_k les valeurs propres généralisées de \mathbf{M}_1 et \mathbf{M}_2 .

Par les mêmes considérations que pour l'Eq. (3.21), on obtient l'expression du segment de géodésique, noté $\mathbf{M}(t)$ joignant deux matrices \mathbf{M}_1 et \mathbf{M}_2 appartenant à $HPD(\mathbb{C})$ par l'équation suivante :

$$\mathbf{M}(t) = \left(\mathbf{M}_1^{1/2}\right)^H \left(\left(\mathbf{M}_1^{-1/2}\right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right)^t \mathbf{M}_1^{1/2}, t \in [0, \dots, 1]. \quad (3.31)$$

Enfin, on généralise de la même manière l'expression de la moyenne géométrique de l'Eq. (3.26) au cas des matrices hermitiennes définies positives par l'expression suivante :

$$\mathbf{P}_{n+1} = \left(\mathbf{P}_n^{1/2}\right)^H \exp \left(-\varepsilon \sum_{i=1}^N \log \left(\left(\mathbf{P}_n^{-1/2}\right)^H \mathbf{M}_i^{-1} \mathbf{P}_n^{-1/2} \right) \right) \mathbf{P}_n^{1/2}, \quad (3.32)$$

Au sein de cette partie, nous avons donc défini la distance au sens de la géométrie de l'information entre deux points de $HPD(\mathbb{C})$, c'est-à-dire deux matrices hermitiennes définies positives. Cette distance pourra donc être appliquée à des matrices de covariance polarimétrique à la place de la distance de Wishart ou de la distance de Box dans des algorithmes de classification, afin d'en étudier les performances. De même, nous avons défini l'expression donnant la moyenne au sens géométrique des matrices de covariance, qui sera utilisée pour calculer les centres de classes. Avant d'appliquer ces résultats sur des données réelles, nous allons mettre en place des simulations afin d'étudier les apports de ces deux techniques pour la classification, comparées aux techniques usuelles.

3.3 Simulations

Une procédure permettant de simuler des données polarimétriques SAR non gaussiennes a été introduite dans [64]. L'idée est de générer des vecteurs complexes gaussiens centrés et ayant pour matrice de covariance la matrice identité \mathbf{I} , et de les multiplier par une variable distribuée selon la loi de texture voulue, puis par la décomposition de Choleski de la matrice qui caractérise leur matrice de covariance, de la manière suivante :

$$\begin{aligned} \mathbf{y} &= \mathbf{L}^H(z), \text{ avec } \mathbf{L} \text{ tel que } \mathbf{M} = \mathbf{L}^H \mathbf{L}, \mathbf{z} = \tau \mathbf{x} \text{ avec } \tau \sim \mathcal{L}(\theta) \text{ et } \mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}), \\ &\Rightarrow \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbb{E}[\mathbf{L}^H \mathbf{z}\mathbf{z}^H \mathbf{L}], \\ &= \mathbf{L}^H \mathbb{E}[\tau \mathbf{x}\mathbf{x}^H] \mathbf{L}, \\ &= \tau \mathbf{L}^H \mathbf{L}, \\ \mathbb{E}[\mathbf{y}\mathbf{y}^H] &= \tau \mathbf{M}. \end{aligned} \quad (3.33)$$

Cette procédure sera utilisée pour générer des nouveaux vecteurs polarimétriques à partir de matrices de covariance connues. Pour obtenir ces matrices, la manière la plus simple de procéder est de les extraire d'une image réelle. Pour cela, on sélectionne manuellement une grande zone de l'image en supposant que les mécanismes polarimétriques varient peu en son sein et on y estime la matrice de covariance. De cette manière, le nombre d'échantillons est très élevé et on a une estimation qui est très proche de la vraie matrice. En répétant cette opération autant de fois que nécessaire sur différentes zones intéressantes de l'image (champs, bâtiments, etc), on obtient plusieurs matrices de covariance de référence. Ces matrices de référence serviront à générer de nouveaux vecteurs polarimétriques qui formeront une image SAR polarimétrique simulée. Les quatre zones choisies comme référence sont indiquées sur l'image de Brétigny sur la Fig. 3.3.

Rappelons que le modèle gaussien est bien adapté au cas où les données SAR sont homogènes mais pas dans



FIGURE 3.3 – Zones d'extraction des matrices de covariance de référence pour la simulation.

le cas de données hétérogènes, comme c'est souvent le cas pour des images SAR haute résolution. D'autres distributions que la distribution gaussienne peuvent être alors considérées pour modéliser le signal, notamment la distribution log-normale, la distribution de Weibull ou la K-distribution. Nous allons ici utiliser la K-distribution décrite au Chapitre 1 pour générer des données SAR polarimétriques. Pour cela, on utilise les matrices de covariance de référence calculées sur les quatre zones indiquées dans la Fig. 3.3 pour générer une image simulée par la procédure décrite dans l'Eq. 3.33. La structure de l'image simulée est présentée sur la Fig. 3.4. Elle est divisée en quatre quadrants en fonction de la matrice de référence qui est utilisée pour générer les données. Ces quadrants sont eux-mêmes subdivisés en quatre sous-quadrants en fonction du paramètre de forme de la K-distribution. Afin d'exacerber les différences en puissance dans chacune des seize parties de l'image, elles sont multipliées indépendamment par une puissance scalaire uniforme. La Fig. 3.5 présente un résultat possible de cette procédure de simulation.

On dispose alors d'une image simulée, de taille 200×200 pixels, sur laquelle on applique un classifieur k-moyennes avec un nombre de classes fixé à quatre, puisque l'on cherche à classifier les différentes matrices de covariance par rapport à leurs propriétés polarimétriques et pas par rapport à leur puissance. La taille du voisinage utilisé pour l'estimation de la matrice de covariance est fixée à 7×7 .

Par ailleurs, on peut choisir plusieurs paramètres au sein du classifieur :

$\sim \mathcal{K}(\mathbf{M}_1, \lambda_1)$	$\sim \mathcal{K}(\mathbf{M}_1, \lambda_2)$	$\sim \mathcal{K}(\mathbf{M}_2, \lambda_1)$	$\sim \mathcal{K}(\mathbf{M}_2, \lambda_2)$
$\sim \mathcal{K}(\mathbf{M}_1, \lambda_3)$	$\sim \mathcal{K}(\mathbf{M}_1, \lambda_4)$	$\sim \mathcal{K}(\mathbf{M}_2, \lambda_3)$	$\sim \mathcal{K}(\mathbf{M}_2, \lambda_4)$
$\sim \mathcal{K}(\mathbf{M}_3, \lambda_1)$	$\sim \mathcal{K}(\mathbf{M}_3, \lambda_2)$	$\sim \mathcal{K}(\mathbf{M}_4, \lambda_1)$	$\sim \mathcal{K}(\mathbf{M}_4, \lambda_2)$
$\sim \mathcal{K}(\mathbf{M}_3, \lambda_3)$	$\sim \mathcal{K}(\mathbf{M}_3, \lambda_4)$	$\sim \mathcal{K}(\mathbf{M}_4, \lambda_3)$	$\sim \mathcal{K}(\mathbf{M}_4, \lambda_4)$

FIGURE 3.4 – Structure de l’image simulée

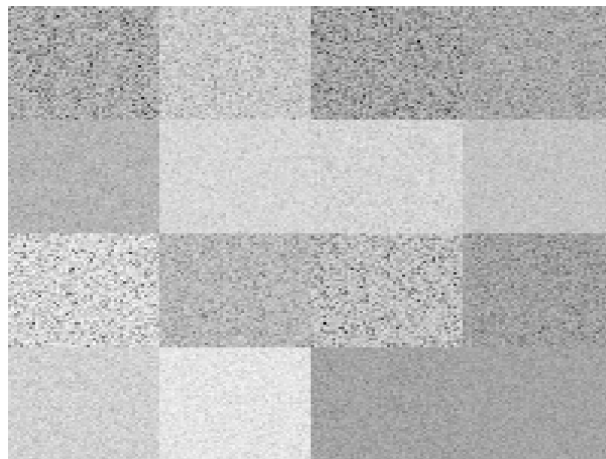


FIGURE 3.5 – Simulation de données SAR polarimétriques k-distribuées.

- l’initialisation : initialisation aléatoire ou par décomposition de Cloude-Pottier
- la distance utilisée : distance de Wishart ou distance géométrique
- l’estimateur de la matrice de covariance : SCM ou FPE
- la méthode de calcul des centres de classe : moyenne arithmétique ou moyenne géométrique

Les Fig. 3.6, 3.7, 3.8 et 3.9 présentent les résultats de classification en fonction de ces différents paramètres, ce qui implique quatre scénarios possibles :

- initialisation par décomposition de Cloude-Pottier et distance de Wishart,
- initialisation par décomposition de Cloude-Pottier et distance géométrique,
- initialisation par décomposition aléatoire et distance de Wishart,

- initialisation par décomposition aléatoire et distance géométrique.

Pour chacun de ces scénarios, les résultats de classification seront présentés pour chacune des combinaisons suivantes sur les matrices de covariance :

- utilisation de la SCM et de la moyenne arithmétique (SCM-AM),
- utilisation de la SCM et de la moyenne géométrique (SCM-GM),
- utilisation du FPE et de la moyenne arithmétique (FPE-AM),
- utilisation du FPE et de la moyenne géométrique (FPE-GM).

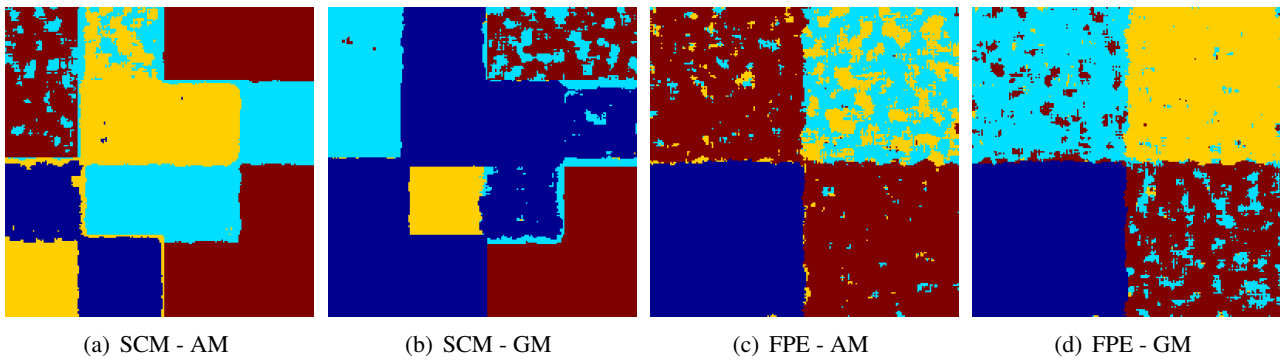


FIGURE 3.6 – Simulations avec initialisation aléatoire et distance de Wishart.

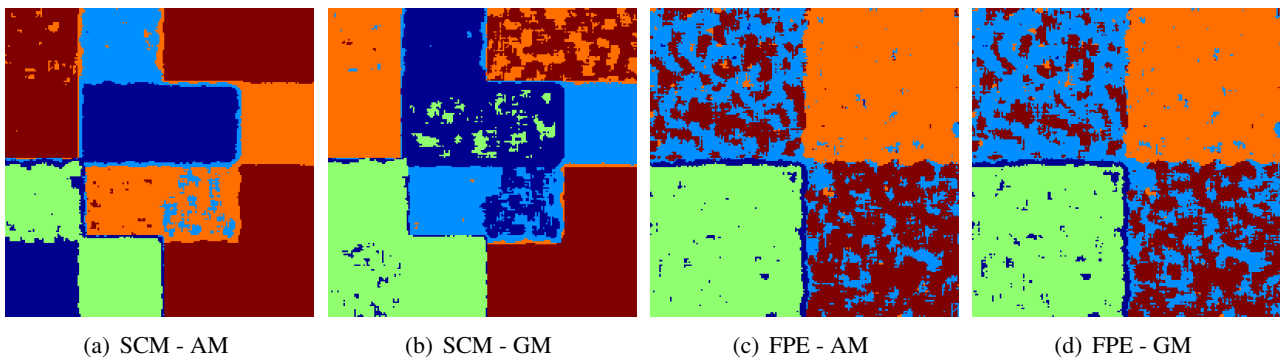


FIGURE 3.7 – Simulations avec initialisation par décomposition de Cloude-Pottier et distance de Wishart.

Dans chacun des cas, on peut une nouvelle fois remarquer la mauvaise classification engendrée par l'utilisation de la SCM. En effet, les classes finales obtenues sont clairement influencées par la puissance de chacun des quadrants. Par exemple, pour la Fig. 3.6(a), on retrouve dans la classe marron les quadrants les plus sombres (donc de puissance plus faible) de l'image de puissance (Fig. 3.5) et les trois autres classes (bleu ciel, bleu et jaune) sont aussi composées des pixels ayant plus ou moins une puissance identique. De même, l'utilisation du FPE permet à nouveau de s'affranchir de cette influence très importante de la puissance puisqu'on obtient une classification beaucoup plus proche de la classification attendue avec une image découpée en quatre quadrants

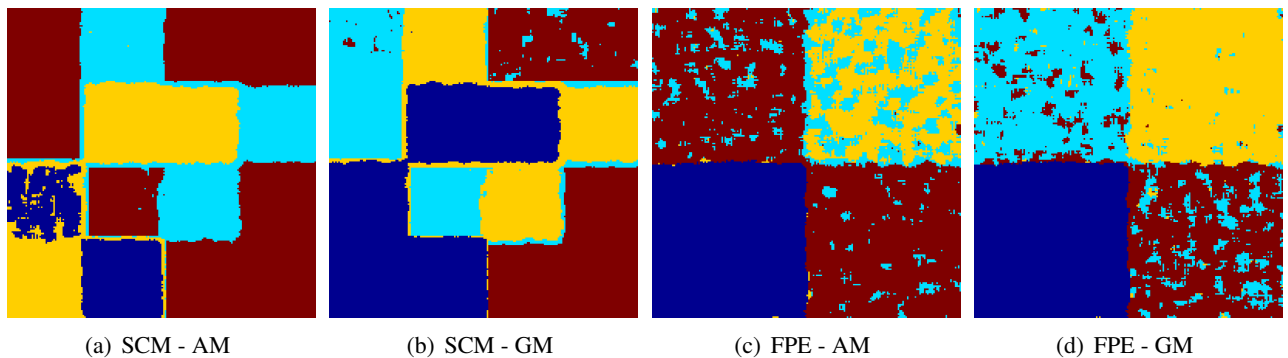


FIGURE 3.8 – Simulations avec initialisation aléatoire et distance géométrique.

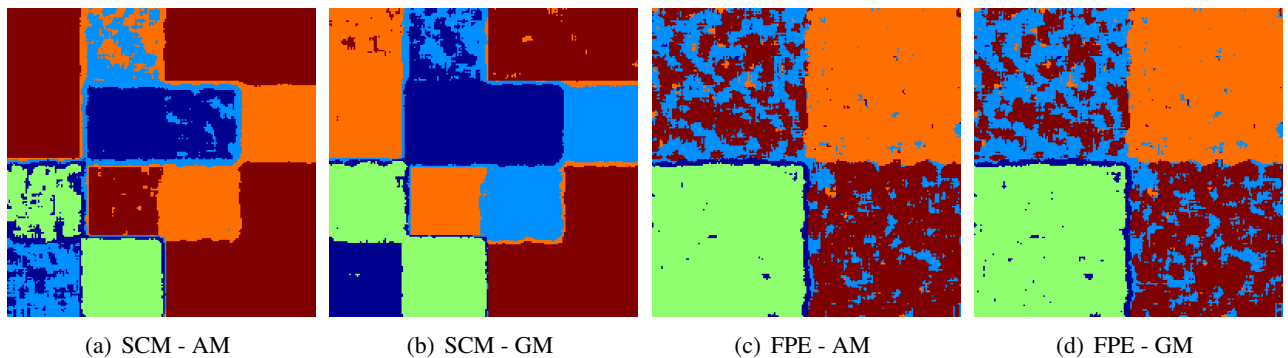


FIGURE 3.9 – Simulations avec initialisation par décomposition de Cloude-Pottier et distance géométrique.

plutôt qu'en seize sous-quadrants pour chacun des quatre scénarios possibles.

Le résultat de la comparaison entre la SCM et le FPE était attendu, au vu des résultats présentés dans le Chapitre 2. Il reste maintenant à comparer les autres choix de paramètres. En ce qui concerne le choix d'une initialisation aléatoire par rapport à une initialisation par la décomposition de Cloude-Pottier, il apparaît que les classes sont mieux séparées en utilisant une initialisation aléatoire, en comparant par exemple les Fig. 3.6(c) et 3.6(d) avec les Fig. 3.7(c) et 3.7(d). La décomposition de Cloude-Pottier introduit un *a priori* sur la répartition des classes qui peut ainsi se révéler néfaste pour la classification. Plusieurs points de départ pour l'initialisation ont été choisis, tous de manière aléatoire, et le comportement observé dans les images présentées ici a été observé pour chacune de ces initialisations aléatoires.

L'utilisation de la moyenne géométrique présente elle aussi un grand intérêt par rapport à l'utilisation de la moyenne arithmétique, comme on peut l'observer sur les Fig. 3.6(d) et 3.8(d) par rapport aux Fig. 3.6(c) et 3.8(c). En effet, dans les Fig. 3.6(c) et 3.8(c) où la moyenne arithmétique est utilisée, deux classes sont identifiées dans les quadrants inférieur gauche et supérieur droit, mais les deux classes restantes sont fusionnées en une seule. Ce comportement n'est pas reproduit lors de l'utilisation de la moyenne géométrique où les quatre classes sont correctement séparées.

Enfin, l'utilisation de la distance géométrique par rapport à la distance de Wishart n'apporte pas d'amélioration significative sur la classification, au vu des Fig. 3.6(d) et 3.8(d).

Ces résultats de simulation permettent de nouveau de valider l'utilisation du FPE, mais aussi l'utilisation d'une initialisation aléatoire par rapport à la décomposition de Cloude-Pottier et surtout de mettre en avant les meilleurs résultats engendrés par l'utilisation de la moyenne géométrique par rapport à la moyenne arithmétique. L'utilisation de la distance géométrique pour remplacer la distance de Wishart n'apporte pas spécialement d'amélioration sur ces simulations. Appliquons maintenant ces mêmes scénarios sur des images réelles.

3.4 Données réelles

Les données utilisées sont encore une fois l'image de Brétigny présentée précédemment (cf. Fig. 1.12). On applique de nouveau les quatre scénarios précédents avec les quatre combinaisons différentes sur cette image. Les résultats sont présentés sur les figures suivantes :

- initialisation aléatoire et distance de Wishart sur la Fig. 3.10,
- initialisation par décomposition de Cloude-Pottier et distance de Wishart sur la Fig. 3.11,
- initialisation aléatoire et distance géométrique sur la Fig. 3.12,
- initialisation par décomposition de Cloude-Pottier et distance géométrique sur la Fig. 3.13.

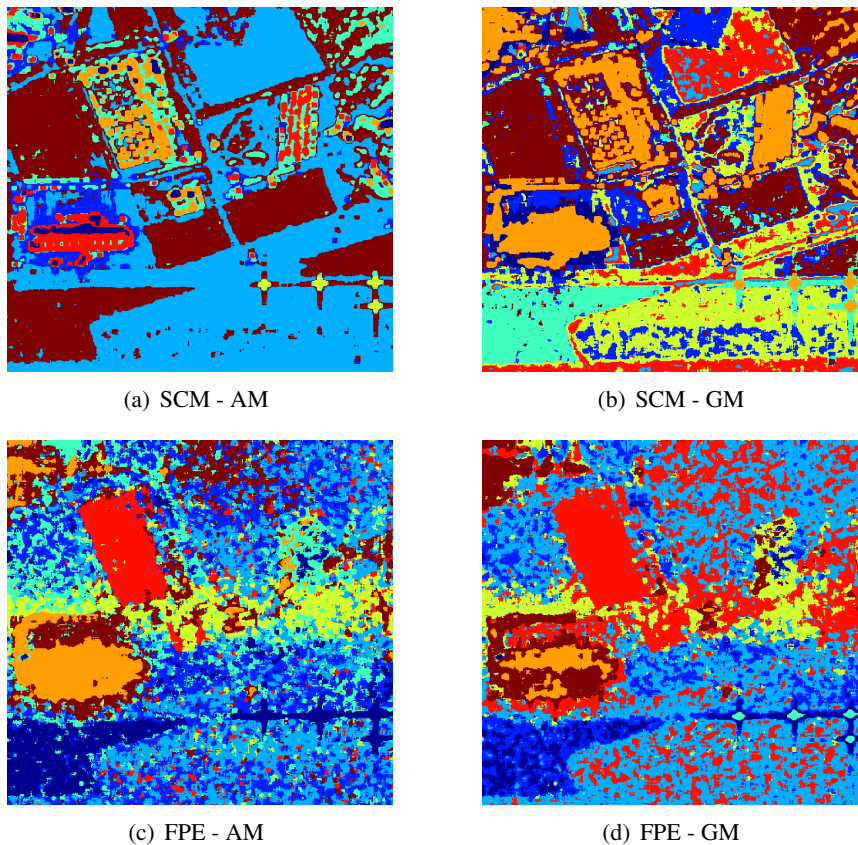


FIGURE 3.10 – Image réelle avec initialisation aléatoire et distance de Wishart.

3.4.1 Utilisation de la distance de Wishart

On retrouve une nouvelle fois une classification en utilisant la SCM très influencée par la puissance avec l'utilisation de la distance de Wishart (cf. Fig. 3.11(a) et Fig. 3.10(a)) avec un effet relativement atténué en

utilisant la moyenne géométrique au lieu de la moyenne arithmétique (cf. Fig. 3.11(b) et 3.10(b)). En effet, le bâtiment de gauche, le parking et la zone urbaine située dans le coin supérieur gauche apparaissent tous les trois moins perturbés par les variations de puissance des pixels les composant.

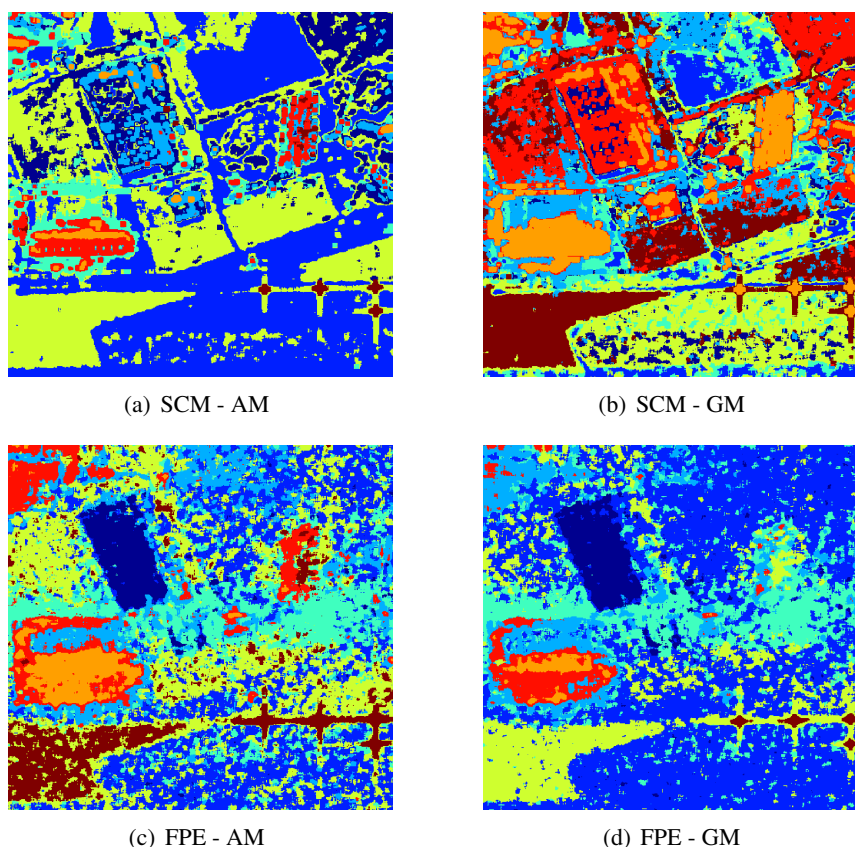


FIGURE 3.11 – Image réelle avec initialisation par décomposition de Cloude-Pottier et distance de Wishart.

Pour le FPE, on remarque peu de différences entre l'utilisation de la moyenne arithmétique et de la moyenne géométrique (cf. Fig. 3.10(c) et 3.10(d) sauf en trois endroits différents : sur le bâtiment de gauche et la zone urbaine, on remarque dans les deux cas principalement deux classes, la classe marron et la classe orange. Dans le cas FPE-AM, les deux classes sont présentes en proportions environ égales sur ces deux zones mais dans le cas FPE-GM, on ne retrouve que la classe marron dans la zone urbaine et une plus forte présence de cette même classe dans le bâtiment. En observant l'image satellite de ce bâtiment (cf. Fig. 1.12(b)), on remarque que celui-ci est entouré d'une bande de bitume qu'on peut associer à la classe orange puisqu'on la retrouve grossièrement au dessus et en dessous de la classe marron sur ce bâtiment. Un autre point plus intéressant est la classification des coins réflecteurs situés dans le coin inférieur droit. Ils sont observables très facilement sur la classification effectuée avec la SCM (à cause de leur forte puissance) mais on ne les retrouve pas sur la classification FPE-AM. Cependant, l'utilisation de la moyenne géométrique permet de retrouver ces coins réflecteurs sur la classification FPE-GM, situés dans une classe ne contenant qu'eux, contrairement aux cas SCM-AM et SCM-GM. Un meilleur calcul des centres de classes permet donc d'obtenir des informations supplémentaires.

Les différences entre l'utilisation de la décomposition de Cloude-Pottier et l'initialisation aléatoire ne sont pas très importantes. On peut remarquer que la classe rouge, correspondant au bâtiment du milieu, dans le cas d'une initialisation aléatoire est répandue un peu partout dans l'image, phénomène qu'on ne retrouve pas

pour le cas de la décomposition de Cloude-Pottier. Par ailleurs, on remarque à chaque utilisation du FPE, une bande horizontale de couleur jaune-vert sur les Fig. 3.10(c) et 3.10(d) et bleu ciel sur les Fig. 3.11(c) et 3.11(d) qui s'étend sur toute la largeur de l'image. La présence de cette bande est inexplicée à l'heure actuelle.

3.4.2 Utilisation de la distance géométrique

En utilisant la distance géométrique, on observe que la technique traditionnelle (SCM-AM) donne des résultats totalement inexploitable. L'utilisation de la moyenne géométrique permet d'obtenir des résultats que l'on peut comparer avec ceux de la distance de Wishart. Dans le cas SCM-GM, le choix de l'initialisation n'a que peu d'impact sur les résultats de classification (cf. Fig. 3.13(b) et 3.12(b)). L'influence du choix de la distance peut s'observer en comparant les Fig. 3.13(b) et 3.12(b) avec les Fig. 3.11(b) et 3.10(b) : on remarque que la classification est meilleure sur les zones homogènes et moins bonne sur les zones hétérogènes en utilisant la distance géométrique.

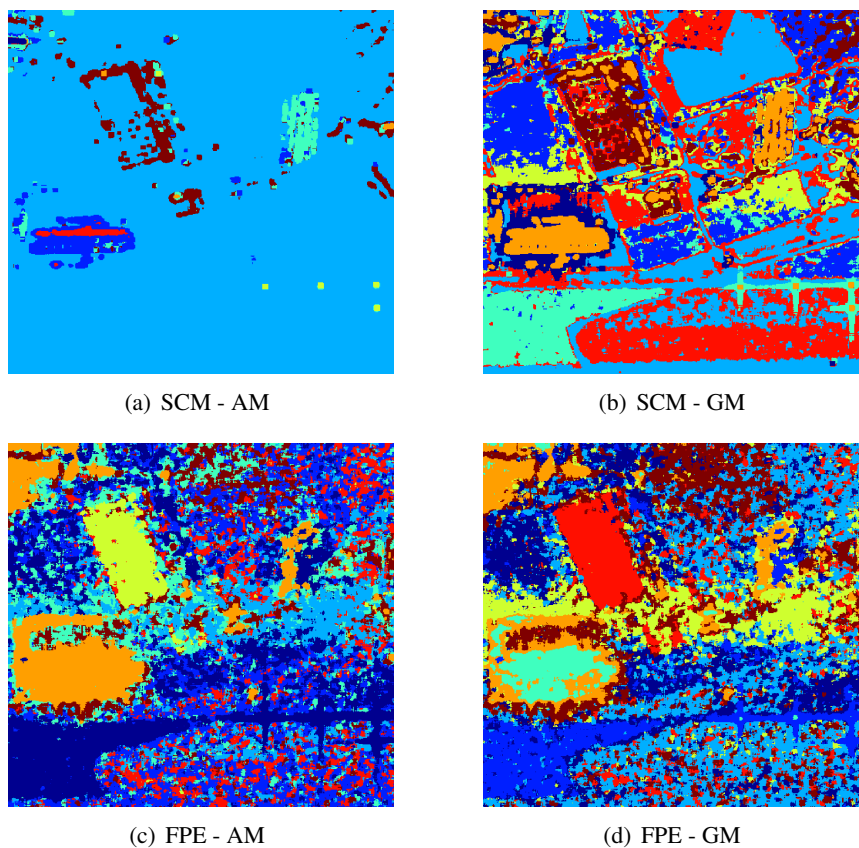


FIGURE 3.12 – Image réelle avec initialisation aléatoire et distance géométrique.

L'utilisation du FPE donne d'une manière générale des meilleurs résultats de classification sur les zones hétérogènes, sauf pour le parking qui apparaît mieux classifié dans le cas SCM-GM. L'attrait de la moyenne géométrique s'observe principalement sur les Fig. 3.10(d) et 3.11(d) puisque les coins réflecteurs situés dans le coin inférieur droit de l'image sont placés dans une seule classe contenant uniquement ces coins réflecteurs. Cependant, le cas FPE-GM avec la distance géométrique et initialisation par décomposition de Cloude-Pottier, observable sur la Fig. 3.13(d), répertorie ces coins réflecteurs dans la même classe que le champ situé dans le

coin inférieur gauche, alors que pour la même configuration en FPE-AM, deux de ces coins réflecteurs sont correctement séparés des autres pixels de l'image, même si d'une manière moins prononcée que sur les Fig. 3.10(d) et 3.11(d). On retrouve ce même comportement au niveau des coins réflecteurs sur la Fig. 3.12(d), ce qui laisse penser que l'utilisation de la distance de Wishart est plus recommandée que celle de la distance géométrique.

La moyenne géométrique semble ainsi mieux conserver la structure des matrices de covariance et permettre une classification plus pertinente.

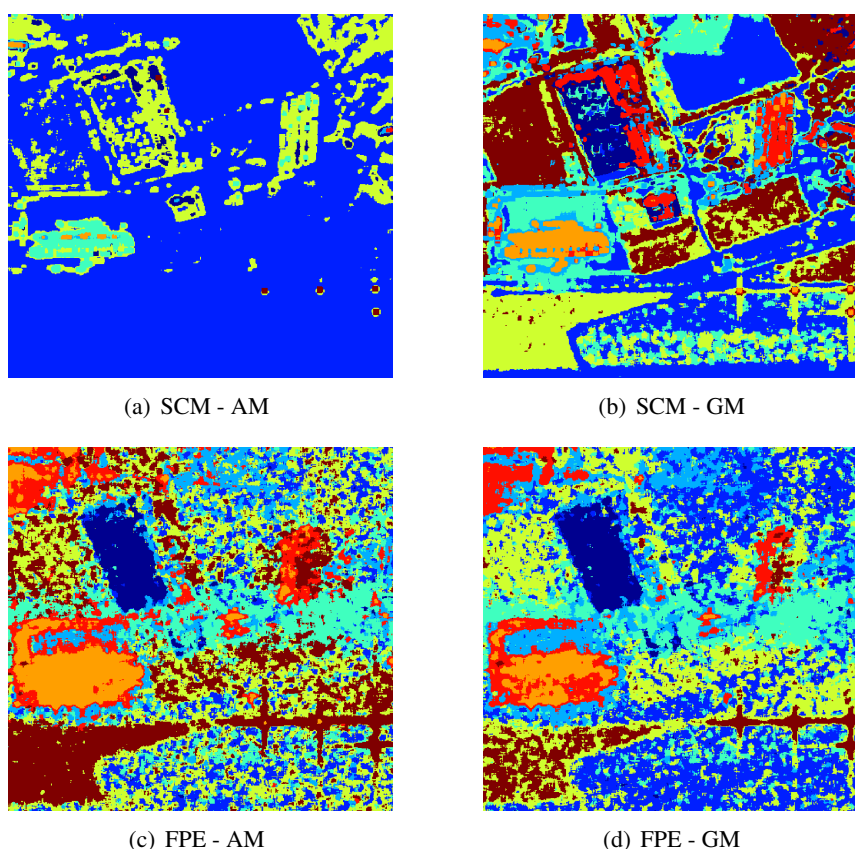
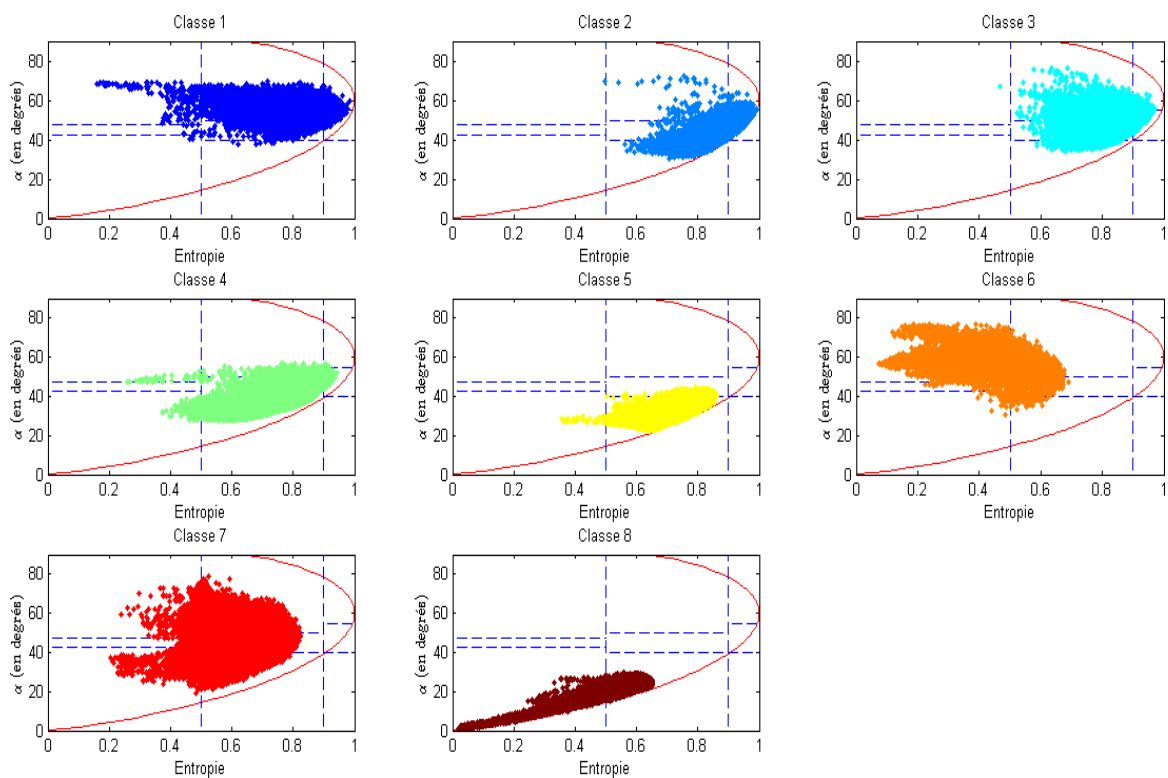


FIGURE 3.13 – Image réelle avec initialisation par décomposition de Cloude-Pottier et distance géométrique.

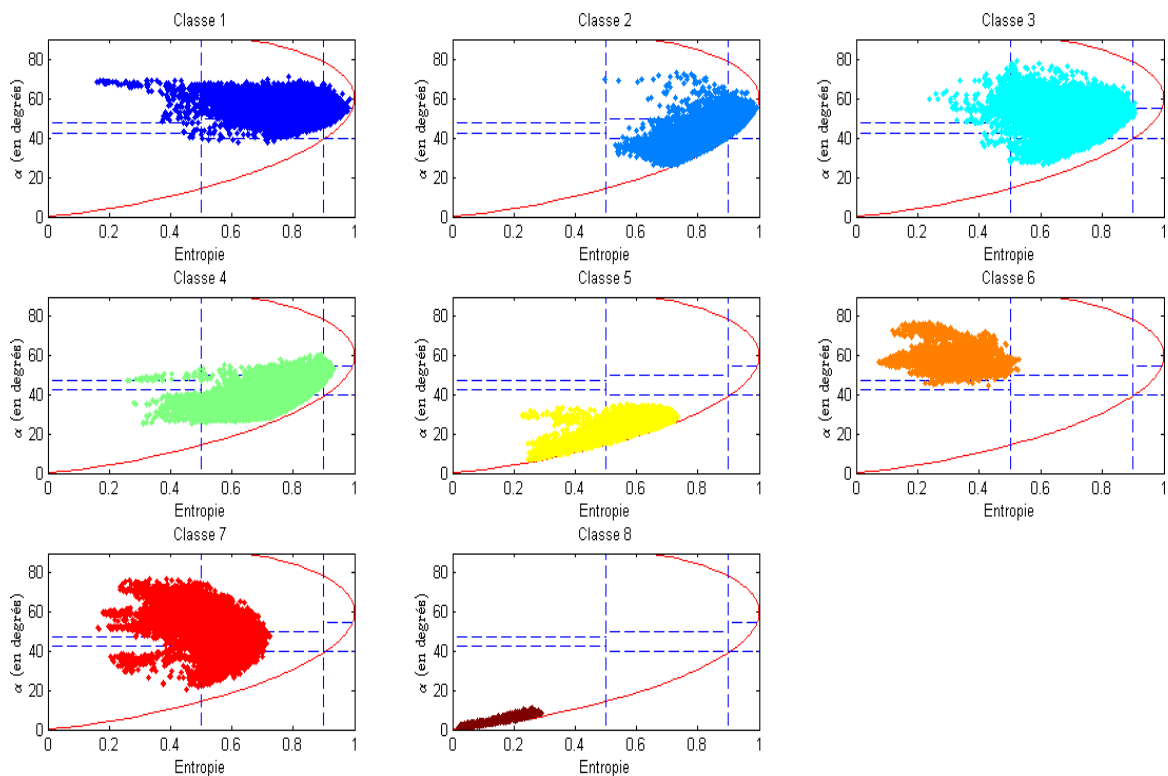
3.4.3 Répartition des pixels dans l'espace $H-\alpha$

On s'intéresse maintenant à la répartition des pixels dans l'espace $H-\alpha$ pour le cas d'une initialisation par décomposition de Cloude-Pottier et distance de Wishart (cf. Fig. 3.11). Les résultats sont présentés sur la Fig. 3.14 pour chaque classe de manière séparée.

On peut observer une grande différence dans la répartition au niveau de la classe marron, puisque c'est celle qui concerne les coins réflecteurs. Dans le cas FPE-AM (Fig. 3.14(a)), les coins réflecteurs sont situés dans la même classe que le champ du coin inférieur gauche, contrairement au cas FPE-GM (Fig. 3.14(b)) où la classe marron les contient uniquement. Il apparaît alors qu'ils se comportent comme une surface, en termes de mécanisme de rétrodiffusion. Ceci n'est pas consistant avec leur nature physique, qui aurait dû les placer avec un α supérieur à 50° . Ce comportement inattendu permet d'expliquer la raison pour laquelle le regroupement



(a) FPE - AM

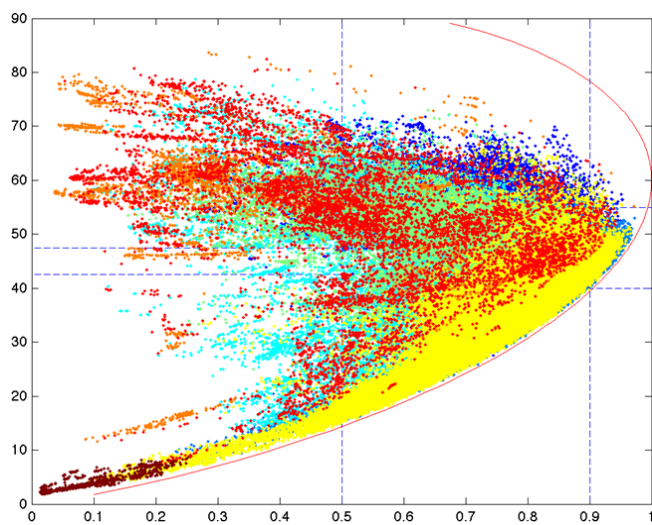


(b) FPE - GM

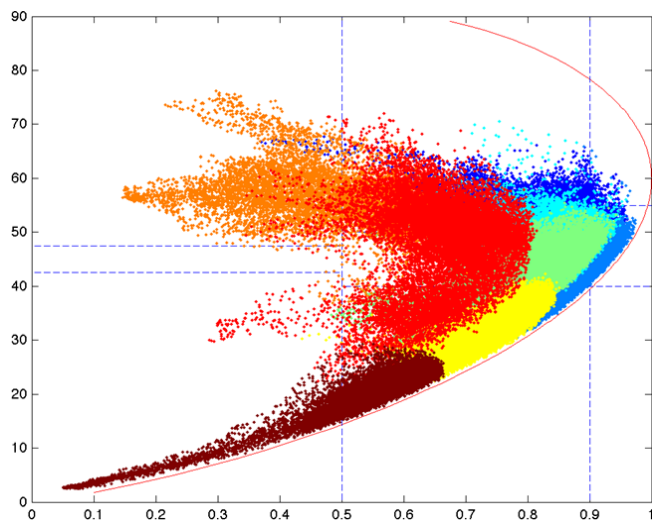
FIGURE 3.14 – Répartition par classe des pixels dans l'espace H- α - FPE, décomposition de Cloude-Pottier et distance de Wishart

des coins réflecteurs et du champ du coin inférieur gauche a lieu en utilisant la moyenne arithmétique puisqu'ils ont, dans ce cas, un comportement polarimétrique proche. Cependant, l'utilisation de la moyenne géométrique permet de mieux séparer les différentes composantes de la classe marron obtenue avec la moyenne géométrique.

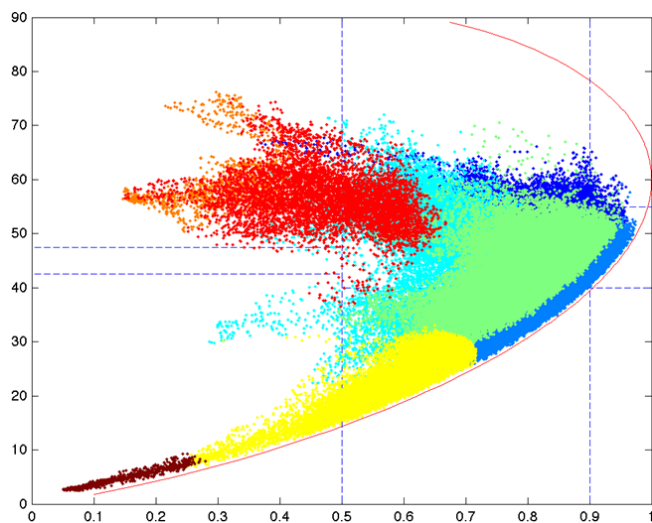
La Fig.3.15 représente la répartition des pixels pour toutes les classes. On peut immédiatement observer l'influence du FPE par rapport à la SCM en comparant les Fig. 3.15(a) et 3.15(b). Pour la SCM, on observe un éclatement total des classes au sein du plan H- α , tandis que pour le FPE, les nuages de points sont beaucoup plus resserrés, ce qui signifie que les pixels au sein de chaque classe sont plus proches au sens polarimétrique. On observe peu de différences en termes de dispersion entre l'utilisation de la moyenne arithmétique et de la moyenne géométrique.



(a) SCM - AM



(b) FPE - AM



(c) FPE - GM

FIGURE 3.15 – Répartition totale des pixels dans l'espace $H-\alpha$ - FPE, décomposition de Cloude-Pottier et distance de Wishart

3.5 Conclusion

Ce chapitre a présenté l'utilisation d'une branche de la géométrie, appelée géométrie de l'information, pour améliorer le calcul de la moyenne de matrices de covariance polarimétriques ainsi que la distance entre différentes matrices. Les calculs permettant d'obtenir une expression de cette moyenne ont d'abord été rappelés pour le cas réel, suivis de leur extension au cas complexe. Une expression de la distance au sens géométrique entre deux matrices est de même définie. Ces techniques sont appliquées d'abord sur des images simulées, sur lesquelles il est possible d'observer une amélioration de la classification, caractérisée par une meilleure séparation entre pixels générés par des matrices de covariance différentes. Ces techniques sont ensuite appliquées sur des données réelles, où l'on observe de nouveau une meilleure séparation des différentes caractéristiques de l'image comparée au cas de l'utilisation de la moyenne arithmétique.

Extension à l'imagerie hyperspectrale

Sommaire

4.1	Principe de l'imagerie hyperspectrale	95
4.1.1	Principe de l'imagerie hyperspectrale	95
4.1.2	Classification pour l'imagerie hyperspectrale	96
4.2	Application : données réelles	98
4.2.1	Jeu de données	98
4.2.2	Algorithmes et résultats	99
4.3	Conclusion	106

Ce chapitre est dédié à la présentation de travaux préliminaires d'extension de ces techniques de classification pour un autre type d'imagerie que l'imagerie radar, à savoir l'imagerie hyperspectrale. Dans un premier temps, nous présenterons le principe de l'imagerie hyperspectrale. Ensuite, un algorithme de classification classique de l'imagerie hyperspectrale sera introduit et comparé aux techniques de classification tirées de l'imagerie SAR polarimétrique et présentées dans les chapitres précédents. Enfin, nous tirerons quelques conclusions sur la pertinence de cette extension.

4.1 Principe de l'imagerie hyperspectrale

4.1.1 Principe de l'imagerie hyperspectrale

L'imagerie hyperspectrale consiste à utiliser une grande partie du spectre électromagnétique (principalement l'infra-rouge et le domaine de la lumière visible) et à illuminer une même scène dans une multitude de bandes de fréquences, étroites et contiguës, ceci afin de pouvoir recréer cette grande partie du spectre pour chaque pixel de l'image. Contrairement à l'imagerie SAR où l'on ne dispose que d'une seule bande, avec éventuellement des polarisations différentes qui augmentent le nombre de données pour chaque pixel de l'image, on se retrouve ici avec des vecteurs de données contenant plusieurs dizaines, voire plus d'une centaine d'acquisitions pour chaque pixel. Ces vecteurs contiennent une grande quantité d'information reliée à la nature physique des matériaux rencontrés par l'onde émise par le système d'acquisition. Cependant, cette information, comme dans le cas de l'imagerie SAR, n'est pas accessible directement et nécessite des traitements afin de pouvoir remonter aux caractéristiques physiques des matériaux.

Des traitements préliminaires sont nécessaires pour corriger un certain nombre d'effets indésirables, comme des défauts du capteur, prendre en compte le relief de la zone imagée ou les effets atmosphériques, notamment les phénomènes d'absorption liés à la vapeur d'eau ou aux différents gaz qui composent l'atmosphère. Certaines longueurs d'ondes sont ainsi presque totalement absorbées au cours du trajet de l'onde dans l'atmosphère avant même d'avoir atteint la surface. Bien qu'on puisse en tirer de l'information sur les couches supérieures de l'atmosphère, l'utilité pour la classification de surfaces est nulle et il est nécessaire de compenser ces effets d'absorption en particulier — et les effets atmosphériques en général — et de retirer les bandes du spectre.

4.1.2 Classification pour l'imagerie hyperspectrale

La difficulté de la classification en imagerie hyperspectrale réside dans plusieurs points :

- un nombre élevé de bandes,
- une variation élevée de la signature spectrale au sein d'une image,
- un coût élevé pour établir une vérité terrain,
- la qualité des données et des traitements correctifs.

Le premier point pose notamment beaucoup de problèmes et est relié au *fléau de la dimension* : le nombre de dimensions spectrales est très élevé donc les points représentés dans cet espace par les pixels de l'image sont très espacés les uns des autres, dans un espace quasiment vide. Il est donc très difficile d'extraire de l'information de manière statistique dans une telle configuration. C'est de plus un problème dans le cadre de l'estimation des matrices de covariance : pour l'imagerie SAR polarimétrique, les vecteurs comportaient trois éléments complexes, donc l'estimation de la matrice de covariance était très rapide, même dans le cas du FPE où une inversion de matrice est nécessaire, comme indiqué dans l'Eq. (1.57). Pour l'imagerie hyperspectrale, ces vecteurs comportent des dizaines, voir plus d'une centaine d'éléments réels. L'estimation de la matrice de covariance par la méthode du point fixe risque donc de prendre un temps beaucoup plus élevé.

On peut classer de manière grossière les algorithmes de classification pour l'imagerie hyperspectrale en deux catégories, comme pour l'imagerie SAR polarimétrique :

- les algorithmes de classification supervisée : on cherche à trouver des règles permettant de classifier les données, à partir d'un jeu de données d'apprentissage déjà classifié. Parmi ces méthodes, on peut notamment citer les réseaux de neurones [13, 23, 80], qui peinent quand le nombre de bandes est élevé. Plus récemment, des algorithmes se basant sur des méthodes à noyaux et en particulier sur les machines à vecteurs de support (ou Support Vector Machines, SVM) ont été utilisés avec succès pour la classification en imagerie hyperspectrale [21, 40, 42]. Cependant, toutes ces méthodes sont très dépendantes de la base d'apprentissage, tirée idéalement de la vérité terrain et qui est donc souvent absente, de petite taille ou de mauvaise qualité. Même dans le cas où cette base d'apprentissage serait de bonne qualité, elle n'est utile que pour des images prises exactement dans les mêmes conditions, ce qui limite énormément leur application.
- les algorithmes de classification non supervisée : ici, on ne dispose pas de base d'apprentissage, c'est la même situation que nous avons utilisée pour l'imagerie SAR polarimétrique. L'algorithme fonctionne uniquement avec l'image que l'on cherche à classifier et essaie de répartir les pixels dans des classes selon un ou des critères spécifiques. Ces méthodes [18, 79, 84] ont été appliquées avec de bons résultats à l'imagerie hyperspectrale, même si les classes données par les algorithmes ne correspondent pas toujours à des vérités physiques et qu'un pré-traitement est nécessaire pour réduire la dimension de l'espace.

On peut aussi citer les méthodes d'apprentissage semi-supervisées qui combinent les deux approches [20].

Etant donnée l'importante quantité d'informations présentes dans le vecteur de données, noté \mathbf{k} , la plupart des algorithmes de classification utilisent directement ce vecteur sous forme de variable aléatoire, contrairement au cas de l'imagerie SAR où la matrice de covariance — moment du second ordre — est beaucoup plus fréquemment utilisée. Il convient alors, pour ces travaux préliminaires, d'utiliser une technique faisant appel à la matrice de covariance du vecteur \mathbf{k} , puisque c'est sur cet objet que les travaux ont été réalisés pour l'imagerie SAR. Pour cela, nous allons utiliser en lieu et place de la distance de Wishart une mesure de distance appelée la distance de Mahalanobis [54] et décrite par l'équation suivante :

$$d_{Maha}(\mathbf{k}, \mathbf{M}) = \sqrt{(\mathbf{k} - \mu)^T \mathbf{M}^{-1} (\mathbf{k} - \mu)^T} \quad (4.1)$$

avec \mathbf{k} un vecteur tiré d'un groupe de valeurs ayant pour moyennes le vecteur μ et pour matrice de covariance \mathbf{M} .

Remarquons que si la matrice de covariance \mathbf{M} est la matrice identité, la distance de Mahalanobis se réduit à la distance euclidienne classique. L'introduction de \mathbf{M}^{-1} dans l'Eq. (4.1) permet de rendre compte de la corrélation entre les différentes composantes du vecteur \mathbf{k} . La Fig. 4.1 illustre les différences entre la distance de Mahalanobis et la distance euclidienne et permet de clarifier les différents termes utilisés dans l'Eq. (4.1). Un échantillon de cent valeurs, représenté par des cercles, a été tiré selon une loi normale bivariée, de moyenne nulle et de matrice de covariance $\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. Quatre nouvelles valeurs, représentées par des étoiles, sont tirées et comparées aux données existantes. Selon la distance euclidienne, ces quatre nouvelles observations sont très proches des données existantes, comme illustré sur la figure de droite (valeurs comprises entre 1.8 et 2.2). Comme la distance euclidienne ne prend en compte que des considérations géométriques, ce comportement est normal mais pas celui désiré. En revanche, la distance de Mahalanobis, illustrée sur la figure de gauche, permet de bien faire la différence entre les deux observations qui sont clairement corrélées aux données (en violet, donc valeurs autour de 5) et les deux observations qui sont en dehors du nuage de points (en jaune, donc valeurs supérieures à 20). La distance de Mahalanobis revient en réalité à transformer les données en données de moyenne nulle et ayant pour matrice de covariance la matrice identité et à appliquer la distance euclidienne sur ces données transformées. Elle permet alors de quantifier la similarité d'une observation par rapport à un groupe de valeurs caractérisées par leurs moyennes et leur matrices de covariance.

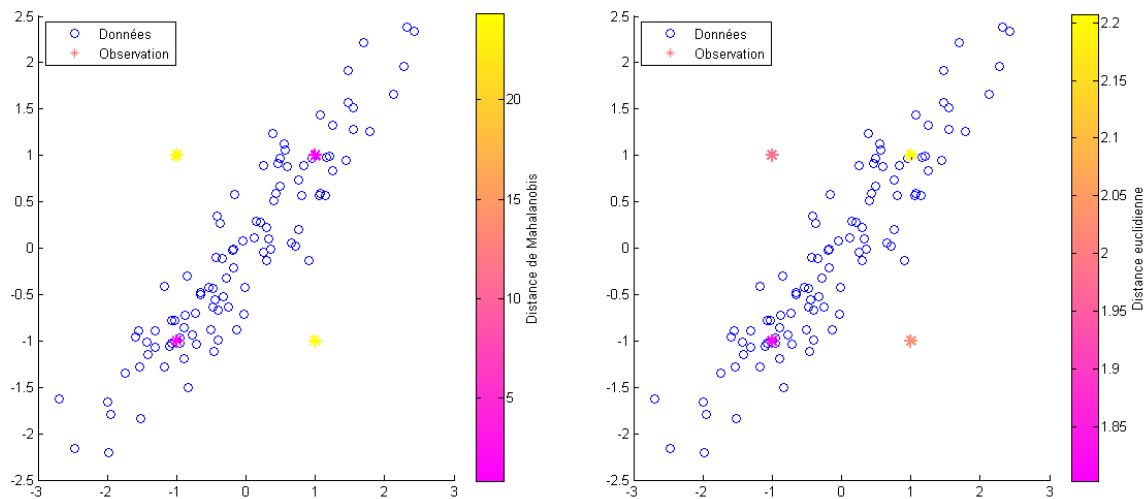


FIGURE 4.1 – Différence entre distance de Mahalanobis et distance euclidienne

Le point le plus intéressant dans le cas présent est effectivement la présence de la matrice de covariance dans l'expression de cette distance. Toutes les considérations étudiées pour le cas de l'imagerie SAR, à savoir le choix de l'estimateur de la matrice de covariance et l'utilisation de la géométrie de l'information pour le calcul des centres de classes peuvent être réutilisées dans le cas de l'imagerie hyperspectrale. Rappelons que dans le chapitre 3, les expressions de la distance géométrique et du barycentre des matrices de covariance ont été d'abord obtenues pour le cas des matrices réelles symétriques définies positives avant d'être étendues au cas des matrices hermitiennes définies positives. On peut donc directement utiliser les Eqs. (3.19) pour la distance et (3.26) pour la moyenne.

Pour comparer les résultats de classification à un algorithme classique, on utilise un algorithme de type k-moyennes utilisant le Spectral Angle Mapper (SAM). Le SAM consiste à comparer deux vecteurs de données \mathbf{k}_1 et \mathbf{k}_2 dans l'espace spectral et de calculer l'angle qui les sépare. On compare généralement les pixels de l'image à des spectres de référence (aussi appelés *endmembers*) obtenus soit par une base de données soit en choisissant des régions d'intérêt de l'image. Dans notre application, ces endmembers sont équivalents aux centres de classe utilisés dans les sections précédentes. L'équation du SAM est la suivante :

$$\delta = \cos^{-1} \left(\frac{\mathbf{k}_1 \mathbf{k}_2}{\|\mathbf{k}_1\| \|\mathbf{k}_2\|} \right), \quad (4.2)$$

avec δ l'angle entre les deux vecteurs \mathbf{k}_1 et \mathbf{k}_2 et $\|\cdot\|$ l'opérateur norme.

4.2 Application : données réelles

4.2.1 Jeu de données

Le jeu de données utilisé a été acquis par le système Hyperion, embarqué sur le satellite EO-1 de la NASA, dans les environs de Singapour en 2002. La résolution au sol est de $30\text{m} \times 30\text{m}$. L'image est composée de 3300×255 pixels et de 242 bandes spectrales. Une zone de 800×250 pixels est extraite de cette image et les bandes correspondant à l'absorption de la vapeur d'eau sont retirées, en raison de leur inexploitabilité. La Fig. 4.2 représente cette zone réduite : la Fig. 4.2(a) correspond à la composition colorée fournie par la NASA et la Fig. 4.2(b) à une composition colorée obtenue en combinant plusieurs bandes à l'aide du logiciel Matlab.

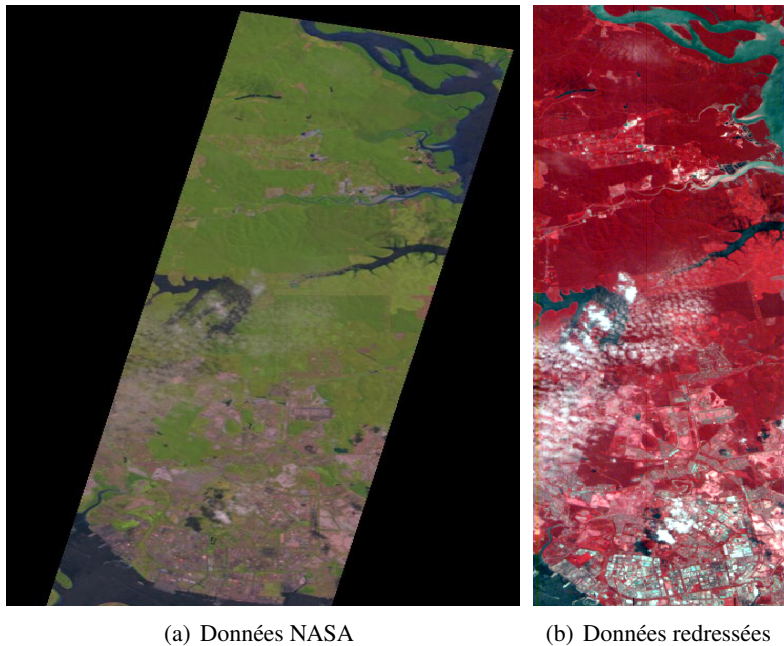


FIGURE 4.2 – Jeu de données hyperspectral Hyperion.

Ce jeu de données présente de même une grande variété de caractéristiques : plusieurs zones urbaines, collines, rivières, marécages, océan, etc.

4.2.2 Algorithmes et résultats

Un algorithme de type K-moyennes est employé pour comparer les différentes combinaisons possibles, qui sont très voisines de celles étudiées dans les chapitres précédents :

- choix de l'estimateur de la matrice de covariance :
 - SCM de l'Eq. (1.35),
 - FPE de l'Eq. (1.57),
- choix de la distance utilisée :
 - distance de Mahalanobis de l'Eq. (4.1),
 - distance de Box de l'Eq. (2.40),
 - distance géométrique de l'Eq. (3.19),
 - SAM de l'Eq. (4.2),
- choix du moyennage :
 - moyenne arithmétique traditionnelle de l'Eq. (1.53),
 - moyenne géométrique de l'Eq. (3.26)

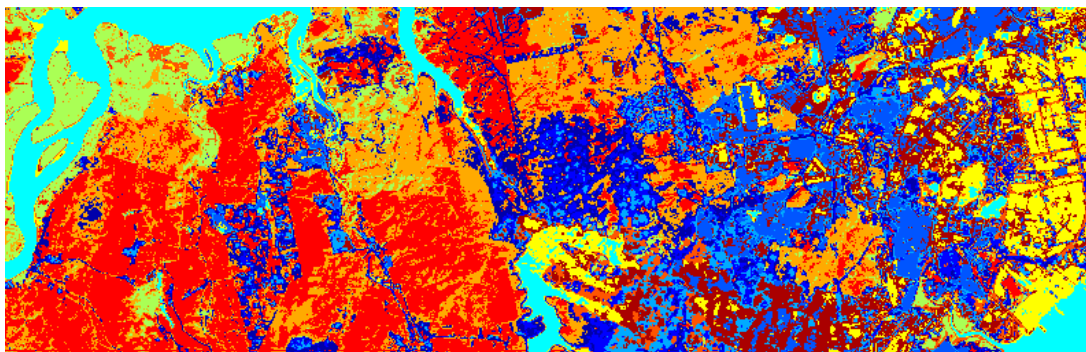
Le nombre de classes de l'algorithme K-moyennes est fixé à douze, de manière arbitraire et dix bandes sont sélectionnées à la main, de manière à représenter la majeure partie du spectre disponible. Les résultats seront comparés à l'application du SAM avec la SCM et la moyenne arithmétique puisqu'il s'agit de la manière traditionnelle de procéder. Les résultats de référence obtenus pour le SAM sont présentés sur la Fig. 4.3. Il est important de noter que dans le cas du SAM, la matrice de covariance n'intervient pas, le seul paramètre est le vecteur de données : il n'y a ainsi aucune influence du choix de l'estimateur ou du type de moyennage. On y repère aisément les rivières et la mer dans la classe bleu ciel, la plupart des zones naturelles dans les classes rouge et orange, les zones naturelles proches des étendues d'eau en vert et les zones plus urbanisées en jaune, marron et bleu foncé. Comparons maintenant ces résultats avec ceux obtenus pour la SCM et la moyenne arithmétique, observables sur la Fig. 4.4.

Cas SCM-AM : on s'aperçoit facilement que toutes les distances proposées entraînent des résultats de classification bien plus mauvais que ceux obtenus avec la distance tirée du SAM. Avec la distance de Mahalanobis, certaines zones caractéristiques apparaissent dans la partie gauche de l'image mais avec une précision beaucoup plus faible que celle obtenue pour le cas du SAM. C'est encore pire avec les distances de Box et géométrique où la classification échoue totalement. On voit néanmoins déjà apparaître un inconvénient de l'utilisation de distances directement reliées à la matrice de covariance puisque la précision est beaucoup plus faible, dû au moyennage nécessaire à l'estimation de la matrice de covariance. En effet, le SAM et la distance de Mahalanobis utilisent uniquement le vecteur de données hyperspectral \mathbf{k} du pixel pour calculer la distance entre un pixel et un centre de classe (pour la distance de Mahalanobis, la matrice située dans l'Eq. (4.1) est le centre de classe, la donnée du pixel est bien le vecteur \mathbf{k}), contrairement à la distance de Box ou la distance géométrique où la donnée utilisée pour calculer effectivement la distance entre un pixel et un centre de classe est la matrice de covariance, qui est estimée grâce à un voisinage du pixel, ce qui entraîne la perte de précision.

Cas SCM-GM : l'utilisation de la moyenne géométrique n'engendre pas de changement pour le SAM (cf. Fig. 4.5(a)), ce qui est logique car le SAM n'utilise pas du tout la matrice de covariance. Pour les autres distances, on aperçoit cependant des changements. Pour la distance de Mahalanobis, on observe une bien meilleure précision sur la partie gauche de l'image, qui est composée principalement de zones naturelles. Les rivières ressortent notamment beaucoup plus clairement. Sur la partie droite de l'image, composée en majorité de zones urbaines, la classification n'est cependant pas satisfaisante puisque l'on aperçoit principalement une



(a) Jeu de données



(b) SCM - SAM - AM

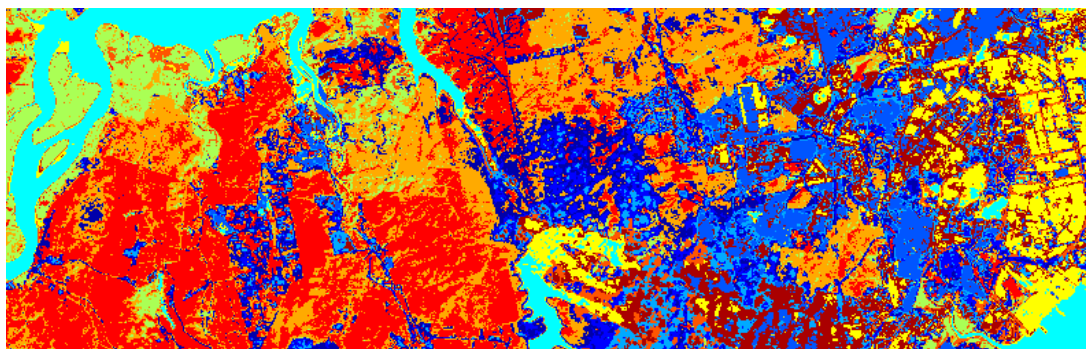
FIGURE 4.3 – Résultats de classification en utilisant la SCM, le SAM et la moyenne arithmétique

large zone bleue, alors que le SAM permettait de faire ressortir les zones de bâtiments et les routes de manière prononcée. Avec la distance géométrique (cf. Fig. 4.4(d)), la classification est une nouvelle fois meilleure que lors de l'utilisation de la moyenne arithmétique, au sens où il est possible d'identifier grossièrement certaines zones caractéristiques de l'image. On retrouve cependant le manque de précision dû à l'estimation de la matrice de covariance. Finalement, l'utilisation combinée de la distance de Box et de la moyenne géométrique fait totalement échouer la classification.

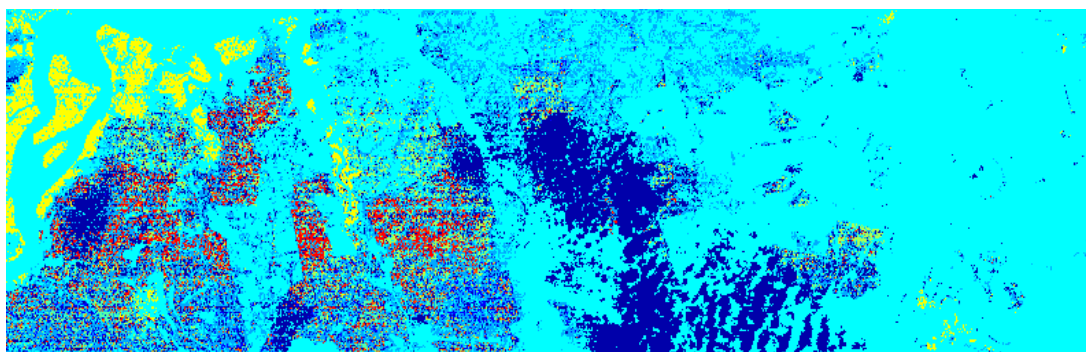
Cas FPE-AM : l'utilisation du SAM avec le FPE (cf. Fig. 4.6(a)) entraîne exactement les mêmes résultats qu'avec la SCM, puisque la matrice de covariance n'intervient pas du tout dans le calcul. Pour la distance de Mahalanobis (cf. Fig. 4.6(b)), le résultat de classification est meilleur qu'avec la SCM-AM (cf. Fig. 4.4(b)) avec une bonne séparation des zones naturelles (en bleu foncé) et des zones urbaines (principalement en orange) mais une mauvaise séparation au sein de ces deux types de zones. De même, les classifications avec la distance de Box (cf. Fig. 4.6(c)) et la distance géométrique (cf. Fig. 4.6(d)) sont meilleures qu'avec la SCM (cf. Fig. 4.4(c) et 4.4(d)) mais restent décevantes comparées aux classifications utilisant le SAM et la distance de Mahalanobis à cause de la mauvaise précision.

Cas FPE-GM : une fois de plus, l'utilisation de la moyenne géométrique n'engendre pas de changement pour le SAM (cf. Fig. 4.7(a)). Pour les autres distances, on observe de nouveau des améliorations par rapport à la moyenne arithmétique. Pour la distance de Mahalanobis (cf. Fig. 4.7(b)), la séparation entre les classes au sein des zones naturelles et des zones urbaines est grandement améliorée. Le résultat reste cependant moins intéressant que le SAM, car trop pixellisé sur la partie gauche de l'image et trop homogène sur la partie droite. Pour la distance géométrique, le même constat que lors de l'utilisation de la SCM-GM peut être fait : classes

mieux dessinées mais avec une précision trop faible. Pour la distance de Box, on retrouve le même scénario que dans le cas de la SCM avec une classification encore moins bonne dans le cas Box-GM que dans le cas Box-AM.



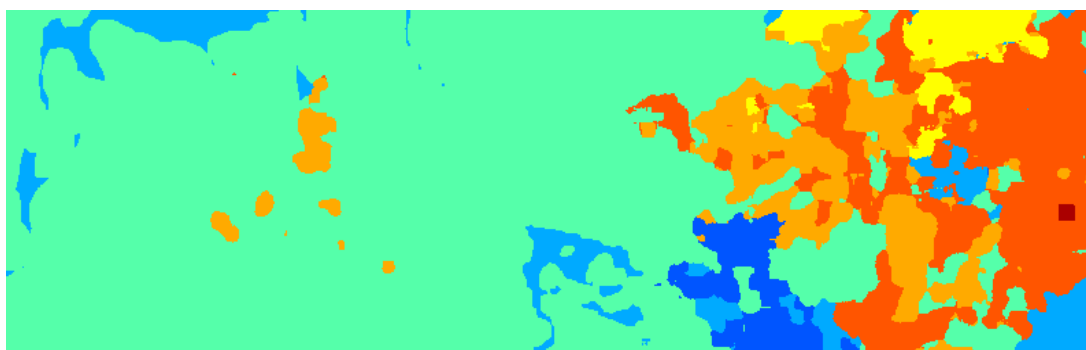
(a) SCM - SAM - AM



(b) SCM - Mahalanobis - AM

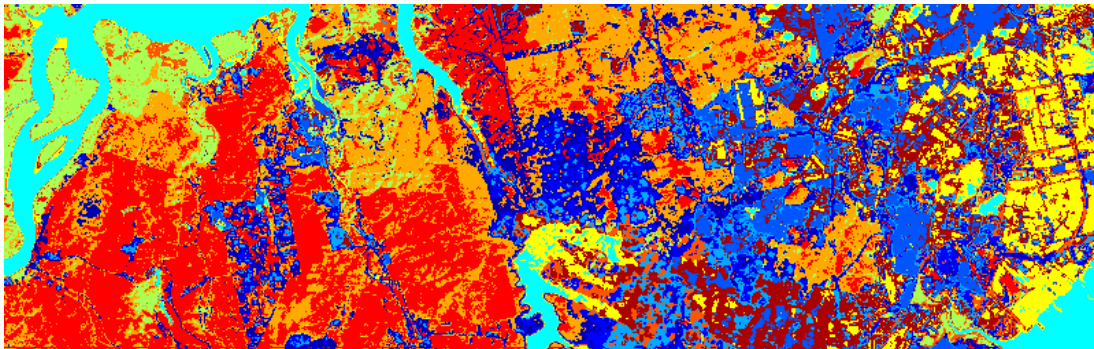


(c) SCM - Box - AM

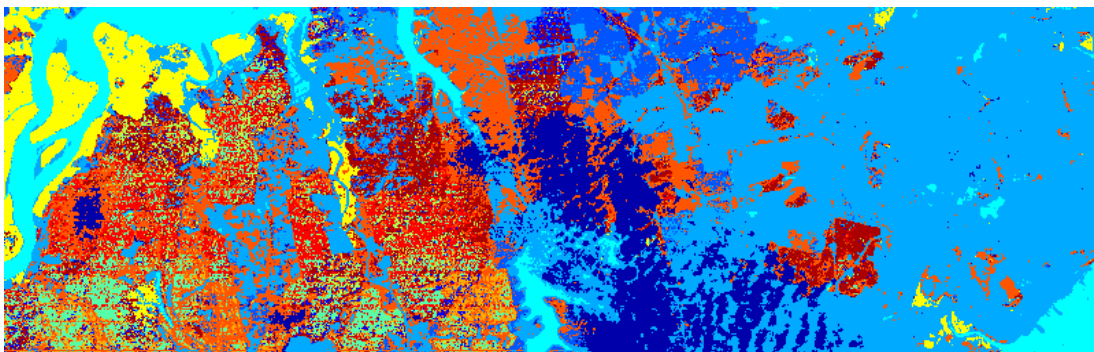


(d) SCM - Géométrique - AM

FIGURE 4.4 – Résultats de classification en utilisant la SCM et la moyenne arithmétique



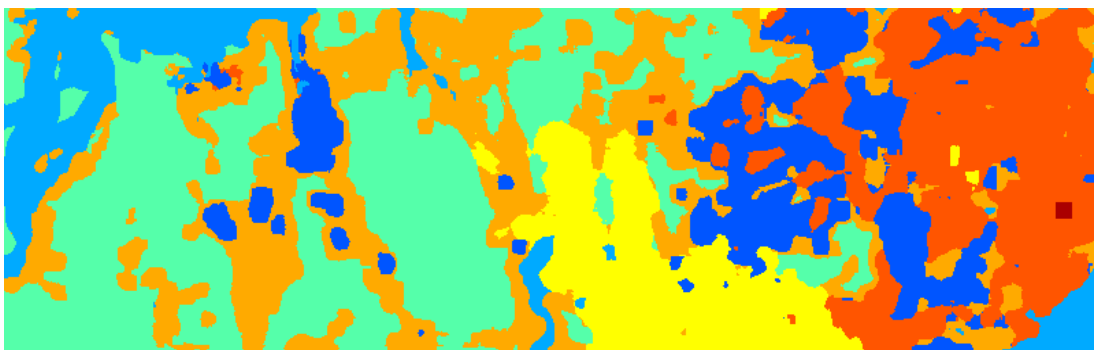
(a) SCM - SAM - GM



(b) SCM - Mahalanobis - GM

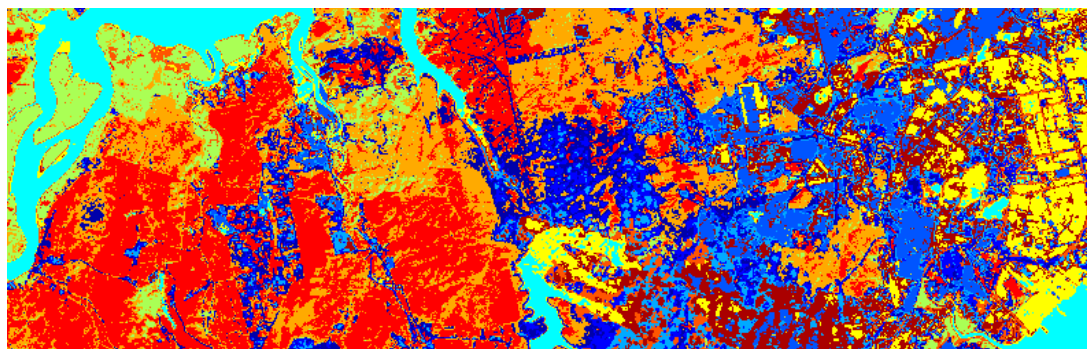


(c) SCM - Box - GM

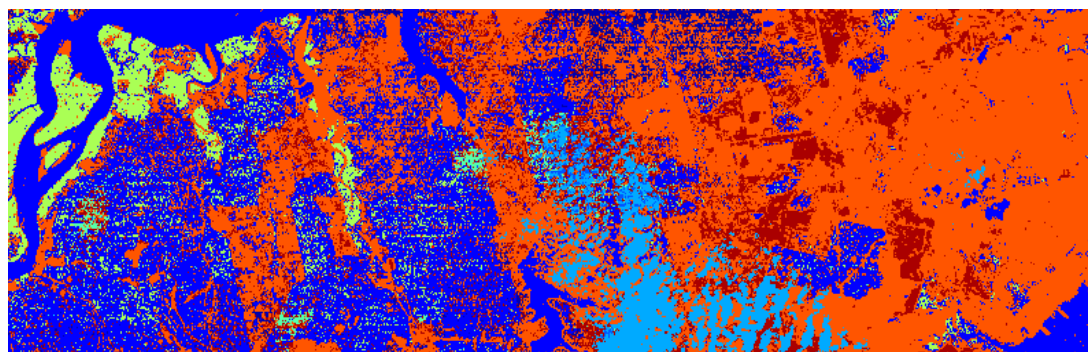


(d) SCM - Géométrique - GM

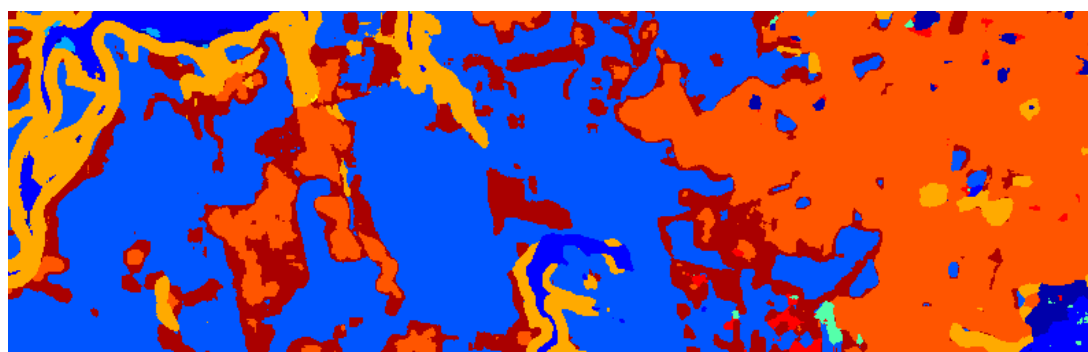
FIGURE 4.5 – Résultats de classification en utilisant la SCM et la moyenne géométrique



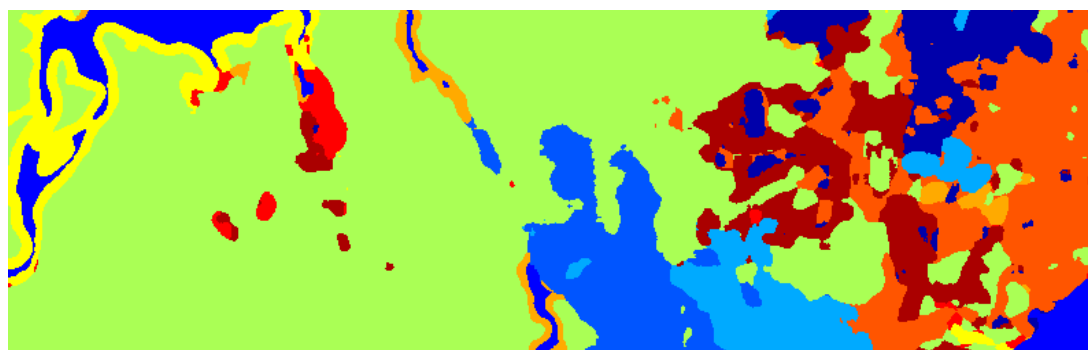
(a) FPE - SAM - AM



(b) FPE - Mahalanobis - AM

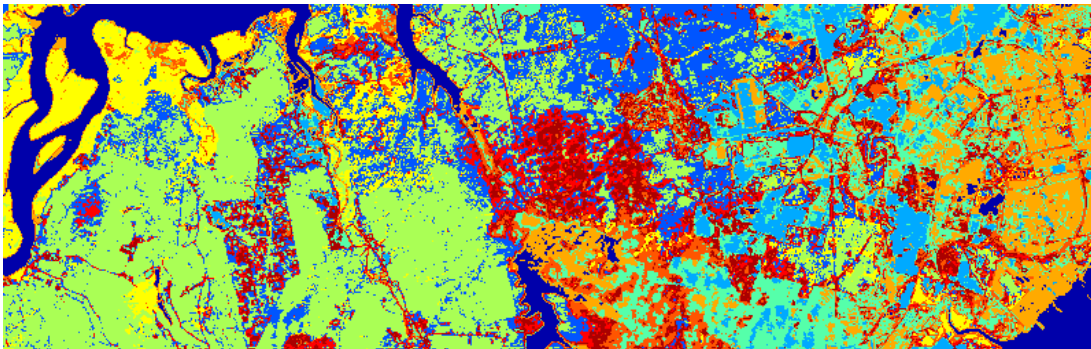


(c) FPE - Box - AM

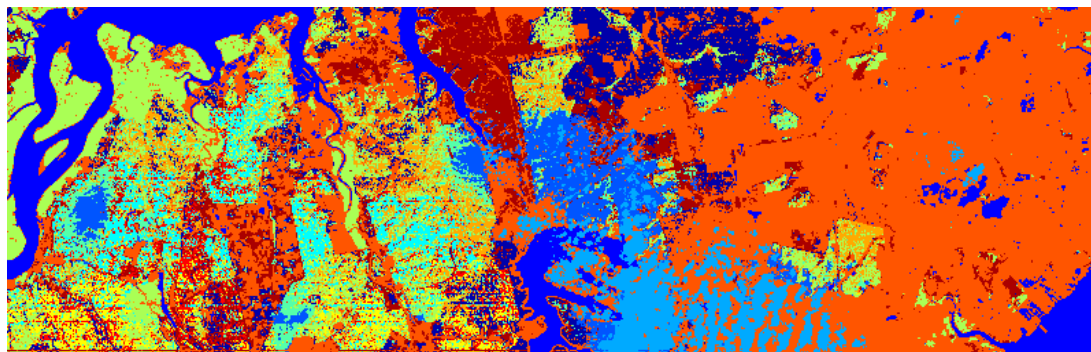


(d) FPE - Géométrique - AM

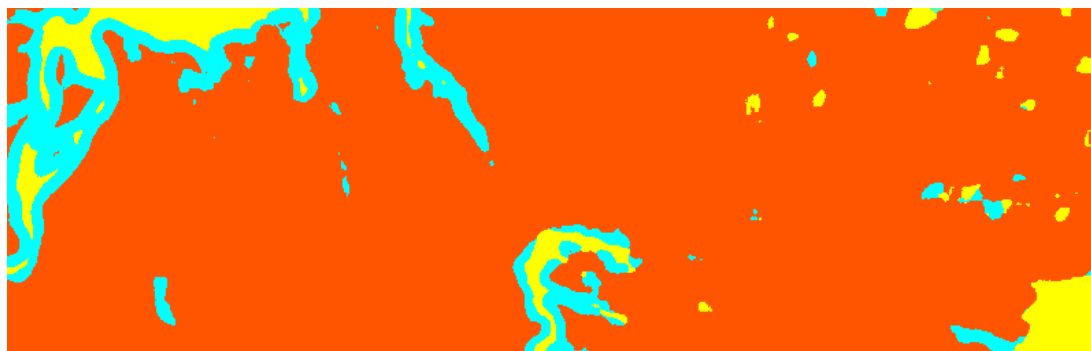
FIGURE 4.6 – Résultats de classification en utilisant le FPE et la moyenne arithmétique



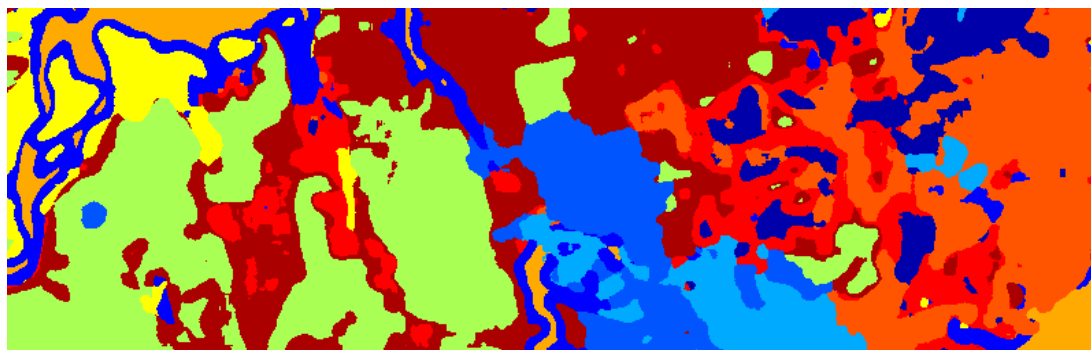
(a) FPE - SAM - GM



(b) FPE - Mahalanobis - GM



(c) FPE - Box - GM



(d) FPE - Géométrique - GM

FIGURE 4.7 – Résultats de classification en utilisant le FPE et la moyenne géométrique

4.3 Conclusion

L'application des techniques issues de la classification d'images SAR polarimétriques paraît donc peu concluante au premier abord, étant donné les résultats présentés précédemment. L'utilisation de distances employant directement la matrice de covariance (distance de Box ou géométrique) engendre une perte de précision drastique à cause du processus d'estimation de la matrice de covariance. Il est possible de réduire cette perte de précision en réduisant la taille du voisinage utilisé pour l'estimation mais on perd alors en qualité d'estimation. L'utilisation de la distance de Mahalanobis semble donner les résultats les plus encourageants, combinée avec le FPE et la moyenne géométrique, mais ils restent néanmoins, sur ce jeu de données et pour cette combinaison de bandes, moins bons qu'en utilisant le SAM. Une meilleure sélection des bandes ou un moyennage sur plusieurs exécutions de l'algorithme avec différents critères d'initialisation permettrait peut-être d'améliorer la classification. Il semble alors conseillé de poursuivre les recherches dans cette direction afin de justifier l'utilisation de ces techniques pour la classification d'images hyperspectrales.

Conclusion

Le travail effectué pendant cette thèse et présenté dans ce manuscrit porte sur des méthodes statistiques et géométriques pour la classification d'images SAR polarimétriques haute résolution. En effet, les algorithmes de classification classiques se basent sur une hypothèse de bruit gaussien comme modèle de signal, qui est rapidement mise en défaut lorsque l'environnement devient impulsionnel ou inhomogène, comme c'est particulièrement le cas dans les images SAR haute résolution, notamment au niveau des zones urbaines. L'utilisation d'un modèle de bruit composé, appelé modèle SIRV, permet de mieux prendre en compte ces phénomènes et de représenter la réalité de manière plus adéquate. Cette thèse s'emploie alors à étudier l'application et l'impact de ce modèle pour la classification des images SAR polarimétriques afin d'améliorer l'interprétation des classifications au sens de la polarimétrie et à proposer des outils adaptés à ce nouveau modèle. En effet, il apparaît rapidement que les techniques classiques utilisent en réalité beaucoup plus l'information relative à la puissance de chaque pixel plutôt qu'à la polarimétrie pour la classification. Un test d'hypothèse est proposé et sa statistique suivant le cas gaussien et le cas SIRV est étudiée. Par ailleurs, les techniques de classification traditionnelles font régulièrement appel à la moyenne de matrices de covariance, calculée comme une moyenne arithmétique. Cependant, étant donnée la nature riemannienne de l'espace des matrices de covariance, cette définition n'est pas applicable et il est nécessaire d'employer une définition plus adaptée à cette structure riemannienne. Ces différentes considérations sont étudiées sur des données simulées mais également sur des données réelles acquises par l'ONERA. Enfin, les prémices d'une extension de ces travaux au cas de l'imagerie hyperspectrale est proposée, de par la proximité de ce type de données avec les données SAR polarimétriques. Les résultats principaux sont rappelés de manière détaillée dans ce chapitre de conclusion puis des perspectives de recherche, dont certaines ont été abordées au cours du manuscrit, sont proposées dans le paragraphe de perspectives.

Conclusions

Le premier chapitre est consacré aux principes fondamentaux de la classification SAR. Le principe du radar à synthèse d'ouverture est présenté, ainsi que le phénomène de polarisation et son intérêt pour la classification. Les techniques classiques de classification, basées sur l'hypothèse gaussienne du signal sont rappelées, avant de mettre en évidence un inconvénient majeur de cette hypothèse gaussienne, à savoir la dépendance très importante à la puissance du pixel, plutôt qu'à l'information polarimétrique proprement dite. En effet, il est montré qu'une classification sur la puissance de chaque pixel donne les mêmes résultats qu'une classification utilisant l'estimateur de la matrice de covariance sous hypothèse gaussienne (appelé Sample Covariance Matrix, ou SCM). Cette forte influence de la puissance permet de s'interroger sur la pertinence du modèle classique vis-à-vis de l'exploitation de l'information polarimétrique. Le modèle non gaussien employé pendant la thèse, le modèle SIRV (pour Spherically Invariant Random Vectors) est alors introduit, ainsi que les outils statistiques associés, notamment l'estimateur de matrice de covariance appelé Fixed Point Estimate (FPE) et ses propriétés. L'une de ces propriétés — à savoir que le FPE se comporte asymptotiquement comme une matrice de covariance estimée sous hypothèse gaussienne, c'est-à-dire qu'il suit asymptotiquement une distribution de Wishart — est particulièrement importante puisqu'elle permet d'étendre facilement des techniques développées pour l'hypothèse gaussienne à l'hypothèse SIRV. Cet estimateur permet de mieux prendre en compte l'hétérogénéité présente dans les images SAR haute résolution et de s'affranchir de l'information de puissance et ainsi de proposer des classifications utilisant uniquement l'information polarimétrique.

Dans le chapitre deux, un test statistique d'égalité de matrices de covariance est proposé afin de pallier certains défauts des techniques de classification classiques, notamment le nombre final de classes comme pa-

Conclusion

ramètre du classifieur ou l'attribution automatique d'un pixel à une des classes existantes. Le test classique est rappelé pour le cas d'un signal distribué selon une loi gaussienne, puis il est étendu au cas d'un signal suivant une loi SIRV. Grâce à la propriété importante de comportement asymptotique du FPE présentée au premier chapitre, cette extension est aisément réalisable. Le test proposé est un test d'hypothèses cherchant à déterminer si les matrices de covariance de deux échantillons sont égales. Deux cas sont considérés, à savoir le cas où l'une des deux matrices de covariance est inconnue et le cas où les deux matrices de covariance sont inconnues. La statistique du test est employée comme une distance au sein d'un algorithme de classification et comparée à la distance de Wishart classique. La région critique du test d'hypothèses permet de rejeter les pixels ne correspondant pas aux classes existantes. Il apparaît que les pixels rejetés correspondent aux zones artificielles de l'image, c'est-à-dire les zones hétérogènes : bâtiment, zones urbaine, parking. L'utilisation de cette distance au sein d'algorithmes de classification existants — classifieur Wishart et segmentation hiérarchique — permet de mettre en évidence ce même phénomène : bonne séparation entre zones naturelles et artificielles mais peu d'amélioration sur la classification au sein de ces deux types de zones. Par ailleurs, les résultats obtenus avec le FPE ont un aspect visuel, *a priori* moins "attirant" que les résultats obtenus avec la SCM mais l'information visible est différente — puissance dans le cas de la SCM et polarimétrie dans le cas du FPE. Ceci entraîne des difficultés d'interprétation car les résultats de classification concordent mal avec les images optiques disponibles.

Le chapitre trois porte sur un élément des algorithmes de classification souvent négligé, à savoir le calcul d'une matrice de covariance moyenne pour chaque classe. En effet, c'est une des étapes du très répandu classifieur Wishart et est classiquement réalisée en utilisant la moyenne arithmétique des matrices de covariance de chaque pixel de la classe. Cependant, la structure de l'espace des matrices de covariance n'est pas euclidienne et des travaux récents issus de la géométrie de l'information, une branche de la géométrie riemannienne, permettent d'obtenir une expression correcte de la moyenne de matrices de covariance. Dans un premier temps, les principes de la géométrie riemannienne sont rappelés, ainsi que la méthode d'obtention du barycentre des matrices hermitiennes définies positives. Une expression d'une distance riemannienne entre deux matrices est pareillement introduite. Dans un deuxième temps, l'impact de ces deux techniques est étudiée sur des données SAR polarimétriques simulées, puis sur des données réelles. Les données simulées suivent une K-distribution avec des paramètres de forme et des matrices de covariance génératrices différentes et mettent en évidence l'intérêt de l'utilisation de la moyenne géométrique par rapport à la moyenne arithmétique puisque la classification est grandement améliorée. Cependant, l'utilisation de la distance riemannienne au lieu de la distance de Wishart n'apporte pas d'amélioration visible de la classification.

Sur les données réelles, on observe le même phénomène sur le jeu de données présenté. L'utilisation de la moyenne géométrique permet de séparer des zones caractéristiques de l'image, qui étaient regroupées avec d'autres zones avec les méthodes traditionnelles, de manière contraire au comportement attendu. De même, l'utilisation de la distance riemannienne n'apparaît pas améliorer la classification, voire tend à la détériorer.

Finalement, au sein du chapitre quatre, le FPE et la moyenne géométrique sont utilisés sur des images hyperspectrales, afin d'étudier l'apport de ces techniques pour sur des données proches de l'imagerie SAR polarimétrique. Après un bref rappel du principe de l'imagerie hyperspectrale, des données issues du capteur Hyperion de la NASA sont utilisées pour appliquer un algorithme de classification proche du classifieur Wishart et comparer l'apport du FPE et de la moyenne géométrique par rapport à un algorithme de base en classification hyperspectrale, le Spectral Angle Mapper (SAM). Les résultats préliminaires laissent supposer que l'utilisation de la matrice de covariance de manière directe n'est pas satisfaisante pour la classification hyperspectrale. En effet, les algorithmes classiques n'utilisent que le vecteur de données pour comparer deux pixels alors qu'en utilisant les matrices de covariance, un moyennage spatial autour de chaque pixel est réalisé pour estimer la matrice, ce qui réduit grandement la précision. Cependant, des travaux de recherche plus poussés doivent être

poursuivis dans ce domaine, notamment avec l'utilisation de la distance de Mahalanobis.

Perspectives

Tout au long de cette thèse, l'accent a été mis sur l'utilisation de l'information polarimétrique pour la classification, contenue dans la matrice de covariance. Il a été montré que l'estimateur classique de la matrice de covariance, la SCM, contenait principalement l'information de puissance relative à chaque pixel, ce qui a conduit à l'utilisation du FPE, qui s'affranchit de cette puissance. Cependant, il apparaît sur les résultats de classification utilisant la SCM que les zones naturelles ressortent de manière beaucoup plus prononcée qu'en utilisant le FPE. Cette information de puissance peut donc être utile dans un algorithme de classification plus complexe pour améliorer le rendu du résultat final.

Par ailleurs, lors de la mise en place du test d'égalité des matrices de covariance, une hypothèse assez forte est employée, à savoir que le FPE se comporte comme une matrice suivant une loi de Wishart. Ce résultat est vrai asymptotiquement et on utilise un nombre de données secondaires relativement élevé pour l'estimation (environ cinquante) mais il serait intéressant d'étudier le comportement réel du FPE à distance finie afin d'adapter, le cas échéant, la statistique du test. Un début de discussion a d'ailleurs été amorcé avec l'étude de l'algorithme de classification hiérarchique.

Il est aussi possible de s'intéresser aux méthodes d'estimation robuste, comme les M-estimateurs, en considérant par exemple l'estimateur d'Huber à la place de la SCM et du FPE. Cet estimateur se comporte comme la SCM pour une certaine proportion des données et comme le FPE pour le reste de ces données, ce qui permettrait de garder une partie de l'information de puissance sans qu'elle "pollue" complètement la matrice de covariance.

L'accent mis sur l'information polarimétrique appelle également à s'interroger sur l'interprétation des résultats de classification par les outils polarimétriques existants. La décomposition de Cloude-Pottier a été utilisée pour proposer des interprétations grossières des divergences entre résultats de classification. D'autres décompositions, comme la décomposition de Freeman-Durden ou de Yamaguchi (surtout vu les travaux présentés régulièrement dans la communauté sur les améliorations apportées à cette décomposition) pourraient s'avérer plus à même de donner une interprétation plus fournie des résultats.

La géométrie de l'information introduit des outils puissants qui ont permis de rendre plus précis les résultats de classification (et plus rigoureux, étant donné que la formule arithmétique de la moyenne est mathématiquement fautive dans le cas des matrices hermitiennes définies positives) mais dont l'implémentation est coûteuse en temps de calcul. Des stratégies d'optimisation pourraient être envisagées pour accélérer ces traitements et les rendre plus opérationnels. Par ailleurs, il n'a été question que de la moyenne de matrices dans cette thèse mais il est possible d'employer la médiane, qui est moins sensible aux éventuels points aberrants présents dans les données.

Enfin, ces techniques peuvent éventuellement être appliquées à d'autres domaines proches de l'imagerie SAR polarimétrique, notamment l'imagerie hyperspectrale, pour laquelle un travail préliminaire a été réalisé mais doit être poursuivi, ou l'interférométrie.

Ces travaux sont une étape de plus vers l'amélioration des méthodes de classification polarimétrique et, plus important, la compréhension plus complète des mécanismes sous-jacents. Ce but n'est, à l'heure actuelle, pas encore atteint et la poursuite des travaux de recherche dans ce domaine me paraît être une nécessité.

Statistical Classification for Heterogeneous Polarimetric SAR Images

Statistical Classification for Heterogeneous Polarimetric SAR Images

Pierre Formont, *Student Member, IEEE*, Frédéric Pascal, *Member, IEEE*, Gabriel Vasile, *Member, IEEE*, Jean-Philippe Ovarlez, *Member, IEEE*, and Laurent Ferro-Famil, *Member, IEEE*

Abstract—This paper presents a general approach for high-resolution polarimetric SAR data classification in heterogeneous clutter, based on a statistical test of equality of covariance matrices. The Spherically Invariant Random Vector (SIRV) model is used to describe the clutter. Several distance measures, including classical ones used in standard classification methods, can be derived from the general test. The new approach provides a threshold over which pixels are rejected from the image, meaning they are not sufficiently “close” from any existing class. A distance measure using this general approach is derived and tested on a high-resolution polarimetric data set acquired by the ONERA RAMSES system. It is compared to the results of the classical $H - \alpha$ decomposition and Wishart classifier under Gaussian and SIRV assumption. Results show that the new approach rejects all pixels from heterogeneous parts of the scene and classifies its Gaussian parts.

Index Terms—Image classification, non-Gaussian modeling, polarimetric synthetic aperture radar, statistical analysis.

I. INTRODUCTION

THE signal returned by a synthetic aperture radar (SAR) is a complex variable, as it measures both amplitude and phase information of the backscattered signal, producing one complex image for each recording. POLarimetric SAR (POLSAR) systems are able to emit and receive in two (or more) orthogonal polarizations. The interactions between the polarized electromagnetic waves emitted by the radar and the target area can be described thanks to the Sinclair matrix [1]. This matrix can be represented in a vectorized version, the scattering vector. From this vector, its covariance matrix can be computed. Random interferences of the waves scattered within each resolution cell are responsible for the speckle noise [2]. Thus, the backscattered signal from polarimetric SAR data is generally modeled by the multivariate, zero-mean, circular

complex Gaussian distribution, which is fully characterized by the covariance matrix.

However, recent POLSAR systems are now capable to acquire very high-resolution images of the Earth’s surface, up to decimetric resolution. This increase in resolution leads to a higher scene heterogeneity, particularly in urban areas, where the clutter can no longer be modeled by a Gaussian process [3]. More precisely, the value of each resolution cell corresponds to the sum of all signals reflected by the scatterers within the cell. With higher resolutions, the number of scatterers varies randomly from cell to cell [4]. Thus, the resulting backscattered signal is heterogeneous, as it is locally Gaussian with random power. To take this heterogeneity into account, one can use the Spherically Invariant Random Vector product model, first introduced by Yao [5] for the information theory. The clutter is modeled as compound-Gaussian process, i.e., the product between the square root of a positive random scalar variable, called the texture, and an independent, zero-mean, complex circular Gaussian vector, called the speckle. Since the texture distribution is not specified, this model encompasses an infinity of distributions, notably the Gaussian one and the K-distribution [6]. Furthermore, this model has been validated by several measurements, see, e.g., [6]–[9]. This model has also been recently used by Rangaswamy in [10] to select homogeneous training data for covariance matrix estimation.

Polarimetric SAR images can be used for several applications, mostly for land cover classification. Therefore, polarimetric SAR images classification is an active area of research. Two main approaches appeared in the corresponding literature. The first approach is to classify the images based on their statistical properties: Kong *et al.* [11] derived a distance measure for single-look complex polarimetric SAR data, which has been extended by Yueh *et al.* [12] and van Zyl *et al.* [13] for normalized complex POLSAR data. A distance measure for the multi-look complex case has been proposed by Lee *et al.* in [14], based on the complex Wishart distribution of the clutter covariance matrix under the Gaussian assumption. The second approach is to classify pixels thanks to their physical characteristics. Several matrix decompositions have been proposed: coherent decompositions, based on the scattering matrix, like the Cameron decomposition [15] or the Krogager decomposition [16], in order to characterize pure scatterers. The distributed scatterers have been studied thanks to incoherent decompositions, i.e., decompositions of the covariance matrix. Freeman *et al.* [17] proposed such a decomposition. In [18], Cloude *et al.* proposed the $H - \alpha$ decomposition, based on the eigendecomposition of the covariance matrix.

Manuscript received April 20, 2010; revised September 23, 2010; accepted December 02, 2010. Date of publication December 23, 2010; date of current version May 18, 2011. This work was supported by the DGA. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lorenzo Bruzzone.

P. Formont and J.-P. Ovarlez are with ONERA DEMR/TSI, 91761 Palaiseau Cedex, France, and also with Supélec/SONDRA, 91192 Gif-sur-Yvette Cedex, France (e-mail: pierre.formont@supelec.fr; jean-philippe.ovarlez@onera.fr).

F. Pascal is with Supélec/SONDRA, 91192 Gif-sur-Yvette Cedex, France (e-mail: frederic.pascal@supelec.fr).

G. Vasile is with the GIPSA-lab DIS/SIGMAPHY, Grenoble INP-BP 46, 38402 Saint-Martin-d’Hères, France (e-mail: gabriel.vasile@gipsa-lab.fr; gabriel.vasile@gipsa-lab.grenoble-inp.fr).

L. Ferro-Famil is with the IETR, 35042 Rennes Cedex, France (e-mail: laurent.ferro-famil@univ-rennes1.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2010.2101579

These two approaches can be combined. Lee *et al.* [19] proposed an unsupervised classification algorithm using both physical and statistical properties of the POLSAR image for the classification. More recently, several methods have been considered for the clustering: fuzzy clustering, EM clustering, neural networks but most of these approaches only consider the Gaussian case. Vasile *et al.* [20] proposed a different approach for the non-Gaussian case, using an adapted model.

The main contribution of this paper is to study the polarimetric SAR classification problem as a statistical test problem on the covariance matrices. First, this approach allows unifying the existing classification methods based on statistical distance measures. Indeed, according to the assumption on the statistical test problem, each distance will be derived again as a test statistic. Second, a more general distance measure will be proposed and will be tested on a polarimetric SAR image. More precisely, the paper is organized as follows: Section II will recap the $H - \alpha$ decomposition and the standard Wishart classifier while Section III presents the heterogeneous model and the improvements on heterogeneous data. Section IV is devoted to the statistical approach based on the theory of statistical tests and validated on real polarimetric SAR image in Section V. Finally, Section VI concludes this paper.

II. STATE-OF-THE-ART ON POLARIMETRIC CLASSIFICATION IN HOMOGENEOUS CLUTTER

Cloude *et al.* proposed a decomposition of the covariance matrix in [18], extracting physical parameters from its eigenvalues, named the $H - \alpha$ decomposition. This decomposition will be used as the initialization of the classification algorithms presented in this paper. It is recapped as follows.

A. $H - \alpha$ Decomposition

Let us denote \mathbf{k} a pixel, i.e., a m -dimensional vector containing complex values. In our context, m will be equal to 3. More precisely, \mathbf{k} can be represented in many bases, the more common ones being the lexicographic basis and the Pauli basis that are defined as follows:

$$\mathbf{k} = \begin{pmatrix} HH \\ VV \\ HV \end{pmatrix} \text{ in the lexicographic basis}$$

$$\mathbf{k} = \frac{1}{\sqrt{2}} \begin{pmatrix} HH + VV \\ HH - VV \\ 2HV \end{pmatrix} \text{ in the Pauli basis}$$

where HH (respectively VV , HV) corresponds to the value acquired with horizontal (respectively vertical, horizontal) polarization in emission and horizontal (respectively vertical, vertical) polarization in reception.

Let us denote \mathbf{T} its covariance matrix, i.e., $\mathbf{T} = E[\mathbf{k}\mathbf{k}^H]$, where H is the conjugate transpose operator and $E[\cdot]$ denotes the statistical mean. From the eigendecomposition of the covariance matrix \mathbf{T} , two main classification parameters can be extracted: the entropy H and the α angle. The underlying assumption is that there is a dominant scattering mechanism in each cell, which can be characterized by these two parameters.

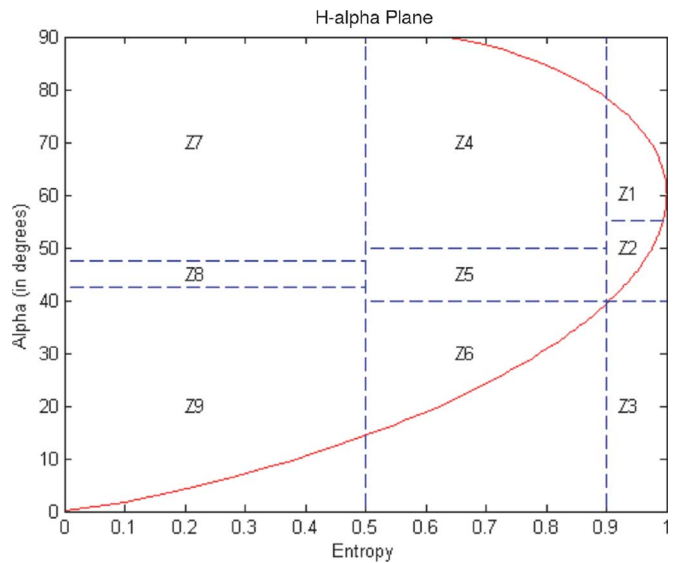


Fig. 1. $H - \alpha$ plane.

H is a function of the eigenvalues λ_k only, while α is a function of both the eigenvalues and the eigenvectors \mathbf{u}_k .

The entropy H measures the chaos inside a resolution cell: if H is close to 0, there is little chaos, meaning one scattering mechanism is much stronger than all the others in the cell. On the contrary, if H is close to 1, no mechanism is sufficiently stronger than the others to be dominant inside the cell.

The α angle characterizes the type of the dominant scattering mechanism: for example, when α is close to 0° , the dominant mechanism is a surface diffusion and when α is close to 90° , the dominant mechanism is a dihedral reflection. The relationship between entropy, α angle and scattering mechanisms is represented in Fig. 1.

The $H - \alpha$ plane is separated into eight areas (Z1 to Z9, except Z3 which is not physically feasible), each one corresponding to a specific scattering mechanism. The red line represents the boundary of physically possible $H - \alpha$ couples.

B. Clustering: Wishart Classifier

One of the main advantages of the $H - \alpha$ decomposition is that it regroups pixels in large clusters. Such information can then be used to classify the image more precisely. Lee *et al.* [19] proposed an algorithm based on the K-means algorithm, using the $H - \alpha$ decomposition as the algorithm initialization and using the covariance matrix distribution to classify the image. The main assumption is that the class center, which is a covariance matrix, can be computed with higher accuracy, because of the large number of samples contained in each class. Once these class centers are known, each pixel can be reattributed to a class, depending on the distance between its covariance matrix and each class center. The distance used is based on the covariance matrix distribution and is detailed in the following.

1) *Wishart Distance*: The pixel under consideration, characterized by its target vector \mathbf{k} , is supposed to follow a zero-mean complex Gaussian distribution denoted $\mathcal{CN}(\mathbf{0}, \mathbf{T})$ with covariance matrix \mathbf{T} . Considering an independent and identically distributed (i.i.d.) N -sample $(\mathbf{k}_1, \dots, \mathbf{k}_N)$ corresponding

to a spatial neighborhood¹ of the pixel \mathbf{k} , the maximum-likelihood (ML) estimator of \mathbf{T} is the well-known sample covariance matrix, given by

$$\hat{\mathbf{T}}_{\text{SCM}} = \frac{1}{N} \sum_{i=1}^N \mathbf{k}_i \mathbf{k}_i^H. \quad (1)$$

$\hat{\mathbf{T}}_{\text{SCM}}$ has a complex Wishart distribution [21]:

$$p(\hat{\mathbf{T}}_{\text{SCM}}) = \frac{N^{m(N-m)} |\hat{\mathbf{T}}_{\text{SCM}}|^{N-m} \exp \left[-N \text{Tr}(\mathbf{T}^{-1} \hat{\mathbf{T}}_{\text{SCM}}) \right]}{K(N, m) |\mathbf{T}|^N}$$

where $K(N, m) = \pi^{m(m-1)/2} \prod_{i=1}^m \Gamma(N - i + 1)$ and $\Gamma(\cdot)$ is the Gamma function [22].

A distance measure, called the Wishart distance, can be derived [14] from this distribution and a Bayes ML procedure

$$d_W(\hat{\mathbf{T}}_{\text{SCM}}, \mathbf{C}_i) = \ln \frac{|\mathbf{C}_i|}{|\hat{\mathbf{T}}_{\text{SCM}}|} + \text{Tr}(\mathbf{C}_i^{-1} \hat{\mathbf{T}}_{\text{SCM}}) \quad (2)$$

where \mathbf{C}_i is the center of the class i .

Algorithm: The subsequent algorithm, based on a classical K-means clustering algorithm is described as follows.

- 1) Initially classify the image using the $H - \alpha$ decomposition, thus creating eight classes.
- 2) For $i = 1 \dots 8$, compute the class centers: $\mathbf{C}_i = (1/N_i) \sum_{k=1}^{N_i} \hat{\mathbf{T}}_{\text{SCM}}^{(k)}$, where N_i is the number of pixels in the class i and where $\hat{\mathbf{T}}_{\text{SCM}}^{(k)}$ is the covariance matrix estimator of the k^{th} pixel in the class i .
- 3) For each pixel, compute the Wishart distances between its covariance matrix and each class center using (2) and assign the pixel to the class whose center minimizes the distance.
- 4) Check the termination criterion. If it is met, stop the algorithm, else go to step 2.

This approach relies heavily on the Gaussian assumption for pixels distribution and can perform poorly when this assumption is not valid anymore. The case of non Gaussian distributions is addressed in the next section.

III. STATE-OF-THE-ART ON POLARIMETRIC CLASSIFICATION IN HETEROGENEOUS CLUTTER

The increase in resolution leads to much smaller resolution cells. This results in a heterogeneous backscattered signal and previous methods may give poor results. It is thus necessary to consider an heterogeneous signal model.

A. SIRV Model

A SIRV is a compound Gaussian vector, defined as the product of a positive scalar random variable τ and a m -dimensional complex circular Gaussian vector, \mathbf{x} . Then the target vector \mathbf{k} can be rewritten as

$$\mathbf{k} = \sqrt{\tau} \mathbf{x}.$$

¹For a rigorous estimation procedure, N independent temporal realizations of the pixel \mathbf{k} should be considered but in practice, those are not available since there is only one image. For that purpose, the spatial independency and the local homogeneity are commonly assumed to have an i.i.d. N -sample.

This model exhibits many advantages such as follows.

- $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{M})$ is called the *speckle* and represents the polarimetric information contained in the clutter.
- τ is called the *texture* and represents the local variations of power from cell to cell, thus the heterogeneity of the clutter.
- As τ 's probability density function (pdf) is not explicitly specified, a whole range of random processes can be described by this model, including of course the Gaussian one. We can also cite the K-distribution for a Gamma distributed texture, Chi, Rayleigh, Weibull, or Rician pdfs [23].
- Measurements campaigns have been realized to show this model matches well with real data [6], [24].

The SIRV model is uniquely defined up to a multiplicative constant with respect to the covariance matrix \mathbf{M} of the Gaussian kernel. Let \mathbf{M}_1 and \mathbf{M}_2 be two covariance matrices such as $\mathbf{M}_1 = \alpha \mathbf{M}_2, \alpha \in \mathbb{R}^{+*}$. The two different sets of parameters $\{\mathbf{M}_1, \tau\}$ and $\{\mathbf{M}_2, (\tau/\sqrt{\alpha})\}$ describe the same SIRV. To solve this identification problem, the covariance matrix \mathbf{M} has to be normalized. We choose to normalize it such that $\text{Tr}(\mathbf{M}) = m$, where $\text{Tr}(\cdot)$ denotes the trace of a matrix. This normalization results in transferring all the power information in the texture parameter, while the speckle parameter contains only information about the polarimetric diversity. As a result, the interesting matrix for our study is the normalized covariance matrix of the Gaussian kernel \mathbf{M} .

Using this model, Gini *et al.* [25] and Conte *et al.* [26] propose the exact ML estimate of \mathbf{M} , when assuming that the textures τ_i of each component of the N -sample of spatial neighbors are deterministic. It is the solution of the following recursive equation:

$$\begin{aligned} \widehat{\mathbf{M}}_{\text{FP}} &= f(\widehat{\mathbf{M}}_{\text{FP}}) = \frac{m}{N} \sum_{i=1}^N \frac{\mathbf{k}_i \mathbf{k}_i^H}{\mathbf{k}_i^H \widehat{\mathbf{M}}_{\text{FP}}^{-1} \mathbf{k}_i} \\ &= \frac{m}{N} \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \widehat{\mathbf{M}}_{\text{FP}}^{-1} \mathbf{x}_i}. \end{aligned} \quad (3)$$

The associated recursive algorithm to compute $\widehat{\mathbf{M}}_{\text{FP}}$, the Fixed Point (FP) of f is given by

$$\widehat{\mathbf{M}}_{\text{FP}}^{(i+1)} = f(\widehat{\mathbf{M}}_{\text{FP}}^{(i)}).$$

To initialize this recursive algorithm, one can choose $\mathbf{M}_{\text{FP}}^{(0)} = \text{Id}$.

The estimate from (3) holds for stochastic τ_i and becomes an approximate maximum-likelihood (AML) estimate. Pascal *et al.* [27] have proven the existence and the uniqueness of the fixed point estimator as the solution of (3). They also prove in [27] the convergence of the recursive algorithm, whatever the initialization.

B. SIRV Distance

This new estimator, adapted to the non-Gaussian case has been used by Vasile *et al.* [20] to derive a distance measure. The estimation scheme is the same as in the Gaussian case, only the estimator used differs. The resulting distance, called the SIRV distance, is given by the following equation:

$$d_S(\widehat{\mathbf{M}}_{\text{FP}}, \mathbf{C}_i) = \ln \frac{|\mathbf{C}_i|}{|\widehat{\mathbf{M}}_{\text{FP}}|} + \frac{m}{N} \sum_{n=1}^N \frac{\mathbf{k}_n^H \mathbf{C}_i^{-1} \mathbf{k}_n}{\mathbf{k}_n^H \widehat{\mathbf{M}}_{\text{FP}}^{-1} \mathbf{k}_n} \quad (4)$$

where \mathbf{C}_i is the center of the class i .



Fig. 2. Amplitude color composition image of the Brétigny region: k_1 - k_3 - k_2 .

C. Results and Discussion

1) *High-Resolution POLSAR Data Set*: The data set considered was acquired by the ONERA RAMSES system in Brétigny, France. The acquisition was made in X-band, with a spatial resolution of 1.32 m in range and 1.38 m in azimuth. The resulting image is 501×501 pixels. Fig. 2 shows a color composition of the data in the Pauli basis

$$\mathbf{k} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} HH + VV \\ HH - VV \\ 2HV \end{pmatrix}$$

k_1 is in red, k_2 in blue, and k_3 in green.

The polarimetric diversity in this image is high as different features are present: several buildings, a parking lot, different kind of fields, forested areas, roads, etc.

2) *H- α Decomposition*: Fig. 3 shows the $H - \alpha$ decomposition applied to the data set using the SCM estimate in Fig. 3(a) and the FP estimate in Fig. 3(b). Both results are similar, in terms of visual interpretation, which means that the FP estimate has the same polarimetric properties as the SCM.

3) *Classical Wishart Classifier*: The Wishart classifier of Section II has been applied on this data set. The results can be seen in Fig. 4.

We can notice that the Gaussian areas are clearly identified: fields in yellow in the bottom part of the image and the blue field in the top part. Urban areas, on the contrary, are a mix on several classes and do not stand out as well. This is due to the fact that the SCM is not a good estimate for heterogeneous areas. Moreover, in [20], it has been shown that the SCM estimate is heavily polluted by the power information.

4) *Fixed Point Estimate*: Replacing the SCM by the FP estimate in the Wishart Classifier algorithm yields much better results in heterogeneous areas, as can be seen in Fig. 5.

The two big buildings stand out much more in this image than in Fig. 4 and are composed of only one class each. The classification remains satisfactory on most of the Gaussian areas, though it is less good than with the SCM estimate, in terms of separation.

5) *SIRV Distance*: When replacing the Wishart distance by the SIRV distance in the algorithm, the results, as seen in Fig. 6, stay very close when using the FP estimate. Notice that the classification improvement comes only from the FP estimate and not from the distances.

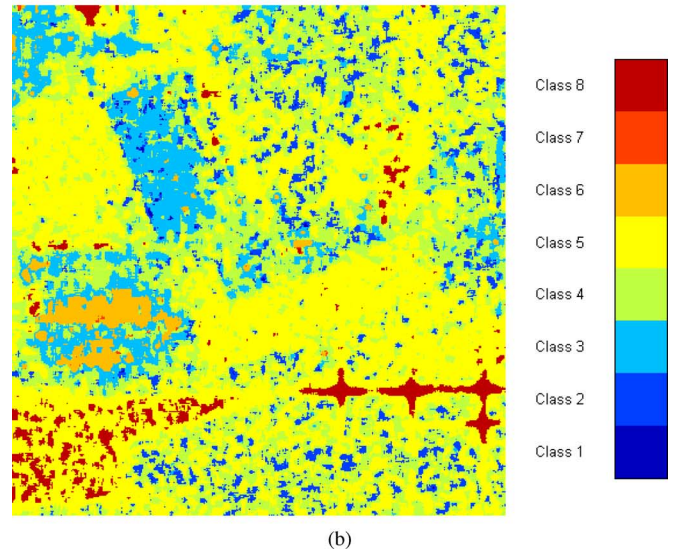
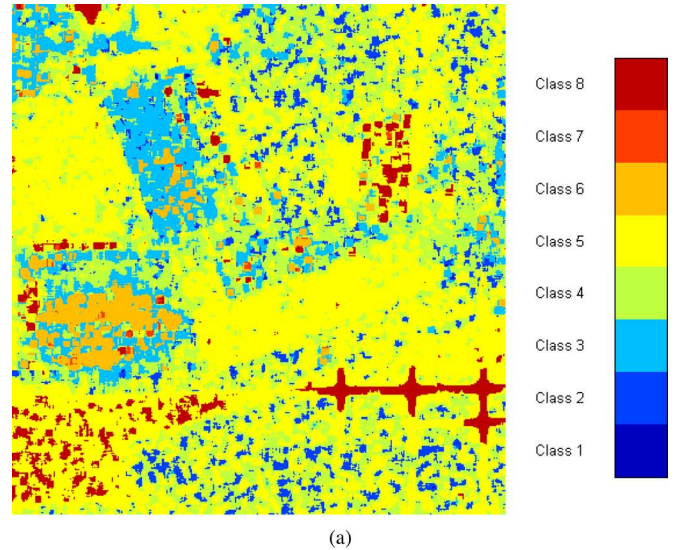


Fig. 3. $H - \alpha$ decomposition. (a) SCM. (b) FP.

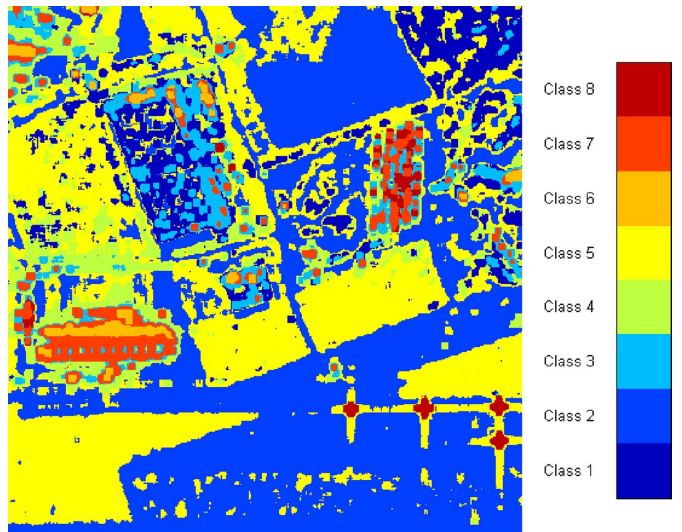
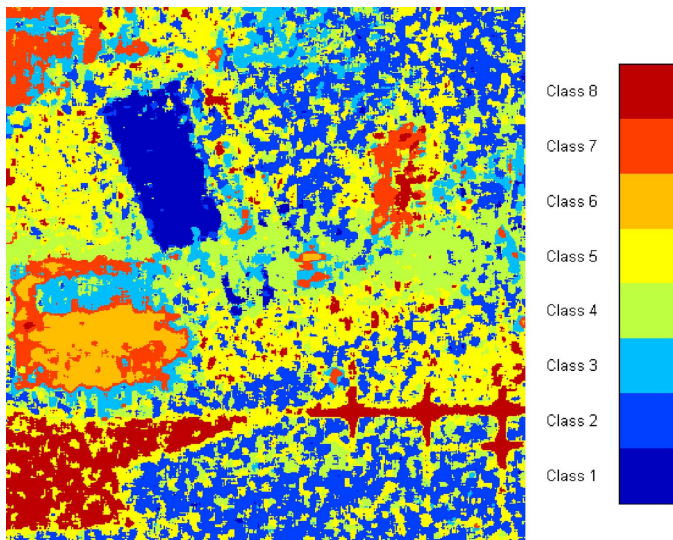
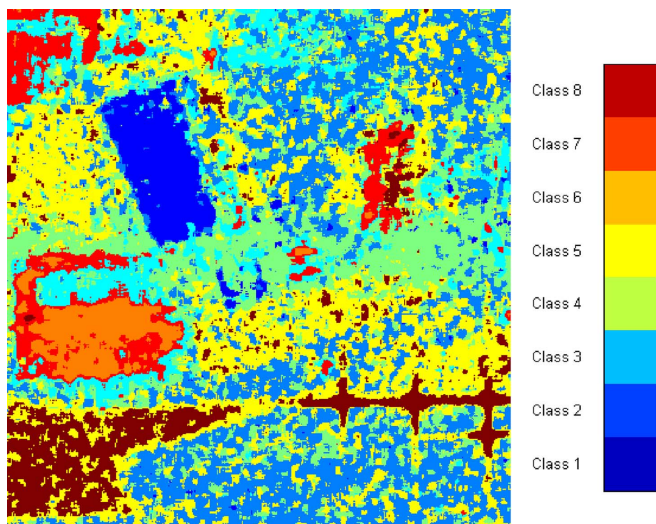


Fig. 4. Wishart classifier, SCM, $H - \alpha$ initialization.

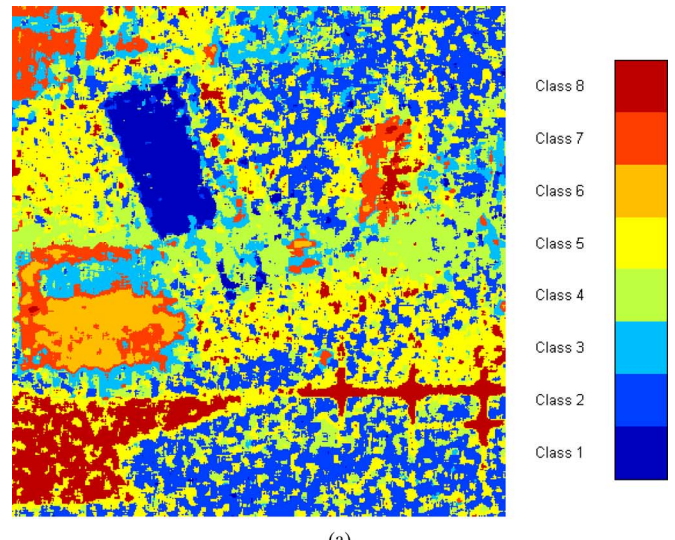
6) *H- α Initialization*: The use of the $H - \alpha$ decomposition for the initialization may create a bias in the final classification

Fig. 5. Wishart classifier, FP, $H - \alpha$ initialization.Fig. 6. SIRV classifier, FP, $H - \alpha$ initialization.

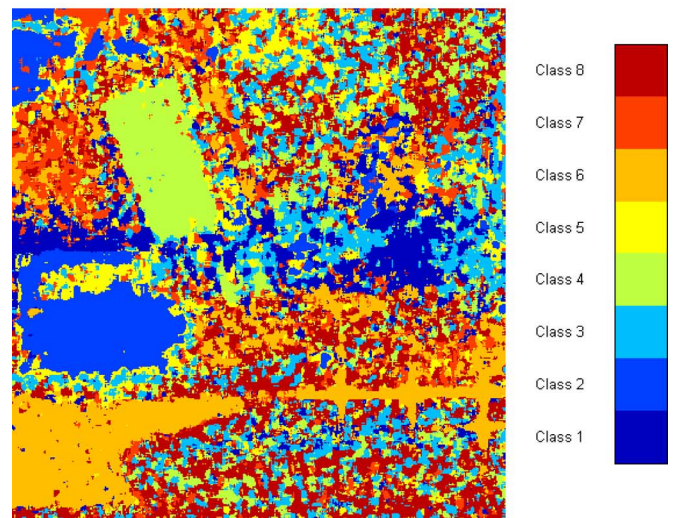
results, as the pixels are initially clustered into classes thanks to their physical properties, while the distances are a result of the statistical distribution of the clutter. We have investigated the initialization by comparing the end results of the classifier, using $H - \alpha$ initialization or a totally random initialization: each pixel is randomly assigned to 1 out of 8 classes at the start of the classifier. Results can be seen in Fig. 7.

Although the colors do not match, the final classification results are very close to each other. We can therefore dispense with the initial $H - \alpha$ decomposition as it does not improve the end results for this data set.

7) *Conclusion*: The SCM estimate gives poor results compared to the FP estimate in a non-Gaussian environment. The physical interpretation given by the $H - \alpha$ prior to the classification is not relevant, as a completely random initialization gives the same classification results. On the other hand, the SIRV distance, while optimal for the SIRV case, does not improve the classification in comparison to the Wishart distance and is



(a)



(b)

Fig. 7. Wishart classifier, FP. (a) $H - \alpha$ initialization. (b) Random initialization.

much more computationally heavy. Besides, these methods do not necessarily take into account the possibility of pixels being too far away from every class center. In that case, it makes no sense to add these pixels to any of the classes. In the classical methods, it is possible to take this possibly into account but the threshold has to be set manually. As we want to classify pixels based on their polarimetric properties, i.e., their covariance matrices, we propose to use a statistical test of equality of covariance matrices, which will give a more general approach to the classification problem. This approach will also allow us to rederive the classical Wishart distance and the SIRV distance of [20] more formally according to the hypotheses used in the statistical test problem. These two distances are simply a particular case of the proposed method.

IV. STATISTICAL APPROACH FOR POLSAR DATA CLASSIFICATION – MAIN CONTRIBUTION

This section is devoted to the study of statistical test for equality of covariance matrices and its application to POLSAR

data classification. A statistical approach to the problem of POLSAR data classification has many advantages.

- It can be applied to both homogeneous and heterogeneous clutter models.
- There is no *a priori* physical interpretation to the classification process.
- It is a very general approach which particularly contain previous methods.

A. Statistical Framework

Let \mathbf{X}_1 and \mathbf{X}_2 be independent random vectors, such that $\mathbf{X}_1 \sim \mathcal{L}(\mathbf{0}, \mathbf{T}_1)$ and $\mathbf{X}_2 \sim \mathcal{L}(\mathbf{0}, \mathbf{T}_2)$, where \mathcal{L} stands for any distribution with the two first moments existing. The goal is to decide if their covariance matrices \mathbf{T}_1 and \mathbf{T}_2 are equal. The resulting binary hypothesis test can be written as

$$\begin{cases} H_0 : \mathbf{T}_1 = \mathbf{T}_2 \\ H_1 : \mathbf{T}_1 \neq \mathbf{T}_2 \end{cases}.$$

Let \mathbf{x}_1 (resp. \mathbf{x}_2) be an observation of \mathbf{X}_1 (resp. \mathbf{X}_2) and $N = N_1 + N_2$. \mathbf{x}_1 and \mathbf{x}_2 are supposed i.i.d. Let \mathbf{x}_1 (resp. \mathbf{x}_2) be of size N_1 (resp. N_2).

The likelihood function is then

$$\Lambda = \prod_{i=1}^2 \text{PDF}(\mathbf{x}_i) \quad (5)$$

where $\mathbf{x}_i^{(k)}$ is the k^{th} element of the N_i -sample \mathbf{x}_i .

This is a more general approach for the pixels classification. Indeed, the idea with previous test is to assign two different pixels in the same class or not.

1) \mathbf{T}_2 Known: Let us assume that \mathbf{T}_2 is known and equal to \mathbf{C}_i , i.e., the center of the class i .

- If \mathbf{X}_1 is Gaussian distributed, the likelihood ratio test statistic derived from (5) easily provides the Wishart distance defined by (2).
- If \mathbf{X}_1 is SIRV distributed, the likelihood ratio test statistic derived from (5) provides the SIRV distance defined by (4).

In both cases, the approach of statistical tests shows that the corresponding distances used for the polarimetric SAR classification are actually specific test statistics. Thus, to decide that two covariance matrices are equal, it is not sufficient to minimize the distance between these two covariance matrices (done in the previous classification procedure). The minimal distance has to be also smaller than a threshold, directly related to the distribution of the test statistic and given by the type I error.

2) \mathbf{T}_2 Unknown: The maximum-likelihood estimator of \mathbf{T}_i is

$$\hat{\mathbf{T}}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{x}_i^{(k)H} \mathbf{x}_i^{(k)}. \quad (6)$$

$\hat{\mathbf{T}}_i$ is the classical SCM defined in (1); hence, it is Wishart-distributed.

Equation (5) needs to be maximized under H_0 , i.e., when $\mathbf{T}_1 = \mathbf{T}_2 = \mathbf{T}$. Equation (5) then becomes

$$\frac{1}{\pi^{mN/2} |\mathbf{T}|^{N/2}} \exp \left(-\frac{1}{2} \sum_{i=1}^2 \sum_{k=1}^{N_i} \mathbf{x}_i^{(k)H} \mathbf{T}_i^{-1} \mathbf{x}_i^{(k)} \right). \quad (7)$$

The optimal estimator of \mathbf{T} is

$$\hat{\mathbf{T}} = \frac{1}{N} (N_1 \hat{\mathbf{T}}_1 + N_2 \hat{\mathbf{T}}_2). \quad (8)$$

By replacing (6) in (8) and \mathbf{T}_i by its estimate $\hat{\mathbf{T}}_i$, the test statistic is

$$\lambda = \frac{|\hat{\mathbf{T}}_1|^{N_1/2} |\hat{\mathbf{T}}_2|^{N_2/2}}{|\hat{\mathbf{T}}|^{N/2}}. \quad (9)$$

Notice the exponents are the size of the samples. Bartlett [28] proposed alternative exponents for the univariate case, replacing the samples size by the degree of freedom of the estimators $\hat{\mathbf{T}}_i$. Equation (9) then becomes

$$t = \frac{|\hat{\mathbf{T}}_1|^{\nu_1/2} |\hat{\mathbf{T}}_2|^{\nu_2/2}}{|\hat{\mathbf{T}}_t|^{\nu_t/2}} \quad (10)$$

where $\nu_i = N_i - 1$ are the degrees of freedom of $\hat{\mathbf{T}}_i$ and $\nu_t = N - 2$, the degree of freedom of $\hat{\mathbf{T}}$.

The statistic t varies between 0 and 1, with values close to 0 rejecting the null hypothesis, while values close to 1 accept the null hypothesis. To illustrate this, let consider (10) as

$$t = \left(\frac{|\hat{\mathbf{T}}_1|}{|\hat{\mathbf{T}}|} \right)^{\nu_1/2} \left(\frac{|\hat{\mathbf{T}}_2|}{|\hat{\mathbf{T}}|} \right)^{\nu_2/2}.$$

When $\hat{\mathbf{T}}_1 = \hat{\mathbf{T}}_2 = \hat{\mathbf{T}}$, $t = 1$. As the difference between $\hat{\mathbf{T}}_1$ and $\hat{\mathbf{T}}_2$ increases, the smaller factor reduces the product more than the higher factor increases it, thus bringing t closer to 0.

Box [29], [30] proposed χ^2 - and F - approximations for the distribution of t . For the χ^2 -approximation, the statistic he proposed is

$$u = -2(1 - c_1) \ln(t) \sim \chi^2 \left(\frac{1}{2}(k-1)m(m+1) \right)$$

$$\text{where } c_1 = \left(\sum_{i=1}^k \frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^k \nu_i} \right) \left(\frac{2m^2 + 3m - 1}{6(k-1)(m+1)} \right)$$

and $\chi^2(a)$ denotes the χ^2 distribution with a degrees of freedom.

The critical region of the test is

$$\left\{ u > \chi^2 \left(\frac{1}{2}(k-1)m(m+1), P_{FA} \right) \right\} \quad (11)$$

where P_{FA} is the type I error, or false-alarm rate. In our case, we set the type I error (accepting the null hypothesis when it is not true, i.e., rejecting the equality of the matrices of the two populations when they are actually equal) to be very low.

3) *SIRV Case*: The same procedure can be applied in the SIRV case, considering that it was demonstrated by Pascal *et al.* in [31] that the FP estimate asymptotically has a Wishart distribution behavior with $(m/(m+1))N$ degrees of freedom. This is a very important property of the FP estimate since all results obtained with the SCM remain valid for the FP estimate due to an asymptotical justification. The SCM estimators $\hat{\mathbf{T}}_1$ and $\hat{\mathbf{T}}_2$ can be replaced by the FP estimates $\hat{\mathbf{M}}_1$ and $\hat{\mathbf{M}}_2$ of the same N -samples \mathbf{x}_1 and \mathbf{x}_2 with the correct degrees of freedom.

Algorithm

The algorithm used is very similar to the previous algorithm of Section II, except for the fact that the number of classes is not specified at the beginning. We start the algorithm with only one class and increase the number of classes at each iteration. To decide if a pixel is joining a class, the test statistic of (10) is used. The N -samples required to compute the estimation of the covariance matrix are simple boxcar neighborhoods, centered on the pixel under consideration. The classification map is initialized with the $H - \alpha$ decomposition: indeed, we need an initial class that has some sense. We tried initializing the algorithm with one pixel randomly chosen across the image but the end results were not satisfying. Indeed, selecting only one pixel for the initialization leads to very different class centers from one execution of the algorithm to the other. A large number of pixels has to be selected to provide a robust initialization. A quick way to do this is to actually use the $H - \alpha$ decomposition. Note that a random initialization with a large number of pixels randomly chosen yields the same results, as discussed in Section V. One of the largest zones of the $H - \alpha$ decomposition is then chosen as the first class of the algorithm. Its center is computed as

$$\mathbf{C}_1 = \frac{1}{N_1} \sum_{k=1}^{N_1} \hat{\mathbf{T}}_{\text{SCM}}^{(k)}.$$

Notice that, for the SIRV case, $\hat{\mathbf{T}}_{\text{SCM}}$ is replaced by $\hat{\mathbf{M}}_{\text{FP}}$. All pixels are then reclassified according to the test of (10), used as a distance measure. Equation (11) gives the critical region of the test. In our case, the number of populations, k , is 2 and the number of variables, m , is 3 so the threshold over which the hypothesis is rejected (i.e., the covariance matrix of the considered pixel is not equal to the covariance matrix of the class) is given by the threshold of a χ^2 -distribution with 6 degrees of freedom and a type I error the user can define (we chose 10^{-3}). This means that pixels are not automatically put in one of the existing classes. A pixel is assigned to the class who minimizes the distance only if this minimal distance is below the threshold. If the minimal distance is over the threshold, the pixel is put in a rejection class, meaning that it is not close to any of the current classes. The number of classes has then to be incremented and the new class is initialized as the previous rejection class.

The selection of training data for the covariance matrix estimation is made with a 5×5 boxcar filter around the pixel under consideration. The Estimated Number of Looks N_e^S in SIRV environment is linked to the Estimated Number of Looks N_e^G in homogeneous Gaussian environment according to the asymptotic statistical behavior of the normalized Fixed Point Estimate [31]: $N_e^S = (m/(m+1))N_e^G$. For the 5×5 boxcar, we get $3/4 \times 25 = 18.75$ looks.

The classification procedure is summarized as follows.

- 1) Initially classify the image using the $H - \alpha$ decomposition. Pick one of the largest zones as the starting class of the algorithm.
- 2) Compute the class centers C_i .
- 3) For each pixel in the image, compute its distance to all the class centers using (10). If the minimal distance is below the threshold, assign the pixel to the corresponding class. Else, put it into the rejection class.
- 4) Once all pixels have been classified, check the termination criterion. If it is not met, increase the number of classes

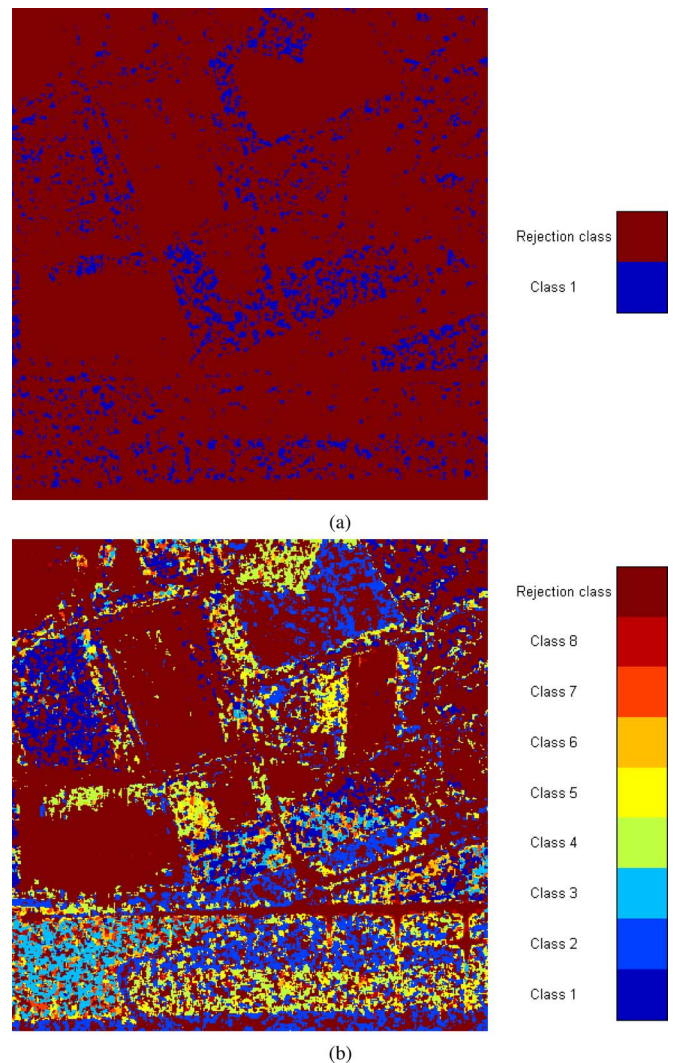


Fig. 8. Box classifier with SCM and associated colormaps. (a) One class + rejection class. (b) Eight classes + rejection class.

by 1 and go to step 2. The new class c_{i+1} is defined as the rejection class computed in iteration (i).

Notice that the introduction of a rejection class is more rigorous than the existing methods with respect to the statistical test procedure. Indeed, previous methods assigned a pixel to the class whose center minimizes the distance measure, even if this distance was very high.

V. RESULTS AND DISCUSSION

The test statistic of (10) has been applied successfully by Conradsen *et al.* in [32] for the Gaussian case in change detection. They were able to detect changes more precisely using this fully polarimetric test statistic.

The algorithm proposed in the previous section has been applied on the data described in Section III for classification. The termination criterion chosen is to stop the algorithm at eight classes (plus one rejection class) in order to compare the classification against a Wishart classifier with the standard eight classes from the $H - \alpha$ decomposition.

1) *SCM Estimate*: Fig. 8 shows the results of the algorithm of Section IV with the SCM estimate of the covariance matrix after 1 iteration on Fig. 8(a) and after eight iterations on Fig. 8(b).

TABLE I
NUMBER OF PIXELS IN EACH CLASS AT EACH ITERATION – SCM

Iteration	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Rejection class
1	23460	0	0	0	0	0	0	0	227541
2	26821	16182	0	0	0	0	0	0	207998
3	23675	30753	6798	0	0	0	0	0	189775
4	22855	36819	12138	5352	0	0	0	0	173837
5	22005	36504	11344	8163	5367	0	0	0	167618
6	21133	34355	10301	14309	5705	5102	0	0	160096
7	20467	32567	10548	16992	7680	4663	4228	0	153856
8	19405	30116	10437	17560	10803	6020	3947	3316	149397

TABLE II
NUMBER OF PIXELS IN EACH CLASS AT EACH ITERATION – FP\

Iteration	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Rejection class
1	134700	0	0	0	0	0	0	0	116301
2	93390	58805	0	0	0	0	0	0	98806
3	74961	48825	39973	0	0	0	0	0	87242
4	63934	50222	30778	29598	0	0	0	0	76469
5	58142	44901	28036	25856	22162	0	0	0	71904
6	52068	39529	27025	28466	19353	17327	0	0	67233
7	44831	35445	27310	28433	22625	13003	16012	0	63342
8	38178	32443	27764	27656	23804	13461	14610	13240	59845

After one iteration, most of the pixels of the image are rejected and no feature of the image is well discernible. When compared to Fig. 3(a), the pixels of the first class appear to belong to the yellow class in the $H - \alpha$ decomposition (which corresponds to a random surface with medium entropy). This is rather accurate since the pixels belong to fields.

After eight iterations, there are still rejected pixels in the fields areas, especially the top field which is still almost entirely rejected. However, the remaining rejected pixels outline for the most part the urban features of the scene: the two big buildings are clearly identified, as well as the small building next to them, the urban area on the top left, the parking lot on the right side and the dihedrals used for calibration. Even the biggest roads can be seen. Concerning the Gaussian parts, the bottom left field is pretty much identified as only one class but the others are a mixture of several classes and do not really stand out. On forested areas on the right part of the image, most of the area is rejected as well.

2) *FP Estimate*: For the FP estimate, the classification result after one iteration is encouraging. The number of rejected pixels is reduced compared to the SCM case (see Tables I and II for numbers of pixels in each class at every iteration) and the man-made features already stand out much more. The reduction of rejected pixels means a reduction in false alarms, which proves that the FP estimate is more adapted to the heterogeneous clutter than the SCM. After eight iterations, Fig. 9(b), the results are pretty close to the results of the SCM concerning the buildings and the dihedrals. On the rest of the image, the classification results are very different from the one obtained with the classical methods. Indeed, the patches are much smaller. This is due to the fact that the classification is not made on the power: as the estimation of the covariance is made with the spatial neighbors of the pixel under consideration, on classical methods, there can be stronger reflectors on nearby cases that will bias the classification results on a large area. It is difficult to provide an interpretation at this time since we do not have the ground truth

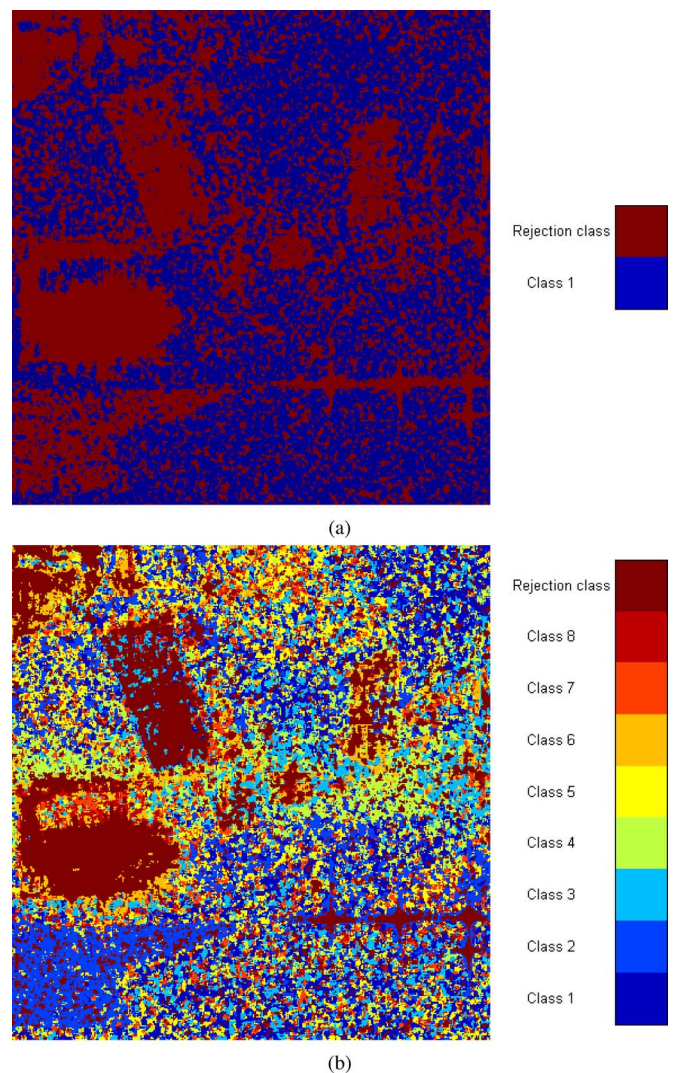


Fig. 9. Box classifier with FP and associated colormaps. (a) One class + rejection class. (b) Eight classes + rejection class.

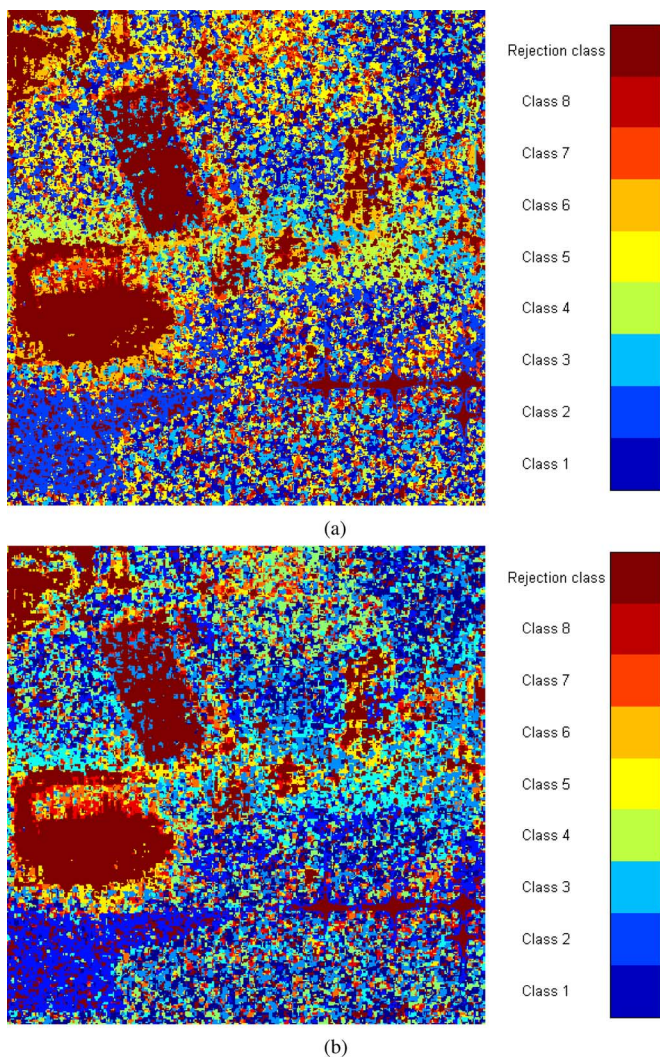


Fig. 10. Box classifier with FP and associated colormaps. (a) $H - \alpha$ initialization. (b) Random initialization.

data. A comparison with other classification methods is difficult as well, since there are very few classification results on the actual polarimetric information.

Furthermore, as mentioned in Section IV, we selected randomly a large number of pixels ($\sim 1/8$ of the total number) for the initialization to compare this method with the $H - \alpha$ initialization. The result of the algorithm with this random initialization is presented on Fig. 10. As in Section III, using a $H - \alpha$ initialization or a random initialization with a large number of pixels provides the same end results, even if the colors differ. It is thus possible to skip the $H - \alpha$ initialization phase for this scheme as well.

VI. CONCLUSION AND PERSPECTIVES

This paper presented a rigorous statistical approach for polarimetric SAR data classification. It relies on a simple binary test of equality of covariance matrices which can be applied to both homogeneous and heterogeneous clutter models.

The heterogeneous clutter in POLSAR data was described by the SIRV model. The Fixed Point estimate of the covariance matrix was used to describe the POLSAR data set. It is independent

of the texture pdf and is an AML estimator for many stochastic processes obeying the SIRV model. Moreover, it is asymptotically Wishart distributed. Performances of a classical classification procedure designed for Gaussian clutter and of a procedure adapted to the heterogeneous assumption were compared on a high resolution polarimetric data set.

The statistical test was presented and classical distance measures derived as particular cases of the test. The main improvement of this approach is to introduce a critical region which properly rejects pixels that are not close enough to the class centers, contrary to existing methods. The test statistic in the general case has been used as a distance measure in a new algorithm, applied on the POLSAR data set. Results show mainly that heterogeneous pixels are rejected from the rest of the image.

The proposed approach is very general and can be applied to other multidimensional SAR techniques such as polarimetric interferometry or multi-frequency polarimetry. Perspectives include the study of the class centers. They are defined as the mean of all covariance matrices inside a class but, although that may hold for the SCM, it does not for the FP estimate as the fixed point of a sum is not the sum of the fixed points. The geometry of covariance matrices have to be considered more accurately to use the correct center of mass. A short term goal is to compare the results of the classification to the ground truth data in order to have a better analysis.

REFERENCES

- [1] W. M. Boerner, Basic Concepts of Radar Polarimetry and its Applications to Target Discrimination, Classification, Imaging, and Identification Univ. of Illinois, Commun. Lab., Chicago, IL, 1982, EMID-CL-82-05-18-0.
- [2] J. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer. A*, vol. 66, pp. 1145–1149, 1976.
- [3] M. S. Greco and F. Gini, "Statistical analysis of high-resolution SAR ground clutter data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 566–575, Mar. 2007.
- [4] K. J. Sangston and K. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 2, pp. 330–340, Apr. 1994.
- [5] K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 5, pp. 600–608, Sep. 1973.
- [6] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Applicat.*, vol. 3, no. 8, pp. 747–768, 1989.
- [7] V. Anastassopoulos, G. A. Lampropoulos, A. Drosopoulos, and M. Rey, "High resolution radar clutter statistics," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 43–60, Jan. 1999.
- [8] E. Conte, A. de Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 903–918, Jul. 2004.
- [9] A. Fusco, C. Galdi, G. Ricci, and M. Tesauro, "Fitting a statistical model to SIR-C SAR images of the sea surface," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 4, pp. 1179–1190, Oct. 2004.
- [10] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2101–2111, Jun. 2005.
- [11] J. A. Kong, A. A. Swartz, and H. A. Yueh, "Identification of terrain cover using the optimal terrain classifier," *J. Electromagn. Waves Applicat.*, vol. 2, pp. 171–194, 1988.
- [12] H. A. Yueh, A. A. Swartz, J. A. Kong, R. T. Shin, and L. M. Novak, "Optimal classification of terrain cover using normalized polarimetric data," *J. Geophys. Res.*, pp. 15 261–15 267, 1993.
- [13] J. J. van Zyl and C. F. Burnette, "Bayesian classification of polarimetric SAR images using adaptive a priori probability," *Int. J. Remote Sens.*, vol. 13, no. 5, pp. 835–840, 1992.

- [14] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, 1994.
- [15] W. L. Cameron, N. Youssef, and L. K. Leung, "Simulated polarimetric signatures of primitive geometrical shapes," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 3, pp. 793–803, May 1996.
- [16] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, no. 18, pp. 1525–1527, Aug. 1990.
- [17] A. Freeman and S. Durden, "A three component scattering model to describe polarimetric SAR data," *Radar Polarimetry*, vol. 1748, pp. 213–225, 1992.
- [18] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.
- [19] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999.
- [20] G. Vasile, J.-P. Ovarlez, F. Pascal, and C. Tison, "Coherency matrix estimation of heterogeneous clutter in high-resolution polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1809–1826, Apr. 2010.
- [21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.
- [22] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. Gaithersburg: National Bureau of Standards, 1964, vol. AMS 55.
- [23] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Computer generation of correlated non-Gaussian radar clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 106–116, Jan. 1995.
- [24] J. B. Billingsley, *Ground Clutter Measurements for Surface-Sited Radar*. Mass. Inst. Technol., Cambridge, 1993, Tech. Rep. 780.
- [25] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy-tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [26] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1908–1915, Aug. 2002.
- [27] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 34–48, Jan. 2008.
- [28] M. S. Bartlett, "Properties of sufficiency and statistical tests," in *Proc. R. Soc. London*, 1937, vol. 160, pp. 268–282.
- [29] G. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, pp. 317–346, 1949.
- [30] G. Box, "Problems in the analysis of growth and linear curves," *Biometrika*, vol. 6, pp. 362–389, 1950.
- [31] F. Pascal, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *Signal Process.*, vol. 56, no. 6, pp. 2206–2216, 2008.
- [32] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 4–19, Jan. 2003.



Pierre Formont (S'10) was born in Villeurbanne, France, in 1986. He received jointly the engineering degree from Ecole Nationale Supérieure de l'Eau, l'Énergie et l'Environnement (ENSE3), Grenoble, France, and the M.S. degree in "Signal, Images, Speech, Telecom" with merit from Institut National Polytechnique de Grenoble (INPG) and Université Joseph Fourier (UJF), Grenoble, France, in 2009. He is currently pursuing the Ph.D. degree in image and signal processing at the French Aerospace Lab (ONERA), Palaiseau, France.



Frédéric Pascal (M'08) was born in Sallanches, France, in 1979. He received the M.S. degree ("Probabilities, Statistics and Applications: Signal, Image et Networks") with merit, in applied statistics from University Paris VII–Jussieu, Paris, France, in 2003 and the Ph.D. degree in signal processing from the University Paris X – Nanterre, advised by Pr. Philippe Forster: "Detection and Estimation in Compound Gaussian Noise" in 2006. This Ph.D. thesis was in collaboration with the French Aerospace Lab (ONERA), Palaiseau, France.

From November 2006 to February 2008, he made a postdoctoral position in the Signal Processing and Information Team, Laboratory Système et Applications des Technologies de l'Information et de l'Énergie (SATIE), CNRS, Ecole Normale Supérieure, Cachan, France. Since March 2008, he has been an Assistant Professor in the SONDRALaboratory, SUPELEC, Gif-sur-Yvette, France. His research interests are estimation in statistical signal processing and radar detection.



Gabriel Vasile (S'06–M'07) received the M.Eng. degree in electrical engineering and computer science and the M.S. degree in image, shapes, and artificial intelligence from the POLITEHNICA University, Bucharest, Romania, in 2003 and 2004, respectively, and the Ph.D. degree in signal and image processing from Savoie University, Annecy, France, in 2007.

From 2007 to 2008, he was a Postdoctoral Fellow with the French Space Agency (CNES) and was with the French Aerospace Laboratory (ONERA), Palaiseau, France. In 2008, he joined the French

National Council for Scientific Research (CNRS), where he is currently a Research Scientist and a member of the Grenoble Image Speech Signal Automatics Laboratory, Grenoble, France. His research interests include signal and image processing, synthetic aperture radar remote sensing, polarimetry, and interferometry.



Jean-Philippe Ovarlez (M'06) was born in Denain, France, in 1963. He received jointly the engineering degree from Ecole Supérieure d'Électronique Automatique et Informatique (ESIEA), Paris, France, and the Diplôme d'Études Approfondies degree in Signal Processing from University of Orsay (Paris XI), Orsay, France, and the Ph.D. degree in physics from the University of Paris VI, Paris, France, in 1987 and 1992, respectively.

In 1992, he joined the Electromagnetic and Radar Division of the French Aerospace Lab (ONERA), Palaiseau, France, where he is currently Principal Scientist and member of the Scientific Committee of the ONERA Physics Branch. In January 2008, he joined the French–Singaporean SONDRALab on a part-time basis to supervise Signal Processing topics. His current activities of research are centered in the topic of statistical signal processing for radar and SAR applications such as time–frequency, imaging, detection, and parameters estimation.



Laurent Ferro-Famil received the Laurea degree in electronics systems and computer engineering, the M.S. degree in electronics, and the Ph.D. degree from the University of Nantes, Nantes, France, in 1996, 1996, and 2000, respectively.

In 2001, he became an Assistant Professor, and, since 2007, he has been an Associate Professor with the University of Rennes 1, Rennes, France, where he is currently a Member of the Radar Polarimetry Remote Sensing Group, Institute of Electronics and Telecommunications of Rennes. His current activities

in education are concerned with analog electronics, digital communications, microwave theory, and polarimetric SAR remote sensing. He is particularly interested in polarimetric SAR signal processing, radar polarimetry theory, and natural media remote sensing using multibaseline PolInSAR data, with application to classification, electromagnetic scattering modeling and physical parameter retrieval, time–frequency analysis, and 3-D reconstruction of environments.

English summary

Large-scale observation of the Earth can be performed mainly by two methods : observation through optical systems or observation through radar systems. Radar systems possess several advantages over optical systems — notably their ability to operate at all times and in all kinds of weather — that contributed to their expansion in a large number of applications, such as surveillance, navigation, foliage penetration, land-cover classification, target detection, etc. One of the most widely used radar systems is the Synthetic Aperture Radar (SAR) which provides high spatial resolution of the observed area. A SAR system emits an electromagnetic wave and measures the amplitude and the phase of the backscattered signal, thus creating a complex image of the illuminated area — called scene. By using the polarization property of electromagnetic waves, polarimetric SAR (PolSAR) systems are able to provide more information on the scene as the pixels of the resulting image contain several measurements, one for each couple of polarization constraints put on the emitting and the receiving antennas. The differences in the measurements for each pixel are directly related to the physical nature of the object (or objects) that are present in this pixel. This information can be exploited by considering the backscattered signal as a random vector and computing its covariance matrix. Traditional methods for PolSAR data processing model the backscattered signal as a complex Gaussian random vector. The covariance matrix of a pixel is then estimated thanks to N secondary data — pixels surrounding the pixel under consideration — with a maximum likelihood approach, which yields the well-known Sample Covariance Matrix (SCM). This matrix follows a distribution known as the Wishart distribution. Polarimetric properties can be extracted from the covariance matrix thanks to a number of decompositions, such as the Cloude-Pottier decomposition, the Freeman-Durden decomposition, the Yamaguchi decomposition, etc. A large number of algorithms have been proposed to perform classification using the covariance matrix. One of the most well-known and most widely used is the Wishart classifier. However, the increase in resolution in recent SAR systems has led to a higher scene heterogeneity, particularly in urban areas, where the backscattered signal can no longer be modeled as a Gaussian process.

The aim of this PhD is then to study statistical and geometrical tools for the classification of high-resolution polarimetric SAR images. Indeed, in this kind of images, especially over urban areas, the traditional techniques derived from the Gaussian assumption fail to properly classify the polarimetric information. The use of a compound model, called the SIRV model (for Spherically Invariant Random Vectors), allows to take into account the heterogeneity and non-Gaussianity in those images. The impact of this model has then been studied for the classification of SAR images and the interpretation of the classification results with respect to the polarimetric information. Indeed, in classical techniques, the polarimetric information has actually much less impact on the classification results than the power information. A statistical test is proposed to improve the classification and its statistic under Gaussiand and SIRV hypothesis are studied. Furthermore, an important step in most traditional techniques is the computation of the mean of covariance matrices, usually done by computing the arithmetic mean. However, the structure of the space of covariance matrices is not Euclidean. This definition of the mean then does not apply and it is necessary to take into account the Riemannian structure of the space in order to provide a suitable definition of the mean. These considerations are studied on simulated data and on real data, acquired by the ONERA. Finally, a tentative extension of this work to hyperspectral classification is proposed, thanks to the relative closeness of the two fields. The thesis is divided in four parts, which are broken down in more detail in the following.

The first part is devoted to the fundamentals of SAR classification. The concept of Synthetic Aperture Radar is presented along with polarization and its interest for classification. Classical classification techniques, based on a Gaussian assumption for the distribution of the backscattered signal from the radar are recalled. A major drawback of this assumption is then brought out : these techniques depend much more on the power information rather than the polarimetric information. Indeed, a classification on the power of each pixel gives the same results than a classification using the estimator of the covariance matrix under Gaussian assumption — the SCM. With such a strong influence of the power information, it is reasonable to wonder about the relevancy of the classical model with respect to a proper utilization of the polarimetric information. The non-Gaussian SIRV model is then introduced, along with the related statistical tools. Under SIRV assumption, the estimator of the covariance matrix is called the Fixed Point Estimate (FPE) and has a range of interesting properties, notably that it behaves asymptotically like a Wishart-distributed matrix, as if the data were Gaussian-distributed. This is an important property because it allows to easily extend techniques developed under Gaussian assumption to the SIRV model. The FPE allows to take into account the heterogeneity of high-resolution SAR images and removes the power information, which leads to classification results depending only on the polarimetric information.

In the second part, a statistical test of equality of covariance matrices is proposed to address some drawbacks of the traditional classification techniques, such as a fixed number of classes or the automatic assignment of a pixel to an already existing class, without taking into account this pixel could be very different from every known class. The classical test is recalled for the Gaussian assumption then extended to the SIRV assumption, thanks to the asymptotic distribution of the FPE. This test is an hypothesis test trying to determine if the covariance matrices of two samples are equal. Two cases are considered : the case where both covariance matrices are unknown and the case where one of those is known. The test statistic is used as a distance between two pixels in an original classification algorithm and compared to the classical Wishart distance. The critical region of the test allows to reject pixels that do not belong to any of the existing classes. It appears that rejected pixels are pixels from artificial areas of the image : buildings, parking lot, etc. The use of this distance in existing classification algorithms — Wishart Classifier and hierarchical clustering — shows the same phenomenon : good separation between natural and artificial areas but not much improvement inside these two types of zones. Besides, the results obtained with the FPE look less “appealing” than those obtained with the SCM but the visible information is different : power for the SCM and polarimetry for the FPE. This leads to a difficult interpretation because the classification results for the FPE are very different from the optical images of the area, unlike the results from the SCM.

The third chapter is dedicated to the introduction of information geometry for a better computation of the mean of covariance matrices. Indeed, classical algorithms often need to compute the mean of covariance matrices, as a way to provide a center for the each different class : for each class, the covariance matrices of all pixels belonging in that class are averaged thanks to the classical Euclidean, arithmetic mean. However, due to the structure of the covariance matrices — Hermitian definite-positive — and their mathematical space, this definition of the mean is not correct. First, the basic principles of information and Riemannian geometries are recalled, as well as the method to derive a geometric distance between two covariance matrices, which leads to the definition of the geometric mean. The impact of using this distance and this expression of the mean is then studied on simulated polarimetric SAR data. The simulated data follows a non-Gaussian distribution, called the K-distribution. Simulations demonstrate a great impact of the geometric mean over the arithmetic mean in terms of classification accuracy. The use of the geometric distance seems however not to be impactful. On real data, the same phenomenon can be observed : the geometric mean allows to separate features of the image that were incorrectly grouped with other features using the arithmetic mean. The geometric distance does not improve the classification and sometimes even worsens it. Its use is therefore not recommended until further study.

Finally, in the last part, the FPE and the geometric mean are used on hyperspectral images, in order to study the impact of these techniques for data close to polarimetric SAR data. Data from the NASA Hyperion sensor are used to apply a classification algorithm, close to the Wishart classifier, and compare the results with a standard hyperspectral classifier, the Spectral Angle Mapper (SAM). Preliminary results show that the direct use of the covariance matrix leads to a large loss of precision compared to standard algorithms for hyperspectral classification that use only the data vector. Indeed, using the covariance matrix means estimating it, which is done on a large neighborhood around the pixel, hence the precision loss. However, a better exploitation of the added data contained in the covariance matrix could prove interesting. Notably, the use of the Mahalanobis distance which uses a matrix to describe the classes yields promising results.

Perspectives

During this PhD, the focus has been the utilisation of the polarimetric information for classification, which is contained in the covariance matrix. It has been shown that the estimator for the Gaussian case, the SCM, contains more power information than polarimetric information, which led to the use of the FPE who does not rely on the power information at all. However, the classification results using the SCM make the natural areas stand out very clearly, much more than when using the FPE. The power information could then be useful to improve the visual aspect of the classification results.

Otherwise, there is a very strong assumption used for the test of equality of covariance matrices, namely that the FPE is asymptotically Wishart-distributed. The size of the sample used for the estimation of the covariance matrix is rather large (around fifty samples) so the “real” behavior is not far from the asymptotic one, but knowing the real distribution of the FPE with a finite number of samples would be better. A discussion on this point has been started with the use of the hierarchical clustering algorithm.

The focus on the polarimetric information leads also to wonder about the interpretation of classification results by existing polarimetric tools. The Cloude-Pottier decomposition has been used to propose rough interpretations of the differences between classification results but other decompositions could be applied : the Freeman-Durden decomposition or the Yamaguchi decomposition for example, especially since there are a lot of papers presented regularly on the Yamaguchi decomposition and its improvements.

Information geometry provides powerful tools which allowed to increase the precision in classification results (and made them correct, since the usual arithmetic mean is not the proper mean for covariance matrices) but whose computational time is much higher. Optimization strategies can be established to quicken this processing. Moreover, only the mean has been mentioned but it is also possible to use the median, which is more robust to possible outliers in the data.

Liste des publications

Publications dans des journaux internationaux à comité de lecture :

- [J1] Pierre Formont, Frédéric Pascal, Gabriel Vasile, Jean-Philippe Ovarlez et Laurent Ferro-Famil, “Statistical Classification for Heterogeneous Polarimetric SAR Images”, *IEEE Journal of Selected Topics in Signal Processing*, vol. 5(3), pp. 567 – 576, 2011.
- [J2] Gabriel Vasile, Frédéric Pascal, Jean-Philippe Ovarlez, Pierre Formont et Michel Gay, “Optimal Parameter Estimation in Heterogeneous Clutter for High-Resolution Polarimetric SAR Data”, *IEEE Geoscience and Remote Sensing Letters*, vol. 8(6), pp. 1046 – 1050, 2011.

Publications dans des conférences internationales avec actes :

- [C1] Pierre Formont, Miguel Angel Veganzones, Joana Maria Frontera-Pons, Frédéric Pascal, Jean-Philippe Ovarlez et Jocelyn Chanussot, “CFAR Hierarchical Clustering of Polarimetric SAR Data”, *IEEE 2013 International Geoscience and Remote Sensing Symposium (IGARSS)*, Melbourne, Australie, 21 – 26 juillet 2013.
- [C2] Pierre Formont, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “Information Geometry for Classification of High Resolution Polarimetric SAR Images”, *IEEE 2012 International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Allemagne, 22 – 27 juillet 2012.
- [C3] Jean-Philippe Ovarlez, Pierre Formont, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “Contribution of information geometry for polarimetric SAR classification in heterogeneous areas”, *Proceedings of the 2011 International Radar Symposium*, Leipzig, Allemagne, 7 – 9 septembre 2011.
- [C4] Pierre Formont, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “On the Extension of the Product Model in POLSAR Processing for Unsupervised Classification Using Information Geometry of Covariance Matrices”, *IEEE 2011 International Geoscience and Remote Sensing Symposium (IGARSS)*, Vancouver, Canada, 24 – 29 juillet 2011.
- [C5] Pierre Formont, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “H/ α Unsupervised Classification for Highly Textured Polinsar Images using Information Geometry of Covariance Matrices”, *2011 PolInSAR Workshop*, Frascati, Italie, 24 – 28 janvier 2011.
- [C6] Pierre Formont, Nicolas Trouvé, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Elise Colin-Koeniguer, “PolSAR Classification Based on the SIRV Model with a Region Growing Initialization”, *2011 PolInSAR Workshop*, Frascati, Italie, 24 – 28 janvier 2011.
- [C7] Pierre Formont, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “A Test Statistic for High Resolution Polarimetric SAR Data Classification”, *IEEE 2010 International Geoscience and Remote Sensing Symposium (IGARSS)*, Honolulu, USA, 25 – 30 juillet 2010.
- [C8] Pierre Formont, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “A New Method for High Resolution Polarimetric SAR Data Classification Based on the M-Box Test”, *8th European Conference on Synthetic Aperture Radar (EUSAR)*, Aachen, Allemagne, 07 – 10 juin 2010.

Publications dans des conférences internationales sans actes :

- [C9] Pierre Formont, Miguel Angel Veganzones, Joana Maria Frontera-Pons, Frédéric Pascal, Jean-Philippe Ovarlez et Jocelyn Chanussot, “CFAR Hierarchical Clustering of Polarimetric SAR Data”, *3ème Workshop SONDRRA*, Hyères, France, 10 – 14 juin 2013.
- [C10] Pierre Formont, Jean-Philippe Ovarlez, Frédéric Pascal, Gabriel Vasile et Laurent Ferro-Famil, “Polarimetric Classification in Highly Textured SAR Images”, *2ème Workshop SONDRRA*, Cargèse, France, 25 – 28 mai 2010.

Publications dans des conférences nationales avec actes :

- [C11] Pierre Formont, Frédéric Pascal, Jean-Philippe Ovarlez, Gabriel Vasile et Laurent Ferro-Famil, “Apport de la Géométrie de l’Information pour la Classification d’Images SAR Polarimétriques”, *Vingt-troisième Colloque GRETSI 2011*, Bordeaux, France, septembre 2011.

Chapitres d’ouvrages :

- [B1] Pierre Formont, Jean-Philippe Ovarlez et Frédéric Pascal, “On the use of Matrix Information Geometry for Polarimetric SAR Image Classification”, *Matrix Information Geometry*, Springer, pp. 257 – 276, 2013.

Bibliographie

- [1] M. Abramowitz, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*,. Dover Publications, Incorporated, 1974. (Cité en pages 14 et 23.)
- [2] V. Anastassopoulos, G. A. Lampropoulos, A. Drosopoulos, and M. Rey, “High resolution radar clutter statistics,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 43–60, 1999. (Cité en page 24.)
- [3] D. Arthur and S. Vassilvitskii, “k-means++ : the advantages of careful seeding,” in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, ser. SODA '07. Philadelphia, PA, USA : Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035. [Online]. Available : <http://portal.acm.org/citation.cfm?id=1283494> (Cité en page 55.)
- [4] C. Atkinson and A. F. Mitchell, “Rao’s distance measure,” *Sankhyā : The Indian Journal of Statistics, Series A*, pp. 345–365, 1981. (Cité en pages 2 et 75.)
- [5] R. Azzam and N. Bashara, *Ellipsometry and Polarized Light*. North Holland, Amsterdam, The Netherlands, 1977. (Cité en page 11.)
- [6] F. Barbaresco, “Innovative tools for radar signal processing based on cartan’s geometry of spd matrices and information geometry,” *IEEE International Radar Conference*, 2008. (Cité en pages 2 et 74.)
- [7] —, “Robust median-based stap in inhomogeneous secondary data :frechet information geometry of covariance matrices,” *2nd French-Singaporean SONDRRA Workshop on EM Modeling, New Concepts and Signal Processing For Radar Detection and Remote Sensing*, 2010. (Cité en pages 2 et 74.)
- [8] —, “Information geometry of covariance matrix : Cartan-siegel homogenous bounded domains, mostow/berger fibration and frechet median,” in *Matrix Information Geometry*. Springer, 2013, pp. 199–255. (Cité en pages 2 et 74.)
- [9] M. S. Bartlett, “Properties of sufficiency and statistical tests,” *Proceedings of the Royal Society of London*, vol. 160, pp. 268–282, 1937. (Cité en page 45.)
- [10] J.-M. Beaulieu and R. Touzi, “Segmentation of textured polarimetric sar scenes by likelihood approximation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2063 – 2072, oct. 2004. (Cité en page 41.)
- [11] P. Beckmann, *The Depolarization of Electromagnetic Waves*. The Golem Press, Boulder, CO, 1968. (Cité en page 11.)
- [12] J. B. Billingsley, “Ground clutter measurements for surface-sited radar,” MIT, Tech. Rep. 780, February 1993. (Cité en page 2.)
- [13] H. Bischof and A. Leonardis, “Finding optimal neural networks for land use classification,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 36, no. 1, pp. 337–341, 1998. (Cité en page 96.)
- [14] L. Bombrun and J.-M. Beaulieu, “Fisher distribution for texture modeling of polarimetric sar data,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, no. 3, pp. 512–516, 2008. (Cité en page 25.)
- [15] G. Box, “A general distribution theory for a class of likelihood criteria,” *Biometrika*, vol. 36, 1949. (Cité en pages 45, 46 et 69.)
- [16] —, “Problems in the analysis of growth and linear curves,” *Biometrika*, vol. 6, pp. 362–389, 1950. (Cité en pages 45 et 46.)
- [17] J. Burbea, “Informative geometry of probability spaces,” DTIC Document, Tech. Rep., 1984. (Cité en pages 2 et 75.)

Bibliographie

- [18] H. Caillol, A. Hillion, and W. Pieczynski, "Fuzzy random fields and unsupervised image segmentation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 31, no. 4, pp. 801–810, 1993. (Cité en page 96.)
- [19] W. L. Cameron, N. Youssef, and L. K. Leung, "Simulated polarimetric signatures of primitive geometrical shapes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 3, pp. 793–803, May 1996. (Cité en page 1.)
- [20] G. Camps-Valls, T. Bados Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 10, pp. 3044–3054, 2007. (Cité en page 96.)
- [21] G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla, J. D. Martín-Guerrero, E. Soria-Olivas, L. Alonso-Chordá, and J. Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 7, pp. 1530–1542, 2004. (Cité en page 96.)
- [22] N. Charon and F. Barbaresco, "Une nouvelle approche pour la détection de cibles dans les images radar basée sur des distances et moyennes dans des espaces de matrices de covariance," *TS. Traitement du signal*, vol. 26, no. 4, pp. 269–278, 2009. (Cité en pages 75 et 78.)
- [23] D. L. Civco, "Artificial neural networks for land-cover classification and mapping," *International Journal of Geographical Information Science*, vol. 7, no. 2, pp. 173–186, 1993. (Cité en page 96.)
- [24] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 68–78, January 1997. (Cité en pages 2, 15, 16 et 18.)
- [25] S. Cloude, "Polarimetry : The characterization of polarimetric effects in em scattering," Ph.D. dissertation, University of Birmingham, Fac. of Eng., UK, 1986. (Cité en page 15.)
- [26] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex wishart distribution and its application to change detection in polarimetric sar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 4–19, January 2003. (Cité en page 41.)
- [27] E. Conte, A. de Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 903–918, 2004. (Cité en page 24.)
- [28] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-gaussian clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 31, no. 2, pp. 617–625, 1995. (Cité en page 26.)
- [29] J. C. Curlander and R. N. McDonough, *Synthetic aperture radar : systems and signal processing*. Wiley New York, 1991, vol. 199, no. 1. (Cité en pages 7 et 8.)
- [30] V. Devlaminck and P. Terrier, "Geodesic distance on non-singular coherency matrix space in polarization optics," *Journal of the Optical Society of America*, vol. 27, no. 3, 2010. (Cité en pages 2 et 74.)
- [31] K. Ersahin, I. Cumming, and R. Ward, "Segmentation and classification of polarimetric sar data using spectral graph partitioning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 164–174, jan. 2010. (Cité en page 41.)
- [32] C. Fang, H. Wen, W. Yirong, and E. Pottier, "Analysis of fully polarimetric sar data based on the cloude-pottier decomposition and the complex wishart classifier," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, july 2007, pp. 168–171. (Cité en page 34.)
- [33] A. Farina, A. Russo, and F. Studer, "Advanced models of targets and disturbances and related radar signal processors'," in *Proceedings of IEEE international radar conference, Washington, DC*, 1985, pp. 151–158. (Cité en page 2.)

- [34] A. Freeman and S. Durden, "A three component scattering model to describe polarimetric SAR data," *Radar Polarimetry*, vol. 1748, pp. 213–225, 1992. (Cité en page 2.)
- [35] A. Fusco, C. Galdi, G. Ricci, and M. Tesauro, "Fitting a statistical model to SIR-C SAR images of the sea surface," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 4, pp. 1179–1190, 2004. (Cité en page 24.)
- [36] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy-tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002. (Cité en page 25.)
- [37] J. Goldman, "Detection in the presence of spherically symmetric random vectors," *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 52–59, 1976. (Cité en page 2.)
- [38] J. W. Goodman, "Some fundamental properties of speckle," *Journal of the Optical Society of America*, vol. 66, no. 11, pp. 1145–1150, Nov 1976. [Online]. Available : <http://www.opticsinfobase.org/abstract.cfm?URI=josa-66-11-1145> (Cité en pages 12 et 13.)
- [39] M. S. Greco and F. Gini, "Statistical analysis of high-resolution sar ground clutter data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 566 –575, march 2007. (Cité en page 25.)
- [40] J. Gualtieri, S. Chettri, R. Crompt, and L. Johnson, "Support vector machine classifiers as applied to aviris data," in *Proc. Eighth JPL Airborne Geoscience Workshop*. Citeseer, 1999. (Cité en page 96.)
- [41] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics : The Approach Based on Influence Functions*, ser. Wiley Series in Probability and Statistics. Wiley, 2011. [Online]. Available : <http://books.google.fr/books?id=XK3uhrVefXQC> (Cité en page 41.)
- [42] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of Remote Sensing*, vol. 23, no. 4, pp. 725–749, 2002. (Cité en page 96.)
- [43] T. Ince, "Unsupervised classification of polarimetric sar image with dynamic clustering : An image processing approach," *Advances in Engineering Software*, vol. 41, no. 4, pp. 636 – 646, 2010. [Online]. Available : <http://www.sciencedirect.com/science/article/pii/S0965997809002567> (Cité en page 34.)
- [44] E. Jakeman, "On the statistics of k-distributed noise," *Journal of Physics A : Mathematical and General*, vol. 13, no. 1, p. 31, 1980. (Cité en page 22.)
- [45] K. Kimura, Y. Yamaguchi, and H. Yamada, "Pi-sar image analysis using polarimetric scattering parameters and total power," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol. 1, july 2003, pp. 425 – 427 vol.1. (Cité en page 47.)
- [46] J. A. Kong, A. A. Swartz, and H. A. Yueh, "Identification of terrain cover using the optimal terrain classifier," *Journal of Electromagnetic Waves and Applications*, vol. 2, pp. 171–194, 1988. (Cité en pages 2 et 18.)
- [47] A. Kostinski and W. Boerner, "On foundations of radar polarimetry," *IEEE Transactions on Antennas Propagation*, vol. 34, no. 12, pp. 1395–1404, 1986. (Cité en pages 10 et 12.)
- [48] E. Krogager, "New decomposition of the radar target scattering matrix," *Electronics Letters*, vol. 26, no. 18, pp. 1525–1527, August 1990. (Cité en page 1.)
- [49] J.-S. Lee, "Refined filtering of image noise using local statistics," *Computer Graphics and Image Processing*, vol. 15, no. 4, pp. 380 – 389, 1981. [Online]. Available : <http://www.sciencedirect.com/science/article/pii/S0146664X81800184> (Cité en page 58.)
- [50] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2249–2258, September 1999. (Cité en pages 2, 19, 47 et 71.)

Bibliographie

- [51] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, 1994. (Cité en pages 2, 18, 20 et 39.)
- [52] J.-S. Lee, M. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 722 – 731, avril 2004. (Cité en pages 34 et 47.)
- [53] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging, From Basics to Applications*. CRC Press, 2009. (Cité en pages 12 et 15.)
- [54] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1. New Delhi, 1936, pp. 49–55. (Cité en pages 27 et 96.)
- [55] M. Moakher, "Differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM Journal of Matrix Analysis and Applications*, vol. 26, no. 3, pp. 735–747, 2005. (Cité en pages 2 et 74.)
- [56] H. Mott, *Antennas for Radar and Communications*. John Wiley & Sons, New York, 1992. (Cité en page 10.)
- [57] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, no. 694-706, pp. 289–337, 1933. [Online]. Available : <http://rsta.royalsocietypublishing.org/content/231/694-706/289.short> (Cité en page 34.)
- [58] L. M. Novak and M. C. Burl, "Optimal speckle reduction in polarimetric sar imagery," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 26, no. 2, pp. 293–305, 1990. (Cité en page 27.)
- [59] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes with errata sheet*. New York, NY, McGraw-Hill Education, 2002. (Cité en page 14.)
- [60] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound gaussian noise : Existence and algorithm analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 34–48, 2008. (Cité en pages 25 et 26.)
- [61] F. Pascal, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *Signal Processing*, vol. 56, no. 6, pp. 2206–2216, 2008. (Cité en pages 26 et 27.)
- [62] W. J. Plant and W. C. Keller, "Evidence of bragg scattering in microwave doppler spectra of sea return," *Journal of Geophysical Research : Oceans (1978–2012)*, vol. 95, no. C9, pp. 16 299–16 310, 1990. (Cité en page 18.)
- [63] E. Pottier and J. Lee, "Unsupervised classification scheme of polsar images based on the complex wishart distribution and the $h/\alpha/a$ polarimetric decomposition theorem," in *Proceedings of EUSAR 2000*, Munich, Germany, 2000. (Cité en page 47.)
- [64] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Computer generation of correlated non-gaussian radar clutter," *IEEE Transactions on Aeronautics and Electronic Systems*, vol. 31, no. 1, pp. 106–116, 1995. (Cité en page 81.)
- [65] A. C. Rencher, *Methods of Multivariate Analysis*. John Wiley and Sons, 2002. (Cité en page 45.)
- [66] K. J. Sangston and K. Gerlach, "Coherent detection of radar targets in a non-gaussian background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 330–340, 1994. (Cité en page 24.)
- [67] H. Schey, *Div, Grad, Curl, And All That : An Informal Text On Vector Calculus*. WW Norton & Company, 2005. (Cité en pages 8 et 9.)
- [68] G. G. Stokes, "On the composition and resolution of streams of polarized light from different sources," *Transactions of the Cambridge Philosophical Society*, vol. 9, p. 399, 1851. (Cité en page 74.)

- [69] J. Stratton, *Electromagnetic Theory*. McGraw-Hill, New York, 1941. (Cité en pages 8 et 9.)
- [70] N. Trouvé, “Comparaison des outils optique et radar en polarimétrie bistatique,” Thèse de doctorat, Ecole Polytechnique X, Nov. 2011. [Online]. Available : <http://tel.archives-ouvertes.fr/tel-00676316> (Cité en page 54.)
- [71] G. Trunk and S. George, “Detection of targets in non-gaussian sea clutter,” *Aerospace and Electronic Systems, IEEE Transactions on*, no. 5, pp. 620–628, 1970. (Cité en page 2.)
- [72] F. T. Ulaby, F. Kouyate, B. Brisco, and T. L. Williams, “Textural information in sar images,” *Geoscience and Remote Sensing, IEEE Transactions on*, no. 2, pp. 235–245, 1986. (Cité en page 23.)
- [73] F. Ulaby, T. Haddock, and R. Austin, “Fluctuation statistics of millimeter-wave scattering from distributed targets,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 26, no. 3, pp. 268–281, 1988. (Cité en page 22.)
- [74] J. J. van Zyl and B. C. F, “Bayesian classification of polarimetric SAR images using adaptive a priori probability,” *International Journal of Remote Sensing*, vol. 13, no. 5, pp. 835–840, 1992. (Cité en page 2.)
- [75] G. Vasile, J.-P. Ovarlez, F. Pascal, and C. Tison, “Coherency matrix estimation of heterogeneous clutter in high resolution polarimetric SAR images,” *IEEE Transactions on Geoscience and Remote Sensing*, 2011. (Cité en page 27.)
- [76] K. Ward, “Compound representation of high resolution sea clutter,” *Electronics letters*, vol. 17, no. 16, pp. 561–563, 1981. (Cité en page 2.)
- [77] S. Watts, “Radar detection prediction in sea clutter using the compound k-distribution model,” in *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 132, no. 7. IET, 1985, pp. 613–620. (Cité en page 2.)
- [78] W. Weibull *et al.*, “A statistical distribution function of wide applicability,” *Journal of applied mechanics*, vol. 18, no. 3, pp. 293–297, 1951. (Cité en page 22.)
- [79] T. Yamazaki and D. Gingras, “Unsupervised multispectral image classification using mrf models and vq method,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 2, pp. 1173–1176, 1999. (Cité en page 96.)
- [80] H. Yang, “A back-propagation neural network for mineralogical mapping from aviris data,” *International Journal of Remote Sensing*, vol. 20, no. 1, pp. 97–110, 1999. (Cité en page 96.)
- [81] K. Yao, “A representation theorem and its applications to spherically-invariant random processes,” *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 600–608, 1973. (Cité en pages 2 et 24.)
- [82] H. A. Yueh, A. A. Swartz, J. A. Kong, R. T. Shin, and L. M. Novak, “Optimal classification of terrain cover using normalized polarimetric data,” *Journal of Geophysical Resources*, pp. 15 261–15 267, 1993. (Cité en page 2.)
- [83] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, “K-distribution and polarimetric terrain radar clutter,” *Journal of Electromagnetic Waves and Applications*, vol. 3, no. 8, pp. 747–768, 1989. (Cité en pages 23 et 24.)
- [84] Y. Zhong, L. Zhang, B. Huang, and P. Li, “An unsupervised artificial immune classifier for multi/hyper-spectral remote sensing imagery,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 2, pp. 420–431, 2006. (Cité en page 96.)

Outils statistiques et géométriques pour la classification des images SAR polarimétriques hautement texturées

Les radars à synthèse d'ouverture (Synthetic Aperture Radar ou SAR) permettent de fournir des images à très haute résolution de la surface de la Terre. Les algorithmes de classification traditionnels se basent sur une hypothèse de bruit gaussien comme modèle de signal, qui est rapidement mise en défaut lorsque l'environnement devient inhomogène ou impulsif, comme c'est particulièrement le cas dans les images SAR polarimétriques haute résolution, notamment au niveau des zones urbaines. L'utilisation d'un modèle de bruit composé, appelé modèle SIRV, permet de mieux prendre en compte ces phénomènes et de représenter la réalité de manière plus adéquate. Cette thèse s'emploie alors à étudier l'application et l'impact de ce modèle pour la classification des images SAR polarimétriques afin d'améliorer l'interprétation des classifications au sens de la polarimétrie et à proposer des outils adaptés à ce nouveau modèle. En effet, il apparaît rapidement que les techniques classiques utilisent en réalité beaucoup plus l'information relative à la puissance de chaque pixel plutôt qu'à la polarimétrie pour la classification. Par ailleurs, les techniques de classification traditionnelles font régulièrement appel à la moyenne de matrices de covariance, calculée comme une moyenne arithmétique. Cependant, étant donnée la nature riemannienne de l'espace des matrices de covariance, cette définition n'est pas applicable et il est nécessaire d'employer une définition plus adaptée à cette structure riemannienne. Nous mettons en évidence l'intérêt d'utiliser un modèle de bruit non gaussien sur des données réelles et nous proposons plusieurs approches pour tirer parti de l'information polarimétrique qu'il apporte. L'apport de la géométrie de l'information pour le calcul de la moyenne est de même étudié, sur des données simulées mais également sur des données réelles acquises par l'ONERA. Enfin, une étude préliminaire d'une extension de ces travaux au cas de l'imagerie hyperspectrale est proposée, de par la proximité de ce type de données avec les données SAR polarimétriques.

Mots clés : classification radar, fouillis non gaussien, polarimétrie, géométrie de l'information

Statistical and geometric tools for the classification of highly textured polarimetric SAR images

Synthetic Aperture Radars (SAR) now provide high resolution images of the Earth surface. Traditional classification algorithms are based on a Gaussian assumption for the distribution of the signal, which is no longer valid when the background is heterogeneous, which is particularly the case for polarimetric SAR images, especially in urban areas. A compound Gaussian model, called the SIRV model, allows to take into account these phenomena. This thesis is then devoted to studying the impact of this model for the classification of polarimetric SAR images in order to improve the interpretation of classification results in a polarimetric sense, and to propose tools better suited to this model. Indeed, classical techniques using the Gaussian assumption actually use the power information of each pixel much more than the polarimetric information. Furthermore, it is often necessary to compute a mean of covariance matrices, usually by taking the standard arithmetical mean. However, the space of covariance matrices has a Riemannian structure, not an Euclidean one, which means this definition of the mean is not correct. We will then present several methods to use the actual polarimetric information thanks to the SIRV model to improve the classification results. The benefit of using a correct, Riemannian definition of the mean will also be demonstrated on simulated and real data. Finally, a preliminary study of an extension of this work to hyperspectral imagery will be presented.

Keywords : radar classification, non-Gaussian clutter, polarimetry, information geometry