



Fusion of heterogeneous remote sensing images by credibilist methods

Imen Hammami

► To cite this version:

Imen Hammami. Fusion of heterogeneous remote sensing images by credibilist methods. Image Processing [eess.IV]. Ecole nationale supérieure Mines-Télécom Atlantique, 2017. English. NNT : 2017IMTA0034 . tel-01814777

HAL Id: tel-01814777

<https://theses.hal.science/tel-01814777>

Submitted on 13 Jun 2018

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

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

**UNIVERSITE
BRETAGNE
LOIRE**

THÈSE / IMT Atlantique

sous le sceau de l'Université Bretagne Loire
pour obtenir le grade de

DOCTEUR D'IMT Atlantique

**Et en cotutelle avec la Faculté des Sciences Mathématiques,
Physiques et Naturelles de Tunis - (Tunisie)**

Spécialité : Signal, Image, Vision

École Doctorale Mathématiques et STIC

Présentée par

Imen Hammami

Préparée dans le département Image & traitement de
l'information

Fusion d'images de télédétection hétérogènes par méthodes crédibilistes

Thèse soutenue le 08 décembre 2017

devant le jury composé de :

Imed Riadh Farah

Professeur, Isamm de la Manouba – Tunisie / président

Arnaud Martin

Professeur, IUT DE Lannion / rapporteur

Jean Dezert

Chercheur scientifique, The French Aerospace Lab Onera - Meudon / rapporteur

Atef Hamouda

Maître assistant, Faculté des sciences de Tunis – Tunisie / examinateur

Boutheina Ben Yaghlane

Professeur, IHEC Carthage - Tunisie / examinateur

Grégoire Mercier

Professeur, eXo maKina – Paris / examinateur

Khaled Bsaies

Professeur, Faculté des sciences de Tunis – Tunisie / co-directeur de thèse

Basel Solaiman

Professeur, IMT Atlantique / directeur de thèse

RAPPORT DE THÈSE
Pour l'obtention du Titre
DOCTEUR EN INFORMATIQUE

Délivré par:
Université Bretagne Loire / IMT Atlantique
Cotutelle internationale avec :
Université de Tunis el Manar / Faculté des Sciences de Tunis (FST)

**Fusion of heterogeneous remote sensing
images by credibilist methods**

Présentée par :
Imen HAMMAMI

Soutenue le **8 décembre 2017** devant le jury composé de :

M. Arnaud Martin - Professeur à l'Université de Rennes 1	Rapporteur
M. I. Riadh FARAH - Professeur à l'ISAMM de la Manouba	Rapporteur
M. Grégoire Mercier - Directeur de la Technologie à eXo maKina	Examinateur
M. Jean Dezert - Maître de Recherches au DTIS, ONERA	Examinateur
Mme. Bouthaina Ben Yaghlane - Professeur à l'IHEC Carthage	Examinatrice
M. Basel Solaiman - Professeur à l'IMT Atlantique	Directeur de thèse
M. Khaled Bsaies - Professeur à la FST	Directeur de thèse

Acknowledgements

This thesis falls within the framework of a cotutelle agreement between the LabS-TICC laboratory of IMT Atlantique, Brest, France and the LIPAH laboratory of the Faculty of Sciences of Tunis, Tunisia. It would not have been possible without persistent help of a large number of peoples to whom I would like to convey my heartfelt gratitude.

First and foremost, I am so grateful to Prof. Arnaud Martin and Prof. Imed Riadh Farah for agreeing to be members of the jury and for dedicating their time and effort to evaluate and examine this work.

Likewise, my special thanks goes to Prof. Boutheina Ben Yaghlaine for being part of my thesis jury.

I would like to offer my sincerest thanks to my supervisor Prof. Basel Solaiman and co-supervisor Dr. Grégoire Mercier for his continuous support, patience, motivation, and most importantly, for its monitoring and availability during this thesis. I am really honoured and pleased to conduct my research work under his co-supervision, and really, I could not have imagined having a better advisor for my thesis. I would like to thank my supervisor Prof. Khaled Bsaies and my co-supervisor Dr. Atef Hamouda from the Faculty of Sciences of Tunis for their valuable advice, guidance and encouragement.

I would like to express my gratitude to Prof. Richard Lepage from École de Technologie Supérieure (ÉTS) of Montréal for welcoming me during one year into his research group in the LIVIA laboratory. His fruitful discussions made enormous contribution to my work. I am very thankful to Dr. Jean Dezert for its insightful and precious comments all along my thesis.

I take this opportunity also to thank all my PhD colleagues at the LIPAH, LabSTICC and LIVIA laboratories for the constructive and pleasant working relationship. My warmest thanks to all my dear friends.

Last but not the least, I would like to express my very profound gratitude to my parents, my sister and my brother for all unconditional love and support. Also, I am so grateful to my husband for always being there for me. I love you all.

Contents

Acknowledgments	i
Abstract	vii
Acronyms	ix
Résumé étendu	1
Introduction	1
Mise en contexte	1
Contributions	2
Rappel sur la théorie des fonctions de croyance	2
Estimation des fonctions de masse pour la classification des images de télédétection	
de grande taille	4
Carte de Kohonen	4
Construction des fonctions de masse	5
Classification jointe des images de télédétection hétérogènes	7
Fusion des fonctions de croyance consonantes basée sur les copules	11
Conclusion et perspectives	13
1 General Introduction	15
1.1 Context and problem statement	15
1.2 Objectives and contributions	16
1.3 Organization	17
2 Credibilist fusion frameworks	21
2.1 Introduction	21
2.2 Basic concepts of Dempster-Shafer Theory (DST)	22
2.2.1 Representation of information	22
2.2.2 Combination of evidence	26
2.2.3 Discounting techniques	30
2.2.4 Distance measures of evidence	31
2.2.5 Decision making	32
2.2.6 Limits of DST	33
2.3 An extension of DST: Towards Dezert-Smarandache Theory (DSmT)	33
2.3.1 Basic foundations	34
2.3.2 DSm models of fusion	37
2.3.3 DSm rules of combination	38

2.4	Conclusion	39
3	Kohonen's map approach for the belief mass modeling	41
3.1	Introduction	41
3.2	Estimation of mass functions in evidence theory	42
3.2.1	Distance-based approaches	42
3.2.2	Likelihood-based approaches	44
3.3	New method to build mass functions	45
3.3.1	Overview on Kohonen's map	45
3.3.2	Feature space for smart basic belief assignment	47
3.3.3	Mass function construction	48
3.4	Simple simulation	51
3.5	Experiments on benchmark data set	54
3.6	Experiments on a real satellite image	56
3.6.1	The classification results in 2^Θ	57
3.6.2	The classification results in D^Θ	61
3.7	Conclusion	63
4	Kohonen-based Credal Fusion of Heterogeneous Data: application to optical and radar joint classification with missing data	65
4.1	Introduction	65
4.2	Generality on data fusion in remote sensing field	66
4.2.1	Definition of data fusion	66
4.2.2	Data fusion levels	67
4.2.3	Fusion of optical and radar data for land cover classification	68
4.3	The proposed method	69
4.3.1	Fusion scenario preview	69
4.3.2	Evaluated features	71
4.3.3	Basic Belief Assignments for Heterogeneous Data	72
4.3.4	Adopted scheme for heterogeneous data fusion	73
4.4	Experimental results	74
4.4.1	Study area and data description	74
4.4.2	Results for joint classification	75
4.4.3	Results for joint classification with missing data	79
4.5	Conclusion	81
5	Copulas-based fusion of consonant belief functions induced by dependent sources of evidences	83
5.1	Introduction	83
5.2	Overview on combination rules of dependent belief functions	84
5.3	Copulas	85
5.3.1	Reminders and notations	85
5.3.2	Copulas as conjunctive aggregation functions	85
5.3.3	K-plot graphical representation	87
5.4	DST and Copulas	88
5.4.1	Random sets and DST	88

5.4.2	Conjunctive combination of dependent consonant belief functions using copula	88
5.5	Credal dependent fusion: disjunctive aggregation rule	90
5.6	The choice of the family of copulas	91
5.7	Experiments	93
5.7.1	Consonant mass functions estimation	94
5.7.2	Results and discussion	94
5.8	Conclusion	100
6	General conclusion	103
	List of Figures	107
	List of Tables	109
	Bibliography	120

Abstract

Abstract: With the advent of new image acquisition techniques and the emergence of high resolution satellite systems, remote sensing data to be exploited have become increasingly rich and varied. Their combination has thus become essential to improve the process of extracting useful information related to the physical nature of the observed surfaces. However, these data are generally heterogeneous and imperfect, which poses several problems in their joint treatment and requires the development of specific methods. It is in this context that falls this thesis that aimed at developing a new evidential fusion method dedicated to heterogeneous remote sensing images processing at high resolution. In order to achieve this objective, we first focus our research, firstly, on the development of a new approach for the belief functions estimation based on Kohonen's map in order to simplify the masses assignment operation of the large volumes of data occupied by these images. The proposed method allows to model not only the ignorance and the imprecision of our sources of information, but also their paradox. After that, we exploit this estimation approach to propose an original fusion technique that will solve problems due to the wide variety of knowledge provided by these heterogeneous sensors. Finally, we study the way in which the dependence between these sources can be considered in the fusion process using the copula theory. For this reason, a new technique for choosing the most appropriate copula is introduced. The experimental part of this work is devoted to land use mapping in case of agricultural areas using SPOT-5 and RADARSAT-2 images. The experimental study carried out demonstrates the robustness and effectiveness of the approaches developed in the framework of this thesis.

Keywords: *Belief function theory, estimation, Kohonen's map, heterogeneous data fusion, optical and radar images, dependencies, copula theory.*

Résumé : Avec l'avènement de nouvelles techniques d'acquisition d'image et l'émergence des systèmes satellitaires à haute résolution, les données de télédétection à exploiter sont devenues de plus en plus riches et variées. Leur combinaison est donc devenue essentielle pour améliorer le processus d'extraction des informations utiles liées à la nature physique des surfaces observées. Cependant, ces données sont généralement hétérogènes et imparfaites ce qui pose plusieurs problèmes au niveau de leur traitement conjoint et nécessite le développement de méthodes spécifiques. C'est dans ce contexte que s'inscrit cette thèse qui vise à élaborer une nouvelle méthode de fusion évidentielle dédiée au traitement des images de télédétection hétérogènes à haute résolution. Afin d'atteindre cet objectif, nous axons notre recherche, en premier lieu, sur le développement d'une nouvelle approche pour l'estimation des fonctions de croyance basée sur la carte de Kohonen pour simplifier l'opération d'affectation des masses des

gros volumes de données occupées par ces images. La méthode proposée permet de modéliser non seulement l'ignorance et l'imprécision de nos sources d'information, mais aussi leur paradoxe. Ensuite, nous exploitons cette approche d'estimation pour proposer une technique de fusion originale qui permettra de remédier aux problèmes dus à la grande variété des connaissances apportées par ces capteurs hétérogènes. Finalement, nous étudions la manière dont la dépendance entre ces sources peut être considérée dans le processus de fusion moyennant la théorie des copules. Pour cette raison, une nouvelle technique pour choisir la copule la plus appropriée est introduite. La partie expérimentale de ce travail est dédiée à la cartographie de l'occupation des sols dans les zones agricoles en utilisant des images SPOT-5 et RADARSAT-2. L'étude expérimentale réalisée démontre la robustesse et l'efficacité des approches développées dans le cadre de cette thèse.

Mots clés : *La théorie des fonctions de croyance, estimation, la carte de Kohonen, fusion des données hétérogènes, images optiques et radars, dépendances, la théorie des copules.*

Acronyms

BBA	Basic Belief Assignment
BPA	Basic Probability Assignment
BR	Bold Rule
CCR	Cautious Conjunctive Rule
CR	Conjunctive Rule
CRC	Conjunctive Rule based on Copula
DC	Distance Classifier
DR	Disjunctive Rule
DRDC	Disjunctive Rule based on Dual Copula
DSmT	Dezert-Smarandache Theory
DST	Dempster-Shafer Theory
ECM	Evidential C-Means
EVCLUS	EVIDential CLUStering
GBBA	Generalized Basic Belief Assignment
GLCM	Gray Level Cooccurrence Matrix
HR	High Resolution
<i>K</i>-NN	<i>K</i> -Nearest Neighbor
PCR	Proportional Conflict Redistribution
RMSE	Root Mean Squared Error
SAR	Synthetic Aperture Radar
SOM	Self Organizing Map

Résumé étendu

Introduction	1
Mise en contexte	1
Contributions	2
Rappel sur la théorie des fonctions de croyance	2
Estimation des fonctions de masse pour la classification des images de télédétection de grande taille	4
Carte de Kohonen	4
Construction des fonctions de masse	5
Classification jointe des images de télédétection hétérogènes	7
Fusion des fonctions de croyance consonantes basée sur les copules	11
Conclusion et perspectives	13

Introduction

Mise en contexte

La grande variété des capteurs (optiques, radar et Lidar) installés sur les satellites ainsi que l'amélioration rapide de leurs caractéristiques spatiales et spectrales ont permis l'acquisition d'une multitude d'images présentant des données de résolution métriques et submétriques extrêmement riches et précises permettant d'atteindre un niveau de détails jamais atteint auparavant. Avec l'avènement de telles images, le contenu d'information à exploiter s'est réellement densifié au cours des dix dernières années. L'extraction des informations plus utiles et complètes liées à la nature physique des surfaces observées est devenue donc de plus en plus convoitée par les diverses applications de la télédétection. Néanmoins, le traitement joint de ces données pose des problèmes particuliers et nécessite par conséquent des méthodes spécifiques. Cela est principalement dû, d'une part, à leur hétérogénéité et, d'autre part, à leur nature imprécise, incomplète, voire erronée.

Dans ce travail de thèse, nous nous sommes intéressés à l'élaboration d'une nouvelle approche crédibiliste pour la fusion des images de télédétection hétérogènes à haute résolution

(HR). Son application a été consacrée à la cartographie de l'utilisation du sol en utilisant des images optiques et radar.

Contributions

Une évaluation de la contribution potentielle des théories de l'évidence à la modélisation et à la fusion de données de télédétection hétérogènes pour concevoir une classification jointe a été faite dans ce travail de recherche. D'un point de vue méthodologique, nous avons étudié la possibilité de mettre en place de nouvelles techniques intervenantes dans les différentes phases de la réalisation du processus de fusion, telle que la modélisation, l'estimation et la combinaison des croyances. Cela nous a conduits à proposer trois contributions innovatrices, qui seront résumées dans les points suivants :

- Les approches classiques et génériques pour la construction des fonctions de masse présentent généralement une complexité de calcul élevée ce qui constitue un obstacle majeur à leur application dans le cas d'images à HR contenant un volume important de données à traiter. Nous définissons alors une nouvelle méthode d'estimation des fonctions de masse à partir des cartes de Kohonen pour rendre cette tâche extrêmement rapide. Cette proposition a aussi l'avantage de définir des valeurs de fonctions de masse pour les différentes formes d'éléments focaux (singletons, ainsi que leurs unions et intersections).
- Notre deuxième contribution porte sur l'adaptation de cette méthode en vue d'une fusion des données hétérogènes radar et optiques acquises à partir d'un paysage agricole. Le cadre crédibiliste introduit est capable de traiter des données optiques complètes et partielles (c.-à-d. manquantes en raison de la présence de nuages).
- Finalement, nous nous sommes intéressés principalement à la façon dont la dépendance entre les sources (observations hétérogènes) peut être prise en compte dans le processus de fusion. Pour ce faire, nous définissons une combinaison des croyances distinctes basée sur la théorie des copules.

Rappel sur la théorie des fonctions de croyance

La théorie de Dempster-Shafer [1, 2] (connue aussi par la théorie des fonctions de croyance (TFC)) est un cadre mathématique robuste permettant le traitement des connaissances imprécises et incertaines à la fois. Ce formalisme repose principalement sur la représentation de la croyance d'une source d'information à travers une fonction de masse m , définie sur l'ensemble de tous les sous-ensembles du cadre de discernement Θ , noté 2^Θ , et à valeurs dans $[0, 1]$. Formellement, m est donnée par :

$$\sum_{A \subseteq \Theta} m(A) = 1. \quad (1)$$

Le cadre de discernement Θ est l'ensemble des réponses possibles au problème de fusion à traiter. Il est composé d'hypothèses exhaustives et exclusives : $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\} = \bigcup_{i=1}^N \{\theta_i\}$. À partir de ce cadre de discernement, on peut construire le power set 2^Θ en incluant toutes les disjonctions d'hypothèses θ_i tel que $\theta_i \cup \theta_j$ ou $\theta_i \cup \theta_j \cup \theta_k \dots$

Il existe d'autres fonctions pour coder la même information contenue dans une fonction de masse m . La fonction de crédibilité $bel(A)$ (appelée aussi fonction de croyance) représente la croyance totale en A . Elle est définie par :

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B), \quad \forall A \subseteq \Theta. \quad (2)$$

La fonction de plausibilité $pl(A)$ quantifie le degré maximal de croyance qui pourrait potentiellement être donné à A . Elle est définie par :

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Theta. \quad (3)$$

L'utilisation du critère du maximum de croyance ou de plausibilité pour la prise de décision correspond aux stratégies les plus simples lorsque nous privilégions une recherche pessimiste ou optimiste d'une solution, respectivement. Le maximum de la probabilité pignistique proposé par Smets est considéré comme une alternative plus prudente. La probabilité pignistique est établie pour tout $A \in 2^\Theta$, avec $A \neq \emptyset$ comme suit :

$$BetP(A) = \sum_{B \in 2^\Theta, B \neq \emptyset} \frac{|B \cap A|}{|B|} m(B), \quad \forall A \subseteq \Theta. \quad (4)$$

Dans le cadre de TFC, plusieurs règles de combinaison ont été introduites pour l'agrégation des croyances dans un contexte multi-sources. Historiquement, la règle de Dempster est la plus ancienne. Soit deux sources d'information S_1 et S_2 émettant des avis représentés respectivement par les fonctions de masse distinctes m_1 et m_2 . Le résultat de leur combinaison par cette règle, notée $m_{1,2}^{DS}$, est donné par la formule suivante :

$$m_{1,2}^{DS}(A) = \frac{1}{1 - \mathbb{K}} \sum_{B \cap C = A} m_1(B) m_2(C), \quad (5)$$

où $\mathbb{K} = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ mesure le degré de conflit entre les fonctions de masse m_1 et m_2 . Cette règle est conçue pour satisfaire l'hypothèse du monde fermé ($m(\emptyset) = 0$). Afin de considérer les problèmes dans l'hypothèse du monde ouvert, la règle conjonctive qui permet la fusion de sources d'information fiables sans aucune normalisation ($\mathbb{K} \neq 0$) peut être utilisée :

$$m_1 \odot_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C). \quad (6)$$

Si au moins l'une des sources combinées est fiable, Dubois et Prade [3] proposent l'utilisation de la règle disjonctive définie comme suit :

$$m_1 \oplus_2(A) = \sum_{B \cup C = A} m_1(B) m_2(C). \quad (7)$$

Dans [4, 5], Dezert et Smarandache proposent une généralisation de la théorie initiale de Dempster-Shafer désignée par le terme anglais Dezert-Smarandache Theory (DSmT) . Dans

leur approche, un raisonnement paradoxal a été introduit en annulant les contraintes d'exclusivité imposées aux hypothèses et la redistribution de la masse conflictuelle aux ensembles non vides en utilisant la règle de normalisation. L'idée principale de DSmt est de travailler sur l'hyper-power set D^Θ du cadre de discernement au lieu de 2^Θ . Cet ensemble est défini comme le treillis de Dedekind construit à partir du Θ avec les opérateurs \cap et \cup . L'attribution des croyances dans DSmt se fait moyennant la fonction de masse généralisée m définie sur D^Θ et à valeurs dans $[0, 1]$:

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in D^\Theta} m(A) = 1. \quad (8)$$

Comme dans TFC, différentes règles de combinaison ont été proposées dans DSmt. Les lecteurs intéressés pourraient se référer à [4] pour plus de détails sur certaines de ces règles. Pour la prise de décision, la probabilité pignistique généralisée peut être utilisée :

$$GPT(A) = \sum_{E \in D^\Theta} \frac{C_M(E \cap A)}{C_M(E)} m(E), \quad \forall A \in D^\Theta, \quad (9)$$

où $C_M(E)$ est la cardinalité de E , définie par Dezert et Smarandache comme le nombre de parties disjointes du diagramme de Venn incluses dans E . La décision est alors prise par le maximum de GPT .

Estimation des fonctions de masse pour la classification des images de télédétection de grande taille

Dans cette section, nous présentons notre nouvelle méthode d'estimation des fonctions de masse.

Carte de Kohonen

Il existe plusieurs versions de la carte auto-organisatrice de Kohonen (Connu aussi par le terme anglais self organizing map (SOM)). Cependant, la philosophie de base est très simple et efficace [6]. L'algorithme SOM permet d'effectuer une projection non linéaire de données de grandes dimensions (définies dans \mathbb{R}^p par exemple) dans un tableau à deux dimensions de $M \times N$ nœuds (voir Fig. 1) [7].

Un vecteur de référence, également appelé vecteur de pondération, $w(i, j) \in \mathbb{R}^p$ est associé au nœud de position (i, j) avec $1 \leq i \leq M$ et $1 \leq j \leq N$. Un vecteur d'entrée $x \in \mathbb{R}^p$ est comparé à chaque $w(i, j)$. La meilleure correspondance est définie comme sortie du SOM : ainsi, les données d'entrée x sont mappées sur le SOM à l'emplacement (i_x, j_x) où $w(i_x, j_x)$ est le neurone le plus proche de x selon une métrique donnée. En pratique, la distance euclidienne est généralement utilisée pour comparer x et $w(i, j)$. Le nœud qui minimise la distance entre x et $w(i, j)$ définit le nœud correspondant le mieux (ou le neurone gagnant), et est désigné par l'indice w_x :

$$\|x - w_x\| = \min_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \|x - w(i, j)\|. \quad (10)$$

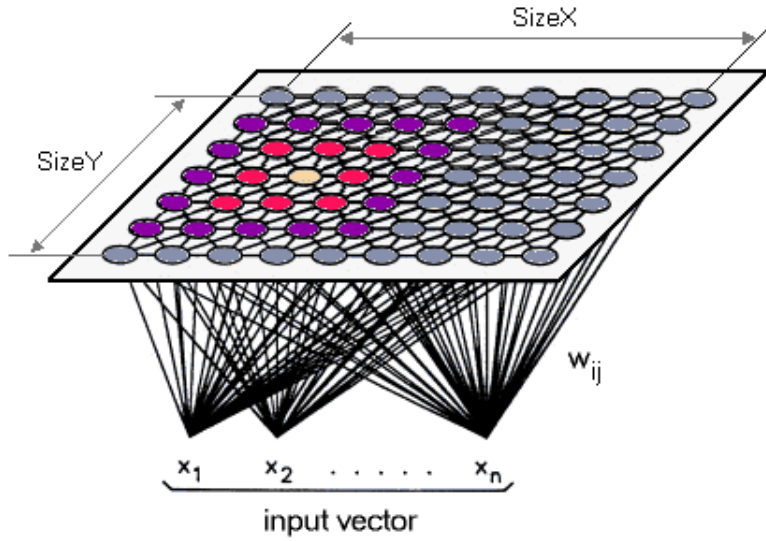


FIGURE 1 – Un schéma représentatif de la carte auto-organisatrice de Kohonen.

On peut également dire que l'SOM réalise une quantification non uniforme qui transforme \mathbf{x} en \mathbf{w}_x en minimisant la métrique donnée. Néanmoins, grâce à la phase d'entraînement les neurones \mathbf{w} sont situés sur la carte en fonction de leur similarité. Alors, en considérant les neurones $\mathbf{w}(i, j)$ situés pas *trop loin* du neurone gagnant \mathbf{w}_x , la distance dans \mathbb{R}^p entre \mathbf{x} et $\mathbf{w}(i, j)$ ne sont pas nettement différents de celle entre \mathbf{x} et \mathbf{w}_x . Cela signifie que dans le voisinage de \mathbf{w}_x sur la carte, se trouvent les neurones gagnants des voisins (dans \mathbb{R}^p) de \mathbf{x} . Par conséquent, une classe dans \mathbb{R}^p est projetée dans la carte au même endroit, restant homogène. De plus, quelle que soit la forme initiale de la classe dans l'espace caractéristique \mathbb{R}^p , la classe projetée est fortement susceptible d'être de forme isotrope sur la carte.

Construction des fonctions de masse

L'affectation intelligente des masses proposée nécessite l'entraînement d'une carte de Kohonen à partir des observations $\mathbf{x} \in \mathbb{R}^p$ à classer et une classification initiale pour définir leurs centres de classe. Donc, deux types de connaissances sont manipulés (voir Fig. 2) pour la construction des croyances : d'une part les observations initiales \mathbf{x} et les centres de classe $\{C_1, C_2, \dots, C_K\}$ dans \mathbb{R}^p et, d'autre part les neurones gagnants \mathbf{w}_x et les centres de classe projetés $\{\mathbf{w}_{C_1}, \dots, \mathbf{w}_{C_K}\}$.

- La masse de chaque hypothèse simple est définie directement sur la carte par :

$$\begin{cases} m(\mathbf{x} \in \theta_k) \simeq 1 & \text{si } \mathbf{w}_x = \mathbf{w}_{C_k} \\ m(\mathbf{x} \in \theta_k) \simeq \frac{d_{\text{map}}(\mathbf{w}_x, \mathbf{w}_{C_k})^{-1}}{\sum_{\ell=1}^K d_{\text{map}}(\mathbf{w}_x, \mathbf{w}_{C_\ell})^{-1}} & \text{sinon,} \end{cases} \quad (11)$$

où $k = 1, 2, \dots, K$ et $d_{\text{map}}(\cdot, \cdot)$ représente la distance utilisée sur la carte de Kohonen. Elle est principalement basée sur la norme euclidienne et elle utilise l'index qui localise les deux vecteurs sur la carte :

$$d_{\text{map}}(\mathbf{w}_1, \mathbf{w}_2) = \sqrt{(i_{w_1} - i_{w_2})^2 + (j_{w_1} - j_{w_2})^2}$$

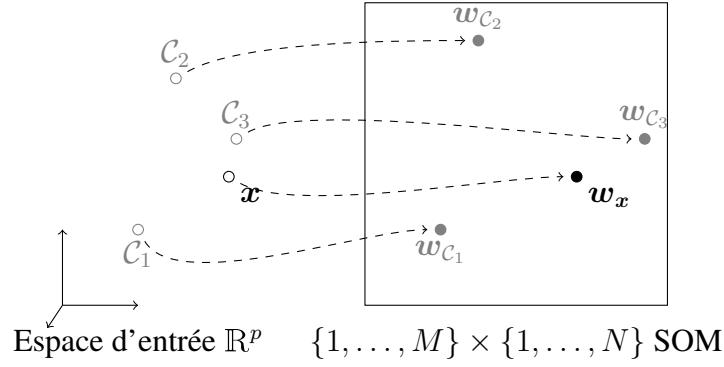


FIGURE 2 – Observations dans l’espace d’entrée et leurs projections dans la carte de Kohonen. Notez que les neurones w_x et w_{C_k} peuvent être localisés sur la carte grâce à leur indice de localisation (m, n) ou dans \mathbb{R}^p avec leur valeur p -composants.

si w_1 (resp. w_2) est situé à la position (i_{w_1}, j_{w_1}) (resp. (i_{w_2}, j_{w_2})) dans la carte.

- Comme la définition des disjonctions d’hypothèses exprime l’absence de discrimination, leur masse est donc définie directement dans l’espace d’entrée. Ensuite, elle est liée à un effet d’échelle entre l’échantillon x à considérer et les deux classes apparentées θ_k et θ_ℓ .

$$m(x \in \theta_k \cup \theta_\ell) \simeq 1 - \tanh(\beta z) \quad (12)$$

avec

$$z = \frac{d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)}{d_{\mathbb{R}^p}(x, \mathcal{C}_k) + d_{\mathbb{R}^p}(x, \mathcal{C}_\ell)} \quad 0 < k, \ell \leq K, k \neq \ell.$$

où β est un paramètre qui représente le niveau d’ambiguïté et $d_{\mathbb{R}^p}(\cdot, \cdot)$ est la distance dans \mathbb{R}^p . Elle peut être définie par la norme euclidienne $\mathcal{L}^2(\mathbb{R}^p)$, mais aussi par une perspective spectrale, telle que la cartographie par l’angle spectral (Spectral Angle Mapper (SAM)) ou la divergence d’information spectrale. Elle peut également être basée sur la divergence de Kullback-Leibler ou l’information mutuelle, en traitant des données radar [8].

L’équation (12) peut être expliquée de cette manière : si un exemple x est très proche de son centre de classe associé \mathcal{C}_k en comparaison avec tout autre centre de classe \mathcal{C}_ℓ , alors il n’y a pas d’ambiguïté dans x appartenant à la classe θ_k . Si ce n’est pas le cas (c-à-d. si les distances entre x et les centres de classes \mathcal{C}_k et \mathcal{C}_ℓ sont de la même échelle), alors il est difficile de discriminer x de la classe θ_k ou θ_ℓ .

- La masse de l’ignorance totale est basée sur la distance d’un échantillon x à la carte. Nous considérons que l’évaluation de la masse d’une observation tombe dans l’ignorance si sa distance à la carte est beaucoup plus importante que la distance de son centre de classe liée à la carte. Donc, elle peut être donnée par :

$$m(x \in \Theta) \simeq 1 - \min \left(\frac{d_{\mathbb{R}^p}(x, w_x)}{d_{\mathbb{R}^p}(\mathcal{C}_x, w_{\mathcal{C}_x})}, \frac{d_{\mathbb{R}^p}(\mathcal{C}_x, w_{\mathcal{C}_x})}{d_{\mathbb{R}^p}(x, w_x)} \right) \quad (13)$$

où \mathcal{C}_x est le centre de la classe de x et $w_{\mathcal{C}_x}$ est sa projection sur la carte.

- La fonction de masse finale doit respecter la contrainte de l'équation (1), donc une étape de normalisation doit être appliquée.

La performance de notre approche de construction des masses est comparée à celle de EVCLUS [9] et de ECM [10] en utilisant l'ensemble de données d'UCI (University California Irvine). Sept jeux de données (voir tableau 1) ont été pris en compte dans ce test. Comme le montre le tableau 2, l'approche basée sur SOM donne des résultats similaires aux deux autres algorithmes. On peut aussi noter que plus le nombre d'échantillons est élevé, plus elle est rapide. Donc, elle semble être une alternative efficace pour gérer les grands volumes de données pour des fins de classification. En fait, la distance dans \mathbb{R}^p est plus exigeante que dans \mathbb{R}^2 . De plus, la forme de la classe dans la SOM est plus isotrope, de sorte qu'aucune considération sur la forme de la variété n'est à considérer. Au contraire, ECM doit se soucier de l'écart-type des classes pour construire la distribution de masse.

TABLE 1 – Caractéristiques des bases de données UCI utilisés pour la comparaison.

Base de données	Caractéristiques	Classes	Échantillons
Banknote authentication	4	2	1372
Pima Indians Diabetes	8	2	768
Seeds	7	3	210
Wine	13	3	170
Statlog (Landsat Satellite)	36	6	6435
Statlog (Image Segmentation)	19	7	2130
Synthetic control chart time series	60	6	600

TABLE 2 – Résultats de la classification avec une estimation des masses par EVCLUS, ECM et l'approche proposée.

Base de données	Banknote authentication	Pima Indians Diabetes	Seeds	Wine	Statlog (Landsat Satellite)	Statlog (Image Segmentation)	Synthetic control chart time series
EVCLUS	61,44 % 1172,2 sec	61,84 % 181,7 sec	74,76 % 34,3 sec	60,58 % 6,7 sec	47,03 % 5857 sec	42,01 % 3657 sec	64,0 % 370 sec
ECM	61,80 % 3,4 sec	65,88 % 3,2 sec	90,0 % 0,3 sec	74,11 % 0,9 sec	69,62 % 480 sec	55,49 % 161 sec	72,5 % 6,9 sec
Notre approche	79,44 % 8,6 sec	71,48 % 6,7 sec	90,95 % 5,8 sec	73,52 % 5,9 sec	69,24 % 163 sec	67,18 % 84 sec	83,5 % 8,0 sec

Des comparaisons avec des méthodes proposées dans la littérature, détaillées dans la partie écrite en anglais, montrent que notre approche permet de construire les fonctions de masse des images de télédétection 150 fois plus rapidement avec des résultats équivalents.

Classification jointe des images de télédétection hétérogènes

De nos jours, les données satellitaires sont de plus en plus accessibles, ce qui nécessite l'élaboration de nouvelles méthodes de traitement intelligent permettant d'extraire des connaissances de haut niveau issues de ces diverses sources d'information. Dans ce contexte, la fusion a maintes fois montré son intérêt dans la résolution de plusieurs problèmes du monde réel en permettant

de profiter au mieux des avantages de chaque source d'information et de surmonter les limitations individuelles de chacune d'elles. Malgré ces avantages, la fusion a été toujours considérée comme une tâche très difficile pour plusieurs raisons y compris, mais non limitée à la complexité du processus de combinaison et l'hétérogénéité des données à agréger. Ce travail introduit une nouvelle approche crédibiliste pour la fusion des données dérivées de capteurs hétérogènes optiques et radar qui est considérée comme l'un des problèmes les plus complexes dans le domaine de la télédétection. Nous nous intéressons particulièrement à la classification jointe des images acquises par les satellites SPOT-5 et RADARSAT-2 dans une zone agricole.

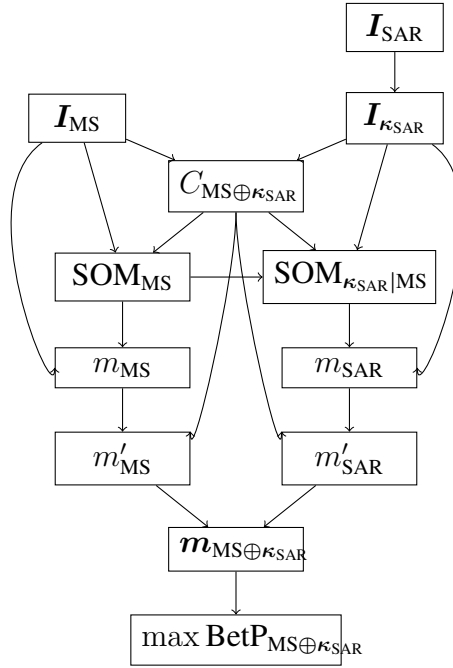


FIGURE 3 – Le schéma général de l'approche proposée.

L'approche proposée décrite dans la figure 3 est principalement constituée des étapes suivantes : premièrement, les descripteurs les plus représentatifs sont extraits à partir de chaque type de donnée d'entrée afin de modéliser les sources d'information utilisées. Puisque nous ne pouvons pas faire confiance à l'information provenant des pixels radar isolés de en raison de la présence du bruit de chatoiement caractérisant ce type d'images, nous optons pour l'utilisation de descripteurs locaux de texture. Notre vecteur de caractéristiques est composé des quatre premiers cumulants $(\mu, \sigma, \beta_1, \beta_2)$, associés respectivement à (la moyenne, l'écart type, l'asymétrie et l'aplatissement) et estimés à partir de l'image radar I_{SAR} moyennant d'une fenêtre d'analyse. De plus, le moment inverse f_5 et la somme moyenne f_6 extraits en utilisant les mesures de texture de Haralick ont été aussi utilisés pour analyser la relation spatiale entre les pixels dans le même voisinage spatial. La combinaison de ces descripteurs génère une image $I_{\kappa_{SAR}}$ à 6 bandes, fournissant l'information locale notée $\kappa = (\mu, \sigma, \beta_1, \beta_2, f_6, f_5)$ qui sera considérée comme l'observation extraite à partir de la source d'information radar. Les p bandes de l'image multispectrale représente l'observation optique, notée I_{MS} .

Étant donné que le processus de fusion est confronté à différents types de caractéristiques, la classification jointe de données hétérogènes doit garantir que les classes sont définies de manière homogène à partir des observations optiques et radar. Ainsi, une première classification jointe

grossière $C_{MS \oplus \kappa_{SAR}}$ est effectuée pour lier les signatures spectrales de I_{MS} et les descripteurs de texture radar de $I_{\kappa_{SAR}}$. Dans cette étude, un simple classificateur K-moyennes est utilisé avec un facteur pour ajuster la dynamique relative entre les deux observations.

Ensuite, le processus de fusion est appliqué à chaque pixel par la théorie de Dempster-Shafer, qui nécessite l'estimation des fonctions de masse m_{MS} et m_{SAR} des sources d'information considérées I_{MS} et $I_{\kappa_{SAR}}$, respectivement. Pour ce faire, l'approche basée sur Kohonen (détaillé ci-dessus) est appliquée, car il a montré sa capacité à gérer les grandes données de télédétection [11]. Les fonctions de masse provenant de l'information multispectrale sont estimées en se basant uniquement sur l'information obtenue à partir du capteur optique (considéré comme une source d'information fiable et complète) pour former SOM_{MS} , tandis que celles associées à la source radar sont estimées tout en tenant en considération l'information optique. En effet, un entraînement hybride asservi de la carte de kohonen a été proposé pour construire $SOM_{\kappa_{SAR}|MS}$. Soit $\mathbf{x} = \{x_1, x_2, \dots, x_p\} \in \mathbb{R}^p$ et $\mathbf{y} = \{y_1, y_2, \dots, y_q\} \in \mathbb{R}^q$ les deux observations hétérogènes fournies par les deux capteurs I_{MS} et $I_{\kappa_{SAR}}$. Les échantillons d'entrée de la carte auto-organisatrice hybride proposée sont effectués à partir des observations co-localisées $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ avec lesquelles une distance doit être associée. Cette distance est une fusion des 2 métriques à appliquer sur chaque type de données initiales :

$$d(\mathbf{z}, \mathbf{z}') = d_{\mathbb{R}^p}(\mathbf{x}, \mathbf{x}') + \alpha d_{\mathbb{R}^q}(\mathbf{y}, \mathbf{y}'), \quad (14)$$

avec $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ et $\mathbf{z}' = (\mathbf{x}', \mathbf{y}')$ étant 2 échantillons dans \mathbb{R}^{p+q} . Le paramètre α est un facteur d'étalonnage croisé, qui tient compte de la dynamique relative entre \mathbf{x} et \mathbf{y} .

Selon cette définition d'un espace de caractéristiques hybride et ses métriques connexes, il est possible d'effectuer un entraînement hybride asservi où les vecteurs de pondération sont définis avec $\mathbf{w}_z = (\mathbf{w}_x, \mathbf{w}_y) \in \mathbb{R}^{p+q}$. Cet entraînement commence par une formation SOM classique des données optiques uniquement, et donne SOM_{MS} . Ensuite, les neurones de SOM_{MS} sont concaténés par des composants q pour s'adapter au \mathbb{R}^{p+q} du traitement joint. L'entraînement de cette carte hybride commence, mais seuls les derniers q -composants (dédiés aux données radar) sont modifiés. Dans ce cas, la partie optique est conservée, tandis que la partie radar suit la partie optique à l'emplacement des classes sur la carte (emplacements des neurones gagnants \mathbf{w}_{C_k}).

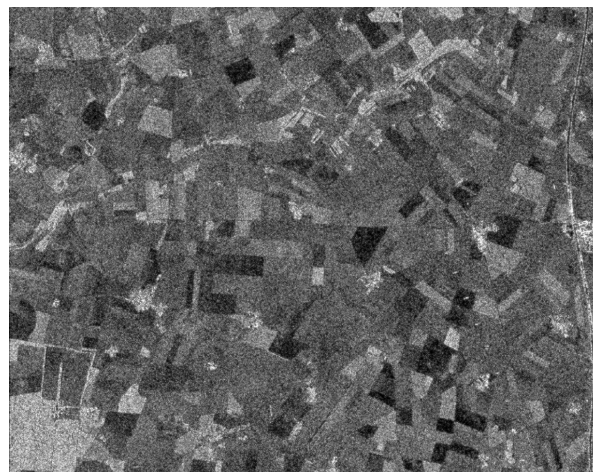
Finalement, pour gérer l'incertitude causée par l'hétérogénéité des données utilisées, certains opérateurs d'affaiblissement sont appliqués avant l'étape de fusion pour donner m'_{MS} et m'_{SAR} . La classification finale de la couverture terrestre est obtenue à partir de la fonction de masse combinée $\mathbf{m}_{MS \oplus \kappa_{SAR}}$ en appliquant le maximum de la probabilité Pignistic $BetP_{MS \oplus \kappa_{SAR}}$.

L'approche proposée a été expérimentée sur une zone d'étude qui couvre une partie de la région de la Beauce, située au sud-ouest de Paris, en France. Cette région est connue pour sa productivité agricole élevée. Elle est aussi essentiellement caractérisée par ses très grands champs dominés par le colza et la céréale (blé, orge, maïs). Une image multispectrale acquise par le satellite français SPOT-5 lors de l'expérience Take-5 et une image radar acquise par le satellite canadien RADARSAT-2 en mode ultra-fin ont été utilisées (les figures 4-(a) et 4-(b)). Les deux images couvrent une superficie d'environ $11.5 \times 9 \text{ km}^2$ et ont les caractéristiques suivantes : l'image SPOT-5 est de taille 1145×903 pixels, une résolution spatiale de 10m, et a quatre bandes (vert (G), rouge (R), Proche Infrarouge (NIR) et moyen Infrarouge (MIR)). L'image RADARSAT-2 est composée de 3850×3010 pixels, avec chaque pixel ayant une résolution spatiale de 3m. En ce qui concerne l'image radar, seuls les canaux qui correspondent

aux polarisations HH et HV sont disponibles. Cependant, seule la polarisation HH a été utilisée dans ce travail par ce qu'elle est mieux adaptée pour la caractérisation des régularités dans la texture des régions agricoles que la polarisation HV.



(a) Composition colorée de l'image SPOT5 acquise le 20 avril 2015. ©CNES



(b) Image RADARSAT-2 acquise le 23 avril 2015. RADARSAT-2 de données et produits ©MacDONALD, DETTWILER et ASSOCIATES LTD – Tous droits réservés

FIGURE 4 – L'image multispectrale (a) et l'image radar (b) acquises sur la région Beauce en France.

Les résultats de la figure 5, dont l'analyse est détaillée dans la partie écrite en anglais, montrent tout l'intérêt de notre approche.

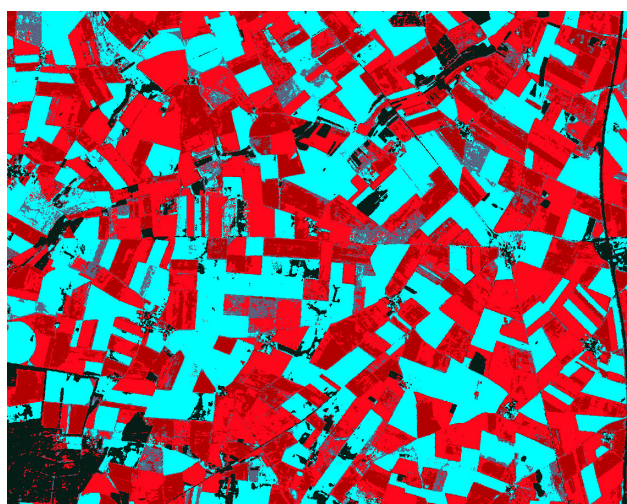


FIGURE 5 – Résultats de la classification jointe calculés par l'application du maximum de la probabilité pignistique sur toutes les classes simples.

Fusion des fonctions de croyance consonantes basée sur les copules

La fusion des données de télédétection optiques et radar est une tâche importante et difficile pour de nombreuses applications telles que la classification multisources, non seulement en raison de la nature très hétérogène des informations qu'ils contiennent, mais aussi des dépendances (corrélations ou information mutuelle) existantes entre les observations. Dans cette partie, nous nous sommes intéressés particulièrement aux problèmes liés à la fusion des sources d'information dépendantes dans le cadre de la théorie de l'évidence en utilisant les copules [12] connus par leur capacité de capturer et de modéliser les structures de dépendance des distributions jointes. L'idée de base était de mettre en évidence la relation entre les ensembles aléatoires et les croyances afin d'étudier la TFC dans le cadre de la théorie des probabilités, mais avec des variables aléatoires ayant des ensembles comme valeurs.

Suite à cette étude, deux opérateurs de combinaison ont été introduits pour réaliser la fusion conjonctive et disjonctive des croyances dépendantes codées par les fonctions de masse consonantes¹.

Combinaison conjonctive basée sur les copules

Soit m_1 et m_2 deux fonctions de masse consonantes et normalisées² définies respectivement dans les cadres de discernement Θ_1 et Θ_2 , leur combinaison suivant la règle conjonctive basée sur la copule (CRC) s'écrit de la manière suivante :

$$m_{1,2}^{CRC}(A) = \sum_{A_1 \cap A_2 = A} m_C(A_1, A_2), \quad \forall A_1 \subseteq \Theta_1, A_2 \subseteq \Theta_2, \quad (15)$$

où $m_C(A_1, A_2) = \sum_{B_1 \subseteq A_1, B_2 \subseteq A_2} (-1)^{|A_1 \setminus B_1| + |A_2 \setminus B_2|} C(\text{bel}_1(A_1), \text{bel}_2(A_2))$ est la masse jointe calculée avec la fonction copule C qui résume le mieux la structure de dépendance existante entre les croyances marginales.

Combinaison disjonctive basée sur les copules

Soit m_1 et m_2 deux fonctions de masse consonantes et normalisées définies respectivement dans les cadres de discernement Θ_1 et Θ_2 , leur combinaison suivant la règle disjonctive basée sur la copule (DRDC) s'écrit de la manière suivante :

$$m_{1,2}^{DRDC}(A) = \sum_{A_1 \cup A_2 = A} m_D(A_1, A_2), \quad \forall A_1 \subseteq \Theta_1, A_2 \subseteq \Theta_2 \quad (16)$$

où $m_D(A_1, A_2) = \sum_{B_1 \subseteq A_1, B_2 \subseteq A_2} (-1)^{|A_1 \setminus B_1| + |A_2 \setminus B_2|} \text{bel}_1(A_1) + \text{bel}_2(A_2) - C(\text{bel}_1(A_1), \text{bel}_2(A_2))$ est la masse jointe calculée avec la fonction copule C qui résume le mieux la structure de dépendance existante entre les croyances marginales.

¹Une fonction de masse m est dite consonante si ces éléments focaux ($A_1 \subseteq \Theta$ ayant une croyance non nulle) sont emboîtés.

²Une fonction de masse m est dite normalisée si $m(\emptyset) = 0$.

Il reste donc maintenant à faire le bon choix de la copule C utilisée dans ces règles de fusion. Habituellement, le choix de la copule dépend de données agrégées. En effet, une copule particulière peut convenir mieux à un ensemble de données qu'à un autre. Au meilleur de nos connaissances, il n'existe pas dans la littérature une méthode efficace pour sélectionner la copule. Généralement, l'utilisation de la copule paramétrique est recommandée, car elle peut être adaptée aux données existantes en estimant correctement ses paramètres. Néanmoins, rien ne peut prouver que ce choix de paramètres garantit la convergence de la copule à la structure réelle de la dépendance sous-jacente des données. Dans ce travail, nous avons choisi d'utiliser la famille des copules archimédiennes qui sont capables de caractériser différentes gammes de dépendances. Le choix de la copule archimédienne la plus adéquate aux observations fusionnées a été fait avec l'interprétation du graphique Kendall plot [13].

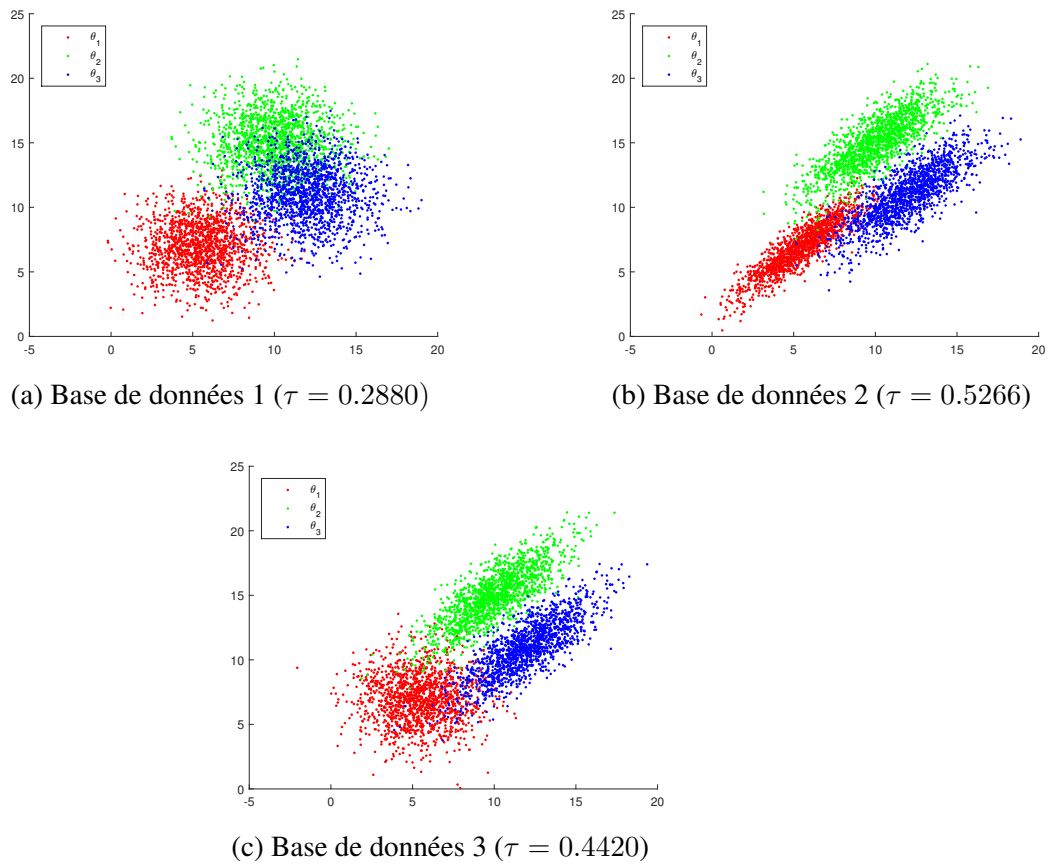


FIGURE 6 – Base des données générées.

Les règles de fusion proposées ont été expérimentées sur trois ensembles de données générées (figure 6) avec différents vecteurs de moyennes et matrices de covariance pour faire varier le degré de dépendance entre les données de test. Comme montré par les résultats du tableau 3 qui sont commentés avec plus de détaille dans la partie rédigé en anglais, les opérateurs de fusion introduits présentent des résultats très prometteurs lorsque l'hypothèse d'indépendance n'est pas vérifiée comparés avec des approches classiques.

TABLE 3 – Résultats de la classification des données simulées.

Règle de combinaison		Valeur de dépendance entre les sources (Calculé avec le taux de Kendall [14])		
		$\tau = 0,2880$	$\tau = 0,4423$	$\tau = 0,5309$
CRC	eq. (15)	88,38%	93,58%	92,56%
La règle prudente	[15]	88,33%	93,36%	91,51%
La règle conjonctive	eq. (6)	88,33%	93,36%	91,51%
DRDC	eq. (16)	85,64%	89,31%	87,47%
La règle hardie	[15]	74,16%	77,31%	76,58%
La règle disjonctive	eq. (7)	74,56%	78,33%	75,04%

Conclusion et perspectives

Dans le cadre de cette thèse, nous avons développé, en premier lieu, une nouvelle méthode de construction des fonctions de masse. L'approche proposée a la particularité de traiter le grand volume de données caractérisant les images de télédétection acquise à haute résolution, ainsi que les données acquises en utilisant d'autres types de capteurs. Une série de comparaisons avec des approches classiques d'estimation des croyances montre que l'utilisation de la carte de Kohonen dans ce type de problème produit des résultats similaires dans un temps plus raisonnable. Ensuite, nous avons proposé une nouvelle méthode crédibiliste de fusion des données hétérogènes avec comme application principale la classification jointe des données optiques et radar. C'est une partie qui fait suite à notre technique d'estimation des fonctions de masse à partir d'observations réelles. L'application de la technique proposée sur un couple d'images SPOT5 et RADARSAT2 acquises à la même date et sur une zone de test qui se trouve dans une région à prédominance agricole montre des résultats très prometteurs en termes de précision de classification et de reconstruction des données optiques manquantes (couverture nuageuse).

De plus, nous avons également introduit deux opérateurs de fusion qui prennent en considération les dépendances (corrélation ou information mutuelle) existant entre les connaissances à combiner. Les résultats dégagés lors de la classification des données synthétiques sont très encourageants.

Comme perspective, l'amélioration des méthodes proposées est envisageable à plusieurs niveaux. Par exemple, l'introduction de conjonctions entre classes au sein de DS_mT lui confère une richesse et une flexibilité particulières pour modéliser les imperfections et le paradoxe des données. Ainsi, il sera intéressant d'adapter notre approche de fusion de données hétérogènes au cadre du DS_mT afin de bénéficier de la sémantique des conjonctions de classe dans la classification jointe des sources très hétérogènes présentant un fort conflit.

La combinaison de fonctions de masse consonantes issues de sources dépendantes porte uniquement sur la fusion de deux sources d'information. L'extension à la fusion de 3 sources ou plus apporte de nouveaux problèmes à résoudre. Aussi, il serait intéressant de l'expérimenter dans le contexte de la classification jointe des données réelles de télédétection.

General Introduction

1.1	Context and problem statement	15
1.2	Objectives and contributions	16
1.3	Organization	17

1.1 Context and problem statement

Since the discovery that the Earth is round some 5 centuries ago, man has not stopped dreaming of photographing the Earth from space. But it was only in 1957 that he realized his dream with the launch of the first artificial satellite that allowed him to acquire the first images of the Earth. Since then, no fewer than 10,000 satellites have been launched into orbit in order to meet scientific, military and economic needs... Earth observation is one of its important applications and is considered as one of the most active fields of research which finds its interest in several applications such as major disasters management, urban areas extent and tropical forests deforestation monitoring, to name a few. The wide variety of sensors installed on these satellites (optical, radar and lidar) and the rapid improvement of their spatial and spectral characteristics have resulted in extremely rich and accurate data with metric and sub-metric resolution, a level of detail never reached before. However, due to the enormous amount of satellite data acquired at this high resolution, increasingly redundant and complementary data are becoming available which complicates their interpretation and extraction of useful information.

Despite the wide variety of existing sensors today, they can be grouped into two main families: 1 - passive sensors capable of recording natural energy as the solar radiation reflected from the earth's surface (available only when the sun illuminates the Earth) 2 - active sensors which, unlike passive sensors, have their own source of illumination and have the advantage of penetrating the clouds and therefore acquiring images in all weather conditions during the day or the night. These different modes of observation are not sensitive to the same information and therefore provide complementary and completely heterogeneous knowledge. Indeed, radar sensors provide information on the roughness and moisture content of the soil which are important parameters often imperceptible by optical sensors. On the other hand, optical sensors, incapable

of penetrating through clouds, have the advantage of producing easy to interpret information compared to radar images, which makes these two sensors two relevant and complementary sources of information.

The use of satellite data in general and the data resulting from their fusion in particular is increasingly propelled by the various applications of remote sensing. It aims at extracting more complete information and truthfully reflecting reality using the different discriminatory elements of these sources of information. However, their joint treatment poses particular problems and therefore requires specific fusion methods that take into consideration this heterogeneity. This is mainly due to the imprecise, incomplete, and even erroneous nature of these data. Moreover, the noise present in these images arises principally from the complex process of satellite images formation and the radiometric, geometric and atmospheric distortions that alter the content of these images, produces ambiguous and difficult to classify areas. This generates inaccurate and uncertain data sources and makes the merging of these data a difficult task. It should also be noted that with the increase in the spatial resolution of these sensors, the sensitivity to the acquisition conditions becomes more acute and the acquired data becomes more and more heterogeneous, which complicates the fusion process.

Several formalisms have been proposed in the literature to model the information provided from a sensor in order to use it in the fusion process, among them we cite: probabilistic Bayesian methods, fuzzy set theory [16], possibilities theory [17–19] and the belief function theory first introduced by Dempster [1], then formalized by Shafer [2]. This last theory is particularly interesting because it proved to have a significant advantage over all other probabilistic approaches in terms of processing heterogeneous information both imprecise and/or uncertain stemming from very varied sources. Furthermore, it can deal with epistemic or subjective uncertainty (*i.e.*, uncertainty resulting from imperfect knowledge) as well as stochastic or objective uncertainty (*i.e.*, uncertainty resulting from data heterogeneity) [20]. The initial theory was modified and ameliorated on several occasions, for example through the work of Dezert-Smarandache [5], a paradoxical reasoning has been proposed.

It is in this context that this work aims at proposing a new evidential method for fusing optical and radar heterogeneous data acquired as High Resolution (HR) remote sensing images in order to improve the joint classification of the studied zones (agricultural areas in our case). The described method must be able to deal with complete optical data as well as missing optical data due to the presence of clouds and/or shadows.

1.2 Objectives and contributions

To meet the general objectives of this thesis, three main axis of research have been fixed:

- Model the different forms of imperfections (imprecision, uncertainty, ambiguity or hesitation between classes) and paradoxes (mixed classes) in the context of remote sensing data acquired using different modalities. Although several methods for estimating mass functions already exist in the literature and are able to perform this task in part, their temporal complexity remains a major obstacle to their application in the case of HR images that contain a large volume of data to be processed.

- Propose a joint classification method for the optical/radar data. The process in question should benefit from the proposed estimation method to construct the mass functions of the heterogeneous data provided by these two types of sensors.
- Model dependencies that can exist between the sources of information during the fusion process. This step is very important to derive a meaningful result. Indeed, since we deal with observations of the same scene, evidences should not be considered as statistically independent.

Methodologically, each of these axes refers to one of the main contributions of our thesis. In the first axis, we propose a new approach for constructing mass functions in a reasonable time from large images. The innovative aspect of this method comes from the fact that we have adopted a geometric viewpoint by projecting the initial representation space of the images into a two-dimensional space, using Kohonen's map in order to simplify the assignment operation of masses for any possible conjunction and/or disjunction of hypotheses.

In the second axis, we propose an original method that aims at tackling the problem of data fusion of heterogeneous sensors such as radar and optical images. Its application focuses on the joint classification of farming landscape images. To this end, Kohonen's unsupervised map classification framework is first used to provide an effective way of handling heterogeneous data, to restore missing parts of optical data and also to estimate the mass functions of these sources of information. Then, some credal discounting techniques from the literature are applied for modelling and handling uncertainty at the pre-fusion phase, in a bid to account for the reliability of the information sources used.

Several rules have been proposed to combine dependent sources of information in the belief function theory. In the third axis, we are particularly interested in approaches which model the dependence using copula theory. In this context, we propose new conjunctive and disjunctive combination rules based on copulas to fuse consonant belief functions. Also, a novel technique for choosing the copula that allows correctly this process is introduced.

The content of this thesis is mainly based on the published papers. Table 1.1 gives the mapping between these publications and the thesis chapters.

1.3 Organization

After this introductory part, the rest of this dissertation is organized as follows:

- The first chapter presents a brief bibliographic overview of the credibilist theories. We explain in particular the set of tools allowing representing and combining imperfect information, as well as decision-making in the framework of belief functions and plausible and paradoxical reasoning theories.
- The second chapter details our contribution regarding the estimation of mass functions using Kohonen's map. The proposed method was compared with the state-of-the-art Basic Belief Assignment (BBA) techniques on a benchmark database and was applied to remote sensing data in the context of image classification. Experimentation shows that our approach gives accurate and reliable results compared to other methods described in the literature, with an ability to handle a large amount of data.

Table 1.1: Contributions of this thesis.

Title	Conference or Journal	Reference	Chapter
Kohonen's map approach for the belief mass modeling	IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2016	[21]	3
The Kohonen map for credal classification of large multi-spectral images	IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2014	[22]	3
On the estimation of mass functions using self organizing maps	Belief Functions: Theory and Applications - Third International Conference (BELIEF), 2014	[11]	3
Kohonen-based credal fusion of heterogeneous data: application to optical and radar joint classification with missing data	IEEE Transactions on Geoscience and Remote Sensing (TGRS), 2016	[23]	4
The Kohonen map for credal fusion of heterogeneous data	IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015	[24]	4
Copulas-based fusion of consonant belief functions induced by dependent sources of evidences	Knowledge Based Systems, 2017	[25]	5

- The third chapter details our contribution concerning the fusion of the optical and radar heterogeneous remote sensing data. The experimental section was dedicated to the land cover classification of the Beauce region in France.
- The fourth chapter is devoted to modelling evidence dependency through the copula theory during the knowledge fusion step. We first describe how copula can be extended in belief function theory and then we introduce our new copula-based rules. To prove the advantage of the latter, a comprehensive experimental study is finally carried out using some benchmark database and simulated Gaussian data.
- A general overview of the work proposed in this thesis and some prospects scheduled for our future works will be exposed in the last chapter.

Credibilist fusion frameworks

2.1	Introduction	21
2.2	Basic concepts of Dempster-Shafer Theory (DST)	22
2.2.1	Representation of information	22
2.2.2	Combination of evidence	26
2.2.3	Discounting techniques	30
2.2.4	Distance measures of evidence	31
2.2.5	Decision making	32
2.2.6	Limits of DST	33
2.3	An extension of DST: Towards Dezert-Smarandache Theory (DSmT)	33
2.3.1	Basic foundations	34
2.3.2	DSm models of fusion	37
2.3.3	DSm rules of combination	38
2.4	Conclusion	39

2.1 Introduction

When it comes to exploit the redundancy and the complementarity of knowledge stemming from very varied sources to give a unique representative information, Dempster-Shafer Theory [1, 2], also known as evidence theory or belief function theory, is considered as an appealing formalism in information fusion domain. Indeed, it offers a robust mathematical framework that allows the processing of both imprecise and uncertain knowledge. Recently, a series of modifications of this theory was suggested, for example through the work of Dezert-Smarandache [4, 5], a paradoxical reasoning has been proposed for dealing with conflicting data sources. The basic concepts of these two credibilist theories are presented in this chapter. The aim is not to give an exhaustive description but to explain some notions in order to lay a foundation for the following chapters.

The remaining of this chapter is structured in two main sections. The first section 2.2 describes the mathematical foundations allowing the representation of imperfect information, their

manipulation and decision-making in belief function theory. It also recalls the limits of this formalism. The second section 2.3, presents a short introduction to Dezert-Smarandache theory.

2.2 Basic concepts of Dempster-Shafer Theory (DST)

This section reviews the basic concepts of belief function theory that contributed to the development of the work presented in this thesis.

2.2.1 Representation of information

2.2.1.1 Frame of discernment

Modelling a fusion problem by DST is mainly based on the definition of the frame of discernment. In general, this frame is denoted by $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\} = \bigcup_{i=1}^N \{\theta_i\}$ and it is consisting of N elements interpreted as hypotheses or propositions. Those elements represent the set of possible answers of the fusion problem under concern and must be:

- Exhaustive: *i.e.*, at least one of the answers θ_i has to be true.
- Exclusive: *i.e.*, the true answer is necessarily unique, $\theta_i \cap \theta_j = \emptyset, \forall i \neq j$.

The constraint of exhaustiveness guarantees that the frame of discernment contains all the possible solutions and is called, in Shafer's model, the closed-world assumption [26]. However, if this condition is not met, Θ is assumed to be incomplete. In such case, we speak about the open-world assumption [27].

From the frame of discernment, a power set denoted 2^Θ can be built, it includes all the subsets A of Θ , more precisely, one has:

$$2^\Theta = \{A, A \subseteq \Theta\} = \{\emptyset, \theta_1, \theta_2, \dots, \theta_N, \theta_1 \cup \theta_2, \dots, \Theta\}.$$

This set serves to allocate parts of belief not only to singleton hypothesis of Θ but also for all possible disjunctions of these hypotheses. This belief can be presented through the mass function defined in the following section.

2.2.1.2 Mass function

The belief of a given source of information (sensor, agent, expert, classifier ...) on imperfect observation is represented as a mass function m , also called Basic Belief Assignment (BBA). Formally, m is the mapping from the power set of Θ to the interval $[0, 1]$ such that:

$$\sum_{A \subseteq \Theta} m(A) = 1. \quad (2.1)$$

The mass of A , denoted $m(A)$, expresses the degrees of belief committed specifically to A that cannot be assigned to any strict subset of A , given the current state of knowledge. Subsets A of Θ verifying $m(A) > 0$ are called focal sets of m . It should be noted that if \emptyset is not a focal set, m is said to be normalized. This condition was originally imposed in the initial Shafer's

model [2], but it may be relaxed in the transferable belief model introduced by Smets in [27]. In such a case, the mass $m(\emptyset)$ is used for representing the conflict between sources. For instance, the case $m(\emptyset) = 1$ corresponds to a total conflict [28].

2.2.1.3 Special classes of mass functions

Definitions of mass functions that benefit from specific denominations in the framework of belief function theory are given in this section. Table 2.1 gives an example of each mentioned particular mass functions on a frame of discernment composed of three hypotheses $\Theta = \{\theta_1, \theta_2, \theta_3\}$.

Definition 1. Subnormal mass function

A subnormal mass function is a function such that \emptyset is a focal set, i.e., $m(\emptyset) > 0$.

Definition 2. Vacuous mass function

A mass function is said to be vacuous or of total ignorance if Θ is the only focal set, i.e., $m(\Theta) = 1$.

Definition 3. Dogmatic mass function

A mass function is said to be dogmatic if Θ is not a focal set, i.e., $m(\Theta) = 0$.

Definition 4. Categorical mass function

The categorical mass function is a non-vacuous mass function that has only one focal set $A \subseteq \Theta$, i.e., $m(A) = 1$ and $m(B) = 0, \forall B \neq A$.

- If A is one of the singletons θ_i of Θ , the knowledge is said to be certain and precise.
- Otherwise (i.e., A is a disjunction of hypotheses), the knowledge is said to be certain and imprecise.

Definition 5. Bayesian mass function

A mass function is said to be bayesian if all its focal sets are singletons of Θ :

$$\begin{cases} m(\theta_i) \geq 0, & \forall \theta_i \in \Theta, \\ m(A) = 0, & \forall A \in 2^\Theta \setminus \{\theta_i\}. \end{cases}$$

In such case the mass function m is equivalent to a probability distribution.

Definition 6. Consonant mass function

The consonant mass function is a function that have nested focal sets.

Definition 7. Simple mass function

A mass function is said to be simple if it has at most two focal sets and, Θ being included:

$$\begin{cases} m(A) = 1 - w, & \forall A \subset \Theta, \\ m(\Theta) = w, \\ m(B) = 0, & \forall B \in 2^\Theta \setminus \{A, \Theta\}, \end{cases}$$

where $w \in [0, 1]$ represents the weight of the ignorance of the simple mass m . Commonly, this mass is denoted A^w .

Definition 8. The negation of a mass function

The negation \bar{m} of a mass function m is defined as the BBA verifying $\bar{m}(A) = m(\bar{A})$, $\forall A \subseteq \Theta$ where $\bar{A} = \Theta \setminus A$ being the complement of A in Θ .

Table 2.1: Example of some special classes of mass functions where the conditions imposed by their definitions are putted in boldface.

Mass function	\emptyset	θ_1	θ_2	$\theta_1 \cup \theta_2$	θ_3	$\theta_1 \cup \theta_3$	$\theta_2 \cup \theta_3$	Θ
Subnormal	0.44	0.22	0.08	0.05	0.05	0.02	0.04	0.1
Dogmatic	0.11	0.22	0.1	0.1	0.37	0.07	0.03	0
Vacuous	0	0	0	0	0	0	0	1
Categorical	0	0	0	0	0	1	0	0
Bayesian	0	0.4	0.2	0	0.4	0	0	0
Consonant	0	0	0.8	0.1	0	0	0	0.1
Simple	0	0.4	0	0	0	0	0	0.6

2.2.1.4 Other functions to represent information

There exists other equivalent representations or formulations of the information encoded by the mass function m . These are the notions of belief bel (also known as credibility), plausibility pl , commonality q and implicability b .

Credibility function

The belief $bel(A)$ represents the total support that can move into the proposition A without any ambiguity and is defined as the sum of the masses of all subsets of A different from \emptyset :

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B), \quad \forall A \subseteq \Theta. \quad (2.2)$$

Plausibility function

The plausibility of A quantifies the maximal degree of belief that could be potentially given to A and is defined as the sum of the masses of all the subsets of Θ that have non-zero intersection with A :

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Theta. \quad (2.3)$$

If all focal sets of the mass function m are non-empty (*i.e.*, m is normalized), the functions bel and pl are dual with means that $pl(A) = 1 - bel(\bar{A})$. Furthermore, clearly one has $bel(A) \leq pl(A)$, for all $A \subseteq \Theta$. The functions $pl(A)$ and $bel(A)$ have the following special properties:

1. $bel(\emptyset) = 0$ and $pl(\emptyset) = 0$.
2. $bel(\Theta) = 1$ and $pl(\Theta) = 1$.
3. The function bel is completely monotone (or monotone of infinite order) *i.e.*,

$$bel\left(\bigcup_{i=1,\dots,n} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} bel\left(\bigcap_{i \in I} A_i\right), \quad (2.4)$$

and by duality, pl is completely alternating (or alternating of infinite order), *i.e.*,

$$pl\left(\bigcap_{i=1,\dots,n} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} pl\left(\bigcup_{i \in I} A_i\right). \quad (2.5)$$

Note that for bayesian mass function, equality holds in equations (2.4) and (2.5).

Commonality function

The commonality $q(A)$ quantifies the sum of the masses allocated to supersets of A and it is defined as the sum of the masses of the sets in which A is included:

$$q(A) = \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \Theta. \quad (2.6)$$

Implicability function

The quantity $b(A)$ is the sum of masses allocated to subsets of A including the mass of the empty set and it is defined as:

$$b(A) = \sum_{B \subseteq A} m(B) = bel(A) + m(\emptyset), \quad \forall A \subseteq \Theta. \quad (2.7)$$

These two last functions are essentially used in the simplification of calculations at the combination level and they verify the following properties:

1. $b(\emptyset) = m(\emptyset)$ and $b(\Theta) = 1$.
2. $q(\emptyset) = 1$ and $q(\Theta) = m(\Theta)$.
3. $\bar{b}(A) = q(\bar{A}), \forall A \subseteq \Theta$.

The functions m , bel , pl , q and b are in one to one correspondence. For instance, the functions of belief, plausibility and commonality can be recovered from the mass function using the so-called Möbius transformation [29, 30] as follows:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} bel(B), \quad (2.8)$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A| - |B| - 1} pl(\overline{B}), \quad (2.9)$$

$$m(A) = \sum_{B \supseteq A} (-1)^{|A| - |B|} q(B), \quad (2.10)$$

where $|B|$ represents the cardinality of $B \subseteq \Theta$ ($|\theta_i| = 1, |\theta_i \cup \theta_i| = 2, \dots, |\Theta| = N$).

2.2.2 Combination of evidence

2.2.2.1 Combination rules of independent evidence

The combination rules are used to fuse several belief functions provided by multiple sources of information in order to synthesize a more reliable global knowledge. Within the framework of DST, several operators have been introduced to aggregate independent evidence. However, the majority of those rules are mainly based on the conjunctive and disjunctive forms of combination which we recall here.

Conjunctive rule

Let us consider two distinct data sources through their mass functions m_1 and m_2 defined on the same frame of discernment Θ . The mass $m_{1,2}^{CR}$ resulting from their combination using the Conjunctive Rule (CR) is defined as:

$$m_{1,2}^{CR}(A) = m_1 \odot_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Theta. \quad (2.11)$$

It can be expressed very simply in terms of commonality functions defined by equation (2.6). Let q_1 and q_2 be the commonality functions associated respectively to m_1 and m_2 . The result of their combination, denoted $q_1 \odot_2$, is expressed as:

$$q_1 \odot_2(A) = q_1(B) q_2(C), \quad \forall A \subseteq \Theta. \quad (2.12)$$

In the form of the equation (2.11), the conjunctive combination can generate a subnormal mass function. In order to satisfy the closed world assumption (*i.e.*, $m(\emptyset) = 0$), a normalized version of this rule has been proposed [1]. It corresponds to Dempster's rule given by:

$$m_{1,2}^{DS}(A) = m_1 \oplus_2(A) = \frac{1}{1 - K} m_1 \odot_2(A), \quad \forall A \subseteq \Theta, \quad (2.13)$$

where $K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ measures the degree of conflict between m_1 and m_2 . This operator is commutative, associative and it admits as neutral element the vacuous mass function. Despite these interesting and advantageous properties, many authors [31–33] show

that this rule cannot be applied to combine highly conflicting input sources since its normalization procedure provides unsatisfactory performances and strange behaviours. As a result, several interesting alternatives to Dempster's rule have been proposed in literature in order to differently redistribute the conflict mass. Interested readers could refer to [3, 34–37] for more details about some of these rules. Here, we present only the PCR6 rule proposed by Martin and Osswald in [38, 39]. This rule allows the redistribution of the conflicting mass only to those elements that are involved in the conflict and proportionally to their individual masses. It is defined by: $m_{PCR6}(\emptyset) = 0$ and for all $A \neq \emptyset \in 2^\Theta$

$$m_{PCR6}(A) = m_1 \odot_2(A) + \sum_{\substack{B \in 2^\Theta \setminus \{A\}, \\ A \cap B = \emptyset}} \left[\frac{m_1(A)^2 \cdot m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 \cdot m_1(B)}{m_2(A) + m_1(B)} \right], \quad (2.14)$$

where all sets involved in the equation are in canonical form and where all fraction having denominators equal to zero is discarded. The PCR6 is commutative and not associative but quasi-associative.

Disjunctive rule

Generally, the conjunctive forms of combination are reserved for the fusion of reliable data sources. If at least one of the combined sources is reliable, Dubois and Prade [3] propose to use its dual, the Disjunctive Rule (DR). The combination of m_1 and m_2 by this rule gives the new mass function $m_{1,2}^{DR}$ defined as follows:

$$m_{1,2}^{DR}(A) = m_1 \odot_2(A) = \sum_{B \cup C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Theta. \quad (2.15)$$

Similarly the implicability functions can be useful to simplify the calculation of this disjunctive form of combination, since:

$$b_1 \odot_2(A) = b_1(B) b_2(C), \quad \forall A \subseteq \Theta. \quad (2.16)$$

DR has the same properties as CR and Dempster's rule.

2.2.2.2 Combination rules of dependent evidence

If the independence assumption of sources is not reasonable, other operators of aggregation are recommended to utilize. In the literature, the cautious and the bold rule of Denœux [15] as well as Kallel's and Le Hégarat-Masclé's rule [40] are most often used.

The canonical decomposition

The conjunctive canonical decomposition consists in decomposing the mass function m , under certain conditions, into a set of simple masses $A^{w(A)}$ combined by the operator \odot :

$$m = \odot_{A \in \Omega} A^{w(A)}, \quad (2.17)$$

where $w(A) \in]0, +\infty[$ is the weight function computed for each $A \in 2^\Theta \setminus \{\emptyset\}$ as follows:

$$\ln(w(A)) = - \sum_{B \supseteq A} (-1)^{|B|-|A|} \ln(q(B)), \quad \forall A \subset \Theta. \quad (2.18)$$

All m having the representation of equation (2.17) are said to be separable belief functions and if m is a non dogmatic, this representation is unique.

The disjunctive canonical decomposition consists in decomposing a subnormal mass function m into a set of complementary simple masses $A^{v(A)}$ combined by the operator \oplus :

$$m = \oplus_{A \neq \emptyset} A^{v(A)}, \quad (2.19)$$

where $A_{v(A)}$ is given by:

$$\begin{cases} A_{v(A)}(A) = 1 - v(A), \\ A_{v(A)}(\emptyset) = v(A), \\ A_{v(A)}(B) = 0, \quad \forall B \in 2^\Theta \setminus \{A, \emptyset\}, \end{cases}$$

and $v(A) \in [0, +\infty[$ is the weight function computed for each $A \in 2^\Theta \setminus \{\emptyset\}$ as follows:

$$\ln(v(A)) = - \sum_{B \subseteq A} (-1)^{|B|-|A|} \ln(Bel(B)), \quad \forall A \subseteq \Theta. \quad (2.20)$$

Cautious conjunctive rule

To combine mass functions coming from dependent sources, Denœux (inspired by Smets) proposed the Cautious Conjunctive Rule (CCR) [15]. This rule is based on the principle of least commitment defined by the following reasoning: Let m_1 and m_2 be two mass functions obtained from reliable sources of information and $m_{1,2}$ their combined belief function. This principle requires the following constraint: $m_{1,2}$ should be more informative than m_1 and m_2 . Let $S_w(m_1)$, (resp. $S_w(m_2)$) be the set of mass functions richer than m_1 (resp. m_2) in the sense of \sqsubseteq_w ¹, then $m_{1,2}$ must belong to the intersection $S_w(m_1)$ and $S_w(m_2)$, so that:

$$m_{1,2} \in \left(S_w(m_1) \cap S_w(m_2) \right). \quad (2.21)$$

The cautious rule consists in determining the less rich mass function searched within the meaning of the partial order \sqsubseteq_w .

Thus, if $A^{w1(A)}$ and $A^{w2(A)}$ are two simple mass functions, their combination by the cautious conjunctive rule is the simple mass function $A^{w1(A) \wedge w2(A)}$. Thereby, the BBA $m_{1,2}$ resulting from the combination of the non dogmatic belief functions $m_1 = \oplus_{A \subset \Theta} A^{w1(A)}$ and $m_2 = \oplus_{A \subset \Theta} A^{w2(A)}$ is then defined as follows:

$$m_{1,2}^{CCR} = m_1 \wedge m_2 = \oplus_{A \subset \Theta} A^{w1(A) \wedge w2(A)}. \quad (2.22)$$

The properties of this rule result from those of the minimum operator (denoted \wedge): commutativity, associativity, idempotence (i.e., $m_1 \wedge m_1 = m_1$) and distributivity of CR with respect to CCR.

¹ $m_{1,2}$ is more informative than m_1 and m_2 in the sense of \sqsubseteq_w if and only if $w_1(A) \leq w_2(A)$, $\forall A \subset \Theta$ (assuming that m_1 and m_2 are non dogmatic).

Bold rule

The bold rule, denoted $m_{1,2}^{BR}$, is based on the principle of maximum commitment that results in the following constraint: the resulting mass $m_{1,2}$ should be less informative than masses to combine m_1 and m_2 . Let A_{v1} and A_{v2} be two complementary simple mass functions, their combination by Bold Rule (BR) [15] is given by:

$$m_{1,2}^{BR} = m_1 \odot_2 = \bigodot_{A \neq \emptyset} A_{v1(A) \wedge v2(A)}. \quad (2.23)$$

This rule has the following properties: commutativity, associativity, idempotence and distributivity of DR with respect to BR.

Kallel's and Le Hégarat-Masclé's rule

In [40], Kallel and Le Hégarat-Masclé propose a variant of the cautious rule, so-called cautious-adaptive rule, that is able to take into account the actual degree of non-distinctness of source through a discounting level. Thus, this new rule varies between the conjunctive rule and the cautious one, depending on the degree of correlation between sources.

Recall that Smet [41] defines the correlation (defined by the commonality factor q_0 associated with a BBA m_0) between two BBAs m_1 and m_2 as:

$$q_0(A) = \frac{q_1(A)q_2(A)}{q_{1 \wedge 2}(A)} = \frac{q_1(A)q_2(A)}{\min(q_1(A), q_2(A))}, \forall A \subseteq \Theta, \quad (2.24)$$

with $q_1, q_2, q_{1 \wedge 2}$ are the commonality functions of the joint conjunctive belief structure underlying, respectively, m_1, m_2 and $m_{1 \wedge 2}$.

Contrary to Smets that requires an in-depth comparison of the origin of the pieces of evidence that have induced m_1 and m_2 when it comes to construct m_0 without knowing $m_{1 \wedge 2}$, the authors propose to compute m_0 by simply replacing the BBA $m_{1 \wedge 2}$ in equation (2.24) respectively by the BBA given by Dencœux's cautious rule when m_1, m_2 are not consonant and by the result of the minimum possibilistic rule when they are consonant. Let $\rho \in [0, 1]$ be the factor that parameterizes the non-distinctness of source with $\rho = 1$ (respectively, $\rho = 0$) means that the evidences are non-distinct (respectively, distinct). The discounting of the correlation m_0 according to ρ is given by:

$$^{(\rho w)}m_0 = \bigodot_{A \subseteq \Theta} A^{^{(\rho A)}w},$$

with $^{(\rho A)}w = (1 - \rho_A)w + \rho_A$.

Then, based on equation (2.24), the authors define the rule ${}^\rho \wedge$ as follows:

$$q_{1 {}^\rho \wedge 2}(A) = \frac{q_1(A)q_2(A)}{{}^{(\rho w)}q_0(A)}, \quad \forall A \subseteq \Theta. \quad (2.25)$$

The limitation of this approach comes from the exploitation of the correlation information computed from the w -least committed joint structure in further computations that is disputable if no evidence is available.

2.2.3 Discounting techniques

Discounting operations have the advantage of alleviating the contradiction existing between the sources and diminishing the influence of unreliable sources during the fusion stage. This always leads to the suppression of the conflict and helps thus to the extraction of the trusted proposition from a set of information sources. Here, some frequently used credal discounting techniques are presented.

Shafer's classical discounting approach

The classical discounting technique introduced by Shafer in [2] consists in proportionally moving part of the belief mass assigned to the focal elements to the set Θ representing the uncertainty. Thus, after quantifying the reliability of the source denoted $\alpha, 0 \leq \alpha \leq 1$, its associated mass functions can be discounted as follows:

$$\begin{aligned} m^\alpha(\Theta) &= (1 - \alpha) \cdot m(\Theta) + \alpha, \\ m^\alpha(A) &= (1 - \alpha) \cdot m(A), \quad \forall A \subseteq \Theta. \end{aligned} \quad (2.26)$$

In the case where $\alpha = 1$, the source is said to be unreliable, which implies a total transfer of the mass to ignorance Θ . Otherwise (*i.e.*, $\alpha = 0$), the source is said to be completely reliable and all the information it provides is accepted.

Priority discounting approach

Let $0 \leq \beta \leq 1$ be the priority of an evidence source calculated using prior knowledge or attributed by an expert or designer fusion. Whereby $\beta = 1$ represents the highest priority assigned to a source and $\beta = 0$ the minimum. The priority discounting approach, as defined in [42], consists in proportionally transferring a part of the masses of focal elements to the empty set \emptyset , compared to Θ used in the discounting approach of Shafer, as follows:

$$\begin{aligned} m^\beta(\emptyset) &= \beta \cdot m(\emptyset) + (1 - \beta), \\ m^\beta(A) &= \beta \cdot m(A), \quad \forall A \subseteq \Theta, A \neq \emptyset. \end{aligned} \quad (2.27)$$

This technique allows the correction (*i.e.*, adjustment) of the initial BBA of the source by considering only its priority. That is how a BBA with sure knowledge retains its full importance in the fusion process.

Contextual discounting approach

Based on the idea that the credibility of the source of information can change depending on the proposition or the object to identify, a contextual discounting has been proposed by Mercier in [43]. This process allows to revise a piece of information represented by a mass function taking into account the reliability of each simple hypothesis $\theta_i \in \Theta, i \in \{1, \dots, N\}$. Let m

be the mass function provided by the source of information, the contextual discounting of this BBA on the context θ_i is given by:

$$m^\lambda = m \odot m_{\theta_i, \lambda}, \quad (2.28)$$

where $m_{\theta_i, \lambda}$ represents the reliability attributed to the simple hypothesis θ_i of the partition Θ . Its mass is defined as:

$$m_{\theta_i, \lambda}(A) = \begin{cases} \lambda & \text{if } A = \emptyset, \\ 1 - \lambda & \text{if } A = \theta_i, \\ 0 & \text{elsewhere.} \end{cases}$$

2.2.4 Distance measures of evidence

Distance measurements have been mainly studied in the framework of DST in order to quantify the degree of dissimilarity between different evidences. They play an important role in many applications such as optimization, clustering analysis, conflict management, etc. In [44], Jousselme and Maupin give a survey of the most available definitions on distance measures of evidence. Extensions of the Euclidean and Bhattacharyya distances are given, respectively, by Cuzzolin [45] and Ristic et al. [46, 47]. Tessem [48] propose to calculate the distance measure between the pignistic probabilities associated to mass functions in question. Some distances are directly defined between different mass functions such as Jousselme's distance [49] that has the advantage of taking into account the cardinality of focal sets. Other distances were studied to define dissimilarity between two BBAs using belief intervals [50]. In the rest of this section, we detail those having been widely used in DST based applications.

Tessem's distance

From a mass function m , such that $m(\emptyset) < 1$, the pignistic probability transformation, denoted $BetP$ [51], can be established as follows:

$$BetP(A) = \sum_{B \in 2^\Theta, B \neq \emptyset} \frac{|B \cap A|}{|B|} m(B), \quad \forall A \subseteq \Theta. \quad (2.29)$$

The idea is to equally distribute the mass assigned to a set A to its elements.

Following [48], Tessem's distance called also, the betting commitment distance is formalized as:

$$d_T(m_1, m_2) = \max_{A \subseteq \Theta} |BetP_1(A) - BetP_2(A)|. \quad (2.30)$$

Jousselme's distance

Jousselme's distance [49] complies with the metric axioms and the structural property [44]. Furthermore, it is considered as an appropriate measure of disagreement between evidences and it is calculated between two BBAs as follows:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2} \cdot (m_1 - m_2)^t \cdot D \cdot (m_1 - m_2)}, \quad (2.31)$$

where D is a symmetric matrix $2^{|\Theta|} \times 2^{|\Theta|}$ defined by:

$$D(A, B) = \begin{cases} 1 & \text{if } A = B = \emptyset, \\ \frac{|A \cap B|}{|A \cup B|} & \text{if } A, B \subseteq 2^\Theta. \end{cases}$$

Euclidean belief interval distance

Let m_1 and m_2 be two BBAs defined on the same frame of discernment Θ . For each focal set $A \in \Theta$ one can calculate the belief interval of A for m_1 and m_2 respectively, which are denoted by $BI_1 = [bel_1(A), pl_1(A)] = [a_1, b_1]$ and $BI_2 = [bel_2(A), pl_2(A)] = [a_2, b_2]$. The distance between these two belief intervals is computed by the Wasserstein distance [52] defined by:

$$d^I([a_1, b_1], [a_2, b_2]) = \sqrt{\left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2}. \quad (2.32)$$

In [50], the authors propose a new distance measure of evidence using the belief interval distance described above. This distance is formalized as:

$$d_{BI}(m_1, m_2) = \sqrt{N_c \sum_{A \in 2^\Theta} [d^I(BI_1(A), BI_2(A))]^2}, \quad (2.33)$$

where $N_c = \frac{1}{2^n - 1}$ is a normalization factor.

2.2.5 Decision making

Decision-making is an essential step in belief function theory, as it operates in an uncertain context. It consists in choosing, among a finite set of potential solutions (choices), the one that best meets the problem under consideration. The maximum of belief or plausibility are the two simplest strategies when we prefer to adopt a pessimistic (prudent) or optimistic (less prudent) attitude of choices, respectively. Another frequently used strategy is the maximum of pignistic probability transformation. It is considered as a balanced strategy between the last two techniques. In this case, the predicted element A_0 is the most probable one:

$$A_0 = \arg \max_{A \in \Theta} BetP(A) \quad (2.34)$$

As a result of the conversion step of the initial mass function to a credibility, plausibility or probability function with respect to the used decision technique, all the information contained in the original BBA will not be fully used. Indeed several mass functions can lead to the same pignistic probability. In order to avoid this problem, Dezert et al. [53] propose to use a decision rule based on a distance measure [54]. This approach consist in calculating the Euclidean belief

interval distance between the mass function m under test and all the categorical mass functions focusing on each of its focal elements in order to choose the one that minimizes this distance:

$$A_0 = \arg \min_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X), \quad (2.35)$$

where $m(X) = 1, X \in 2^\Theta$ is the categorical BBA and d_{BI} is the Euclidean belief interval distance. This decision-making technique has the advantage to make decision on singleton elements as well as any other type of elements of the frame of reasoning. For more information about decision-making, interested readers are invited to consult reference [55].

2.2.6 Limits of DST

Although the theory of belief functions seems very attractive in terms of the various basic tools that it offers for the rich and flexible modelling of uncertain and imprecise information, nevertheless two major limitations have been observed in its applicability in real fusion problems. The first defect of this theory is its framework of reasoning Θ whose refinement becomes inaccessible in some problems because of the vague, relative and imprecise nature of its elements. In such case, DST cannot be applied since its formalism does not take into account the paradoxical (conflicting) nature of the information to be treated. The second defect is Dempster's rule of combination, which is often questionable in the literature because of:

- Its lack of complete theoretical justification. The debate about this point has emerged in many research works [56–58], but, to the best of our knowledge, none of these justifications is convincing since it can be hard to verify the condition of independence of sources, and, then to give a unique meaning of "independence" here.
- These weaknesses revealed with Zadeh's famous example [31] when the source reliability assumption is not correct. Indeed, it has been demonstrated that this rule has a counter-intuitive behaviour when the conflict between the sources is high as well if it is weak [33].

To remedy these shortcomings, a series of modifications of the normalization step of Dempster's rule was suggested and thereafter several interesting and valuable alternative rules of combination were created for dealing with highly conflicting sources of evidences. Among them we can mention, without being exhaustive, Yager's rule [35], Dubois and Prade's rule [3] and Smet's rule [26]. Lefevre et al. have proposed, in [37], a way to unify these different approaches of combination using a weight operator that allows the redistribution of the conflict. This approach of unification makes it possible also to define other new rules of combinations responding to a specific objective. Dezert-Smarandache theory framework is also considered as a serious alternative to overcome the limitations of the belief function theory. This framework provides new foundations for the plausible and paradoxical reasoning and proposes new combination rules, as it will be shown in the next section.

2.3 An extension of DST: Towards Dezert-Smarandache Theory (DSmT)

DSmT formalism was firstly proposed by Dezert [5] as a generalization of the classical Dempster-Shafer theory. It constitutes a rigorous mathematical framework to deal with the fusion of

uncertain, highly conflicting and imprecise sources. Indeed, it provides a new interesting combination mechanism that can solve static, complex or dynamic fusion problems. The basic idea is to allow elements of the frame of discernment to overlap. Thus, the mutual exclusivity constraint imposed upon the hypotheses in DST is not assumed in general. In this section, the basic foundations of DSmT, its models of fusion and its most important rules of combination are reviewed.

2.3.1 Basic foundations

Let $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ be the frame of discernment composed of finite sets of hypotheses which are exhaustive but not necessarily exclusive. DSmT works on the hyper-power set D^Θ of this frame defined as the set of all composite possibilities built from Θ which \cap and \cup operators such that:

1. $\emptyset, \theta_1, \dots, \theta_N \in D$.
2. $\forall E \in D^\Theta, F \in D^\Theta, (E \cup F) \in D^\Theta, (E \cap F) \in D^\Theta$.
3. No other elements belong to D^Θ , except those, obtained by using rules 1 or 2.

The following example gives an idea of the construction of the hyper-power set for dimensions of Θ equal or less than 3.

Example:

- If $\Theta = \{\theta_1\}$, then one has $D^\Theta = \{\alpha_0 \triangleq \emptyset, \alpha_1 \triangleq \theta_1\}$.
- If $\Theta = \{\theta_1, \theta_2\}$, then one has $D^\Theta = \{\alpha_0 \triangleq \emptyset, \alpha_1 \triangleq \theta_1 \cap \theta_2, \alpha_2 \triangleq \theta_1, \alpha_3 \triangleq \theta_2, \alpha_4 \triangleq \theta_1 \cup \theta_2\}$.
- If $\Theta = \{\theta_1, \theta_2, \theta_3\}$, then one has $D^\Theta = \{\alpha_0, \alpha_1, \dots, \alpha_{18}\}$ where

$$\begin{array}{ll}
 \alpha_0 \triangleq \emptyset & \\
 \alpha_1 \triangleq \theta_1 \cap \theta_2 \cap \theta_3 & \alpha_{10} \triangleq \theta_2 \\
 \alpha_2 \triangleq \theta_1 \cap \theta_2 & \alpha_{11} \triangleq \theta_3 \\
 \alpha_3 \triangleq \theta_1 \cap \theta_3 & \alpha_{12} \triangleq (\theta_1 \cap \theta_2) \cup \theta_3 \\
 \alpha_4 \triangleq \theta_2 \cap \theta_3 & \alpha_{13} \triangleq (\theta_1 \cap \theta_3) \cup \theta_2 \\
 \alpha_5 \triangleq (\theta_1 \cup \theta_2) \cap \theta_3 & \alpha_{14} \triangleq (\theta_2 \cap \theta_3) \cup \theta_1 \\
 \alpha_6 \triangleq (\theta_1 \cup \theta_3) \cap \theta_2 & \alpha_{15} \triangleq \theta_1 \cup \theta_2 \\
 \alpha_7 \triangleq (\theta_2 \cup \theta_3) \cap \theta_1 & \alpha_{16} \triangleq \theta_1 \cup \theta_3 \\
 \alpha_8 \triangleq (\theta_1 \cap \theta_2) \cup (\theta_1 \cap \theta_3) \cup (\theta_2 \cap \theta_3) & \alpha_{17} \triangleq \theta_2 \cup \theta_3 \\
 \alpha_9 \triangleq \theta_1 & \alpha_{18} \triangleq \theta_1 \cup \theta_2 \cup \theta_3.
 \end{array}$$

Contrary to DST, we note that DSmT also allows assigning beliefs to all unions and intersections of hypotheses.

The cardinality of the hyper-power set conjointly increases with the cardinality of the frame of discernment Θ on which it is based and it is generally majored by 2^{2^N} , where N denotes the cardinality of Θ . The problem of determining the cardinality of D^Θ is similar in nature with the famous Dedekind's problem [59, 60] on computing the number of isotone Boolean functions. Indeed, we use the sequence of Dedekind's numbers to find the sequence of cardinalities. Table 2.2 gives the Dedekind's sequence according to the cardinality of Θ .

Table 2.2: The sequence of Dedekind's numbers.

Cardinality of Θ	0	1	2	3	4	5	6	7	8
cardinality of D^Θ	1	2	5	19	167	7580	7828353	$241 * 10^{10}$	$561 * 10^{20}$

In order to enumerate the various elements constituting the hyper-power set and to simplify the implementation of most useful operations in DSmT, some codifications of Venn diagram have been proposed. If $|\Theta| = n$, the one proposed by Smarandache [4] allows to codify the $2^n - 1$ distinct parts of this diagram as follows:

- Each $\langle i \rangle$ represents the part of θ_i without overlap with the others $\theta_j, i \neq j$
- Each $\langle ij \rangle$, or $\langle ijk \rangle$ represents respectively the intersection of the part $\langle i \rangle$ and $\langle j \rangle$ only, or $\langle i \rangle$ and $\langle j \rangle$ and $\langle k \rangle$, etc.

Figure 2.1 represents an illustration of this codification for the case of the frame of discernment of 3 dimensions. For instance, for θ_1 and $\theta_1 \cap \theta_2$, Smarandache's codification gives respectively $\{\langle 1 \rangle, \langle 12 \rangle, \langle 13 \rangle, \langle 123 \rangle\}$ and $\{\langle 12 \rangle, \langle 123 \rangle\}$. Although his method works well, the authors raise a problem if the cardinality size of the frame of discernment is equal or more than 10. In [61], Martin introduces a simpler and more practical codification which consists in attributing only one integer number of $[1 : 2^n - 1]$ to each disjoint parts of this diagram.

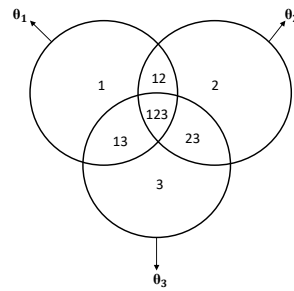


Figure 2.1: Venn diagram of a free model for a 3D frame.

The vector d_n of all the elements of D^Θ can be obtained then by solving the following system of simple linear equations:

$$d_n = D_n \cdot u_n, \quad (2.36)$$

where D_n is the binary matrix of Dedekind and u_n is the chosen codification basis. The following example shows how those element are constructed using Smarandache's codification.

$$\begin{pmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
 0 & 1 & 1 & 0 & 1 & 1 & 1 \\
 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
 1 & 0 & 1 & 0 & 1 & 1 & 1 \\
 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
 1 & 1 & 1 & 1 & 1 & 1 & 1
 \end{pmatrix}
 \begin{pmatrix}
 < 1 > \\
 < 2 > \\
 < 12 > \\
 < 3 > \\
 < 13 > \\
 < 23 > \\
 < 123 >
 \end{pmatrix}
 =
 \begin{pmatrix}
 \alpha_0 \\
 \alpha_1 \\
 \alpha_2 \\
 \alpha_3 \\
 \alpha_4 \\
 \alpha_5 \\
 \alpha_6 \\
 \alpha_7 \\
 \alpha_8 \\
 \alpha_9 \\
 \alpha_{10} \\
 \alpha_{11} \\
 \alpha_{12} \\
 \alpha_{13} \\
 \alpha_{14} \\
 \alpha_{15} \\
 \alpha_{16} \\
 \alpha_{17} \\
 \alpha_{18}
 \end{pmatrix}$$

The attribution of beliefs in DS_mT is analogous to the classical Dempster-Shafer theory, but it is done on hyper-power set D^Θ , instead of the power set Θ of the basic frame of reasoning. For instance, the generalized mass function m , also called Generalized Basic Belief Assignment (GBBA) is defined as the mapping from the hyper-power set D^Θ to the interval $[0, 1]$ such that:

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in D^\Theta} m(A) = 1. \quad (2.37)$$

For decision-making from the combined GBBA, the generalized pignistic transformation [4] can be used:

$$GPT(A) = \sum_{E \in D^\Theta} \frac{C_M(E \cap A)}{C_M(E)} m(E), \quad \forall A \in D^\Theta \quad (2.38)$$

where $C_M(E)$ is the cardinality of E , defined within DS_mT framework as the number of the disjoint parts of Venn diagram included in E . The decision is then taken by the maximum of GPT .

2.3.2 DSm models of fusion

DSmT offers the possibility of working with three different fusion models depending on the intrinsic nature of hypotheses of the fusion problem under consideration. Here we give a brief definition of each of these models.

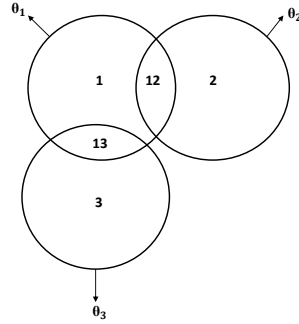


Figure 2.2: Venn diagram of a hybrid model for a 3D frame.

- **Free model:**

The free DSm model, denoted $\mathcal{M}^f(\Theta)$, represents the situation where the whole hyper-power set D^Θ is considered. As shown in Table 2.3, the main drawback of this model is its complexity of implementation due to the high memory size required for storing the elements of D^Θ . Indeed, it is almost impossible for our current computers to store all the elements of $\mathcal{M}^f(\Theta)$ when $|\Theta| > 6$.

Table 2.3: Memory size requirements for D^Θ .

$ \Theta $	2	3	4	5	6	7
Size of θ_i	1 bytes	1 bytes	2 bytes	4 bytes	8 bytes	16 bytes
Number of elements	4	18	166	7579	7828352	$\approx 2.4 * 10^{12}$
Size of D^Θ	4 bytes	18 bytes	0.32 Kb	30 Kb	59 Mb	$3.6 * 10^4$ GB

- **Hybrid model:**

The hybrid DSm model, denoted $\mathcal{M}(\Theta)$, represents the situation where some sets of the hyper-power set D^Θ are not possible due to one or more integrity constraints. There exists three types of these constraints:

1. Exclusivity constraint: the exclusivity constraint arises when we force certain intersections of Θ to be empty.
2. Non-existential constraint: the non-existential constraint arises when we force certain unions of Θ to be empty.
3. Hybrid constraint: the hybrid constraint is a combination of constraints 1 or 2.

Figure 2.2 shows the Venn diagram of the hybrid model given for the constraint $\theta_2 \cap \theta_3 = \emptyset$ and Θ composed of three hypotheses. As we can notice, this model is a restricted set that contains fewer elements than in the hyper-power set case. It is generally the most suitable and faithful model to the real problem of fusion.

- **Shafer's model:**

The Shafer's model, denoted $\mathcal{M}^0(\Theta)$, represents the situation where all the exclusivity constraints are imposed. In such case, the hyper-power set D^Θ is reduced to the classical power set 2^Θ used in DST.

2.3.3 DSm rules of combination

In this section, we detail the two most commonly used combination rules in the framework of DSmT.

2.3.3.1 The classic DSm rule of combination

Let us consider two distinct paradoxical or rational sources of information through their mass functions m_1 and m_2 defined on the same frame of discernment. The classic DSm rule of combination, also called Dezert-Smarandache's rule, corresponds to the conjunctive consensus of these sources operating under D^Θ and it is given by:

$$m_{\mathcal{M}(\times)}(C) \equiv m(C) = \sum_{\substack{A, B \in D^\Theta \\ A \cap B = C}} m_1(A) m_2(B), \quad \forall C \in D^\Theta. \quad (2.39)$$

This rule of combination is commutative and associative.

2.3.3.2 The hybrid DSm rule of combination

The combination DSm Hybrid rule is designed to take into account all possible integrity constraints of the chosen fusion problem. So, it can work for all fusion models. This rule is given for two independent sources of information by:

$$m_{DSmH}(A) = m_{\mathcal{M}(\Theta)}(A) \triangleq \phi(A) \left[S_1(A) + S_2(A) + S_3(A) \right], \quad (2.40)$$

where the function $\phi(A)$ is a binary function equal to 1 if A is a non-empty set and 0 otherwise.

$$S_1(A) = \sum_{\substack{X_1, X_2 \in D^\Theta \\ X_1 \cap X_2 = A}} m_1(X_1) \cdot m_2(X_2), \quad (2.41)$$

$$S_2(A) = \sum_{\substack{X_1, X_2 \in \emptyset \\ [\mathcal{U}(X_1) \cap \mathcal{U}(X_2) = A] \vee [(\mathcal{U}(X_1) \cap \mathcal{U}(X_2) = \emptyset) \wedge (A = I_t)]}} m_1(X_1) \cdot m_2(X_2), \quad (2.42)$$

$$S_3(A) = \sum_{\substack{X_1, X_2 \in D^\Theta \\ X_1 \cup X_2 = A \\ X_1 \cap X_2 = \emptyset}} m_1(X_1) \cdot m_2(X_2). \quad (2.43)$$

In equation (2.42), $\mathcal{U}(X)$ represents the union of all θ_i that compose X , I_t represents the total ignorance and $S_2(A)$ represents the mass of all sets of \emptyset assigned to ignorance. $S_1(A)$ of equation (2.41) corresponds to the classic DSm rule defined in the previous section and $S_3(A)$ of equation (2.43) transfers the sum of the sets of \emptyset to the non-empty sets.

2.4 Conclusion

In this chapter, the main concepts of DST are presented as well as its extension (*i.e.*, DSMT). This theory seems to be very attractive as it offers very strong properties and functions allowing the representation and combination of imperfect data. It also offers several measures to model all forms of uncertainty (outliers, conflicting data, reliability of sources, etc.). However, the estimation of basic belief assignments has always been a difficulty for applying belief function theory efficiently in real-world applications. Although the different approaches presented in literature has received substantial attention in several research disciplines, their uses in remote sensing applications present some limits when processing large remote sensing images due to the unreasonable execution time. This point will be addressed in the next chapter, in which, an overview of some state-of-art approaches is given. Then, our proposed method in the case of representing knowledge of large quantity of multi-variate data is presented.

Kohonen's map approach for the belief mass modeling

3.1	Introduction	41
3.2	Estimation of mass functions in evidence theory	42
3.2.1	Distance-based approaches	42
3.2.2	Likelihood-based approaches	44
3.3	New method to build mass functions	45
3.3.1	Overview on Kohonen's map	45
3.3.2	Feature space for smart basic belief assignment	47
3.3.3	Mass function construction	48
3.4	Simple simulation	51
3.5	Experiments on benchmark data set	54
3.6	Experiments on a real satellite image	56
3.6.1	The classification results in 2^Θ	57
3.6.2	The classification results in D^Θ	61
3.7	Conclusion	63

3.1 Introduction

Despite, the fact that belief function theory excels in extracting the most truthful proposition from a multisource context, the estimation of basic belief assignments has always been a difficulty for applying belief functions efficiently in applications. In this chapter, we propose to define a new approach for estimating mass functions in the case of representing knowledge in complex systems, where the quantity of information is important (*i.e.*, a complex feature space \mathbb{R}^p). The construction of mass function can be done through Kohonen's map [6] that allows to approximate the feature space dimension into a projected 2D space (so called map). Thus, the use of Kohonen's map simplifies the process of assigning mass functions on conjunctions and disjunction of hypotheses when considering the relative distance of an observation to the map.

In the feature space (in \mathbb{R}^p), operations on basic belief assignment can be much more complex and may not be feasible due to computing time or accuracy consideration.

This chapter is organized as follows. The second section 3.2 briefly survey some existing methods for estimating mass functions. Then section 3.3 introduces the main ideas of the proposed approach and explains the underlying methodology. Section 3.4 provides simple examples to illustrate the methodology. In section 3.5, the results obtained by the proposed approach are compared to some state-of-the art methods on a set of benchmark database. Then, section 3.6 presents a deeper analysis of the classification results on a large SPOT image. Finally, section 4.5 concludes.

3.2 Estimation of mass functions in evidence theory

The development of an effective and operational decision-making system necessarily involves the correct modelling of mass functions, which in turn reflects the belief of a given source of information. The more reliable the modelling, the closer the decision is to reality and vice versa. Several methods have been proposed in the literature, and their choice must be made depending on the nature of data and the application. In general, we distinguish two main family approaches. Likelihood based approaches [2, 62], require the knowledge, or the estimation, of the conditional probability density for each class. The second family is the distance based approaches [9, 10, 63–65]. However, these two types of estimation present some limits: among them we can mention the need of the *a priori* knowledge on the hypotheses which is not always easy to know, especially, for compound hypotheses. In this section some approaches of these categories are browsed in detail.

3.2.1 Distance-based approaches

The distances-based approaches correspond to models where masses relative to data depend on distances calculated in the feature space. Here, the three most-known models in the literature are presented. One is based on the algorithm of K -Nearest Neighbor (K -NN) [63], the other is based on the clustering method C-means [10], and finally the EVCLUS (Evidential CLUStering) [9] algorithm of proximity data that assigns a BBA to each object from the matrix of dissimilarities between objects. In the rest of this chapter, we use the notation $m(\mathbf{x} \in A)$ that stands for $m(A)$ when there is no ambiguity.

3.2.1.1 The evidential classification algorithms

BBA with a K -NN algorithm In this estimation approach, only the singleton θ_n and the whole frame of discernment Θ are considered. Focal elements and the mass functions are estimated from a learning set $\mathcal{L} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$ for which their corresponding class is known: \mathbf{x}_ℓ is assigned to class θ_{x_ℓ} among $\{\theta_1, \theta_2, \dots, \theta_N\}$. For each instance \mathbf{x} to be classified, the K -NN is used to retain only the closest vectors of \mathbf{x} . Let $\mathcal{N}_K(\mathbf{x})$ be the set of the K -nearest neighbors of \mathbf{x} in \mathcal{L} . This set can be considered to pieces of evidence regarding the class of \mathbf{x} . For each element \mathbf{x}_k in $\mathcal{N}_K(\mathbf{x})$ (\mathbf{x}_k being assigned to class θ_{x_k}), the strength of this evidence

decreases with the distance $d(\mathbf{x}, \mathbf{x}_k)$. The BBAs are then given by the following expression:

$$\begin{cases} m_k(\mathbf{x} \in \theta_{x_k}) = \alpha \varphi_{\theta_{x_k}}(d(\mathbf{x}, \mathbf{x}_k)), \\ m_k(\mathbf{x} \in \Theta) = 1 - \alpha \varphi_{\theta_{x_k}}(d(\mathbf{x}, \mathbf{x}_k)), \end{cases} \quad (3.1)$$

where $0 < \alpha < 1$ is a constant and $d(\mathbf{x}, \mathbf{x}_k)$ being the distance between the vector \mathbf{x}_k and \mathbf{x} . φ_{θ_n} is a decreasing function verifying $\varphi_{\theta_n}(0) = 1$ and $\lim_{d \rightarrow \infty} \varphi_{\theta_n}(d) = 0$. The φ_{θ_n} function might be an exponential function following this form:

$$\varphi_{\theta_n}(d) = \exp(-\gamma_n d^2), \quad (3.2)$$

where γ_n is a positive parameter determined separately for each class $\theta_n \in \{\theta_1, \theta_2, \dots, \theta_N\}$. Typically in DST framework, the combined belief function m is obtained by the application of Dempster combination operator on each sources of evidence (*i.e.*, partial information) m_k .

$$m = \oplus_{k \in [1, \dots, K]} m_k, \quad (3.3)$$

The described method defines the Distance Classifier (DC) [63]. Despite its promising results, this approach has a major shortcoming because it cannot deal with new (exploratory) data. This point may be explained by the cost of this algorithm which is quite high because it has to calculate the Euclidean distance to each of the neighbors, and sort them to find the nearest K . This task has a computational complexity of $O(L \times p)$ for each new BBA, where p is the space dimension.

3.2.1.2 The evidential clustering algorithms

BBA with ECM algorithm In [10], Denœux and Masson propose a new automatic classification method called ECM (Evidential C-Means). Let $\mathcal{L} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$ be a collection of vectors in \mathbb{R}^p describing the L observations. Let K be the desired number of classes. Each cluster is represented by a prototype or a center $\mathbf{v}_k \in \mathbb{R}^p$. Let V denotes a matrix of size $(K \times p)$ composed of the coordinates of the cluster centers such that $V_{k,q}$ is the q th component of the cluster center \mathbf{v}_k . ECM looks for matrices $M = (m_{\ell,k})$ (mass functions matrix of dimension $(L \times K)$ with elements $m_{\ell,k} = m(\mathbf{x}_\ell \in \theta_k)$) and V by minimizing the following objective function:

$$J_{\text{ECM}}(M, V) = \sum_{\ell=1}^L \sum_{\substack{k=1 \\ \theta_k \subseteq \Theta, \theta_k \neq \emptyset}}^K c_k^\alpha m_{\ell,k}^\beta d^2(\mathbf{x}_\ell, \mathbf{v}_k) + \sum_{\ell=1}^L \delta^2 m_{\ell,\emptyset}^\beta, \quad (3.4)$$

subject to the constraint:

$$\sum_{\substack{k=1 \\ \theta_k \subseteq \Theta, \theta_k \neq \emptyset}}^K m_{\ell,k} + m_{\ell,\emptyset} = 1, \quad \forall \ell \in \{1, \dots, L\}, \quad (3.5)$$

where $m_{\ell,\emptyset}$ stands for $m(\mathbf{x}_\ell \in \emptyset)$, δ controls the amount of data considered as outliers, β is a weighting exponent that controls the imprecision of the partition and α is a parameter to control the degree of penalization. The c_k^α coefficient is a penalty factor that prevents from high cardinality class.

This algorithm holds a great importance in processing complex and imprecise data since it allows the allocation of the masses to the different subsets of the frame of discernment. Unfortunately, it has an exponential complexity relative to the number of classes and linear complexity relative to the number of samples.

BBA with EVCLUS algorithm Let us consider two BBAs m_i and m_j regarding the class membership of two observations \mathbf{x}_i and \mathbf{x}_j . The aim of EVCLUS BBA estimation is: the more similar the observations, the lower the degree of conflict between their mass function and the higher plausible that they belong to the same class. As shown in [9], this idea can be explained as follows. Let R_{ij} be the following proposition samples \mathbf{x}_i and \mathbf{x}_j belong to the same class corresponding to the following subset of the Cartesian product $\Theta^2 = \Theta \times \Theta$:

$$R_{ij} = \{(\theta_1, \theta_1), (\theta_2, \theta_2), \dots, (\theta_K, \theta_K)\}.$$

The plausibility $Pl_{i \times j}$ of the proposition R_{ij} can be shown to be equal to:

$$\begin{aligned} Pl_{i \times j}(R_{ij}) &= \sum_{\substack{A \times B \in \Theta^2 \\ (A \times B) \cap R_{ij} \neq \emptyset}} m_{i \times j}(A \times B) \\ &= \sum_{A \cap B \neq \emptyset} m_i(A) m_j(B) \\ &= 1 - \sum_{A \cap B = \emptyset} m_i(A) m_j(B) = 1 - \mathcal{K}_{ij}, \end{aligned}$$

where $m_{i \times j}(A \times B)$ is the BBA that describes ones beliefs regarding the class membership of both samples and \mathcal{K}_{ij} is the degree of conflict between m_i and m_j .

Let us assume that the available data consist of a $L \times L$ dissimilarity matrix $D = (d_{ij})$, EVCLUS looks for $M = (m_1, m_2, \dots, m_L)$ the credal partition of $\mathcal{L} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$ a set of L observations to be classified in Θ by minimizing an stress function inspired from multidimensional scaling (MDS) methods [66] such that the degree of conflict \mathcal{K}_{ij} represents a form of distance between the observations and reflects the dissimilarities d_{ij} . The stress function to be minimized is given by:

$$J_{\text{EVCLUS}}(M, a, b) = \frac{1}{Ct} \sum_{i < j} \frac{(a\mathcal{K}_{ij} + b - d_{ij})^2}{d_{ij}}, \quad (3.6)$$

where a and b are two coefficients, d_{ij} is the dissimilarity between \mathbf{x}_i and \mathbf{x}_j and Ct is a constant defined for normalization as:

$$Ct = \sum_{i < j} d_{ij}.$$

Thus, EVCLUS can be thought of as an iterative optimization, with respect to M , a and b , under the criterion of equation (3.6) to be minimized by using a gradient-based procedure. The major drawback of this algorithm is its computational complexity, and thus, it is limited to data sets of a few thousand elements and less than 20 classes.

3.2.2 Likelihood-based approaches

Among the several probabilistic models that have been proposed in the literature, we present here the Appriou's approach [67] that considers each class k with $k \in \{1, \dots, K\}$ as a particular source of information. The mass is defined through the transfer of the bayesian probability function to the total ignorance and the complementary class. The BBA associated to the hypothesis

θ_k is then defined through the source of information S_k with the following mass:

$$\begin{cases} m_k(\mathbf{x} \in \theta_k) = \alpha_k \frac{Rp(\mathbf{x} \in \theta_k)}{1 + Rp(\mathbf{x} \in \theta_k)}, \\ m_k(\mathbf{x} \in \bar{\theta}_k) = \alpha_k \frac{1}{1 + Rp(\mathbf{x} \in \theta_k)}, \\ m_k(\mathbf{x} \in \Theta) = 1 - \alpha_k, \end{cases} \quad (3.7)$$

where α_k with values in $[0, 1]$ is a discounting factor associated with the reliability of the model to the class θ_k and $R = 1/\max_k p(\mathbf{x} \in \theta_k)$ is a positive normalized coefficient less or equal to 1. The classes $\bar{\theta}_k$ are defined in 2^Θ such as $\theta_k \cap \theta_k^c = \emptyset$. From these K belief functions, each elementary sources are fused by using the orthogonal sum given in equation (3.3), yielding a complete BBA on 2^Θ . In [68] a transfer model is introduced to distribute the initial masses over the compound hypotheses (disjunction of classes).

3.3 New method to build mass functions

To estimate the mass functions, we adopt an essentially geometrical viewpoint by projection of the initial representation space of the data to a two-dimensional space only, using a Kohonen's map. These geometric considerations allow a smart mass belief assignment, not only for simple hypotheses but also for disjunctions and conjunctions of hypotheses. Thus, it can model at the same time ignorance, imprecision, and paradox. In the rest of this section, first we give an overview on Kohonen's map, also called Self Organizing Map (SOM). Then, we present the feature space that is defined to help the estimation of mass functions. The BBA itself is detailed in Subsection 3.3.3.

3.3.1 Overview on Kohonen's map

There exist many versions of the SOM. However, the basic philosophy is very simple and already effective [6]. A SOM defines a mapping from the input feature space (say \mathbb{R}^p) onto a regular array of $M \times N$ nodes (see Figure 3.1) [7].

A reference vector, also called weight vector, $\mathbf{w}(i, j) \in \mathbb{R}^p$ is associated to the node at each position (i, j) with $1 \leq i \leq M$ and $1 \leq j \leq N$. An input vector $\mathbf{x} \in \mathbb{R}^p$ is compared to each $\mathbf{w}(i, j)$. The best match is defined as output of the SOM: thus, the input data \mathbf{x} is mapped onto the SOM at location (i_x, j_x) where $\mathbf{w}(i_x, j_x)$ is the neuron the most similar to \mathbf{x} according to a given metric. The SOM performs a non linear projection of the probability density function $p(\mathbf{x})$ from the high-dimensional input data onto the 2-D array.

In practical applications, the Euclidean distance is usually used to compare \mathbf{x} and $\mathbf{w}(i, j)$ in \mathbb{R}^p , so that $d(\mathbf{x}, \mathbf{w}(i, j)) = \|\mathbf{x} - \mathbf{w}_x\|$. The node that minimizes the distance between \mathbf{x} and $\mathbf{w}(i, j)$ defines the best-matching node (or the so-called winning neuron), and is denoted by the subscript \mathbf{w}_x :

$$d(\mathbf{x}, \mathbf{w}_x) = \|\mathbf{x} - \mathbf{w}_x\| = \min_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \|\mathbf{x} - \mathbf{w}(i, j)\|. \quad (3.8)$$

An optimal mapping would be the one that maps the probability density function $p(\mathbf{x})$ in the most faithful fashion, preserving at least the local structures of $p(\mathbf{x})$.

It can be considered also that the SOM achieves a non-uniform quantization that transforms \mathbf{x} to \mathbf{w}_x by minimizing the given metric. Nevertheless, thanks to the training phase (detailed

below) the neurons \mathbf{w} are located on the map according to their similarity. Then, when considering neurons $\mathbf{w}(i, j)$ located not *too far* from the winning neuron \mathbf{w}_x , the distance in \mathbb{R}^p between \mathbf{x} and $\mathbf{w}(i, j)$ is not dramatically different from the one between \mathbf{x} and \mathbf{w}_x . That means that in the neighborhood of \mathbf{w}_x on the map (*i.e.*, with closed location i and j), are located the winning neurons of the neighbors of \mathbf{x} in \mathbb{R}^p . Hence, a class in the feature space \mathbb{R}^p is projected into the map at the same area, remaining homogeneous. Moreover, whatever the initial shape of the class in the \mathbb{R}^p feature space, the projected class is highly likely to be of isotropic shape in the map.

3.3.1.1 Training Phase

The learning phase may be thought of as a classification phase, such as a K-means classification algorithm. Neurons are first sampled (in \mathbb{R}^p) randomly and then, iteratively in a similar way as in the K-means algorithm, they are modified to fit a training sample $\mathcal{L} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$. One of the main differences from the K-means algorithm is that the nodes which are close to the best-matching node in the map will learn from the same input \mathbf{x} also.

While the initial values of the \mathbf{w} may be set randomly, they will converge to a stable value at the end of the training process, by using equation (3.9):

$$\mathbf{w}(t+1) = \mathbf{w}(t) + h_{\mathbf{w}, \mathbf{w}_x}(t) (\mathbf{x} - \mathbf{w}(t)), \quad (3.9)$$

where t is the iteration index.

During one iteration of the training phase, every input \mathbf{x}_ℓ , taken from the training set \mathcal{L} , is processed according to equation (3.9). $h_{\mathbf{w}, \mathbf{w}_x}(t)$ is called *neighborhood kernel*: it is a function defined over the lattice points of Kohonen's map, usually $h_{\mathbf{w}, \mathbf{w}_x}(t) = h(d(\mathbf{w}, \mathbf{w}_x), t)$ where $d(\mathbf{w}, \mathbf{w}_x)$ stands for the distance between the location of \mathbf{w} and \mathbf{w}_x on the map. While increasing $d(\mathbf{w}, \mathbf{w}_x)$, or increasing t , $h_{\mathbf{w}, \mathbf{w}_x}(t)$ decreases monotonically to 0. The average width and the form of $h_{\mathbf{w}, \mathbf{w}_x}(t)$, defines the stiffness of the elastic surface to be fitted to the data set. Let

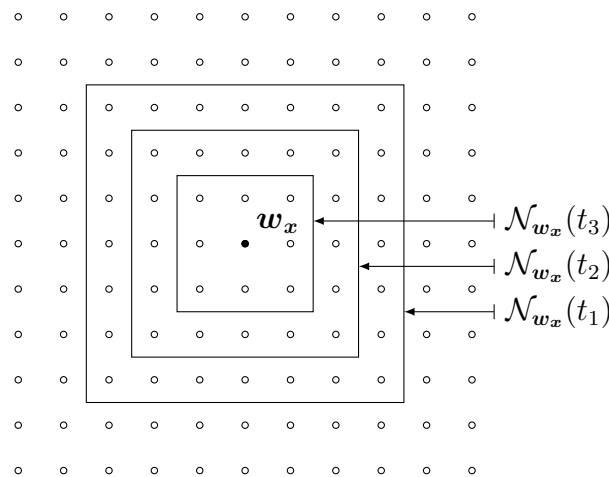


Figure 3.1: Schematic of a 11×11 Kohonen's map. Several topological neighborhood $\mathcal{N}_{\mathbf{w}_x}(t_i)$ of the winning neuron \mathbf{w}_x are drawn. The size is decreasing with the number of iterations ($t_1 < t_2 < t_3$) during the training phase, according to (3.10).

their index in the neighborhood of w_x be denoted by the set $\mathcal{N}_{w_x}(t)$ (see Figure 3.1).

$$h_{w,w_x}(t) = \begin{cases} \alpha(t) \exp\left(-\frac{d(w,w_x)}{2\sigma^2(t)}\right) & \text{if } w \in \mathcal{N}_{w_x}(t), \\ 0 & \text{if } w \notin \mathcal{N}_{w_x}(t). \end{cases} \quad (3.10)$$

The value of $\alpha(t)$ is then identified with a *learning-rate factor* ($0 < \alpha(t) < 1$). Both $\alpha(t)$ and the support of $\mathcal{N}_{w_x}(t)$ are usually decreasing monotonically in time (during the ordering process). $\sigma(t)$ is the width of the neighborhood that corresponds to the radius of the neighborhood of w_x in $\mathcal{N}_{w_x}(t)$. In practice, $\alpha(t)$ and $\sigma(t)$ vanish with time. Typically, linearly decreasing functions are defined such as: $\alpha(t) = \alpha_0 \times \frac{T-t}{T}$ and $\sigma(t) = \sigma_0 \times \frac{T-t}{T}$, where T stands for the number of iterations.

3.3.1.2 Projection

Once the SOM has been trained, it acts as a similar way to as a set of clusters yielded by a K-means algorithms. Here, the index of each w is defined in 2-D, and each w located in the same area of the map has similar value in \mathbb{R}^p .

For each sample x to be processed, it is projected on the map by using equation (3.8) to find its corresponding neuron w_x . The SOM may be considered to as a nonuniform quantization of the feature space [69]. This nonuniform quantization performed by Kohonen's map has the advantage to make the class definition on the map (*i.e.*, through the quantization index) more isotropic than in \mathbb{R}^p . Then, the map may be considered to as an approximation in $\{1, \dots, M\} \times \{1, \dots, N\}$ of the initial manifold of \mathbb{R}^p , while preserving its topology.

3.3.2 Feature space for smart basic belief assignment

The proposed smart BBA intends to evaluate the mass of each class in 2^Θ or D^Θ according to the topology of the observed manifold. Then, two sets of data may be handled (see Figure 3.2): on the first hand the initial observations x and class centers $\{C_1, C_2, \dots, C_K\}$ in \mathbb{R}^p and, on the other hand the so-called *winning neurons* w_x and the projected class centers w_{C_k} . It is worth noting that there is no link between the training of the classifier that defines $\{C_1, C_2, \dots, C_K\}$ in \mathbb{R}^p and the SOM that defines the set of neurons $w(i, j)$ in \mathbb{R}^p , $1 \leq i \leq M, 1 \leq j \leq N$, except that both are trained by using the same training samples (or a part of those).

w_x is determined following equation (3.8) and for $k \in \{1, \dots, K\}$, w_{C_k} is determined in a similar way as stated in the following equation:

$$w_{C_k} = \arg \min_{\substack{w(i,j) \\ 1 \leq i \leq M, 1 \leq j \leq N}} \|C_k - w(i, j)\|. \quad (3.11)$$

Then, Kohonen's map can be used to build easily BBA and to balance between conjunction and disjunction when considering relative distance of an observation to the map. Moreover, the use of Kohonen's map simplifies the evaluation of the masses since operations on the maps require calculation on index only, while operations on the feature space (in \mathbb{R}^p) may be much more complex (when dealing with stochastic divergence for instance). So two kinds of distances will be considered and their related difference will induce uncertainty.

1. $d_{\mathbb{R}^p}(\cdot, \cdot)$ which is the distance in \mathbb{R}^p . It can be defined through the Euclidean norm $\mathcal{L}^2(\mathbb{R}^p)$ but also through a spectral point of view such as the spectral angle mapper or the spectral information divergence [70]. It may also be based on the Kullback-Leibler divergence or the mutual information when dealing with Synthetic Aperture Radar (SAR) [8].
2. $d_{\text{map}}(\cdot, \cdot)$ which is the distance along Kohonen's map. It is mainly based on the Euclidean norm and uses the index that locates the two vectors on the map: $d_{\text{map}}(\mathbf{w}_1, \mathbf{w}_2) = \sqrt{(m_1 - m_2)^2 + (n_1 - n_2)^2}$ if \mathbf{w}_1 (resp. \mathbf{w}_2) is located at position (m_1, n_1) (resp. (m_2, n_2)) on the map.

3.3.3 Mass function construction

This section details our proposed method for building a BBA by using Kohonen's map and an initial classifier on \mathbb{R}^p .

3.3.3.1 Mass of simple hypotheses

The definition of masses of focal elements could be based on the distance on the feature space. Nevertheless, an appropriated definition should take into account the variance of the classes to weight each of them, as it is the case in a likelihood point of view. This weighting is already performed by the projection onto Kohonen's map so that, the mass of focal class is defined as:

$$\begin{cases} m(\mathbf{x} \in \theta_k) \simeq 1 & \text{if } \mathbf{w}_x = \mathbf{w}_{C_k}, \\ m(\mathbf{x} \in \theta_k) \simeq \frac{d_{\text{map}}(\mathbf{w}_x, \mathbf{w}_{C_k})^{-1}}{\sum_{\ell=1}^K d_{\text{map}}(\mathbf{w}_x, \mathbf{w}_{C_\ell})^{-1}}, & \text{otherwise,} \end{cases} \quad (3.12)$$

where $k = 1, 2, \dots, K$, \mathbf{w}_{C_k} is the projected class, \mathbf{w}_x is the winning neurons.

According to equation (3.12), we consider that the more the distance $d_{\text{map}}(\mathbf{w}_x, \mathbf{w}_{C_k})$ (relatively to the other distances between \mathbf{x} and C_ℓ on the map) the less the mass $m(\mathbf{x} \in \theta_k)$.

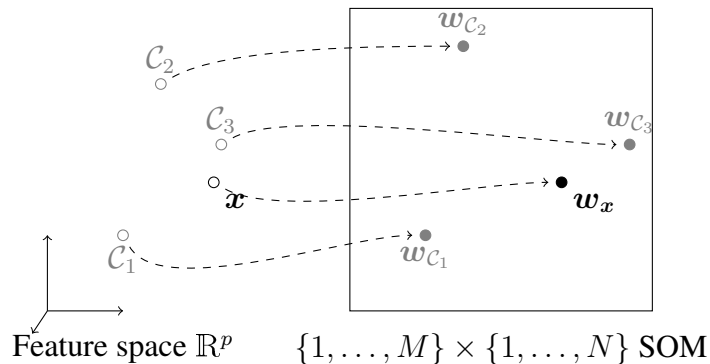


Figure 3.2: Observations in the feature space and their projections into Kohonen's map. Note that the neurons \mathbf{w}_x and \mathbf{w}_{C_k} can be located on the map through their location index (m, n) or in \mathbb{R}^p with their p -component value.

3.3.3.2 Mass of the full ignorance

From the feature space, we consider that the mass evaluation of an observation falls into ignorance if its distance to the map is much more important than the distance of its related class center to the map. Then, it can be expressed as follows:

$$m(\mathbf{x} \in \Theta) \simeq 1 - \min \left(\frac{d_{\mathbb{R}^p}(\mathbf{x}, \mathbf{w}_x)}{d_{\mathbb{R}^p}(\mathcal{C}_x, \mathbf{w}_{\mathcal{C}_x})}, \frac{d_{\mathbb{R}^p}(\mathcal{C}_x, \mathbf{w}_{\mathcal{C}_x})}{d_{\mathbb{R}^p}(\mathbf{x}, \mathbf{w}_x)} \right), \quad (3.13)$$

where \mathcal{C}_x is the class center of \mathbf{x} , $\mathbf{w}_{\mathcal{C}_x}$ is its projection on the map.

3.3.3.3 Mass of the conjunction between two classes

In the set D^Θ , the conjunction between two classes may be defined into the feature space as the space in-between the two classes. But, one has to account for the variance of each classes that increases the complexity of this measure. Once again, it is much more convenient to define the $\theta_k \cap \theta_\ell$ mass directly into Kohonen's map, as:

$$m(\mathbf{x} \in \theta_k \cap \theta_\ell) \simeq e^{-\gamma(z-1)^2}, \quad (3.14)$$

with

$$z = d_{\text{map}}(\mathbf{w}_x, \frac{\mathbf{w}_{\mathcal{C}_k} + \mathbf{w}_{\mathcal{C}_\ell}}{2}) \quad 0 < k, \ell \leq K, \ell \neq k.$$

Equation (3.14) stipulates that the value of $m(\mathbf{x} \in \theta_k \cap \theta_\ell)$ becomes maximal when \mathbf{x} reaches the middle of $[\mathbf{w}_{\mathcal{C}_k}, \mathbf{w}_{\mathcal{C}_\ell}]$ segment. Equation (3.14) yields a value of $m(\mathbf{x} \in \theta_k \cap \theta_\ell)$ closed to 1 in the middle. Moreover, $m(\mathbf{x} \in \theta_k \cap \theta_\ell)$ vanishes when \mathbf{x} is far away from the $[\mathbf{w}_{\mathcal{C}_k}, \mathbf{w}_{\mathcal{C}_\ell}]$ segment. The γ parameter tunes this vanishing behavior. For example, if we want equation (3.14) be over $\frac{1}{2}$ between the 1st and the 3rd quartile of $[\mathbf{w}_{\mathcal{C}_k}, \mathbf{w}_{\mathcal{C}_\ell}]$ segment, then γ should be equal to $2\sqrt{2}$. For a smaller domain around the median of $[\mathbf{w}_{\mathcal{C}_k}, \mathbf{w}_{\mathcal{C}_\ell}]$ segment, γ should be greater (see Figure 3.4).

This conjunctive mass estimation does not apply in the classical Dempster-Shafer framework (*i.e.*, when working in 2^Θ only assuming Shafer's model of the frame Θ).

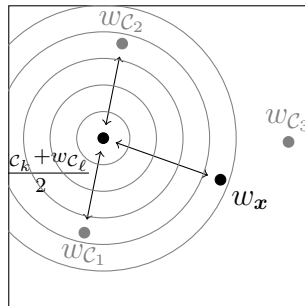
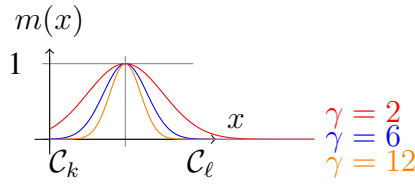


Figure 3.3: Simple case of conjunction between two class in the map.


 Figure 3.4: Behavior of $m(\mathbf{x} \in \theta_k \cap \theta_\ell)$ with γ , according to equation (3.14).

3.3.3.4 Mass of disjunction between two classes

The ignorance in the decision-making between two classes \mathcal{C}_k and \mathcal{C}_ℓ may be considered as the dual of equation (3.14), but here by considering distances in the feature space. When a sample \mathbf{x} is not *too far* from class \mathcal{C}_k or \mathcal{C}_ℓ , it is not *too difficult* to decide if it has to be associated to the class k or ℓ . But if \mathbf{x} is *far* from \mathcal{C}_k and \mathcal{C}_ℓ , it comes the disjunction as related in Figure 3.5. That corresponds to a context where the distances between \mathbf{x} and the classes are of the same scale: $d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_k) \approx d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_\ell)$. But such criteria is not enough since it includes also the case where \mathbf{x} is located in-between \mathcal{C}_k and \mathcal{C}_ℓ . So it has to be weighted by the distance between the two classes $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)$. If $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell) \ll d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_k)$ and $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell) \ll d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_\ell)$, \mathbf{x} falls in the disjunctive case since \mathbf{x} is considered *far* to \mathcal{C}_k and \mathcal{C}_ℓ . Then, the criteria defined in equation (3.15) is based on the ratio between $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)$ and $d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_\ell)$.

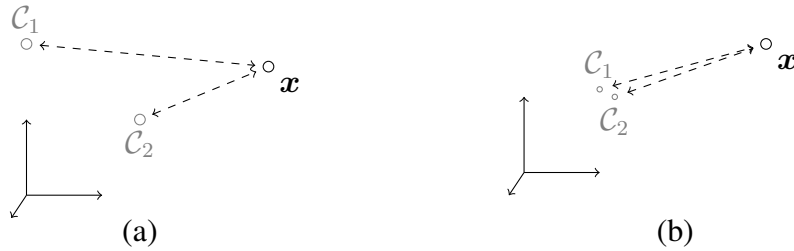


Figure 3.5: Disjunction between two class: (a) non ambiguous case, (b) ambiguous case.

Then, the mass of the disjunction $\theta_k \cup \theta_\ell$ is modeled by:

$$m(\mathbf{x} \in \theta_k \cup \theta_\ell) \simeq 1 - \tanh(\beta z), \quad (3.15)$$

with

$$z = \frac{d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)}{d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_\ell)} \quad 0 < k, \ell \leq K, k \neq \ell.$$

Here, the β parameter stands for the level of ambiguity. When \mathbf{x} is close, in \mathbb{R}^p , to the segment $[\mathcal{C}_k, \mathcal{C}_\ell]$, $d(\mathcal{C}_k, \mathcal{C}_\ell) \simeq d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\mathbf{x}, \mathcal{C}_\ell)$ so that z is close to 1, and $m(\mathbf{x} \in \theta_k \cup \theta_\ell)$ has to vanish. Then, the areas where equation (3.15) vanishes are shown on curves of Figure 3.6. The more the β , the less the ambiguous mass.

3.3.3.5 Conjunction and disjunction for more than 2 classes

This construction that takes into consideration the ratio of distance between 2 classes or the distance to the middle of 2 classes can be extended to more than 2 classes. For instance, equation (3.14) can be based on the centroid of more than 2 class. Equation (3.15) can be generalized

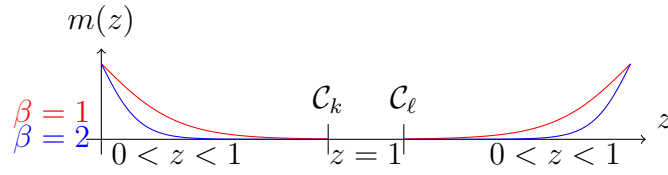


Figure 3.6: Shape of equation (3.15) for some value of β .

by the composition of one against one class from a set of K classes, divided by the sum of distance of \mathbf{x} to each of the K class centers. Nevertheless, this method of construction has not been deeper investigated since those compositions should not have significative impact on the fusion or the classification results.

3.3.3.6 Normalized BBA

The complete BBA has to respect constraint of equation (2.1) in DST and of equation (2.37) in DSMT so that is it necessary to apply a normalization step to the unnormalized BBA obtained by separately calculating the belief masses on simple and compound hypotheses, presented in sections 3.3.3.1–3.3.3.4.

3.3.3.7 Determination of parameters β and γ

The determination of the parameters β and γ can be found automatically by minimizing the following constraints, defined in [63, 71]:

$$E = \sum_{i=1}^{N_S} \sum_{n=1}^N (BetP(\mathbf{x}_i \in \theta_n) - \Upsilon(\mathbf{x}_i \in \theta_n))^2$$

where N_S is the number of samples, $BetP(\mathbf{x}_i \in \theta_n)$ stands for the pignistic probability of \mathbf{x}_i (vector to classify) according to the simple hypothesis θ_n and $\Upsilon(\mathbf{x}_i \in \theta_n)$ is a function that is equal to 1 if the sample \mathbf{x}_i does belong to the simple hypothesis θ_n (as stated *a priori* from the learning base), and 0 otherwise.

3.4 Simple simulation

This section presents a simulation dedicated to a simple 4-class problem. Although the SOM is more appropriated to be used to perform a non linear projection from \mathbb{R}^p to $\{1, \dots, M\} \times \{1, \dots, N\}$ with $p > 2$, this naive case of study has been defined in \mathbb{R}^2 for a better visualisation. Figure 3.7 shows, with black circles, a data set in \mathbb{R}^2 that is decomposed into 4 clusters. Each of those clusters have a Gaussian shape with different covariance matrices.

The classification yielded by a K-means gives 4 clusters \mathcal{C}_1 to \mathcal{C}_4 , which appear in Figure 3.7 (green bullets). Their locations are approximatively: \mathcal{C}_1 : (0.18, 0.18), \mathcal{C}_2 : (0.6, 0.18), \mathcal{C}_3 : (0.25, 0.4) and \mathcal{C}_4 : (0.45, 0.5) which corresponds to the center of each Gaussian sampling.

When performing a Kohonen's map of size 8×8 , it yields the map characterized in Figure 3.7 (red bullets). As drawn in the \mathbb{R}^2 feature space, the map is seen dramatically deformed according to the density of data samples. The more the density of samples (in black circles), the

more the density of the neurons (in red bullets), which is a characteristic of a non-uniform quantization. The location of the red bullets corresponds of the value of the weight of the neurons in \mathbb{R}^2 .

Then, when a sample (black circle) is *projected* into the map, it is associated to its winning neuron according to equation (3.8), *i.e.*, associated to the closest red bullet according to the Euclidean distance in \mathbb{R}^2 . Figure 3.8-(a) shows the same figure as Figure 3.7 highlighting some areas. The ellipses in blue highlight the areas between the different clusters, while the red ellipse at the top right of the figure points out an outlier. On Figure 3.8-(b) is shown Kohonen's map into its natural geometry in $\{1, \dots, N\} \times \{1, \dots, M\}$: the distance between neurons corresponds to the distance along the edges of the map, *i.e.*, considering the indexes. The green bullets in this map shows the winning neurons w_{C_k} of the class centers C_k . The neurons shown in blue correspond to the neurons rounded in Figure 3.8-(a)(blue ellipses). Those neurons are located

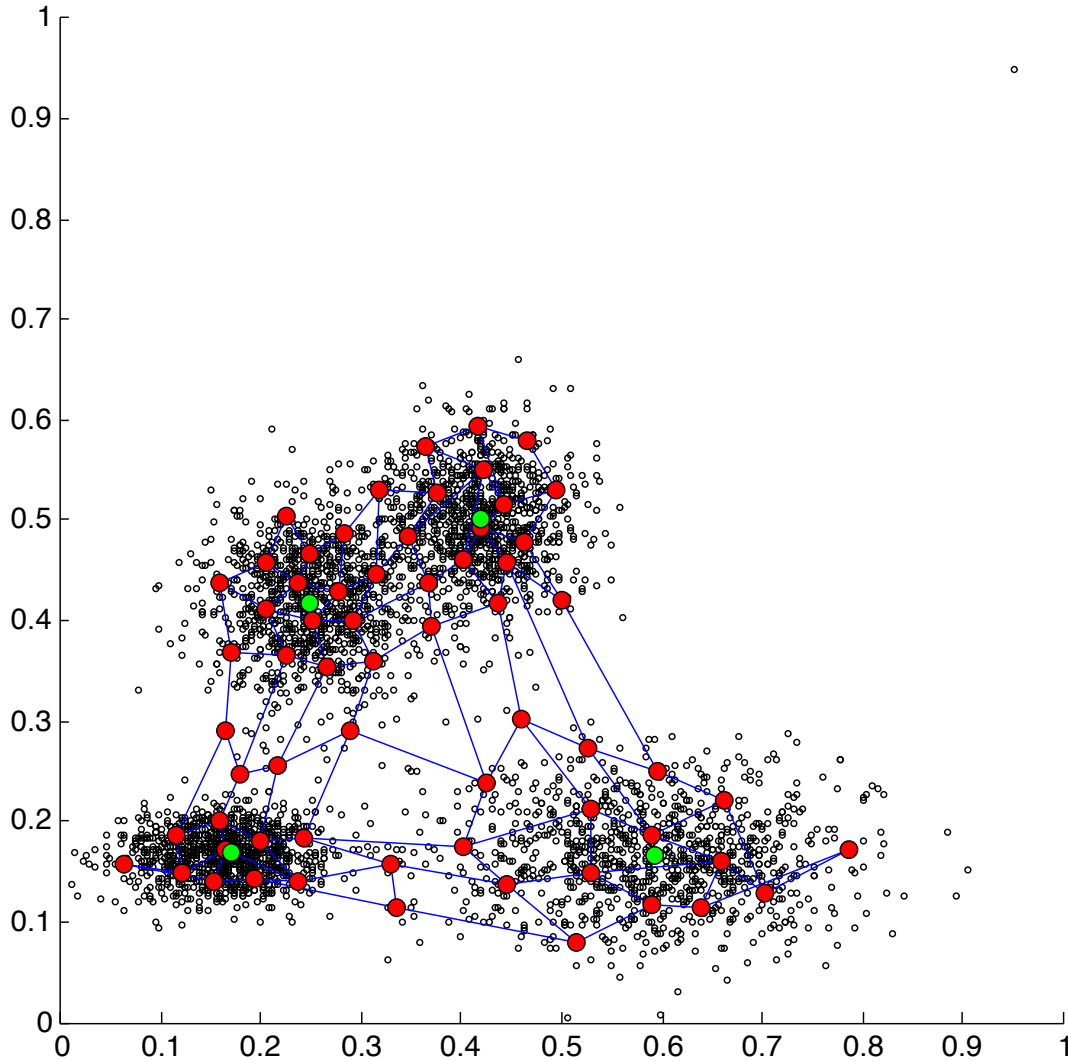


Figure 3.7: Simple simulation of a four-class manifold with an outlier. Black circles: samples of the data set in \mathbb{R}^2 . Red bullets: locations of the neurons of Kohonen's map. Blue lines: SOM projected in the feature space. Green bullets: K-means class centers.

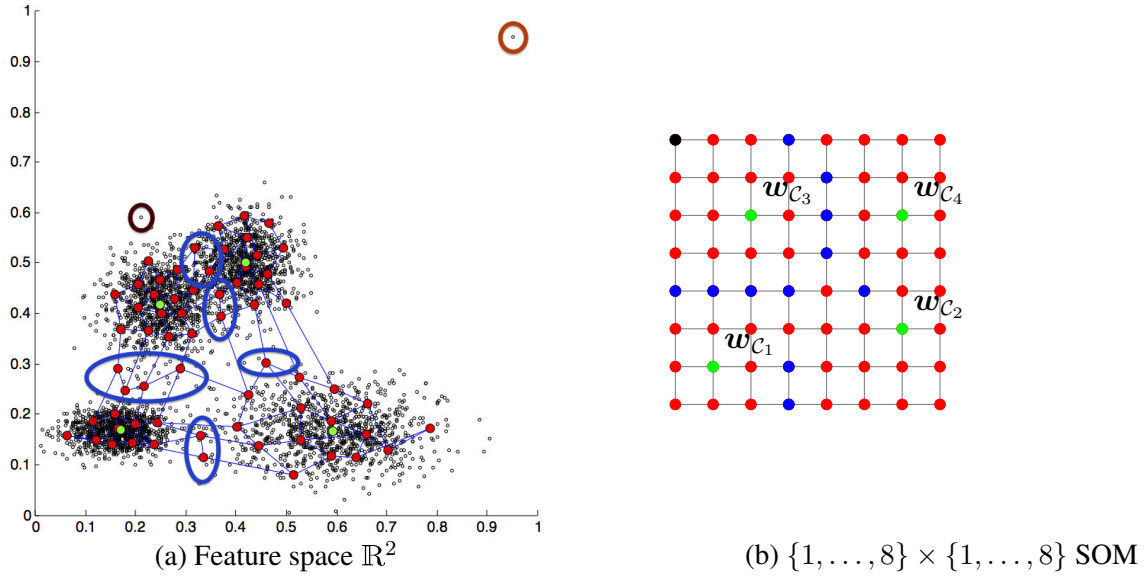


Figure 3.8: Simple simulation of Figure 3.7 with its equivalent in the SOM geometry. Green bullets on the map correspond to the winning neurons w_{C_k} of the class centers C_k , blue bullets, on the map in (b), correspond to the location of the neurons that are located in between classes in \mathbb{R}^2 . The black neuron at the top left of the map corresponds to the winning neuron of the sample rounded with a brown ellipse at location (0.2, 0.6).

between classes in \mathbb{R}^2 and also between the corresponding class centers w_{C_k} and w_{C_ℓ} . This point illustrates the topological preservation of Kohonen's map.

Let us focus on the sample at location (0.2, 0.6), which is rounded with a brown ellipse. A first look at Figure 3.8-(a) points out that this sample is located very near class C_3 but a little bit outside the main concentration of the data set. The winning neuron associated to this sample is drawn in Figure 3.8-(b)(black bullet), at the top left of the map (with index location (1, 8)). It is clear that this neurons is closed to w_{C_3} and far from the winning neurons of the other classes. Then, the second maximum of the BBA reaches the mass $m(\mathbf{x} \in \theta_3 \cap \theta_4)$ (with a value of 0.1135) and the third is devoted to $m(\mathbf{x} \in \Theta)$ (at 0.1005). Considering equation (3.15) Following equation (3.13), it appears that $d_{\mathbb{R}^p}(\mathbf{x}, \mathbf{w}_x)$ is of significant value in comparison of $d_{\mathbb{R}^p}(C_3, \mathbf{w}_{C_3})$ so that $m(\mathbf{x} \in \Theta)$ has also a significant value.

Let us focus on the outlier located at (0.95, 0.95) in Figure 3.7 (top right). This sample is located in Figure 3.8-(a)(top right). It is far from the rest of the data set and also far from Kohonen's map. Its winning neuron is located at position (8, 8) (*i.e.*, at the top right of the map in Figure 3.8-(b)). Since this neuron is closed to w_{C_4} it is expected that the mass $m(\mathbf{x} \in \theta_4)$ be significant. It is the case with a value of 0.1798. Nevertheless, the maximum value of the BBA is reached with $m(\mathbf{x} \in \Theta)$ with 0.2660 which underlines the outlier behavior of this sample. The resulting BBA is very informative because the rest of the masses vanish below 0.09.

Let us focus now in a sample located at position (0.2, 0.3) in Figure 3.8-(a). This point is in the middle of two classes C_1 and C_3 . A little bit closer to class C_1 . Its winning neuron falls in Figure 3.8-(b) (blue bullets) at location (3, 4). Then the mass of $m(\mathbf{x} \in \theta_1 \cap \theta_3)$ traps a significant value as high as 0.1313. Nevertheless, the second highest value of this BBA is reached by $m(\mathbf{x} \in \theta_1 \cap \theta_4)$ with 0.1102. The fact is that, considering the location of the

winning neuron in Kohonen’s map, it is near to the middle of w_{c_1} , w_{c_2} and w_{c_4} . The third maximum falls to $m(\mathbf{x} \in \theta_2 \cup \theta_4)$ (value 0.0831).

This simple example shows that the aim of this BBA modeling technique that induces simple consideration on the distance from samples to clusters in the feature space \mathbb{R}^p and in Kohonen’s map $\{1, \dots, N\} \times \{1, \dots, M\}$.

3.5 Experiments on benchmark data set

In order to highlight some advantages and possible drawbacks of the proposed SOM-based BBA modeling, the performance of the SOM-based BBA is compared to EVCLUS and ECM ones by using data set provided by the University of California - Irvine (UCI) Machine Learning Repository. Seven data sets out of 270 have been taken into consideration with various amount of features (that corresponds to the feature space dimension \mathbb{R}^p) and number of classes (from 2 to 7) as detailed in Table 3.1.

Table 3.1: Characteristics of the UCI data sets used for comparison.

Data set	Features	classes	samples
Banknote authentication	4	2	1372
Pima Indians Diabetes	8	2	768
Seeds	7	3	210
Wine	13	3	170
Statlog (Landsat Satellite)	36	6	6435
Statlog (Image Segmentation)	19	7	2130
Synthetic control chart time series	60	6	600

In this section, the experimental results are based on the classical Dempster-Shafer framework (*i.e.*, we work with 2^Θ only). Indeed, ECM, EVCLUS are only working in this framework.

It is worth noting that the Matlab programs of ECM and EVCLUS have been downloaded from the official webpage page of Thierry Denœux for those experiments¹. Most of the internal parameters have been let to their default value. The distance δ to the empty set has been changed to 100 in ECM and the regularization parameter has been changed to 0.5 in EVCLUS. The number of clusters in ECM and EVCLUS has been fixed according to Table 3.1, depending on the data set.

Kohonen’s map has been trained with the following parameters: a size of 20×20 neurons (except for Seeds and Wine a size of 10×10 neurons), trained with 200 iterations. An initial neighborhood size $\mathcal{N}_w(t_0)$ of 10 neurons and a learning rate $\alpha(t_0)$ of 0.9. These values were carefully selected in order to guarantee convergence of the map with appropriate number of

¹Thierry Denœux’s webpage is available at <https://www.hds.utc.fr/~tdenoex/dokuwiki/en/software>.

neurons to well balance the tradeoff between quantization error and manifold approximation, so as to improve results. The quantization error through the Root Mean Squared Error (RMSE) is used here as criterion to evaluate the quality of Kohonen convergence

$$\text{EQM} = \left(\frac{1}{N_S} \sum_{\ell=1}^{N_S} \|\mathbf{x}_\ell - \mathbf{w}_{x_\ell}\|^2 \right)^{\frac{1}{2}}.$$

Table 3.2: Classification results of SOM-based BBA in 2^Θ for different value of β .

Data set	$\beta = 1$	$\beta = 2$	$\beta = 6$
Seeds	87.6190%	90.9524%	89.0476 %
Wine	71.1765%	73.5294 %	71.7647%

In this section, the values of the parameter β has been selected based on the results shown in Table 3.2. In this experiment, $\beta = 2$ yields the best classifications results. Also, it can be noticed that the proposed method is not so sensitive to the value of β .

Table 3.3: Classification results in 2^Θ of EVCLUS, ECM and SOM-based BBA with decision by the maximum of pignistic probability.

Data set	Banknote authentication	Pima Indians Diabetes	Seeds	Wine	Statlog (Landsat Satellite)	Statlog (Image Segmentation)	Synthetic control chart time series
EVCLUS	843	475	157	103	3027	895	384
	61.44 %	61.84 %	74.76 %	60.58 %	47.03 %	42.01 %	64.0 %
	1172.2sec	181.7sec	34.3sec	6.7 sec	5857 sec	3657 sec	370 sec
ECM	848	506	189	126	4480	1282	453
	61.80 %	65.88 %	90.0 %	74.11 %	69.62 %	55.49 %	72.5 %
	3.4sec	3.2sec	0.3sec	0.9sec	480sec	161sec	6.9sec
SOM-based	1090	549	191	125	4456	1431	501
	79.44 %	71.48 %	90.95 %	73.52 %	69.24 %	67.18 %	83.5 %
	8.6sec	6.7sec	5.8sec	5.9sec	163sec	84sec	8.0sec

It appears that the SOM-based BBA yields most of the time the highest classifications results (put in boldface in the results of Table 3.3). In each row of Table 3.3, the first line corresponds to the number of correctly classified samples, the second line corresponds to the proportion of samples correctly classified, and the last line shows the computation time. It is worth noting that when ECM performs better, the SOM-based approach is close to the best accuracy (73.52 % versus 74.11 % for the benefit of ECM with the Wine database, and 69.24 % versus 69.62 % with the Statlog Landsat satellite images database). Equivalent results prove that SOM-based BBA is just a simplified (*i.e.*, quantized) version of the feature space ECM work with. Better

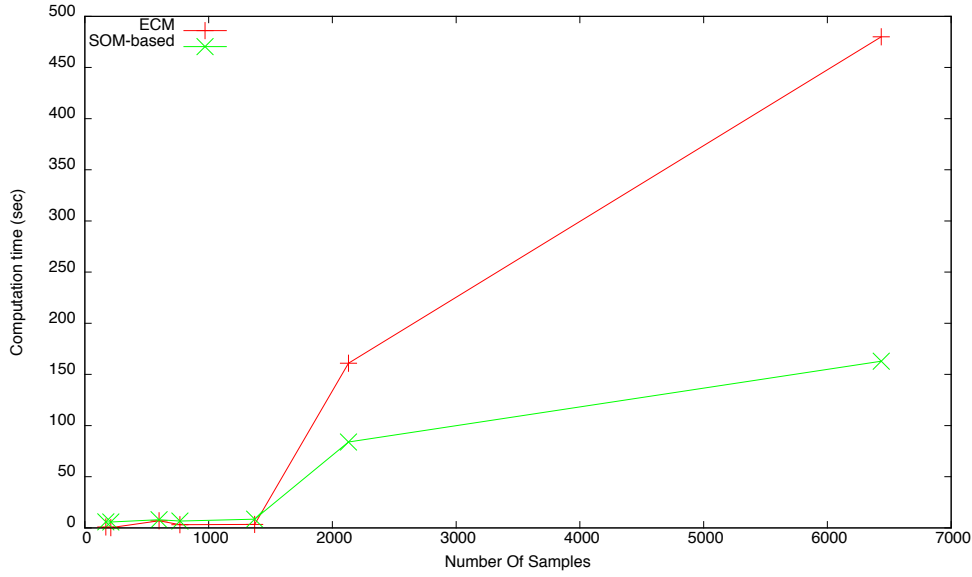


Figure 3.9: Computation time depending on the feature space dimension. SOM-based approach is more appropriated for processing large amount of data than ECM.

results are due to the fact that distances on the map (in 2D) are more appropriated for complex (or non isotropic) class (in pD). EVCLUS is always below. It seems that the performance ranking between ECM and SOM-based BBA is not depending on the feature space dimension nor the number of classes since the Wine and Statlog Landsat satellite image data bases are very different to each other. Since the SOM-based approach considers a projected feature space of dimension 2, it may induce on those cases a too coarse approximation of the manifold in comparison to ECM. Nevertheless, it is worth noting that the benefit in using a SOM-based approach for BBA is related to the number of samples to be handled. Figure 3.9 shows that the more the number of sample the fastest the SOM-based approach in comparison to the ECM while yielding the same level of accuracy. Then the SOM-based approach appears to be a valuable alternative to handle large data set such as real images for classification purpose. In fact, distance in \mathbb{R}^p is more computational demanding than in \mathbb{R}^2 . Indeed, the form of the class in the SOM is more isotropic, so that no consideration on the shape of the manifold is to be considered. On the contrary, ECM has to care of the standard deviation of the classes to build the mass distribution. Then the SOM-based approach appears to be a valuable alternative to handle large data set such as real images for classification purpose.

3.6 Experiments on a real satellite image

The proposed methodology is now applied on a SPOT image ($1318 \times 2359 = 3$ Mega pixels) taken in 2000 for classification purpose. From the variety of objects constituting this image, five clusters may be distinguished: Covered Fields (CF) light-red area, Bare soil (BS) red area, Wooded Area (WA) dark-red area, Water or Wet area (WWA) green area and Bare Soil and Wet Area (BSWA) bright-green area (see Figure 3.10). Those five classes will constitute our frame of discernment $\Theta = \{CF, BS, WA, WWA, BSWA\}$. This 3-band multispectral image

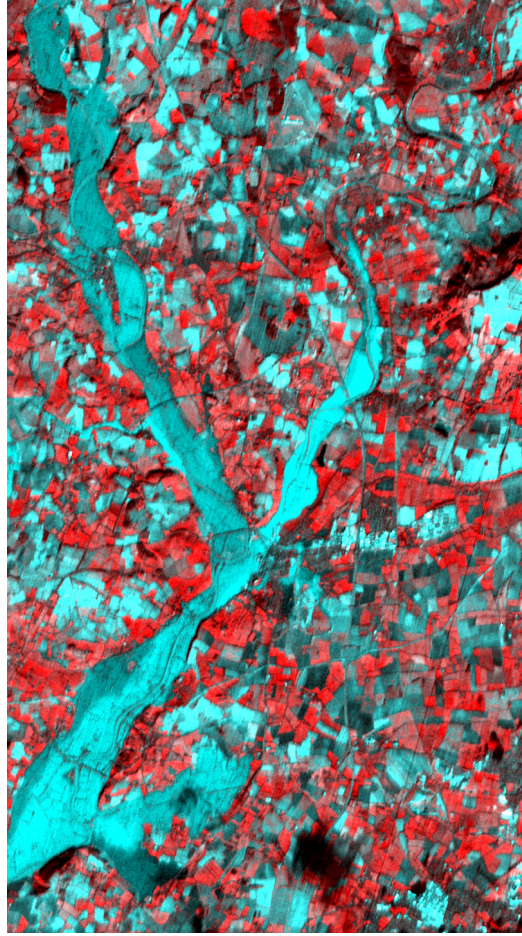


Figure 3.10: False color composite of the SPOT image. ©CNES.

represents a single source of information in \mathbb{R}^3 , so that there is no fusion process within the components of each pixel for BBA (except DC which uses equation (3.3) to perform a fusion rule class by class). In this experiment, DC, ECM and the SOM-based methods are tested. Kohonen's map has been trained with the same parameters as in Section 3.5.

3.6.1 The classification results in 2^Θ

In order to generate mass function on the disjunction of hypotheses in DC, Dempster's combination rule given by equation (3.3) has been replaced by the disjunctive rule given by equation (2.15). Figure 3.11 shows the classification of the original image with DC approach and the proposed approach by using the criterion of the maximum of pignistic probability for decision-making on simple hypotheses (classes). Figure 3.12 shows the classification results all over simple classes and all disjunctions of classes. The performance of DC and SOM-based classifiers is shown through the confusion matrices form in Table 3.5 and Table 3.6, respectively. The test has been done over 16692 pixels where 3273 represent Covered Fields, 2273 Wooded Area, 3013 Bare Soil, 6005 Water or Wet area and 10 Bare Soil and Wet Area. The legend (colors of decision classes in the images classification), is given in Table 3.4.

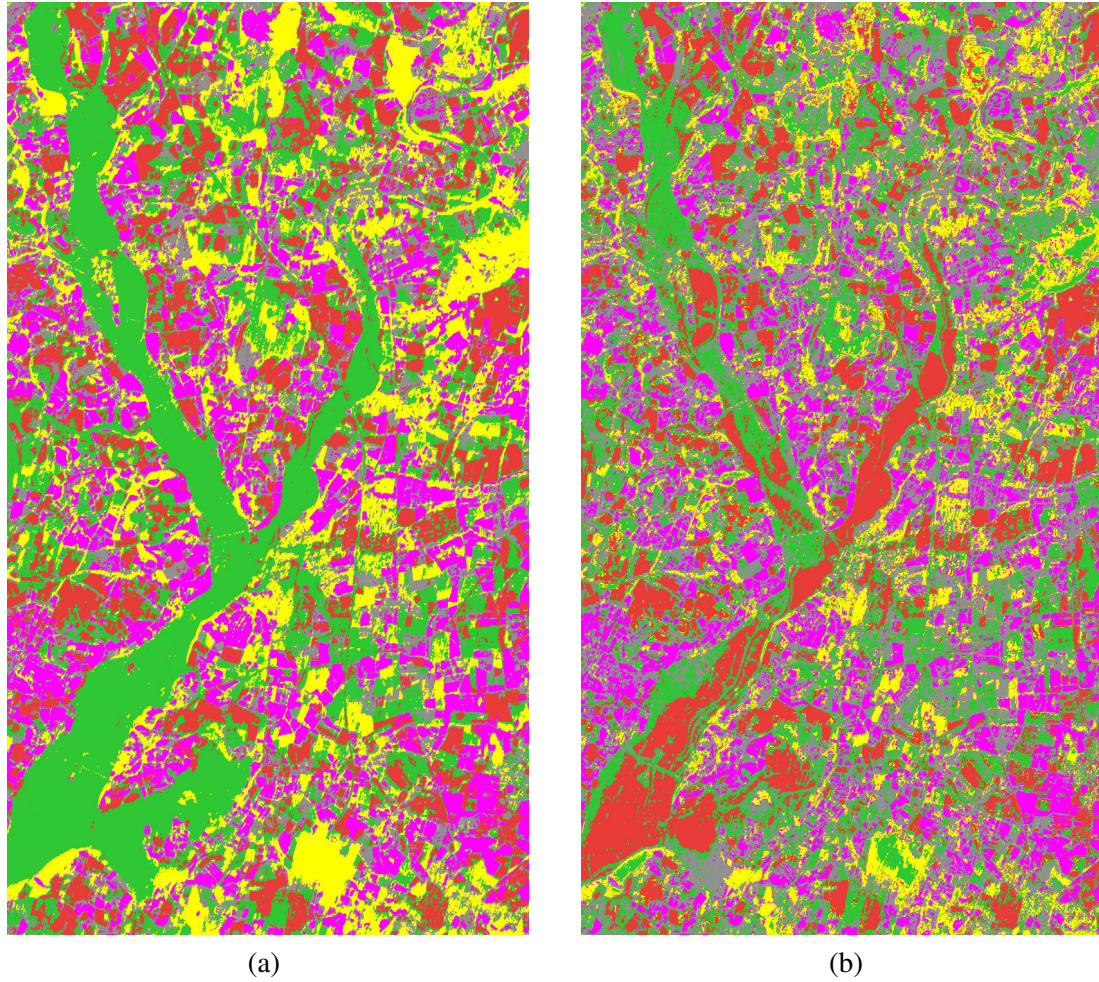


Figure 3.11: Classification results in 2^Θ with decision by maximum of pignistic probability over all simples hypotheses: (a) SOM-based BBA. (b) DC results.

Table 3.4: DST legend used on classification results of Figure 3.12.

■ WWA	■ BSWA \cup BS	■ BSWA \cup CF
■ BSWA	■ BSWA \cup WA	■ BS \cup WWA
■ BS	■ BS \cup WA	■ BS \cup CF
■ WA	■ WWA \cup WA	■ WWA \cup CF
■ CF	■ BSWA \cup WWA	■ WA \cup CF

As Table 3.6 shows, the SOM-based approach presents promising results. Indeed by comparing our approach to the DC approach (see Table 3.5), it can be noticed that class detection has been improved. In Figure 3.11-(a), the river is well discriminated in comparison to other classes while in Figure 3.11-(b) a great conflict appears when those classes WW and BSW have to be discriminated. Figure 3.12 demonstrates that our approach reduces the number of decision class (8 classes), whereas DC approach yields multiple classes. For example the whole river is

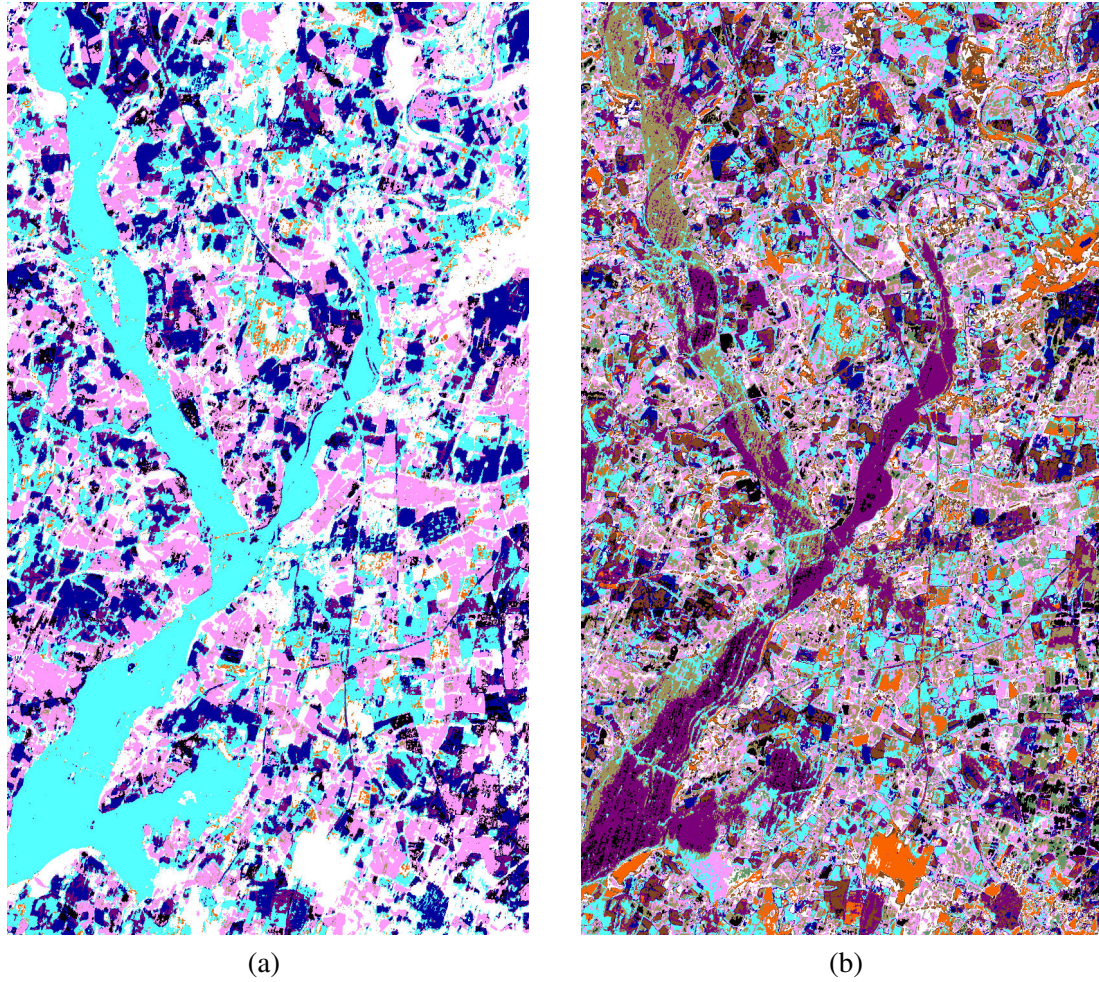


Figure 3.12: Classification results in 2^Θ with decision by maximum of pignistic probability over all simple hypotheses and all disjunctions of hypotheses: (a) SOM-based BBA. (b) DC results.

Table 3.5: Quantitative results in 2^Θ obtained using the confusion matrix for DC approach.

	BSWA	BS	WWA	CF	WA
BSWA	2102	0	0	0	26
BS	87	2106	397	3	420
WWA	2393	326	3092	194	0
CF	180	206	819	1068	0
WA	4	614	155	45	2455

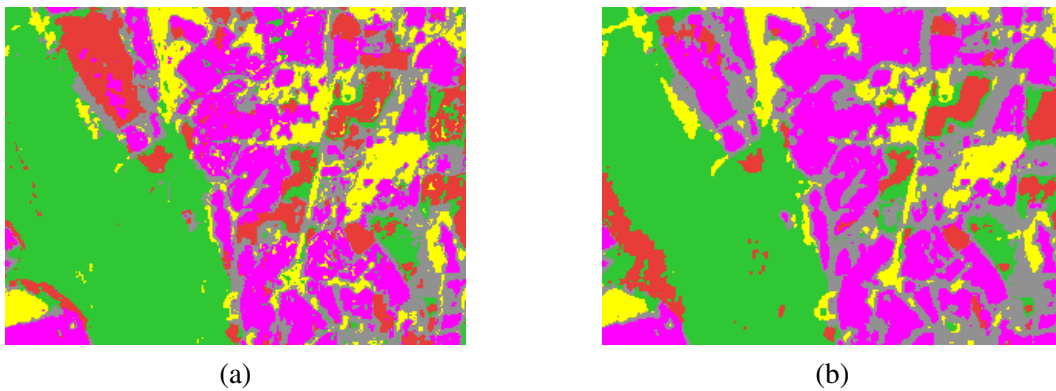
almost attributed to a single class; this reflects more what we have in the reality, while with the other approaches the river is classified into various class.

Table 3.6: Quantitative results in 2^Θ obtained using the confusion matrix for the proposed SOM-based approach.

	BSWA	BS	WWA	CF	WA
BSWA	1631	282	215	0	0
BS	206	2423	16	93	215
WWA	0	45	5359	601	0
CF	0	103	53	2117	0
WA	165	86	3	1	3018

After having exploring the performance of the SOM-based approach, this part focuses on the ability of the SOM-based approach to deal with a large amount of multi-variate data. To evaluate this, the unsupervised clustering method ECM has been used for its simplicity in generating the BBA in the case of exploratory data analysis. This algorithm requires a great amount of computing time for processing the large images. Here, a crop of the original image (300 by 220 pixels) has been processed so that the computation time remains acceptable. The classification results (see Figure 3.13) show that the SOM-based method gives higher performances than the ECM algorithm, while remarkably reducing the computational cost. Indeed, SOM-based method has a linear computational complexity depending to the number of classes for each new calculated BBA. These results prove that the proposed approach provides a very significant advantage in the case of processing large images.

All the algorithms in the experiments were coded in MATLABTM without specific optimization and run on a machine with 3.4 GHz Intel Core i7-3770M processor and 8 GB memory running the Windows 7 Server operating system. The execution times for these algorithms are: 20 minutes and 12 seconds for the SOM-based BBA shown in Figure 3.13-(a), and 2 days and 6 hours and 45 seconds for the ECM algorithm shown in Figure 3.13-(b). It corresponds to an increase in computation speed of 150.

Figure 3.13: Classification results in 2^Θ with decision by maximum of pignistic probability: (a) SOM-based approach. (b) ECM.

The computation of the complete scene at Figure 3.11-(a) took 4 hours and 21 minutes and 36 seconds.

3.6.2 The classification results in D^Θ

In this experiment, the result of DC is given by replacing Dempster's combination rule given by equation (3.3) by the conjunctive rule given by equation (2.11). Figure 3.14 shows the classification of the original image by using maximum of generalized pignistic probability over all simple classes and all conjunctions of classes. The performance of DC and SOM-based classifiers is shown through the confusion matrices form in Table 3.8 and Table 3.9, respectively. Table 3.7 represents the colors assigned to each conjunctions of classes in classification. The colors assigned to simple classes and to disjunctions of classes are the same as those defined in Table 3.4.

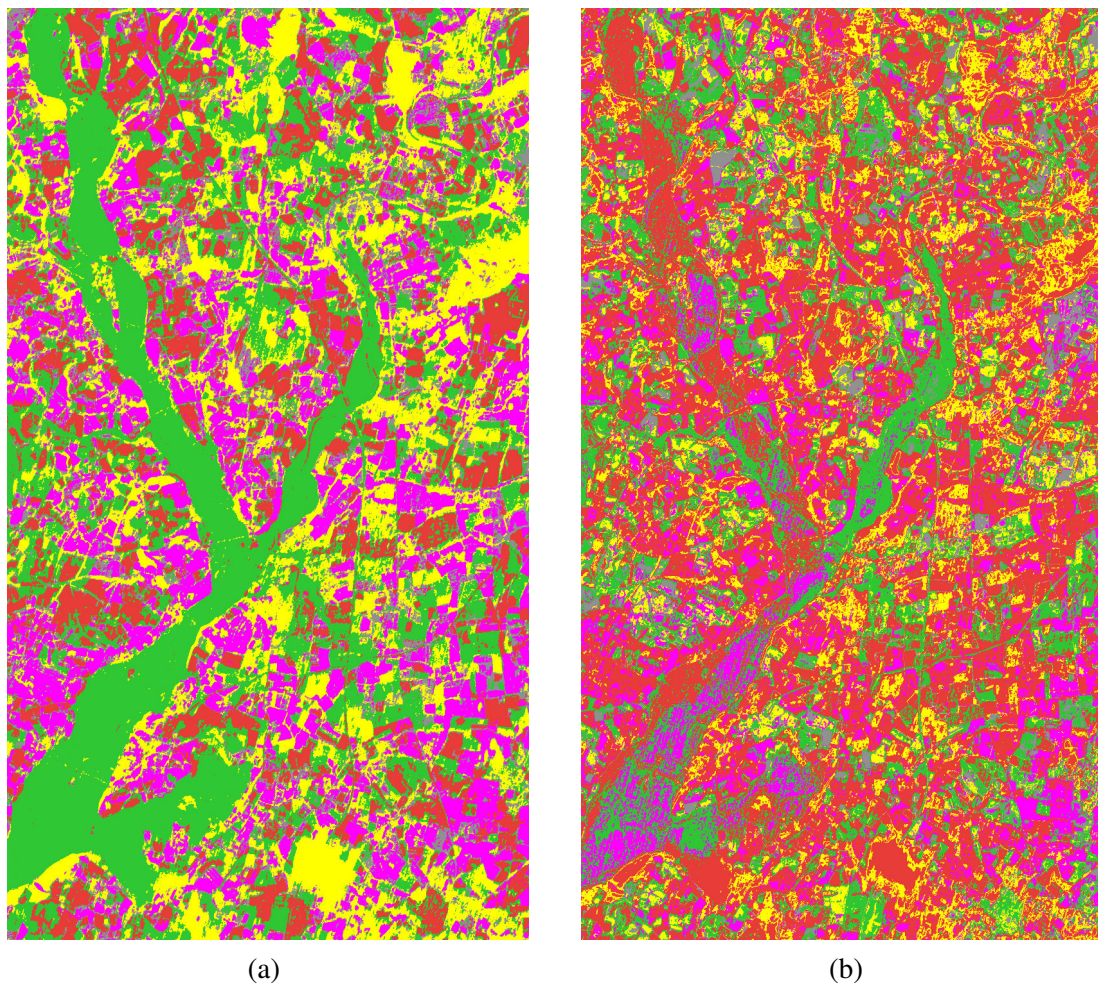


Figure 3.14: Classification results in D^Θ with decision by maximum of generalized pignistic probability over all simple hypotheses and all conjunctions of hypotheses: (a) SOM-based approach. (b) DC results.

As shown in Table 3.8, the generation of the masses on the conjunctions of hypotheses has degraded remarkably the DC result. This is due to the conflicting nature of the conjunctive rule

Table 3.7: Legend used for classification D^Θ shown in Figure 3.14.










	$BSWA \cap BS$		$BSWA \cap CF$
	$BSWA \cap WA$		$BSA \cap WWA$
	$BSA \cap WA$		$BSA \cap CF$
	$WWA \cap WA$		$WWA \cap CF$
	$BSWA \cap WWA$		$WA \cap CF$

Table 3.8: Quantitative results in D^Θ obtained using the confusion matrix for DC approach.

	BSWA	BS	WWA	CF	WA
BSWA	0	1053	317	382	376
BS	2056	90	86	6	255
WWA	2249	2	2081	24	1649
CF	1536	0	0	737	0
WA	1733	1	151	45	1334

Table 3.9: Quantitative results in D^Θ obtained using the confusion matrix for the proposed SOM-based approach.

	BSWA	BS	WWA	CF	WA
BSWA	1913	0	215	0	0
BS	2080	1507	368	313	545
WWA	0	0	5404	601	0
CF	0	1	2272	0	0
WA	165	29	8	3	30068

when unreliable sources are combined. The SOM-based approach (see Table 3.9) can overcome this problem by calculating the masses of conjunctions from Kohonen's map.

Figure 3.15 shows the classification of the original image by using maximum of generalized pignistic probability over all simples classes, all conjunctions of classes and all disjunctions of classes. As seen in DST-based experiment, it appears that the SOM-based approach yields promising results with a very reasonable computation time in such situations even with large number of classes.

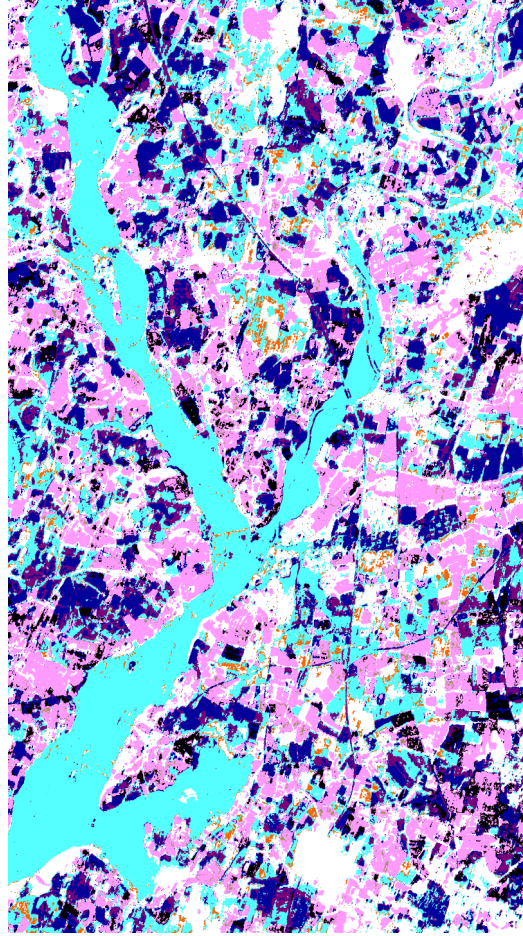


Figure 3.15: Credal classification results in D^\ominus through the SOM-based approach: maximum of generalized pignistic probability.

3.7 Conclusion

The interest of evidence theory came from its ability to deal with uncertain and paradox data through the mass functions. Nevertheless, to the best of our knowledge, rare are the estimating mass functions approaches that consider the belief masses on compound hypotheses directly. In this chapter, a new method for mass function construction through Kohonen's map has been proposed, and some experiments of the proposed method has been dedicated to image classification. The comparison with state-of-the art UCI database showed the accuracy of the SOM-based approach and its capability to deal with large amount of data. A further advantage can be added which is the possibility to perform the assignment of belief masses on the conjunctive and disjunctive hypotheses directly.

In this part of our research work, we focus on the application of the proposed Kohonen's map based BBA on SPOT images only, which is based on a quadratic distance evaluation. The extension to the problem of optical and SAR remote sensing images for joint classification will be investigated in the following chapter.

Kohonen-based Credal Fusion of Heterogeneous Data: application to optical and radar joint classification with missing data

4.1	Introduction	65
4.2	Generality on data fusion in remote sensing field	66
4.2.1	Definition of data fusion	66
4.2.2	Data fusion levels	67
4.2.3	Fusion of optical and radar data for land cover classification	68
4.3	The proposed method	69
4.3.1	Fusion scenario preview	69
4.3.2	Evaluated features	71
4.3.3	Basic Belief Assignments for Heterogeneous Data	72
4.3.4	Adopted scheme for heterogeneous data fusion	73
4.4	Experimental results	74
4.4.1	Study area and data description	74
4.4.2	Results for joint classification	75
4.4.3	Results for joint classification with missing data	79
4.5	Conclusion	81

4.1 Introduction

Today, data is becoming even more available and accessible, which in turn calls for a smart processing regime allowing complete and useful information to be extracted from various sources.

However, optimizing the decision step in this regime through an efficient fusion process is a challenging task, especially in terms of merging data available from heterogeneous sensors. The aim of this chapter is to introduce a new credal algorithm to fuse data derived from heterogeneous sensors, such as optical and radar data, which represents one of the most important issues faced in the field of remote sensing. SAR (Synthetic Aperture Radar)/optical information fusion is investigated in this study for joint classification of agricultural areas with missing data.

This chapter is organized as follows: Section 4.2 presents an overview of data fusion in the field of remote sensing. Section 4.3 introduces the proposed credal algorithm for merging optical and SAR information. The results obtained and the experimental validation are presented in section 4.4. And finally, section 4.5 concludes.

4.2 Generality on data fusion in remote sensing field

Having appeared initially in the military domain to manage very large amounts of information, data fusion has today become an important field of research in multiple domains such as robotics, biomedicine and image analysis, to name a few. The objective of this section is to provide a general overview of the basic concepts of data fusion as well as its use particularly in the field of remote sensing for interpreting optical and radar data.

4.2.1 Definition of data fusion

It is difficult to formulate a precise and a consensual definition of the term "data fusion". Consequently, several definitions have been proposed (for review and discussion of many of these definitions, reference [72] is recommended) and each of them reflects the perception of the domain which varies from one scientist to another according to his research discipline. In the following, we present some of the most used definitions in the field of remote sensing:

"Data fusion is the joint use of heterogeneous information for the assistance with the decision-making." [73]

"Data fusion is the set of methods, tools, means using data coming from various sources of different nature, in order to increase the quality (in a broad sense) of the requested information." [74]

"Data fusion is a formal framework in which are expressed means and tools for the alliance of data of the same scene originating from different sources. It aims at obtaining information of greater quality; the exact definition of greater quality will depend upon the application." [75]

"Fusion consists of combining information originating from several sources in order to improve decision-making." [76]

Independently of the formal framework or not in which data fusion is defined, we note that these definitions emphasize the simple principle of fusion by focusing on these two interesting points: 1) the fusion process operates on heterogeneous data from different sources. Therefore, the information provided is either of different natures (multisource fusion), or of the same natures, but taken at different conditions to provide additional knowledge (multidate fusion). 2)

the merge process aims to improve the quality of the resulting information for better decision-making. Indeed, it makes it possible to remedy the imperfections of the information collected by exploiting their redundancy and/or their complementarity. These imperfections can be of several natures [3], such as uncertainty and imprecision, incompleteness and ambiguity, conflict and contradiction, etc.

4.2.2 Data fusion levels

In general, fusion information in the remote sensing field can be performed at three different levels [77]: pixel, object possessing characteristics or attributes and decision level. The decision level [78] operates directly on individual decisions found by applying a proper processing for each image. Although this level is considered the most robust among the three, its solution is not globally optimal, since it seeks to optimize each source individually. The object level [79] is mainly based on the extraction of one or more characteristic maps by computing the relevant

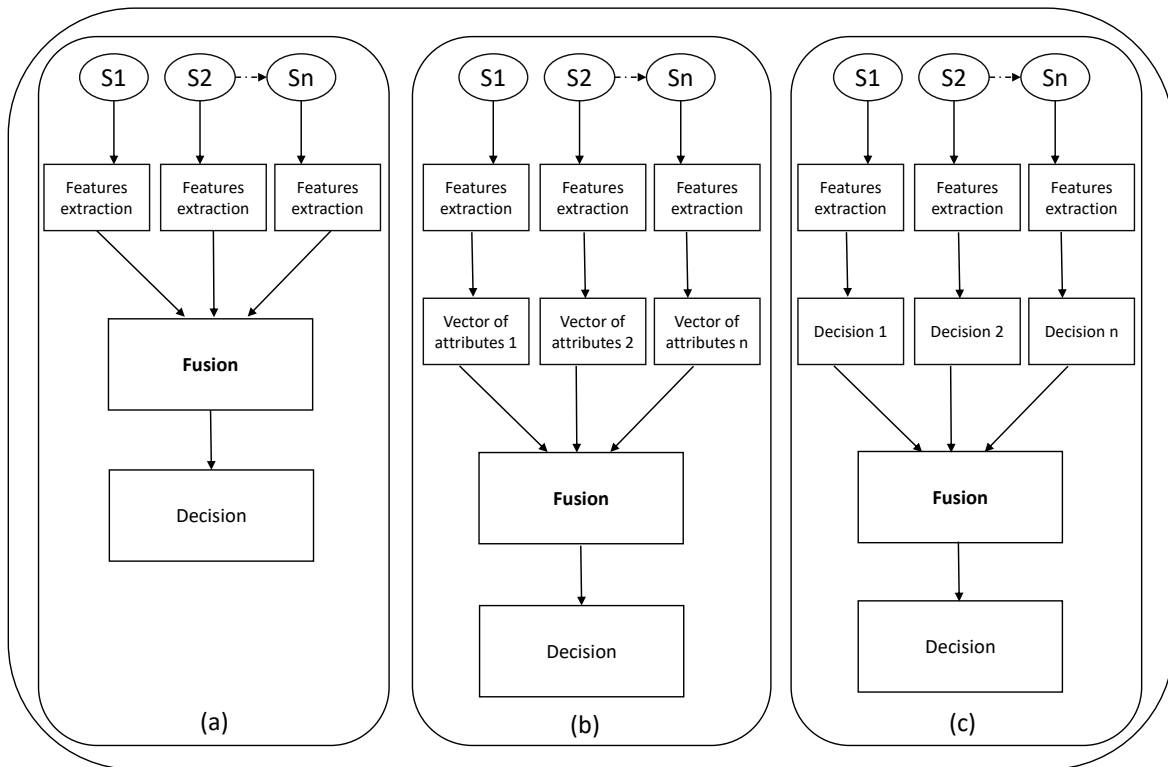


Figure 4.1: Data fusion levels (a) Low-level fusion or pixel fusion (b) Attributes fusion (c) High-level fusion or decision fusion.

descriptors from an input image. As in the case of high level fusion, this approach induces a loss of information inherent to the replacement of original data by the attributes extracted in subsequent processing. Ideally, all the data should be merged at the lowest level or the pixel level [80], in which raw data extracted from each pixel, such as spectral or temporal information of the considered sources, are used. However, it should be noted that the design of an appropriate approach is very difficult due to the complexity of joint processing of heterogeneous data,

such as optical and radar data. This pixel-based heterogeneous fusion requires the use of accurate co-registered images that are often derived from a resampling process in a pre-processing step. Figure 4.1 gives an idea about those three different levels of fusion.

4.2.3 Fusion of optical and radar data for land cover classification

Continuous development of acquisition techniques of satellite images has led to the emergence of new challenges in the field of remote sensing. One of the most discussed challenges is related to the joint interpretation of optical and radar data for land cover classification [81]. Although the use of optical data has been the subject of several in-depth studies and has produced very promising results in this field [82–85], their high sensitivity to certain atmospheric conditions, such as cloudy weather, has prompted researchers to integrate other types of data.

Radar sensors make it possible to provide complementary information to those generated using optical sensors, which make them a very interesting alternative. Indeed, in optical sensors, the information is mainly influenced by the reflectance properties of the mapped object surface. Their response is thus related to the chemical, physical and biological characteristics of the target. While, radar sensors detect the backscattered signal, which is essentially, conditioned by the structural (e.g., the size, shape and orientation) and dielectric properties of the target. This variety of information provided offers a great discriminatory potential in land cover classification [86]. Several studies have shown the benefits of fusing optical and radar data to improve the accuracy of the applied classification techniques [87–90].

The work of Joshi et al. [81] presents a bibliography of nearly 112 references on the fusion of optical and radar data, the majority of which are related to our problem. These studies usually follow a methodology mainly composed of two phases: 1- Extract relevant features from the optical and the radar image; 2- Fuse the resulting features using supervised or unsupervised classifiers. Maximum-likelihood decision rule was used by Idol et al. in [91] to determine if radar texture measures combined to optical imagery influence land-cover/use classification accuracies. The described approach is mainly based on a maximum-likelihood decision rule for the classifications of spectral signatures obtained from multiple landscape features. They found that Sensor fusion of optical and radar obtained an accuracy of 93% compared to the optical ASTER overall accuracy of 81%, and combining the original radar and a variance texture measure increased the Radarsat-2 overall accuracy to 78% and PALSAR to 80%.

Artificial Neural Network (ANN) and Support Vector Machine (SVM) have also been successfully used for the classification of multimodal data sets [92] and [93]. In [93], the problem of multitemporal synthetic aperture radar data and optical imagery is addressed. Each data source is classified separately using a SVM, after that, the original outputs of each SVM discriminant function are fused using another SVM, which is trained on the a priori outputs. A multilayer feedforward networks devoted to multisensor remote-sensing image classification is applied on [94]. The obtained results show the efficiency of ANN compared to traditional statistical parametric methods such as maximum likelihood. However, some well-known drawbacks such as how to define the network architecture or how to fix the number of hidden layers persist.

The work of [95] also shows the advantage of using ANN in general and Convolutional Neural Networks (CNN) in particular for the fusion of heterogeneous data. The authors applied a deep learning algorithm mainly based on automatically extracted features to generate a map indicating the changes between the two optical and radar images. This algorithm gives good

performances compared to techniques already proposed in the literature but the problem of considerable learning time of this family of classifiers makes their use very difficult especially in presence of a small satellite image database. A more recent technique for fusing RADARSAR-2 data and optical multispectral data for Land Use Land Cover extraction from a tropical agricultural zone is described in [96]. In this work, several fusion strategies including the Brovey transform, the wavelet transform, Ehlers and Layer Stacking have been applied to merge the results of a pixel wise classification with an object-based classification. The obtained classification errors especially in built-up area and bare ground are justified by the fact that the optical and radar data used in the experimentation were acquired at two different periods of the year.

The usefulness of using radar data to fill missing data in optical data due to the presence of clouds that cover the mapped agricultural area is explored by Betbeder et al. in [97]. These authors combine the results of a polarimetric decomposition obtained from a series of TerraSAR-X images and several indices extracted from Spot 4 images: the Leaf Area Index, the Fraction of Vegetation Cover and the Fraction of Absorbed Photosynthetically Active Radiation. The theory of belief functions has been thereafter used as a fusion operator, but the article lacks details about the applied operator and the technique used to manage the conflict between the different sources of information. Instead of focusing on the choice of the applied fusion method or the extracted features, other works have concentrated on the impacts of feature normalization on radar and optical data fusion. Zhang et al. [98] propose a novel approach for feature normalization suitable for optical and SAR fusion. They resolve this problem by normalizing the extracted features into three different scales $[-1, 1]$, $[0, 255]$, and $[0, 1]$ to handle negative values of HH and HV backscattering coefficients. They conclude that distribution-dependent classifiers (e.g., a maximum likelihood classifier) are independent of feature normalization; moreover, advanced classifiers (e.g., a support vector machine) with built-in normalization are also not influenced by feature normalization. In contrast, a minimum distance classifier and an artificial neural network (ANN) depend on the input values of optical and SAR features and thus can be influenced by feature normalization.

4.3 The proposed method

4.3.1 Fusion scenario preview

An overview of the fusion of heterogeneous data for joint classification purposes may be given as follows. Without a lack of generality, the description focuses on the optical and the radar image fusion problem, and the overall analysis is dedicated to land cover classification in a farming area. Fusion of heterogeneous data flowchart is given in Figure 4.2.

Two kinds of images are considered: a first optical image I_{MS} , which is multispectral with p spectral bands, is considered, as well as a SAR image I_{SAR} . The optical image may be considered to as a reliable observation as soon as no clouds or shadows affect the data. The SAR image is characterized by the speckle noise [99]. Moreover, with no polarimetric capability (which is the case in this study), the information that can be extracted from this SAR image is much less reliable than the one from the optical data.

Since no confidence may be given to a single SAR pixel due to the presence of speckle noise on a homogeneous area, a local texture descriptor is used. The descriptor is based on the first four cumulants $(\mu, \sigma, \beta_1, \beta_2)$, namely (mean, standard deviation, skewness and kurtosis)

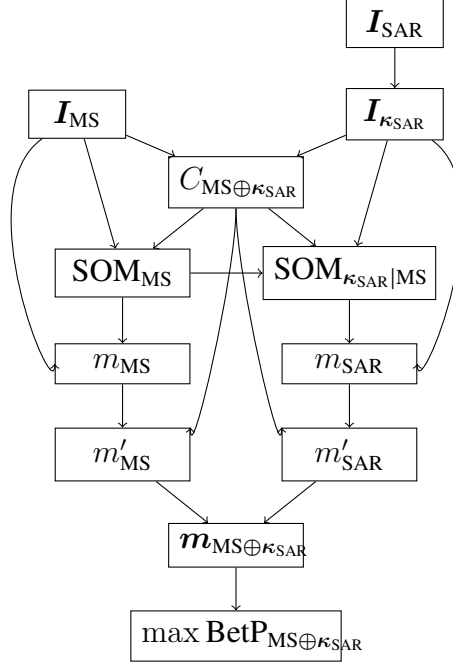


Figure 4.2: Credal fusion framework between a reliable optical multispectral image and a SAR observation. A first coarse joint classification is performed, yielding $C_{MS \oplus \kappa_{SAR}}$ which guarantees class homogeneity between the 2 sensors. Kohonen's map is trained from optical multispectral data I_{MS} to yield SOM_{MS} , and then from parameters extracted from the radar data with an enslaved constraint on the location of the neurons of $SOM_{\kappa_{SAR}|MS}$. From Kohonen's maps, BBA is performed and then specific discounting operators are applied on the mass depending on their reliability (that yields m'_{MS} and m'_{SAR}). Fusion is performed by PCR6 rule and decision-making is ensured with the maximum of Pignistic probability to yield a joint land cover classification.

estimated from I_{SAR} with a sliding window [100]. In addition, some parameters extracted from the Haralick texture analysis [101] (the sum average f_6 and the inverse different moment f_5) are used to account for the co-occurrence of the pixels in a neighborhood. The next subsection gives in detail the definitions of all used texture parameters.

They yield a 6-band image $I_{\kappa_{SAR}}$, holding local parameters $\kappa = (\mu, \sigma, \beta_1, \beta_2, f_6, f_5)$ that will be considered as the SAR information for the rest of the chapter. The p -band multispectral image represents the spectral information. The fusion process is carried out, in each pixel, between the BBA calculated from the two sources of information considered.

Given that the fusion process is facing different kinds of features, the joint classification of heterogeneous data needs to guarantee that the classes are defined in a homogeneous fashion both from the optical and from the SAR observations. Thus, a first joint classification $C_{MS \oplus \kappa_{SAR}}$ is performed to link the spectral signatures of I_{MS} and the SAR texture descriptors of $I_{\kappa_{SAR}}$. In this study, a simple K-means classifier is used with an appropriate distance that accounts for the heterogeneity of the joint observation and cross-calibration factor, which in turn accounts for the relative dynamics between the two observations.

Then, the fusion of heterogeneous data is performed through belief function theory, which requires a BBA for each information source (m_{MS} and m_{SAR}). A Kohonen-based BBA [11] is applied since it has shown its capacity to handle large remote sensing data. m_{MS} is estimated by

considering only the optical information (it is considered to be a reliable source of information), while m_{SAR} is enslaved to the framework of the optical information.

Once the BBA of these two sources of information is established, belief function theory is applied to combine the two pieces of information, which explains the uncertainty caused by the data heterogeneity; this in turn involves different degrees of reliability of sources and data imperfection. The final classification is obtained by the maximum of the Pignistic probability, as detailed in the following sections.

4.3.2 Evaluated features

In general, features used in remote sensing images classification are based on spectral, statistical, temporal or textural information contained in a pixel or a group of pixels. This latter type is particularly the most important in the interpretation and the analysis of SAR data [102, 103]. In this work, two families of texture features are used for the SAR descriptor: Haralick texture measurements and statistical moments computed in a neighboring of each pixel. For the optical data, only spectral features are used. These features include the surface reflectance of the different bands that compose the used image.

Haralick texture measurements

As introduced by Haralik [101], the Gray Level Cooccurrence Matrix (GLCM) gives the occurrences number of the relationship of a reference pixel with its neighboring pixel located at a given displacement d and according to the direction ϑ . Four orientations can be considered: 0° , 45° , 90° and 135° degrees. GLCM is of size $N_g \times N_g$, where N_g is the number of gray levels in the analyzed image. Each element (i, j) of this matrix is defined by the number of pixels with gray level j located at d of a pixel with gray level i . Several characters descriptive of the textures can be calculated from this matrix. As indicated above, we only use in this work the sum average and the inverse different moments. Table 4.1 gives the definitions of these chosen texture measures.

Table 4.1: Haralik features computed from the SAR image.

Texture name	Equation	Description
F_5 : Inverse difference moment	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i, j)$ <p>$P(i, j)$ is the value in the cell (i, j) in the matrix.</p>	It is a measure of local similarity in the image.
F_6 : Sum average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$ <p>where x and y are the coordinates (row and column) of an element in the cooccurrence matrix and $p_{x+y}(i)$ is the probability summing $x + y$.</p>	

Statistical moments

First order statistics are generally used to describe the randomness aspect of texture, *i.e.*, without taking into consideration the spatial dependence between pixels. They include the first

four statistical moments: mean, standard deviation, skewness and kurtosis. The computations of these characteristics are carried out on a region of interest based on its probability distribution (histogram h) of the luminance. In our case, the local histogram extracted according to a sliding window of odd size $w \times w$, is modelled using Edgeworth expansion [8].

Let $p(i)$ be the probability density of the intensity levels occurrence, calculated by dividing the values $h(i)$ in the total number of pixels in the sliding window:

$$p(i) = h(i)/w.w, \quad i = \{0, 1, \dots, N_g - 1\} \quad (4.1)$$

The definitions and the descriptions of the computed local moments are given in Table 4.2.

Table 4.2: Local moments measures computed from the SAR image.

Texture name	Equation	Description
Mean	$\mu = \sum_{i=0}^{N_g-1} ip(i)$	It defines the average level of intensity of the region or texture.
Variance	$\sigma^2 = \sum_{i=0}^{N_g-1} (i - \mu)^2 p(i)$	It describes the variation of intensity around the mean.
Skewness	$\mu^3 = \sigma^{-3} \sum_{i=0}^{N_g-1} (i - \mu)^3 p(i)$	It describes how symmetric the intensity distribution is about the mean.
Kurtosis	$\mu^4 = \sigma^{-4} \sum_{i=0}^{N_g-1} ((i - \mu)^4 p(i)) - 3$	It measures of the flatness of the distribution.

4.3.3 Basic Belief Assignments for Heterogeneous Data

The core of this study arises in the belief mass assignment in a heterogeneous context with or without missing information using our SOM-based BBA approach introduced in the previous chapter.

4.3.3.1 Enslaved Kohonen-based BBA

This section describes how to establish a BBA from a non-reliable source of information (*i.e.*, a SAR image) in the perspective of fusion with a heterogeneous reliable piece of data (*i.e.*, an optical image). To this aim, a hybrid SOM is defined through a hybrid neuron definition that takes the spectral signature of the optical data (in \mathbb{R}^p , p being the number of spectral bands) and the texture descriptor of the SAR data (in \mathbb{R}^q , here $q = 6$, to hold the first 4 cumulants and 2 Haralick parameters).

Let $\mathbf{x} = \{x_1, x_2, \dots, x_p\} \in \mathbb{R}^p$ and $\mathbf{y} = \{y_1, y_2, \dots, y_q\} \in \mathbb{R}^q$ be the two heterogeneous observations provided by two heterogeneous sensors. The input samples of the proposed hybrid SOM are done through the co-located observations $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ with which a distance must be associated. This distance is a fusion of the 2 metrics that are to be applied on each type of initial data:

$$d(\mathbf{z}, \mathbf{z}') = d_{\mathbb{R}^p}(\mathbf{x}, \mathbf{x}') + \alpha d_{\mathbb{R}^q}(\mathbf{y}, \mathbf{y}'), \quad (4.2)$$

with $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ and $\mathbf{z}' = (\mathbf{x}', \mathbf{y}')$ being 2 samples in \mathbb{R}^{p+q} . The parameter α is a cross-calibration factor, that accounts for the relative dynamics between \mathbf{x} and \mathbf{y} .

According to this definition of a hybrid feature space and its related metrics, it is possible to perform a training of a joint SOM where the weighting vectors are defined with $\mathbf{w}_z = (\mathbf{w}_x, \mathbf{w}_y) \in \mathbb{R}^{p+q}$. Nevertheless, this joint processing of our heterogeneous data does not account for source reliability. Optical and SAR data interfere in the same manner in the location

of each class center on the map (the class center of winning neurons w_{c_k} , $k \in \{1, 2, \dots, K\}$), while SAR data is much less reliable than optical data in the land cover classification accuracy. Then, instead of a joint processing, an enslaved processing is set up to perform SOM training and a SOM-based BBA of the SAR data only.

Enslaved SOM training starts with a classical SOM training of the optical data only, and yields SOM_{MS} . Then, the neurons of SOM_{MS} are concatenated by q components to fit the \mathbb{R}^{p+q} of the joint processing. The training of this hybrid map begins, but only the last q -components (dedicated to the SAR data) are modified. In this case, the optical part is preserved, while the SAR part follows the optical part in the location of classes on the map (locations of the winning neurons w_{c_k}). This defines $SOM_{\kappa_{SAR}|MS}$ and then m_{SAR} , as shown in Figure 4.2.

4.3.3.2 Joint Kohonen-based BBA for missing data

When the optical sensor acquires a scene in the presence of clouds, two kinds of missing data must be considered: the parts of the data that are hidden by the clouds themselves, and the parts that are affected by the shadow of the clouds. A mask allows the training of Kohonen's map with valid data only.

In order to process the images with missing data, two perspectives may be adopted when the optical pixels are affected by clouds and shadows, and are considered to be missing.

- 1) When no information is brought by the optical part, its related mass function may express a total ignorance:

$$\begin{aligned} \forall \theta \in 2^\Theta, \theta \neq \Theta \quad m_{MS}(\theta) &= 0, \\ m_{MS}(\Theta) &= 1. \end{aligned} \tag{4.3}$$

This first point of view does not however take into consideration the joint observation between optical and radar sensors.

- 2) The optical pixel may be recovered by using the joint Kohonen's map $SOM_{\kappa_{SAR}|MS}$ which models the links between optical and radar parts in the observation.

When a pixel x_{MS} is considered missing in the optical image due to the presence of clouds or shadow, the co-located radar observation y_{SAR} is considered. Its winning neuron in the radar restriction of $SOM_{\kappa_{SAR}|MS}$ allows us to consider the optical part of Kohonen's map. This spectral signature is substituted for x_{MS} to recover the missing information.

4.3.4 Adopted scheme for heterogeneous data fusion

Once the problem of mass construction of heterogeneous data is resolved, the purpose of this section is to discuss the strategy adopted for combining these different pieces of evidence.

4.3.4.1 Uncertainty management

After computing the masses of evidence m_{SAR} and m_{MS} of our heterogeneous sensors data, some existing discounting techniques are firstly applied to manage the uncertainty before the fusion process. On the one hand, the contextual discounting described by equation (2.28) is applied on m_{MS} to render the modelling very flexible by transferring a part of the mass of simple hypotheses to the masses of the appropriate disjunctions of hypotheses. On the other hand, the

priority discounting approach described by equation (2.27) is applied on m_{SAR} , in order to prune the radar source (considered, here to be of lower quality than the optical source). This technique is generally used to rank pieces of evidence according to their priority using prior knowledge obtained from a fusion designer. Section 4.4.2 provides the means of estimating the contextual discounting weighting λ_i of each context and the priority discounting weighting β .

4.3.4.2 Knowledge fusion

Once the mass functions of each source are updated, a fusion step is required in order to synthesize the final information that describes belonging to the set of possible classes. However, as a priority discounting is applied, Dempster's rule of combination cannot be applied since it doesn't respond to the discounting of sources towards the empty set [42]. So, only combination rules allowing the redistribution of conflict should be considered. The PCR6 [38, 39], given by equation (2.14), is applied here to calculate $m_{\text{MS} \oplus \kappa_{\text{SAR}}}$. Indeed, it allows the redistribution of the possible conflict between the information brought by the optical sensor or the radar sensor. From this combined mass function, the joint classification must be done based on the maximum of the pignistic probability as decision criteria.

4.4 Experimental results

In order to assess the performance of the proposed heterogeneous data fusion algorithm, some experimental studies on a SPOT-5 and RADARSAT-2 images are carried out in this section. After a presentation of the study area in section 4.4.1, results of optical and radar joint classification with complete data, along with a validation step, are presented in section 4.4.2. Finally, section 4.4.3, focuses on the effectiveness of the proposed approach in dealing with (simulated) missing data.

4.4.1 Study area and data description

Our study area covers a part of the Beauce region, located in the south-west of Paris, France. This region is known for its high agricultural productivity. It is also essentially characterized by its very large fields dominated by rape and cereal (wheat, barley, corn) crops. A multispectral image acquired by the SPOT-5 French satellite during the Take-5 experiment and a radar image acquired by the RADARSAT-2 Canadian satellite in Ultra-Fine mode are used in this experiment. The two images cover an area of approximately $11.5 \times 9 \text{ km}^2$ and have the following features: the crop of the SPOT-5 image is characterized by a size of 1145×903 pixels, a spatial resolution of 10m, and has four bands (Green (G), Red (R), Near InfraRed (NIR) and Medium InfraRed (MIR)). The RADARSAT-2 image is composed of 3850×3010 pixels, with each pixel having a spatial resolution of 3m. Regarding the radar image, only HH (horizontal transmit and horizontal receive) and HV ((horizontal transmit and vertical receive) polarization channels are available. However, HH-polarization was only used since it interacts more efficiently than HV-polarization with agricultural crops. Figure 4.3-(a) and Figure 4.3-(b) show, respectively, the false color composite of the SPOT image and its registered radar image.

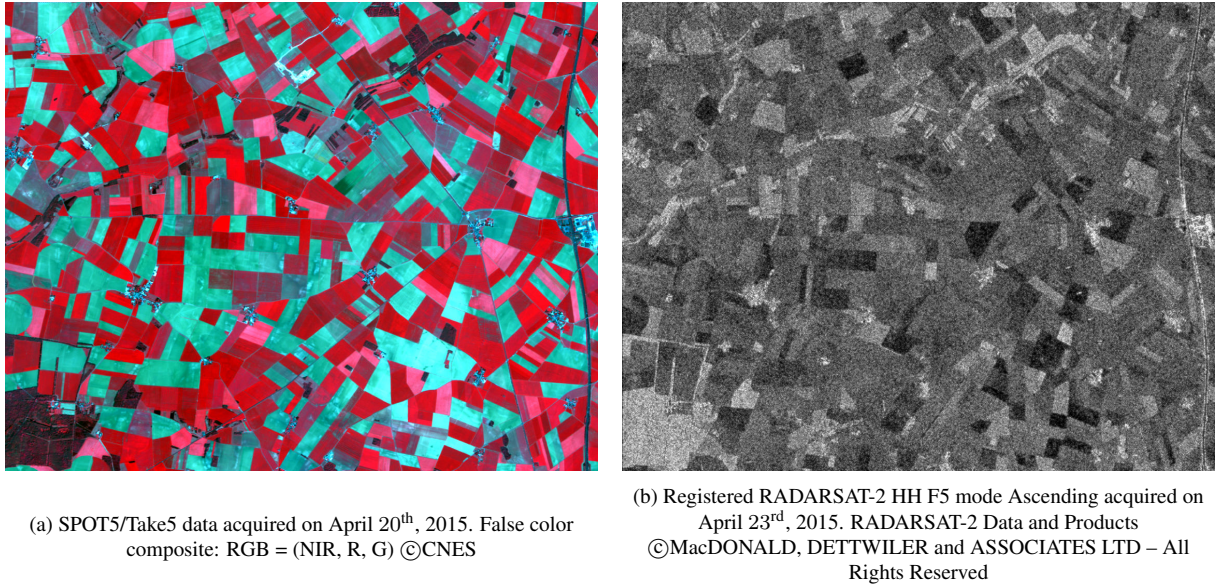


Figure 4.3: Multispectral (a) and radar (b) images acquired over Beauce, France, used in the experiments.

4.4.2 Results for joint classification

The fusion process is achieved at the coarser resolution of both images, that is, at a resolution of 10m, which is that of the SPOT image. To this end, the RADARSAT-2 image is first processed in order to extract the local statistical parameters $(\mu, \sigma, \beta_1, \beta_2, f_6, f_5)$; the processing is done through a sliding window of 51×51 for $(\mu, \sigma, \beta_1, \beta_2)$ and 15×15 for the Haralick texture parameters (f_6, f_5) estimation.

In order to prevent bordering effects between parcels, a naive map extracted from the multispectral image serves as a mask in the local parameter estimation of radar data. This guarantees a parameter estimation on effective homogeneous areas and preserves the borders of each parcel. The choice of analysis window size is based on our object dimensions of interest. Therefore, our analysis windows size is proportional to those field dimensions. The 3m-resolution feature image is then downsampled to a 10m-image, and then registered to the SPOT geometry. Figure 4.4 shows a false color composition of the radar information at a 10m resolution. The color composition is shown with $RGB=(\mu, \sigma, \beta_1)$.

In order to merge the belief degrees associated with each pixel from the two input images, a unified frame of discernment is required. The simple classes of this frame are defined using the K-means unsupervised classifier, where the parameter K is set to 5. It is applied to a stack image collecting the spectral information of the SPOT image and the texture information of the SAR image: $I_{MS} \oplus \kappa_{SAR}$ image. The weight factor of equation (4.2) is set to $\alpha = 5.10^{-3}$ which corresponds to the average ratio between the mean value of the optical and radar data. Five different land cover types are identified:

- C1 A brown class in Figure 4.3-(a) and yellow areas in Figure 4.4, which correspond to wooded areas;
- C2 Dark red fields in Figure 4.3-(a), which do not have an explicit signature in the radar image, and which correspond mainly to durum wheat (planted in winter);

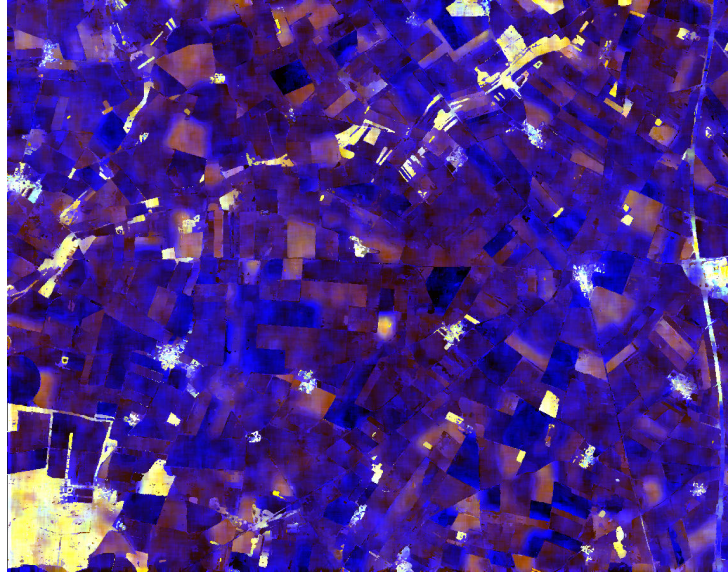


Figure 4.4: $I_{\kappa_{\text{SAR}}}$ image corresponding to the SAR texture information with false color composition: $\text{RGB}=(\mu, \sigma, \beta_1)$.

- C3 Light red fields in Figure 4.3-(a) and light brown fields in Figure 4.4, which correspond mainly to rape;
- C4 Cyan fields in Figure 4.3-(a) which correspond to bare soils and appear dark in Figure 4.3-(b). In fact, they correspond mainly to corn and cereal seedling;
- C5 Grey fields in Figure 4.3-(a) which have no significant signature in the radar image, and which correspond mainly to barley (planted in early spring).

Ground truth was collected in July, while the data were acquired in April, and as a result, any ambiguity between different kinds of crops could not be resolved, as many fields were still in the seedling state. Hence, it was decided that only 5 classes could be discriminated. The results of the joint classification are shown in Figure 4.5. This classification is used as reference data in the following.

Figure 4.6 presents the results of the two Kohonen maps trained with the optical and the SAR information. The 65×65 neuron maps were trained with 5000 samples per class. The initial learning rate and neighborhood size were set respectively to 1 and 60.

The multispectral map, in Figure 4.6-(a), shows the distribution of spectral signatures representing cover soils and bare soils in this farming area. In the marginal zone between these two types of spectral signatures, there is a location dedicated to man-made structures (buildings and roads) and a forest area (dark, at the bottom left area of the map). The enslaved map, dedicated to radar data in Figure 4.6-(b), shows the same kind of neurons at the same location on the map, viewed by the textural parameters extracted from SAR data. It can be seen that the area in the middle of the map, $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$, extending from the right to the left, appears homogeneous, while we have 2 different areas in SOM_{MS} . This illustrates the fact that the optical sensor is mainly sensitive to the presence of chlorophyll in this farming area, while the radar is sensitive to the surface roughness. However, surface roughness may appear similar, from a radar observation, in bare soil and also in cover fields, depending on the plantation. Nevertheless, the wooded area is clearly discriminated from SAR sensor. The wooded area appears in brown-

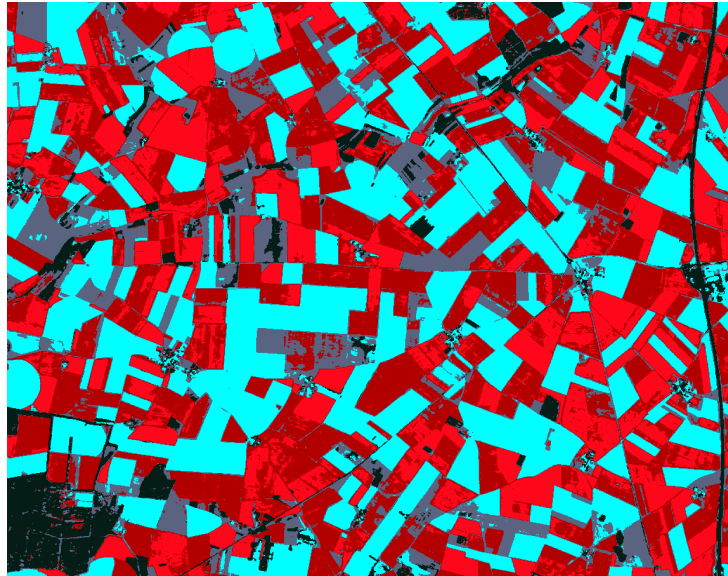
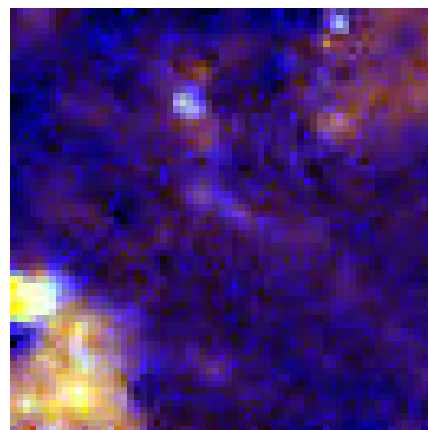


Figure 4.5: Unsupervised K-means classification results (with $K = 5$ classes) applied jointly on multispectral and SAR information.



(a) SOM_{MS}
RGB=(NIR, R, G)



(b) $SOM_{\kappa_{SAR}|MS}$
RGB=(μ, σ, β_1)

Figure 4.6: SOM maps of size 65×65 neurons. The map (a) has been trained with the optical data only, while the map (b) has been enslaved to map (a) and trained with radar data. Hence, co-located neurons bring the same ground information between the 2 maps.

yellow on the image in Figure 4.4 and at the bottom left of $SOM_{\kappa_{SAR}|MS}$ in Figure 4.6-(b). The SAR sensor does not help in the discrimination between bare soil and cover soil; nevertheless, it easily discriminates the 2 kinds of cover fields that appear in red from a SPOT point of view (e.g., Figure 4.5), as an area at the top left of the $SOM_{\kappa_{SAR}|MS}$ map appears in brown. It helps to do discriminate between corn seed in winter and in early spring. The use of the enslaved radar part of $SOM_{\kappa_{SAR}|MS}$ allows the Credal classification to tackle this ambiguity in order to improve the final joint classification.

By using Kohonen's maps and the joint classification, the estimation of BBA is performed next [21]. Then, two different discounting techniques are integrated within the fusion system in

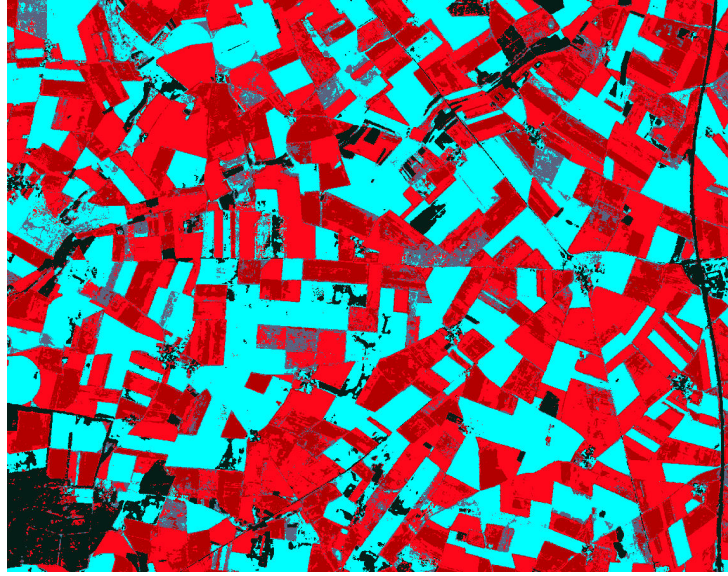


Figure 4.7: Joint classification results with decision by maximum of pignistic probability over all simples hypotheses.

order to manage the uncertainty and the contradiction (conflict) of our sources. Regarding the m_{MS} evidence, the contextual discounting weights, λ of equation (2.28), were calculated using results given by the confusion matrix [104] derived from the cross decisions given by the optical source and the decisions of the reference data only. In our case, the simple hypotheses (classes) are the contexts. Each weight is calculated using the percentage of correct classifications of the target class. The contextual reliability factors for the five classes are: $\lambda_1 = 0.85$, $\lambda_2 = 0.86$, $\lambda_3 = 0.52$, $\lambda_4 = 0.9$, $\lambda_5 = 0.87$. Regarding the m_{SAR} evidence, as the SAR source has a lower priority than the optical source in our proposed fusion method, the priority discounting factor, β of equation (2.27), must be less than 1, and the higher its value, the more the information it provides is taken into consideration. Later, this is set to 0.4 in this work, based on the subjective attribute of this source.

Figure 4.7 illustrates the classification resulting from the fusion of optical and SAR information using the PCR6 rule. The decision criteria are based on the maximum of the pignistic probability $BetP$ defined in equation (2.29). The validation step is carried out through the confusion matrix shown in Table 4.3. It is worth noting that the proposed approach provides an interesting accuracy with a Correct Classification Rate (CCR) of 77.25% and an Index Kappa of 0.74. These rates were computed over all pixels belonging to the ground truth.

The overall classification is quite similar to the reference one in Figure 4.5. Nevertheless, the accounting for source reliability for each class allows the decision-making to mitigate ambiguous fields such as those which are slightly covered, but may still be considered to be bared, or the seedling fields which may not have the same surface roughness. This impacts mainly the class C5 fields (mainly barley), which can be relocated to class C2 or C3, depending on the roughness signature and density of chlorophyll.

Table 4.3: Quantitative results obtained using the confusion matrix.

	C1	C2	C3	C4	C5
C1	95.35	0.00	1.94	2.69	0.02
C2	0.02	60.65	34.83	0.11	4.39
C5	0.00	4.87	95.13	0.00	0.00
C4	4.59	0.00	0.04	95.36	0.01
C3	1.75	9.04	17.88	31.55	39.78

4.4.3 Results for joint classification with missing data

In order to evaluate the effectiveness of the proposed method in the case of missing data from the optical sensor, some cloud-free regions belonging to the initial data set are manually masked. Two masked regions, namely, zone 1 and zone 2, were selected in order to hide different kinds of ground cover. The mask of zone 1 mainly covers a non-agricultural area composed of 18,056 pixels, while the mask of zone 2 covers some covered and some bare soils composed of 6,300 pixels, as shown in Figure 4.8.

Figure 4.8-(b) shows the results of the joint classification obtained by applying simulated cloud cover. For better visualization, Figure 4.9 gives a more detailed view of this classification. It is in fact the Credal classification yielded by the SAR observation only. A quantitative analysis gives an overall classification accuracy of 73.94 % which is very close to the accuracy yielded by the complete $MS \oplus \kappa_{SAR}$ data.

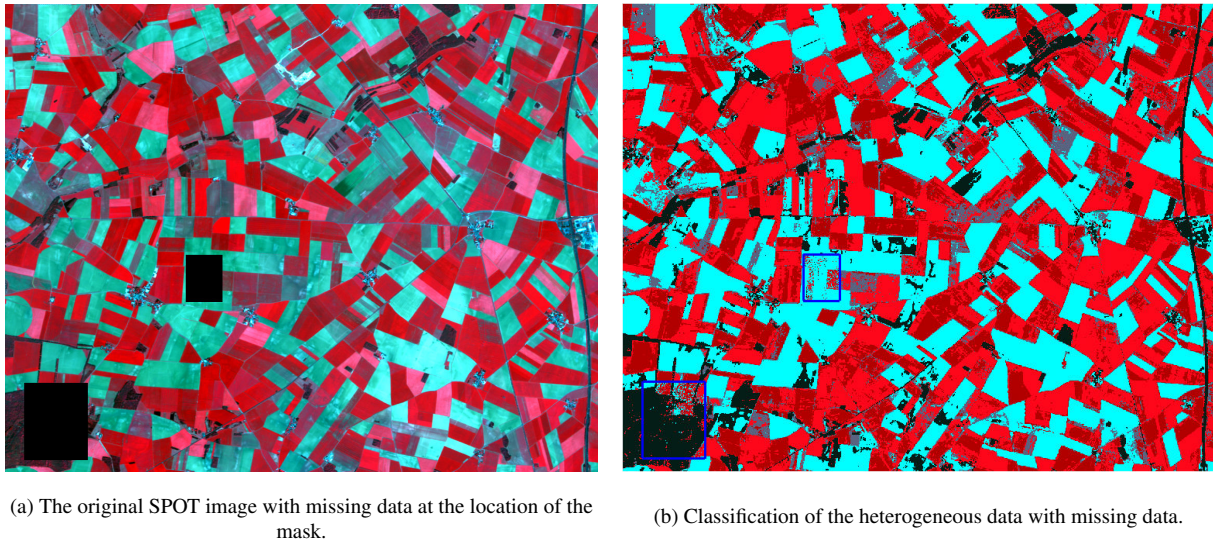
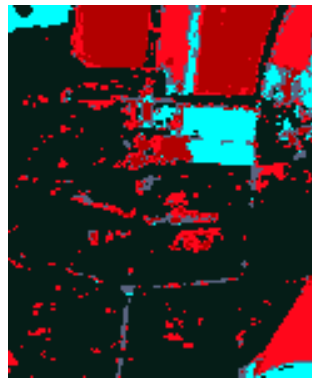


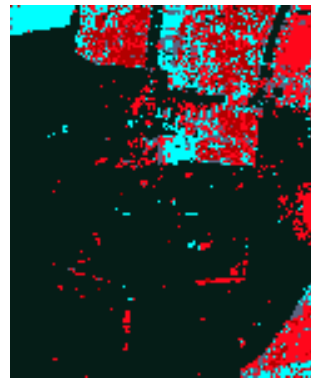
Figure 4.8: Results for joint classification with simulated cloud cover. The decision is performed by the maximum of pignistic probability over all simple hypotheses.

The classification errors that can be found in Figure 4.9 result from the lack of information due to radar observations in spring for fields classification. The wooded area bears a specific

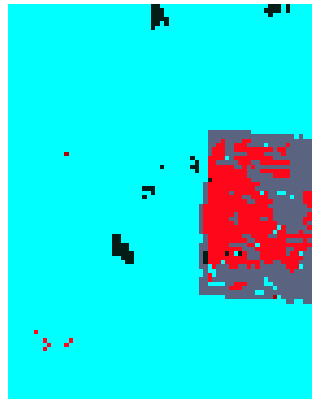
signature from the radar image such that almost no errors are to be found, for this class, between the images of Figures 4.9-(a) and (b). The difference in resolution between the SPOT image and the SAR-based texture information image explains the missing of chlorophyll-type patches in Figure 4.9-(b). Nevertheless, the fields at the top of Figure 4.9-(a), which are likely to be bare soil and durum wheat, are not discriminated from the SAR observation. Ground truth collected in July shows that the bare soils are being prepared for vegetable seedlings, and as such, the band-C surface roughness is not discriminated. The same remark may be made regarding the results of Figures 4.9-(c) and (d). Here the bare soil, in (c), is estimated, in (d), with grey and red pixels which correspond to rape and barley signatures. Ground truth shows that peas have been planted on this strip of land, and beans on the rest. Although still looking at bare soil, the SAR observation discriminated two kinds of land cover. Then, despite noise, the estimated classification with missing data remains consistent.



(a) Zone 1, original classification



(b) Zone 1, estimated classification



(c) Zone 2, original classification



(d) Zone 2, estimated classification

Figure 4.9: Zoom of Figure 4.8-(b). Correct Classification Rate (CCR) of 72.28 % for masked zone 1 and 75.60% for masked zone 2.

4.5 Conclusion

In this chapter, a new credal fusion algorithm has been proposed that aims at integrating complementary information derived from optical and radar remote sensing data for land cover mapping in agricultural areas. The proposed approach refers to the direct combination of heterogeneous data at the pixel level and considers the image registration problems as resolved. The aim of this strategy is mainly based on hybrid training of Kohonen's map using heterogeneous data for mass functions estimation. This step helps to deal with the heterogeneity of data sources by representing those in the same semantic meaning through co-located observations. The methodology benefits from this joint training of heterogeneous data to restore missing parts of optical data. It is worth noting that the enslaved processing, described in this chapter, is relevant when one of two sources of information is considered to be more accurate than the other. If the cloud coverage is becoming too significant, the related source of information may no longer be considered accurate. In that case, a joint processing should be preferred in order to recover missing data.

The experimental part of this study showed the benefit of heterogeneous joint processing in the analysis of a complex farming area, even in the case of missing data.

Copulas-based fusion of consonant belief functions induced by dependent sources of evidences

5.1	Introduction	83
5.2	Overview on combination rules of dependent belief functions	84
5.3	Copulas	85
5.3.1	Reminders and notations	85
5.3.2	Copulas as conjunctive aggregation functions	85
5.3.3	K-plot graphical representation	87
5.4	DST and Copulas	88
5.4.1	Random sets and DST	88
5.4.2	Conjunctive combination of dependent consonant belief functions using copula	88
5.5	Credal dependent fusion: disjunctive aggregation rule	90
5.6	The choice of the family of copulas	91
5.7	Experiments	93
5.7.1	Consonant mass functions estimation	94
5.7.2	Results and discussion	94
5.8	Conclusion	100

5.1 Introduction

Conventionally, Bayes and Dempster-Shafer reasoning frameworks are considered as the two most important approaches that deal with uncertainty representation. Although both theories have many similarities since they have their origins in the probability theory, they have some differences, the most important is the rule of aggregation. Dempster's rule assumes that the

sources of evidence to combine are independent which is always questionable in the practice. In this chapter, we focus on this insufficiency and we propose a new strategy to combine dependent consonant belief functions thanks to statistical copulas analysis [12]. Our approach consists in identifying the copula that best summarizes the existing dependency structure between the sources and then to combine the marginal belief functions accordingly.

The present chapter is organized as follows. The second section gives an overview of approaches allowing the combination of dependent evidences. Section 5.3 recalls basics of copula theory. In section 5.4, we explain how copulas can be investigated in DST to combine in a conjunctive way the beliefs. Section 5.5 introduces our new copulas-based disjunctive rule. Section 5.6 gives our strategy to pick the copula that best fits the problem at hand. Section 5.7 illustrates the effectiveness of the proposed method. Finally, section 5.8 concludes.

5.2 Overview on combination rules of dependent belief functions

DST formalism is often presented as a generalization of a Bayesian model because it can handle the distinction between uncertainty and ignorance. This point of view is, however, disputable as soon as Dempster's rule is under concern as pointed out by several authors [31–33, 105–109]. Dempster's rule presents the advantages to be commutative and associative, but also it has two main limitations: 1) its normalization procedure provides unsatisfactory performances and strange behaviours even in low conflicting cases [109], and 2) it requires the independency of sources of evidence to combine which is rarely satisfied in the real-world applications. This second limitation has encouraged researchers to work on it and if we take a look at the various solutions (operators) they propose for combining dependent evidence, we can classify them into three families of approaches:

- The first family seeks to satisfy the idempotence assumption of the combination rule. This propriety ensures that our belief on imperfect observation is not modified if the two used dependent sources of information provide the same knowledge. Thus, some authors [110, 111] have proposed to extend some existing idempotent rules coming from other theories of uncertainty to the evidence theory.
- The second family of methods uses the least commitment principle in the choice of the combination operators [15, 112]. As well as idempotent rules, this kind of merging rules seek to minimize the conflict by adopting a cautious attitude (which is guaranteed by their ability to handle the redundancy between evidences) when dependence between sources is doubtful. The weakness of these rules appears in their similar treatment for any degree of dependence between beliefs.
- The third family [40, 41, 113] consists in combining the marginal belief structures where the dependence structure is assumed to be known or to identify.

Subsection 2.2.2.2 details the main rules of those families. In the following, we focus on the third category of approaches and we are interested specially on rules allowing the combination of dependent evidence using copulas [12] that have been successfully used to model dependency for multivariate distributions in the framework of probability theory. Although several

studies [114–116] on the feasibility of its adaptation in the belief functions framework exists, to the best of our knowledge, they are rare that deals with the copula choice problem.

5.3 Copulas

Copulas represent the most attractive tools for characterizing any joint random variables. They were firstly introduced by Sklar [117, 118]. However, an analogous concept for capturing and modelling the dependency structures of joint distributions had independently previously appeared in the works of Hoeffding [119, 120]. In this section, we review the most important concepts of bivariate copulas.

5.3.1 Reminders and notations

We recall here the definition of the univariate distribution function of a uniform random variable on $[0, 1]$ which is closely linked to copulas.

Let $F(x) = P(X \leq x)$ be the cumulative distribution function (cdf) of the random variable X . By convention, its generalized inverse is defined by $F^{-1}(y) = \inf(x | F(x) \geq y)$. The variable $U = F(x)$ is then of uniform law on $[0, 1]$, since its distribution function is given by:

$$Pr(U \leq u) = P(X \leq F^{-1}(u)) = \begin{cases} 0 & \text{if } u \leq 0, \\ u & \text{if } 0 \leq u \leq 1, \\ 1 & \text{if } 1 \leq u. \end{cases}$$

5.3.2 Copulas as conjunctive aggregation functions

Informally, an aggregation operator is a mathematical tool that gives a unique representative object belonging to a given type from multiple objects of the same type, simply we speak about combination of information. For example, in the mathematical framework, aggregation operators handle only numbers.

Definition 9. (from [121]) *An aggregation operator is a function*

$$Ag : [0, 1]^n \rightarrow [0, 1]$$

that satisfies:

1. $Ag(0, \dots, 0) = 0$ and $Ag(1, \dots, 1) = 1$.
2. $Ag(x_1, \dots, x_n) \leq Ag(y_1, \dots, y_n)$
if $(x_1, \dots, x_n) \leq (y_1, \dots, y_n)$.

Dubois and Prade [122] distinguish four main aggregation function family approaches:

- Conjunctive aggregation function: all Ag that verifies $Ag \leq \min(x_1, \dots, x_n)$.
- Disjunctive aggregation function: all Ag that verifies $Ag \geq \max(x_1, \dots, x_n)$.

- Average aggregation function: any idempotent Ag .
- Mixed aggregation function: a particular combination of previous ones.

Particular kinds of aggregation functions are copulas. In fact, all conjunctive aggregation functions Ag that fulfill the n -increasing condition¹ are copulas. A copula [12] is a function which joins univariate marginal distribution functions to their multivariate distribution function. As n -dimensional copulas are notoriously hard to estimate except some specific cases (e.g., Gaussian copulas) and may pose some issues, we will only consider in the following the case of bivariate copulas which will be used later. Formally, a 2-dimensional copula is a function C from $[0, 1]^2$ to $[0, 1]$ such that:

1. C is grounded, *i.e.*, $C(u_1, 0) = C(0, u_2) = 0$ for all u_1 and $u_2 \in [0, 1]$.
2. The one-dimensional margins are uniform, *i.e.*, $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$ for all u_1 and $u_2 \in [0, 1]$.
3. C is 2-increasing, *i.e.*, the following inequality holds for all $(u_1, u_2), (v_1, v_2) \in [0, 1]^2$ such that $0 \leq u_1 \leq v_1 \leq 1$ and $0 \leq u_2 \leq v_2 \leq 1$:

$$C(u_1, u_2) + C(v_1, v_2) \geq C(u_1, v_2) + C(v_1, u_2). \quad (5.1)$$

Important examples of the copula are:

- The product copula that characterizes totally independence

$$C^\perp(u_1, u_2) = u_1 u_2. \quad (5.2)$$

- The comonotonicity copula that characterizes the complete positive dependence

$$C^-(u_1, u_2) = \min(u_1, u_2). \quad (5.3)$$

- The countermonotonicity copula that characterizes the complete negative dependence

$$C^+(u_1, u_2) = \max(u_1 + u_2 - 1, 0). \quad (5.4)$$

The copulas C^- and C^+ represent the lower bound (respectively the upper bound) of Fréchet-Hoeffding. So we have for each copula C that $C^- \leq C \leq C^+$. In addition, every copula C almost everywhere admits partial derivatives $\partial C / \partial u_1$ and $\partial C / \partial u_2$ [12]. Moreover, the density of the copula (which corresponds to the density of probability) is given by: $\partial^2 C / \partial u_1 \partial u_2$. Sklar's theorem [117] represents the cornerstone of the copulas theory; indeed it offers a link between the joint distribution of random variable and copula.

¹*i.e.*, taking the case $n = 2$ it holds that for all $x_1, x_2, y_1, y_2 \in [0, 1]$, with $0 \leq x_1 \leq y_1 \leq 1$ and $0 \leq x_2 \leq y_2 \leq 1$, we have $Ag(x_1, x_2) + Ag(y_1, y_2) \geq Ag(x_1, y_2) + Ag(y_1, x_2)$.

Theorem 1. Consider two arbitrary random variable X_1 and X_2 with marginal cdfs F_1 , F_2 and joint cumulative distribution function (j-cdf) F , then there exists a copula C such that:

$$F(x_1, x_2) = C\left(F_1(x_1), F_2(x_2)\right). \quad (5.5)$$

Let $R(x)$ denotes the range of the cdf x , C is uniquely determined on $R(F_1) \times R(F_2)$ if F_1 and F_2 are continuous [123].

One can also rewrite equation (5.5) for $(u_1, u_2) \in [0, 1]^2$ as

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)). \quad (5.6)$$

5.3.3 K-plot graphical representation

Kendall plot, also called K-plot, is a goodness-of-fit technique for copulas introduced recently by Genest and Boies in [13]. It is a rank-based graphical tool inspired by the underlying concept of Q-Q plot (so called Quantile-Quantile plot) for detecting dependencies in a bi-variate data. Let $\{(x_{1,1}, x_{2,1}), (x_{1,2}, x_{2,2}), \dots, (x_{1,N}, x_{2,N})\}$ be a set of observations of the joint random variables X_1 and X_2 , this method consists in transforming this pair of data into $\{(W_{1:N}, H_1), \dots, (W_{N:N}, H_N)\}$ by following these steps.

1. For all i in $\{1, \dots, N\}$, calculate H_i as follows:

$$H_i = \frac{1}{N-1} \text{card}\{j \neq i : x_{1,j} \leq x_{1,i}, x_{2,j} \leq x_{2,i}\}. \quad (5.7)$$

2. Sort the H_i , such that $H_{(1)} \leq H_{(2)} \leq \dots \leq H_{(N)}$ to obtain the rank statistics of the observations that corresponds to the quantile-sample.
3. Calculate the theoretical quantiles $W_{i:N}$ that corresponds to the expectation of the i -th order statistics of a sampling of a random variable of cdf $F_{X_1, X_2}(X_1, X_2)$, considered to be equal to $F_{X_1}(X_1)F_{X_2}(X_2)$. That is to say (X_1, X_2) are considered to be independent. Then, its order statistics is given for all $1 \leq i \leq N$ by:

$$W_{i:N} = N \binom{N-1}{i-1} \int_0^1 t(K(t))^{i-1} (1-K(t))^{N-i} dK(t) \quad (5.8)$$

where $K(t) = Pr(X_1 X_2 \leq t) = Pr(UV \leq t) = t - t \log(t)$.

Finally, K-plot is obtained by plotting the pair $(W_{i:N}, H_i)$. Two important particular cases that can be identified from this graphic: the case of positive dependency if the points are located above the diagonal $K(t) = t$ (which corresponds to the perfect independency case) and vice versa for negative dependency.

5.4 DST and Copulas

As previously mentioned, the belief function theory is well-known for its ability to model uncertainty and imprecision. A comprehensive study on the feasibility of adapting copulas as part of the belief function theory to combine dependent evidence was done separately by Nguyen [124] and Schmelzer [116, 125]. The basic idea is to point out the relation between random set and evidence in order to study DST within the framework of probability theory but with random variables having sets as values. In the following of this part, the most important results of this extension as well as the copulas-based conjunctive rule for the combination of consonant belief functions induced by dependent sources of evidences are presented.

5.4.1 Random sets and DST

Let (Ω, \mathcal{A}, P) be a probability space and let $(\mathbf{U}, \mathcal{U})$ be a measurable space where \mathbf{U} is the power set of Θ . One can define a finite random set $S : \Omega \rightarrow \mathbf{U} = 2^\Theta$, with the probability distribution function given by $f : 2^\Theta \rightarrow [0, 1]$ such that:

$$f(A) = P[S = A], \quad \forall A \in 2^\Theta. \quad (5.9)$$

Although this probability distribution fully characterizes the finite random set S , sometimes, it is more convenient to determine its distribution using the containment functional [126] defined as follows:

$$F(A) = P[S \subseteq A] = \sum_{B \subseteq A} f(B), \quad \forall A \in 2^\Theta. \quad (5.10)$$

This set function can be considered as the counterpart of the cumulative distribution function $P(X \leq x)$ of a random variable X , where \leq on \mathbb{R} is replaced by the inclusion relation \subseteq with which \mathbf{U} is partially ordered. Note also that one can obtain the function f from F using the so-called Möbius inversion formula:

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B). \quad (5.11)$$

If $f(\emptyset) = 0$, i.e., S is a non empty finite random set, then F is mathematically isomorphic to belief function. As a result, a close link can be established between Dempster-Shafer theory and random sets theory [127, 128]: any Basic Probability Assignment (BPA) m can be then represented by a finite random set S characterized by the couples $(A_i, m(A_i))$ such that $A_i \in 2^\Theta$ and whose distribution is given by the belief function.

5.4.2 Conjunctive combination of dependent consonant belief functions using copula

Traditionally, in order to combine two BPAs m_1 and m_2 defined respectively in the frames of discernment Θ_1 and Θ_2 , we must firstly find their joint basic probability assignment $m : 2^{\Theta_1} \times 2^{\Theta_2} \rightarrow [0, 1]$ that encode the dependency present between both. In [124], Nguyen presents results for modelling this dependency using copula. In addition, he proves that given a joint basic probability assignment m there exists a copula linking m to its margins as follows:

$$m_1(.) = \sum_{A_2 \in 2^{\Theta_2}} m(., A_2) \quad \text{and} \quad m_2(.) = \sum_{A_1 \in 2^{\Theta_1}} m(A_1, .). \quad (5.12)$$

Its approach benefits from the canonical random set representations of belief functions. Let $A_i^1, A_i^2, \dots, A_i^{n_i}$ be the enumeration of the power set 2^{Θ_i} , $i = 1, 2$, this representation consists in subdividing the probability spaces $(0, 1]$ into n_i subintervals whose lengths correspond to probability weights of m_i , $i = 1, 2$. See Alvarez [114] for further details. The marginal distribution functions can be then defined on \mathbb{R} using the above enumerations as follows:

$$F_1(x_1) = \sum_{j_1 \leq x_1} m_1(A_1^{j_1}) \quad \text{and} \quad F_2(x_2) = \sum_{j_2 \leq x_2} m_2(A_2^{j_2}). \quad (5.13)$$

These functions are piecewise constant functions increasing by $m_i(A_i^{j_i})$ at $x_i = j_i$ where $i = 1, 2$. As well, the joint distribution function can be given by:

$$F(x_1, x_2) = \sum_{j_1 \leq x_1, j_2 \leq x_2} m(A_1^{j_1}, A_2^{j_2}). \quad (5.14)$$

By Sklar's theorem, there exists a copula C' such that $F(x_1, x_2) = C'(F_1(x_1), F_2(x_2))$. However, as different orderings of the power set 2^{Θ_i} , $i = 1, 2$ can be found, the joint distribution function and thus also the copula C' depends on the used pair of enumerations.

Nguyen proves that the joint density m can be expressed in terms of the copula C' :

$$m_{C'}(A_1, A_2) = \mu_{C'}((F_1(k_1 - 1), F_1(k_1)] \times (F_2(k_2 - 1), F_2(k_2)]), \quad (5.15)$$

where k_1 and k_2 are the indices for which $A_1 = A_1^{k_1}$ and $A_2 = A_2^{k_2}$, respectively, and $\mu_{C'}$ is the C-volume of the rectangular event $[u_1, v_1] \times [u_2, v_2] \in [0, 1]^2$, $u_i \leq v_i$, $i = 1, 2$ associated to the joint focal element (A_1, A_2)

$$\mu_{C'}((u_1, v_1] \times (u_2, v_2]) = C'(v_1, v_2) - C'(v_1, u_2) - C'(u_1, v_2) + C'(u_1, u_2). \quad (5.16)$$

In order to select a single copula, Nguyen proposes to choose among all the pairs of enumeration the one who maximizes the entropy. In [125], Schmelzer proves that if the marginal belief functions bel_1 and bel_2 are minitive (*i.e.*, their associated random sets are consonant), the joint belief function

$$bel_{12}(A_1, A_2) = \sum_{B_1 \subseteq A_1, B_2 \subseteq A_2} m(B_1, B_2) \quad (5.17)$$

is then biminitive², and there exists a single copula C such that for all $A_1 \subseteq \Theta_1, A_2 \subseteq \Theta_2$ it holds that

$$bel_{12}(A_1, A_2) = C(bel_1(A_1), bel_2(A_2)). \quad (5.18)$$

²(from [125]) A function bel is called biminitive or minitive in each component if for all $A_1, B_1 \in 2^{\Theta_1}$ and $A_2, B_2 \in 2^{\Theta_2}$ such that $A_1 \cap B_1 \in 2^{\Theta_1}$ and $A_2 \cap B_2 \in 2^{\Theta_2}$ it holds that $bel(A_1 \cap B_1, A_2 \cap B_2) = \min\{bel(A_1, A_2), bel(A_1, B_2), bel(B_1, A_2), bel(B_1, B_2)\}$.

Following equations (5.18) and (5.15), two copulas C' and C can be used to describe the relation between the marginal and the joint probabilistic information. In the case of consonant belief functions, this two latter coincide only if the enumerations $2^{\Theta_1} = \{A_1^1, A_1^2, \dots, A_1^{n_1}\}$ and $2^{\Theta_2} = \{A_2^1, A_2^2, \dots, A_2^{n_2}\}$, are chosen in such a way that the focal sets of the marginal belief functions are ordered in increasing order (with respect to set inclusion). In the sequel, we are interested, in particular, to the problem of choosing copula in the fusion process of such belief functions. Suppose that there are two sources of information S_1 and S_2 and one has an information that quantifies the dependency relation between their consonant beliefs functions m_1 and m_2 . Then the joint basic probability assignment can be defined by:

$$m_C(A_1, A_2) = \sum_{B_1 \subseteq A_1, B_2 \subseteq A_2} (-1)^{|A_1 \setminus B_1| + |A_2 \setminus B_2|} C(\text{bel}_1(A_1), \text{bel}_2(A_2)), \quad (5.19)$$

Using this definition, the Conjunctive Rule based on Copula, denoted CRC, can be given in the following manner:

$$m_{1,2}^{CRC}(A) = \sum_{A_1 \cap A_2 = A} m_C(A_1, A_2), \quad (5.20)$$

According to the choice of the copula, the CRC rule allows to combine dependent beliefs, independent beliefs and intermediate cases. For example, by applying the product copula in equation (5.20), we find the conjunctive rule equation (2.11). Thus, some conjunctive-based fusion rules (such as Yager's rule [35] and Dempster's rule) may be rewritten with such dependent formalism. Dempster's rule for independent evidence is given by:

$$\begin{aligned} m_{1,2}^{DS}(A) &= \frac{1}{1-K} m_1 \odot_2(A), \quad \text{as stated in equation (2.13),} \\ &= \frac{1}{1-K} \sum_{A_1 \cap A_2 = A} m_1(A_1) \times m_2(A_2) \\ &= \frac{1}{1-K} \sum_{A_1 \cap A_2 = A} m_{C^\perp}(A_1, A_2), \end{aligned} \quad (5.21)$$

where K is defined as:

$$\begin{aligned} K &= m_1 \odot_2(\emptyset) \\ &= \sum_{A_1 \cap A_2 = \emptyset} m_1(A_1) \times m_2(A_2) \\ &= \sum_{A_1 \cap A_2 = \emptyset} m_{C^\perp}(A_1, A_2). \end{aligned} \quad (5.22)$$

5.5 Credal dependent fusion: disjunctive aggregation rule

As it is conceived, the evidential disjunctive rule allows the preserving of the set of beliefs of each source independently to its reliability. Like most combination rules, this rule does not allow aggregation of dependent sources of information. Several alternatives [129] have been proposed to overcome this deficiency by ensuring a prudent attitude and condone the real

dependence modelling. The aim of this section is to define a new aggregation rule that keeps the disjunctive behaviour for various values of dependency.

Recall that in aggregation domains [130], a typical relationship between disjunctive aggregation functions Di and conjunctive aggregation functions Co are by means of order reversing mapping (negative function) N such that

$$Di(x_1, \dots, x_n) = N\left(Co\left(N(x_1), \dots, N(x_n)\right)\right) \quad (5.23)$$

This relation is called duality and will be used here to define the dual aggregation operator to copula that offers our solution for capturing dependence relationship between sources of information to be combined disjunctively. Usually a strong negation $N(x) = 1 - x$, $x \in [0, 1]$ is used to define such an operator:

$$Di(u, v) = 1 - C(1 - u, 1 - v) \quad (5.24)$$

In the framework of copula theory [12], the dual copula D is not defined in the sense of equation (5.24). Let U, V be two random variables uniformly distributed over $[0, 1]$, if they are linked by a copula C , $C(u, v) = Pr(U \leq u, V \leq v) = Pr(\{(U \leq u) \cap (V \leq v)\})$, then the corresponding D is given by the following formula:

$$\begin{aligned} D(u, v) &= u + v - C(u, v) \\ &= Pr(\{(U \leq u) \cup (V \leq v)\}) \end{aligned} \quad (5.25)$$

For certain copulas, D coincides with Di (see section 5.6). In this case, we can say that D allows the disjunctive aggregation of dependent marginal u, v and we can define the Disjunctive Rule based on Dual Copula (DRDC) in the following manner: Suppose that, there are two sources of information S_1 and S_2 and we have information that quantifies the dependence relation between their belief functions m_1 and m_2 . Then DRDC is given by:

$$m_{1,2}^{DRDC}(A) = \sum_{A_1 \cup A_2 = A} m_D(A_1, A_2), \quad (5.26)$$

where

$$\begin{aligned} m_D(A_1, A_2) &= \sum_{B_1 \subseteq A_1, B_2 \subseteq A_2} (-1)^{|A_1 \setminus B_1| + |A_2 \setminus B_2|} D(bel_1(A_1), bel_2(A_2)) \\ &= \sum_{B_1 \subseteq A_1, B_2 \subseteq A_2} (-1)^{|A_1 \setminus B_1| + |A_2 \setminus B_2|} bel_1(A_1) + bel_2(A_2) \\ &\quad - C(bel_1(A_1), bel_2(A_2)). \end{aligned} \quad (5.27)$$

5.6 The choice of the family of copulas

As we have shown in the foregoing section, the copula fits to Dempster-Shafer framework in the case of combining non distinct consonant belief functions. So it remains now to make the right choice of the copula in the credal fusion process from the various existing copulas. Usually, the choice of the copula depends on the data set used. Indeed, a particular copula may be fits

better for one data set than for another. To the best of our knowledge, there does not exist in the literature an efficient method for selecting copula. Generally, the use of parametric copula is recommended because it can be adapted to existing data by estimating its parameters properly. Nevertheless, nothing can prove that this choice of parameters guarantees the convergence of copula to the real structure of the underlying dependency of the data. In this work, we choose to use the family of Archimedean copulas defined as follows:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)), \quad (5.28)$$

where $\varphi : [0, 1] \rightarrow [0, +\infty)$ is a continuous, strictly decreasing function with $\varphi(1) = 0$.

In [131], Alsina et al. consider that any commutative associative copulas are t-norms; or equivalently any t-norms which satisfy the 1-Lipschitz condition³ are copulas. So it is clear that Archimedean copulas are also t-norms since they belong to the overlap of copulas and t-norms (see Figure 5.1). As a result, it is easy to demonstrate that CRC shares all interesting properties of t-norms, including associativity. This point constitutes the first reason for choosing this family of copulas.

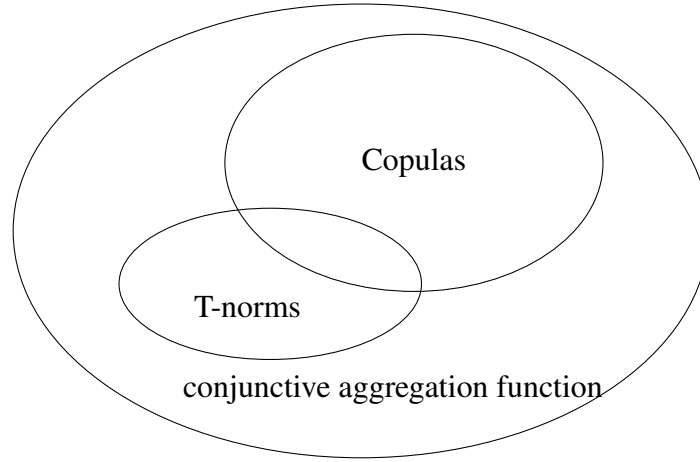


Figure 5.1: Copulas and t-norms as aggregation functions.

Furthermore, Archimedean copulas are able of capturing and modelling various ranges of dependencies. Indeed, depending on their generating functions (see for instance an overview in [12]), several copulas can be easily derived. Here, we recall only those bi-dimensional of Frank, Clayton and Gumbel which are the most widely used in applications.

- Gumbel copula (1960) is an Archimedean copula which exhibits greater dependence in the positive tail than in the negative so it represents an appropriate choice to modelling strongly correlated marginal at high values but less correlated at low values. Gumbel copula is given by its generator function

$$\varphi_r(t) = (-\ln(t))^r \quad (5.29)$$

to yield

$$C_r^G(u, v) = \exp \left[- \left((-\ln(u))^r + (-\ln(v))^r \right)^{\frac{1}{r}} \right], \quad (5.30)$$

³i.e., for any $(u_1, u_2, v_1, v_2) \in [0, 1]^4$ it holds that $C(u_1, v_1) - C(u_2, v_2) \leq |u_1 - u_2| + |v_1 - v_2|$.

where $r \in [1, +\infty)$ is the parameter of the copula. The value $r = 1$ reflects the perfect independency between the marginal of the distribution and the great values of r reflects that Gumbel copula approaches Fréchet-Hoeffding upper bound [12].

- Clayton copula (1978) [132] is well known for its ability to capture lower tail dependency. This copula is given by its generator

$$\varphi_r(t) = \frac{1}{r}(t^{-r} - 1) \quad (5.31)$$

to yield

$$C_r^C(u, v) = \frac{uv}{1 - r(1 - u)(1 - v)}, \quad (5.32)$$

where $r \in [0, +\infty)$ is the parameter of the copula. $r = 0$ occurs when the marginal distributions are independent. $r \rightarrow +\infty$ makes Clayton copula approximating Fréchet-Hoeffding upper bound.

- Frank copula (1979), introduced in [133], plays a particular role in the evaluation of conjunctions under dependency between marginal and it is defined with

$$\varphi_r(t) = -\ln \left(\frac{\exp(-rt) - 1}{\exp(-r) - 1} \right)$$

as follows:

$$C_r^F(u, v) = \frac{-1}{r} \ln \left(1 + \frac{(\exp(-ru) - 1)(\exp(-rv) - 1)}{\exp(-r) - 1} \right), \quad (5.33)$$

where $r \in (-1, +\infty)$. The value $r = 0$ reflects the perfect dependency, the great values of r reflects opposite dependency and independency corresponds to the value $r = 1$.

The link between Kendall's τ and the parameter of any Archimedean copula of generator function φ is given by:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du.$$

For the DRDC, only Frank copula can be used to retain the property of a t-conorm. In fact, it has been demonstrated in [133] that is the only one that verifies the matching between the dual copula and the dual aggregation operator to copula for Archimedean class of copulas.

5.7 Experiments

In order to prove the effectiveness of using copula in combining dependent consonant belief functions, a credal classification problem is used here. The experiments have been done on benchmark and generated data sets. The benchmark data set is provided by the University of California - Irvine (UCI) Machine Learning Repository⁴. The simulated data set consists of three overlapped Gaussian distributions. For each data set, only two features (*i.e.*, sources

⁴The data set is available at <http://archive.ics.uci.edu/ml>.

of information) are used to discriminate its classes. The first part of this section explains the strategy to be followed for estimating consonant belief functions from the used data sets and the second part is devoted to the results of the application of the appropriated copula in the fusion process, as well as to their interpretation.

5.7.1 Consonant mass functions estimation

Let Θ be a referential and π be a possibility distribution, which assigns to each singleton $\theta \in \Theta$ a possibility degree of its occurrence $\pi(\theta) \in [0, 1]$ with:

- $\pi(\theta) = 0$ means that θ is rejected; it is totally impossible;
- $\pi(\theta) = 1$ means that θ is completely possible.

From this distribution, a possibility Π and necessity N measures can be defined as follows:

$$\begin{aligned}\Pi(A) &= \max_{\theta \in A} \pi(\theta) \\ N(A) &= 1 - \Pi(\bar{A})\end{aligned}$$

In [134], Dubois and Prade pointed out that π can be modelled as a consonant random set, since the measure N is a special case of the credibility measure. Let $\pi_1 = 1 > \pi_2 > \dots > \pi_N$ be the distinct values taken by π and by convention $\pi_{N+1} = 0$. Let A_i denote the π_i -cut of the possibility distribution π . Then, we have, for any non-empty subset A of Θ :

$$m(A) = \begin{cases} \pi_i - \pi_{i+1}, & \text{if } A = A_i, i = 1, \dots, r \\ 0, & \text{otherwise} \end{cases}$$

Now, the problem of estimating consonant mass function can be reduced to the estimation of a possibility distribution to which we apply this transformation.

5.7.2 Results and discussion

5.7.2.1 Benchmark data set

In this work, the dependence between sources is assumed to be the dependence between their data. Let $(x_{1,1}, x_{2,1}), (x_{1,2}, x_{2,2}), \dots, (x_{1,N}, x_{2,N})$ be a set of observations of the joint random variables (RV) X_1 and X_2 that characterize respectively the sources S_1 and S_2 where each observation represents the object to be classified in the data set. The dependence between sources can be computed using Kendall's rank correlation coefficient [14] defined as a concordance versus discordance measure of order statistics:

$$\tau = Pr((X_1 - X_1^*)(X_2 - X_2^*) > 0) - Pr((X_1 - X_1^*)(X_2 - X_2^*) < 0) \quad (5.34)$$

where X_1^* and X_2^* are 2 RVs following the same laws as X_1 and X_2 (respectively) but considered as independent. An empirical estimator exists for Kendall's τ from a set of N joint observations

of X_1 and X_2 :

$$\tau = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N-1} x_{1,ij} x_{2,ij}}{\binom{N}{2}} \quad (5.35)$$

with

$$x_{1,ij} = \begin{cases} 1, & \text{if } x_{1,i} \leq x_{1,j} \\ -1, & \text{otherwise} \end{cases} \quad x_{2,ij} = \begin{cases} 1, & \text{if } x_{2,i} \leq x_{2,j} \\ -1, & \text{otherwise} \end{cases}$$

After having calculated Kendall's τ needed in the estimation of the parameter of the copula. Now it remains to determine the best-fitting copula for the corresponding data set. This

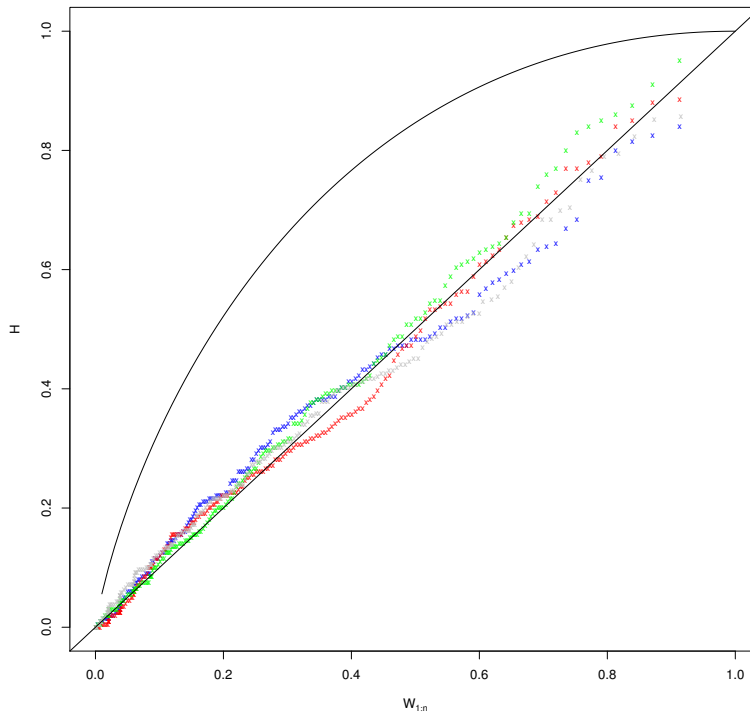


Figure 5.2: K-Plot in a quasi-independent case. Sample from the seeds training base, with $\tau = 0.0097$, are shown in grey, simulated models with $\tau = 0.0097$ is shown in blue for Frank copula, in red for Clayton copula and Green for Gumbel copula. Best goodness-of-fit is given by Frank copula.

goodness-of-fit is implemented via K-plot graphical method [13]. It consists of plotting two rank statistics in a similar way as the Quantile-to-Quantile plot which is known for 1D distribution model validation. Here, rank statistics are computed from the sample of the data set (y -axis H of the graph) and compared to the rank statistics of the joint observation but with an independent hypothesis (x -axis $W_{i:n}$ of the graph).

Some comparisons may be performed by simulating data set with the same dependency parameter (here the same Kendall's τ) by using a specific copula model. Here, Frank, Clayton

and Gumbel copulas have been investigated. The copula whose plot best fits the sample plot is considered to be the one that describes the best the mutual dependency of the two considered sources. Frank, Clayton and Gumbel copulas come from the same family for Archimedean copulas and are described in section 5.6. The following two examples are extracted from the seeds training base⁵.

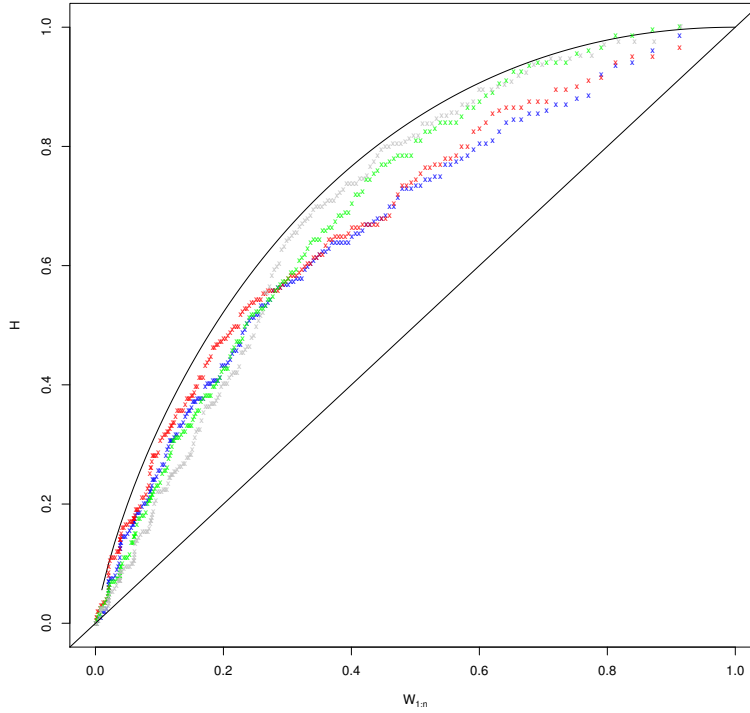


Figure 5.3: K-Plot in a more dependent case. Sample from the seeds training base, with $\tau = 0.7132$, are shown in grey, simulated models with $\tau = 0.7132$ is shown in blue for Frank copula, in red for Clayton copula and Green for Gumbel copula. Best goodness-of-fit is given by Gumbel copula.

Figure 5.2 shows an example of a K-plot that characterizes the dependency between 2 sources by comparing its joint order statistics to the equivalent order statistics through an independent hypothesis. The case shown in Figure 5.2 corresponds to an almost independent case, as $\tau = 0.0097$ (the case $\tau = 0$ stands for perfect independency). The sample of this K-plot follows the diagonal that characterizes the area of independent sources. With this low value of τ , all the considered copulas also follow the diagonal line as those copulas can handle the independent case. Nevertheless, the numerical distance between the curve of the considered data set and the copula-based simulated sample shows that Gumbel copula achieves the best goodness-of-fit (distance to sample equals 0.0017 for Gumbel copula, 0.0028 for Frank copula and 0.0048 for Clayton copula, with parameter adjusted for $\tau = 0.0097$).

In Figure 5.3, a more dependent case is shown with $\tau = 0.7132$. Here, Frank copula shows the best goodness-of-fit (distance to sample of 0.0024 for Frank copula, 0.0038 for Clayton

⁵The seeds database is available at <http://archive.ics.uci.edu/ml/datasets/seeds>.

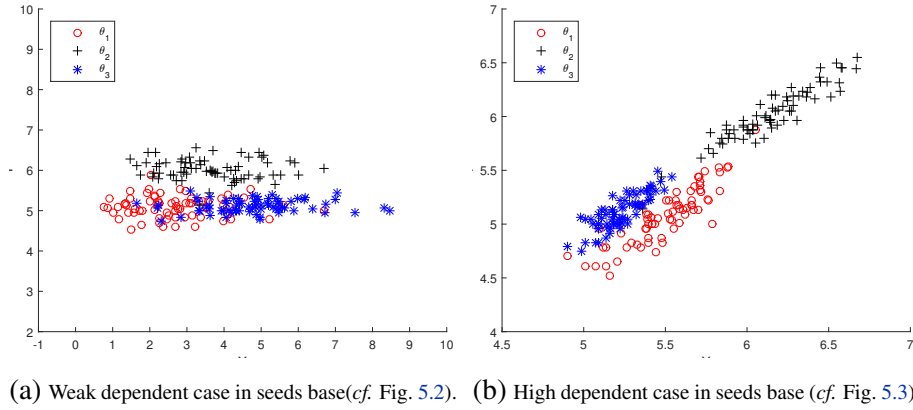


Figure 5.4: Reference data.

copula and 0.0057 for Gumbel copula) which shows that the dependency structure does not have the same behaviour as the one of Figure 5.2. In Figure 5.3, the plot of observations and all investigated Archimedean models of dependency are located on the upper curve that represents the case of perfect dependency. In this example, the so-called Gumbel Copula (defined in equation (5.30)) has been found to fit the best dependence between dependent sources of information.

Figures 5.5 and 5.6 illustrate the classification results for the different elements of seeds training base studied according respectively to the weak dependent sources with $\tau = 0.0097$ and to the high dependent sources with $\tau = 0.7132$. Figure 5.4 shows the reference data for each of those cases. The method for estimating the possibility distribution transformed to mass functions used in this section is the one described by Klir in [135]. For decision making on simple classes the criterion of the maximum of pignistic probability is used.

For the weak dependent case, we notice that the results of the classification of the copulas-based fusion rules CRC and DRDC are identical to those given by CR and DR, respectively. This is expected because the behaviour of Archimedean copulas for the weak dependency tends to the one of the product copula (so that for example equation (5.20) is becoming similar to equation (2.11)). The Bold rule [129] presents major differences from the two other disjunctive rules. Indeed, we can notice in Figure 5.5-(e) that it has the highest false classification. Obviously, such a result is a logical consequence of the non-specific treatment of the dependence structure between the sources.

For the high dependent case, the best accuracy is given by the proposed combination rules as well as the other employed rules. However, it seems that the nature of disjoint classes (see Figure 5.4-(b)) used in this test do not appear the advantage of expressing or modelling dependence in the fusion step. It is on this point that the second test, given below, was designed.

5.7.2.2 Generated data set

In this second test, three data sets as shown by Figure 5.7 are applied. Each one of them corresponds to a three 2D overlapped Gaussian distributions $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Table 5.1 displays the different means vectors and covariance matrices used in order to vary the dependence degree of

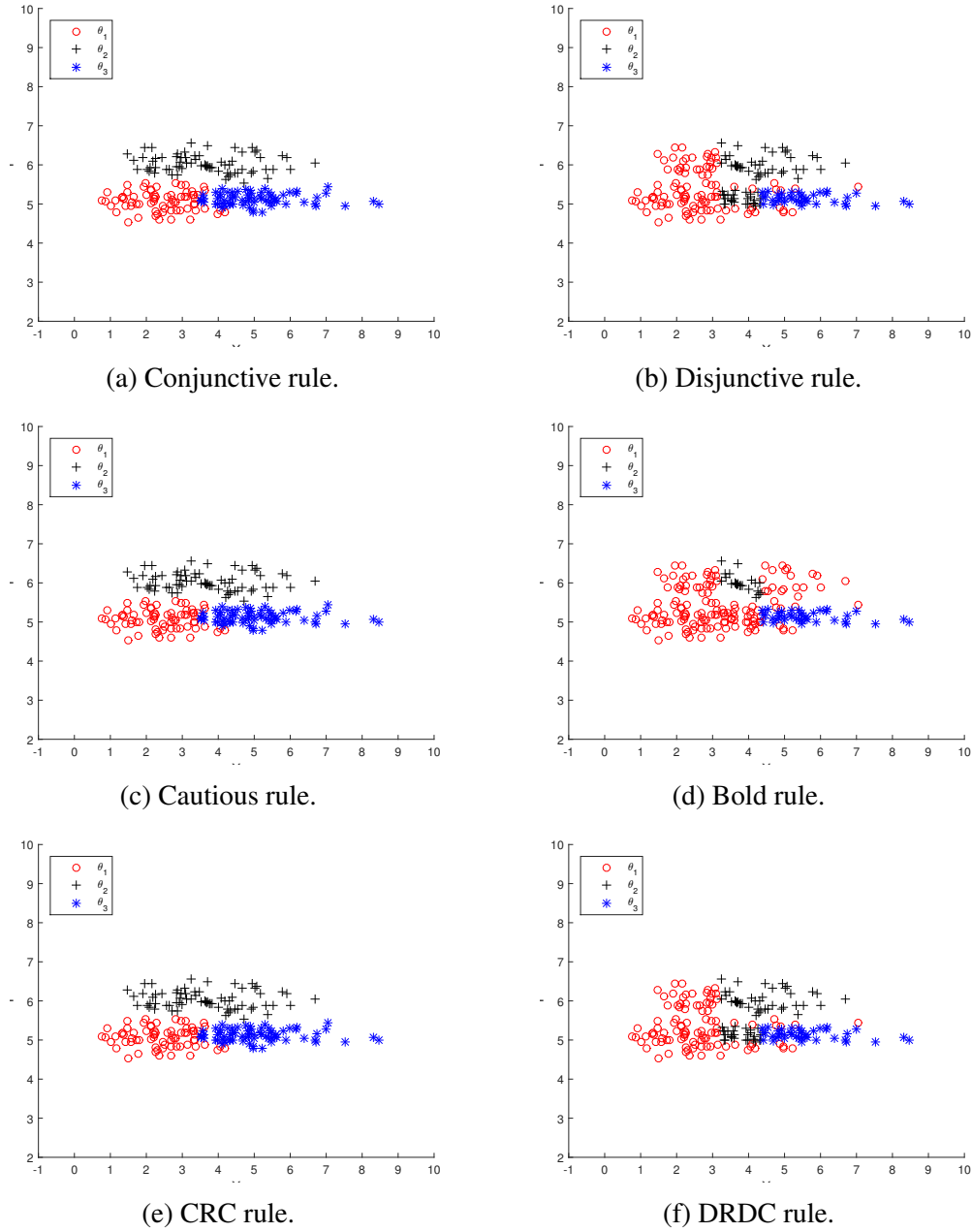


Figure 5.5: Classification results for the weak dependent case in seeds base (*cf.* Figure 5.2). X stands for the “asymmetry coefficient” and Y for the “length of kernel groove” components of the seeds database.

data. We have generated 1500 random samples for each of the three classes θ_1 , θ_2 and θ_3 .

As in the first test, a credal classification problem is considered in order to evaluate the proposed copula-based strategy of fusion. The rates of correct classification are given in Table 5.2 using the confusion matrix. As we can see, CRC and DRDC present promising results. Indeed by comparing our CRC rule to the cautious and the conjunctive rules, it can be noticed that all class detections have been improved (see in bold the first line of this table). Moreover, it is worth noting that the more the value of dependency, the more the improvement of the results

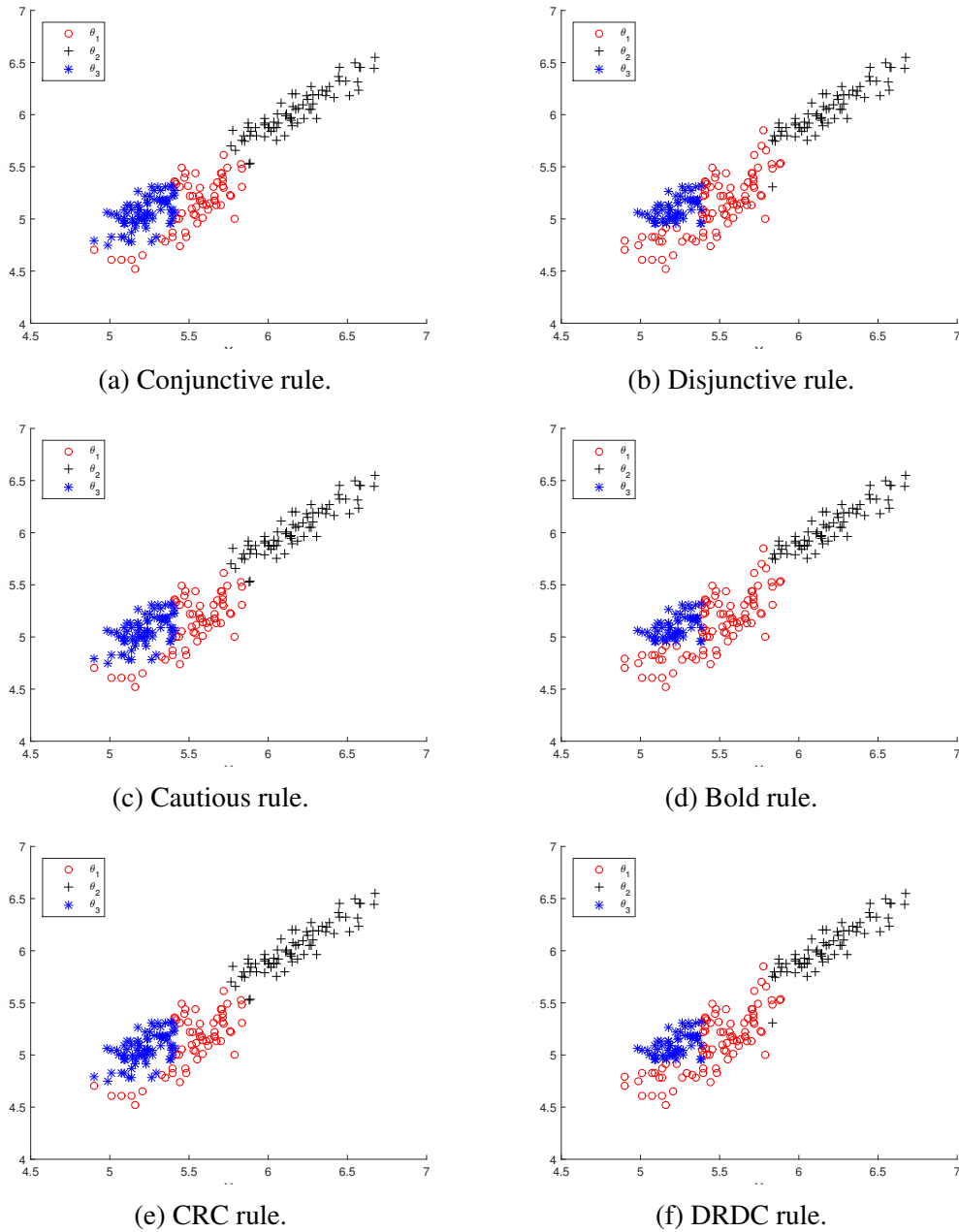


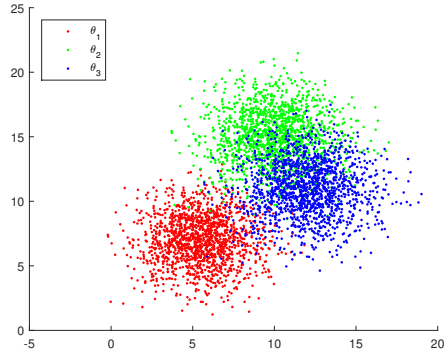
Figure 5.6: Classification results for the high dependent case in seeds base (*cf.* Figure 5.3). X stands for the “length of kernel” and Y for the “length of kernel groove” components of the seeds database.

of CRC and DRDC rules. The cautious rule is always below and it gives the same result of CR even if the independence hypothesis is not verified.

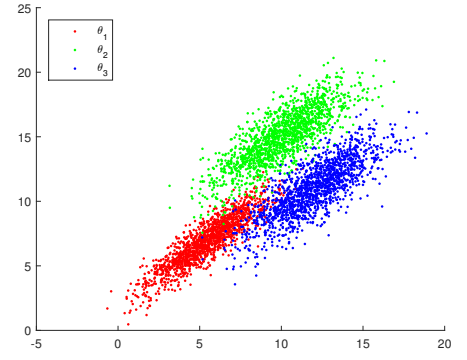
Similarly, the DRDC appears always better than the bold and the disjunctive rules. In addition, it seems that BR degrade the accuracy of the classification independently of the degree of dependence (74.16% versus 74.56% for the data set 1 ($\tau = 0.2880$) and 77.31% versus 78.33% for the data set 3 ($\tau = 0.4423$)).

Table 5.1: Means and covariances of generated data sets.

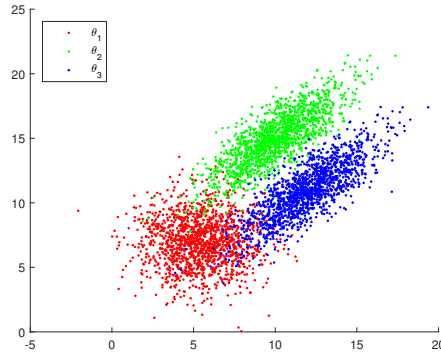
Class	Data set 1		Data set 2		Data set 3	
	μ	Σ	μ	Σ	μ	Σ
θ_1	[5.5 7]	$1.9^2 * [1 \ 0; 0 \ 1]$	[5.5 7]	$1.9^2 * [1 \ 0.9; 0.9 \ 1]$	[5.5 7]	$1.9^2 * [1 \ 0; 0 \ 1]$
θ_2	[10 15]	$2.1^2 * [1 \ 0; 0 \ 1]$	[10 15]	$2.1^2 * [1 \ 0.8; 0.8 \ 1]$	[10 15]	$2.1^2 * [1 \ 0.8; 0.8 \ 1]$
θ_3	[12 11]	$2.1^2 * [1 \ 0; 0 \ 1]$	[12 11]	$2.1^2 * [1 \ 0.8; 0.8 \ 1]$	[12 11]	$2.1^2 * [1 \ 0.8; 0.8 \ 1]$



(a) Data set 1 ($\tau = 0.2880$)



(b) Data set 2 ($\tau = 0.5266$)



(c) Data set 3 ($\tau = 0.4420$)

Figure 5.7: Generated data sets.

5.8 Conclusion

Evidence theory has often been interpreted as a generalization of probability theory thanks to the introduction of belief functions. This point of view has been widely disputed in the literature because of certain differences, the most important of which is Dempster's popular aggregation rule that assumes the distinction between beliefs to be combined. In this paper, we are interested in the extension of the copula that provides an efficient means for the dependency modelling to the framework of belief functions. One of the biggest challenges that we encountered in the works present in the literature is the choice of the copula. The work presented in this

Table 5.2: Classification results.

Combination rules		Dependency value between sources		
		$\tau = 0.2880$	$\tau = 0.4423$	$\tau = 0.5309$
CRC	eq. (5.20)	88.38%	93.58%	92.56%
Cautious Rule	eq. (2.22)	88.33%	93.36%	91.51%
CR	eq. (2.11)	88.33%	93.36%	91.51%
DRDC	eq. (5.26)	85.64%	89.31%	87.47%
Bold Rule	eq. (2.23)	74.16%	77.31%	76.58%
DR	eq. (2.15)	74.56%	78.33%	75.04%

study provides a solution for the disjunctive and conjunctive combination in the case where the information assessed by a source of information is encoded in the form of consonant mass functions. From the experiments performed on benchmark and generated data sets, we have shown the efficiency of our copulas-based combination rules. Indeed they give more coherent results than the ones given by the other rules.

General conclusion

In this thesis, we were interested in evaluating the potential contribution of the credibility theory to the modelling and fusion of heterogeneous remote sensing data. More precisely, our objective is to combine the information provided by high spatial resolution optical and radar images in order to achieve a joint classification. From a methodological point of view, we studied the possibility of setting up new techniques intervening in the different phases of this fusion process realization, such as the modelling, the estimation and the combination of beliefs. This led us to propose three original contributions, which will be summarized in the following.

Synthesis of the works undertaken

As a first step, namely in chapter 3, a novel method dedicated to the estimation of mass functions is introduced. The proposed approach has the particularity of processing the large volume of data characterizing high-resolution remote sensing images, as well as data acquired using other types of sensors. Based on the Kohonen map, a simplification of the input space was applied to intelligently manage the assignment of the masses. Contrary to existing approaches, our method exploits the whole conceptual power of credibilist theories by allowing to deal with uncertain and paradoxical data. Indeed, it calculates the supports of confidence for the singletons, conjunctive and disjunctive classes. In this way, we guarantee accurate and faithful modelling of the imperfect data used for the interpretation of the observed scene. The effectiveness and accuracy of the proposed method were confirmed by a series of comparisons with literature methods on benchmark databases and satellite data.

Then, chapter 4 is dedicated to discuss the problem of the fusion of data derived from optical and radar sensors. The SAR /optical information fusion is explored in this study for the joint classification of agricultural areas. The developed method is mainly based on the adaptation of the assignment of mass functions already introduced in the previous chapter to handle heterogeneous data. Indeed, a hybrid training of the Kohonen map is proposed. We have constructed two variants of this approach to treat missing data due to the presence of clouds and shadows as well as cloudless optical data. A pair of SPOT-5 and RADARSAT-2 images is used in the experimentation and the application of the proposed technique on the Beauce region in France shows very promising results in terms of classification precision and reconstruction data that miss the optical images.

Finally, in chapter 5, we have tackled the problem of combining information induced by belief functions in the case where their sources of information do not necessarily satisfy the independence hypothesis. To achieve this purpose, we propose to calculate the joint mass function using the copula that best describes the dependence structure between the two marginal mass functions to be combined. Then, two combination operators have been constructed by taking into account this information about the existing dependence (*i.e.*, the joint mass) to fuse the dependent beliefs encoded by consonant belief functions in a conjunctive and disjunctive way. The use of the proposed rules in a classification problem show a significant increase in precision compared to the prudent and the bold rules of Denœux usually used.

Perspectives

Many interesting research tracks are possible as a result of this work, among them we cite:

- The introduction of conjunctions between classes within DS_mT gives it a particular richness and flexibility to model the imperfections and the paradox of the data. Thus, it will be interesting to adapt our approach presented in chapter 4 to DS_mT framework in order to benefit from the semantics of the class conjunctions in the joint classification of highly heterogeneous sources presenting a strong conflict.
- Let us recall that the approach proposed in chapter 4 is entirely satisfied with the dissemination of the evidential knowledge derived from the textural information registered by the radar sensors. It will be interesting to study the advantage of analyzing the polarimetric capabilities of SAR data that have found wide applications, especially in crops discrimination in a farming area.
- The study presented in chapter 5 focuses on the fusion of two dependent sources only. The extension to the fusion of 3 or more sources of information brings new problems that have to be tackled. In fact, accounting for the dependence in nD (*i.e.*, defining a copula $C(u_1, u_2, \dots, u_n)$), the 2-by-2 dependency structure may not be equivalent between couple of sources so that it induces a conjunctive rule that is not associative anymore. This point constitutes one of our main future works.
- Likewise in chapter 5, our method was mainly tested on benchmarks and synthetic data sets. Hence, it is promising to investigate the performance of the CRC rule of combination to different remote sensing applications like optical and radar data fusion.
- The introduced CRC rule gives a conflictual mass which seems to describe the disaccord between the dependent sources of information well. Hence, it will be interesting to study the behaviour of this mass depending on the number of combined BPAs. Moreover, it is possible to study new rules of combination that allow the conflict redistribution.

List of Figures

1	Un schéma représentatif de la carte auto-organisatrice de Kohonen.	5
2	Observations dans l'espace d'entrée et leurs projections dans la carte de Kohonen. Notez que les neurones w_x et w_{C_k} peuvent être localisés sur la carte grâce à leur indice de localisation (m, n) ou dans \mathbb{R}^p avec leur valeur p -composants.	6
3	Le schéma général de l'approche proposée.	8
4	L'image multispectrale (a) et l'image radar (b) acquises sur la région Beauce en France.	10
5	Résultats de la classification jointe calculés par l'application du maximum de la probabilité pignistique sur toutes les classes simples.	10
6	Base des données générées.	12
2.1	Venn diagram of a free model for a 3D frame.	35
2.2	Venn diagram of a hybrid model for a 3D frame.	37
3.1	Schematic of a 11×11 Kohonen's map. Several topological neighborhood $\mathcal{N}_{w_x}(t_i)$ of the winning neuron w_x are drawn. The size is decreasing with the number of iterations ($t_1 < t_2 < t_3$) during the training phase, according to (3.10).	46
3.2	Observations in the feature space and their projections into Kohonen's map. Note that the neurons w_x and w_{C_k} can be located on the map through their location index (m, n) or in \mathbb{R}^p with their p -component value.	48
3.3	Simple case of conjunction between two class in the map.	49
3.4	Behavior of $m(x \in \theta_k \cap \theta_\ell)$ with γ , according to equation (3.14).	50
3.5	Disjunction between two class: (a) non ambiguous case, (b) ambiguous case.	50
3.6	Shape of equation (3.15) for some value of β	51
3.7	Simple simulation of a four-class manifold with an outlier. Black circles: samples of the data set in \mathbb{R}^2 . Red bullets: locations of the neurons of Kohonen's map. Blue lines: SOM projected in the feature space. Green bullets: K-means class centers.	52
3.8	Simple simulation of Figure 3.7 with its equivalent in the SOM geometry. Green bullets on the map correspond to the winning neurons w_{C_k} of the class centers C_k , blue bullets, on the map in (b), correspond to the location of the neurons that are located in between classes in \mathbb{R}^2 . The black neuron at the top left of the map corresponds to the winning neuron of the sample rounded with a brown ellipse at location $(0.2, 0.6)$	53

3.9	Computation time depending on the feature space dimension. SOM-based approach is more appropriated for processing large amount of data than ECM.	56
3.10	False color composite of the SPOT image. ©CNES.	57
3.11	Classification results in 2^Θ with decision by maximum of pignistic probability over all simples hypotheses: (a) SOM-based BBA. (b) DC results.	58
3.12	Classification results in 2^Θ with decision by maximum of pignistic probability over all simples hypotheses and all disjunctions of hypotheses: (a) SOM-based BBA. (b) DC results.	59
3.13	Classification results in 2^Θ with decision by maximum of pignistic probability: (a) SOM-based approach. (b) ECM.	60
3.14	Classification results in D^Θ with decision by maximum of generalized pignistic probability over all simples hypotheses and all conjunctions of hypotheses: (a) SOM-based approach. (b) DC results.	61
3.15	Credal classification results in D^Θ through the SOM-based approach: maximum of generalized pignistic probability.	63
4.1	Data fusion levels (a) Low-level fusion or pixel fusion (b) Attributes fusion (c) High-level fusion or decision fusion.	67
4.2	Credal fusion framework between a reliable optical multispectral image and a SAR observation. A first coarse joint classification is performed, yielding $C_{MS \oplus \kappa_{SAR}}$ which guarantees class homogeneity between the 2 sensors. Kohonen's map is trained from optical multispectral data I_{MS} to yield SOM_{MS} , and then from parameters extracted from the radar data with an enslaved constraint on the location of the neurons of $SOM_{\kappa_{SAR} MS}$. From Kohonen's maps, BBA is performed and then specific discounting operators are applied on the mass depending on their reliability (that yields m'_{MS} and m'_{SAR}). Fusion is performed by PCR6 rule and decision-making is ensured with the maximum of Pignistic probability to yield a joint land cover classification.	70
4.3	Multispectral (a) and radar (b) images acquired over Beauce, France, used in the experiments.	75
4.4	$I_{\kappa_{SAR}}$ image corresponding to the SAR texture information with false color composition: $RGB=(\mu, \sigma, \beta_1)$.	76
4.5	Unsupervised K-means classification results (with $K = 5$ classes) applied jointly on multispectral and SAR information.	77
4.6	SOM maps of size 65×65 neurons. The map (a) has been trained with the optical data only, while the map (b) has been enslaved to map (a) and trained with radar data. Hence, co-located neurons bring the same ground information between the 2 maps.	77
4.7	Joint classification results with decision by maximum of pignistic probability over all simples hypotheses.	78
4.8	Results for joint classification with simulated cloud cover. The decision is performed by the maximum of pignistic probability over all simples hypotheses.	79

4.9	Zoom of Figure 4.8-(b). Correct Classification Rate (CCR) of 72.28 % for masked zone 1 and 75.60% for masked zone 2.	80
5.1	Copulas and t-norms as aggregation functions.	92
5.2	K-Plot in a quasi-independent case. Sample from the seeds training base, with $\tau = 0.0097$, are shown in grey, simulated models with $\tau = 0.0097$ is shown in blue for Frank copula, in red for Clayton copula and Green for Gumbel copula. Best goodness-of-fit is given by Frank copula.	95
5.3	K-Plot in a more dependent case. Sample from the seeds training base, with $\tau = 0.7132$, are shown in grey, simulated models with $\tau = 0.7132$ is shown in blue for Frank copula, in red for Clayton copula and Green for Gumbel copula. Best goodness-of-fit is given by Gumbel copula.	96
5.4	Reference data.	97
5.5	Classification results for the weak dependent case in seeds base (<i>cf.</i> Figure 5.2). X stands for the “asymmetry coefficient” and Y for the “length of kernel groove” components of the seeds database.	98
5.6	Classification results for the high dependent case in seeds base (<i>cf.</i> Figure 5.3). X stands for the “length of kernel” and Y for the “length of kernel groove” components of the seeds database.	99
5.7	Generated data sets.	100

List of Tables

1	Caractéristiques des bases de données UCI utilisés pour la comparaison.	7
2	Résultats de la classification avec une estimation des masses par EVCLUS, ECM et l’approche proposée.	7
3	Résultats de la classification des données simulées.	13
1.1	Contributions of this thesis.	18
2.1	Example of some special classes of mass functions where the conditions imposed by their definitions are putted in boldface.	24
2.2	The sequence of Dedekind’s numbers.	35
2.3	Memory size requirements for D^\ominus	37
3.1	Characteristics of the UCI data sets used for comparison.	54
3.2	Classification results of SOM-based BBA in 2^\ominus for different value of β	55
3.3	Classification results in 2^\ominus of EVCLUS, ECM and SOM-based BBA with decision by the maximum of pignistic probability.	55
3.4	DST legend used on classification results of Figure 3.12.	58
3.5	Quantitative results in 2^\ominus obtained using the confusion matrix for DC approach.	59
3.6	Quantitative results in 2^\ominus obtained using the confusion matrix for the proposed SOM-based approach.	60
3.7	Legend used for classification D^\ominus shown in Figure 3.14.	62
3.8	Quantitative results in D^\ominus obtained using the confusion matrix for DC approach.	62
3.9	Quantitative results in D^\ominus obtained using the confusion matrix for the proposed SOM-based approach.	62
4.1	Haralik features computed from the SAR image.	71
4.2	Local moments measures computed from the SAR image.	72
4.3	Quantitative results obtained using the confusion matrix.	79
5.1	Means and covariances of generated data sets.	100
5.2	Classification results.	101

Bibliography

- [1] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [2] G. Shafer. *A mathematical theory of evidence*. Princeton university press, Princeton, New Jersey, 1976.
- [3] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264, 1988.
- [4] F. Smarandache and J. Dezert (Eds.). *Advances and Applications of DSmT for Information Fusion (Collected Works)*. American Research Press, Rehoboth, 2004.
- [5] J. Dezert. Foundations for a new theory of plausible and paradoxical reasoning. *Information and Security*, 9:13–57, 2002.
- [6] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.
- [7] M. A. Kraaijveld, J. Mao, A. K. Jain, et al. A nonlinear projection method based on kohonen’s topology preserving maps. *IEEE Transactions on Neural Networks*, 6(3):548–559, 1995.
- [8] J. Inglada and G. Mercier. A new statistical similarity measure for change detection in multitemporal sar images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1432–1445, 2007.
- [9] T. Denœux and M.-H. Masson. EVCLUS: EVidential CLUSstering of proximity data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):95–109, 2004.
- [10] M.-H. Masson and T. Denœux. ECM: An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384–1397, 2008.
- [11] I. Hammami, J. Dezert, G. Mercier, and A. Hamouda. On the estimation of mass functions using self organizing maps. In *Belief Functions: Theory and Applications - Third International Conference, BELIEF, Oxford, UK*, pages 275–283, 2014.
- [12] R. B. Nelsen. *An Introduction to Copulas*. Lecture notes in statistics. Springer, 1999.
- [13] C. Genest and J.-C. Boies. Detecting dependence with Kendall plots. *The American Statistician*, 57(4):275–284, 2003.

-
- [14] M. G. Kendall. *Rank correlation methods*. Griffin, London, 1948.
- [15] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.
- [16] L. A. Zadeh. Fuzzy sets. *Fuzzy Sets and Systems*, 8(3):338–353, 1965.
- [17] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.
- [18] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Kluwer Academic/Plenum Publishers, 1988.
- [19] D. Dubois. Possibility theory and statistical reasoning. *Computational Statistics and Data Analysis Vol.*, 57:47–69, 2006.
- [20] Q. Dong and X. Liu. Risk assessment of water security in haihe river basin during drought periods based on D-S evidence theory. *Water Science and Engineering*, 7(2):119–132, 2014.
- [21] I. Hammami, G. Mercier, A. Hamouda, and J. Dezert. Kohonen’s map approach for the belief mass modeling. *IEEE Transaction on Neural Networks and Learning Systems*, 27(10):2060–2071, 2016.
- [22] I. Hammami, G. Mercier, and A. Hamouda. The Kohonen map for credal classification of large multispectral images. In *IEEE Geoscience and Remote Sensing Symposium, IGARSS, Quebec City, QC, Canada*, pages 3706–3709, 2014.
- [23] I. Hammami and G. Mercier. Kohonen-based credal fusion of heterogeneous data: application to optical and radar joint classification with missing data. *IEEE Transactions on Geoscience and Remote Sensing*, 2016. (under revision).
- [24] I. Hammami, G. Mercier, and A. Hamouda. The Kohonen map for credal fusion of heterogeneous data. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS, Milan, Italy*, pages 2947–2950, 2015.
- [25] I. Hammami, G. Mercier, and J. Dezert. Copulas-based fusion of consonant belief functions induced by dependent sources of evidences. *Knowledge Based Systems*, 2017. (to be submitted).
- [26] P. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.
- [27] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 6(2):191–234, 1994.
- [28] A. Martin. About conflict in the theory of belief functions. In *Denoeux T., Masson MH. (eds) Belief Functions: Theory and Applications. Advances in Intelligent and Soft Computing*, volume 164. Springer, Berlin, Heidelberg, 2012.

- [29] R. Kennes. Computational aspects of the Möbius transformation of graphs. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(2):201–223, 1992.
- [30] R. Kennes and P. Smets. Uncertainty in artificial intelligence. In *Chapter: Computational Aspects of the Möbius Transformation*, pages 401–416. Elsevier Science Publishers, 1991.
- [31] L. A. Zadeh. *On the validity of Dempster’s Rule of combination of evidence*. Electronics Research Laboratory Memo M 79/24. University of California, Berkeley, 1979.
- [32] F. Voorbraak. On the justification of Dempster’s rule of combination. *Artificial Intelligence*, 48:171–197, 1991.
- [33] J. Dezert, P. Wang, and A. Tchamova. On the validity of Dempster-Shafer Theory. In *15th International Conference on Information Fusion (FUSION), Singapore*, pages 655–660, 2012.
- [34] C. K. Murphy. Combining belief functions when evidence conflicts. *Decision Support Systems*, 29(1):1–9, 2000.
- [35] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41:93–137, 1987.
- [36] M. C. Florea, J. Dezert, P. Valin, F. Smarandache, and A.-L. Jousselme. Adaptative combination rule and proportional conflict redistribution rule for information fusion. In *Proceedings of the COGNitive systems wiith Interactive sensors (COGIS)*, 2006.
- [37] E. Lefevre, O. Colot, and P. Vannoorenberghe. Belief function combination and conflict management. *Information Fusion*, 3:149–162, 2002.
- [38] A. Martin and C. Osswald. Human expert fusion for image classification. *Information and Security : An International Journal, Special issue on Fusing Uncertain, Imprecise and Conflicting Information*, volume 20, may 2006.
- [39] A. Martin and C. Osswald. A new generalization of the proportional conflict redistribution rule stable in terms of decision. In F. Smarandache and J. Dezert, editors, *Advances and Applications of DSMT for Information Fusion, (Collected Works, Vol. 2)*. American Research Press, Jul 2006.
- [40] A. Kallel and S. Le Hegarat-Masclé. Combination of partially non-distinct beliefs: The cautious-adaptive rule. *International Journal of Approximate Reasoning*, 50(7):1000–1021, 2009.
- [41] P. Smets. The concept of distinct evidence. *Proceedings of the 4th conference on Information Processing and Management of Uncertainty in knowledge-based systems, (IPMU)*, pages 789–794, 1992.
- [42] F. Smarandache, J. Dezert, and J.-M. Tacnet. Fusion of sources of evidence with different importances and reliabilities. In *13th Conference on Information Fusion (FUSION), Edinburgh, UK*, pages 1–8, 2010.

-
- [43] D. Mercier, B. Quost, and T. Denœux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258, 2008.
- [44] A.-L. Jousselme and P. Maupin. Distances in evidence theory: Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53(2):118–145, 2012.
- [45] F. Cuzzolin. A geometric approach to the theory of evidence. *IEEE Transactions on Systems, Man, and Cybernetics, Part C :Applications and Reviews*, 38(4):522–534, 2008.
- [46] B. Ristic and P. Smets. The TBM global distance measure for the association of uncertain combat ID declarations. *Information Fusion*, 7(3):276–284, 2006.
- [47] B. Ristic and P. Smets. Global cost of assignment in the TBM framework for association of uncertain ID reports. *Aerospace Science And Thecnologie*, 11(4):303–309, 2007.
- [48] B. Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2):315–329, 1993.
- [49] A.-L. Jousselme, D. Grenier, and É. Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2(2):91–101, 2001.
- [50] D. Han, J. Dezert, and Y. Yang. New distance measures of evidence based on belief intervals. In *Belief Functions: Theory and Applications - Third International Conference, BELIEF, Oxford, UK*, pages 432–441, 2014.
- [51] P. Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.
- [52] A. Irpino and R. Verde. Dynamic clustering of interval data using a wasserstein-based distance. *Pattern Recognition Letters*, 29(11):1648–1658, 2008.
- [53] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, and Y. Yang. Decision-making with belief interval distance. In *Belief Functions: Theory and Applications - 4th International Conference, BELIEF, Prague, Czech Republic*, pages 66–74, 2016.
- [54] A. Essaid, A. Martin, G. Smits, and B. Ben Yaghlane. A distance-based decision in the credal level. In *Artificial Intelligence and Symbolic Computation - 12th International Conference, AISC, Seville, Spain*, pages 147–156, 2014.
- [55] P. Smets. Decision making in a context where uncertainty is represented by belief functions. In *Srivastava, R.P., Mock, T.J. (Eds.), Belief Functions in Business Decisions*, pages 17–61. Physica-Verlag, Heidelberg, Germany, 2002.
- [56] N. Wilson. The assumptions behind Dempster’s Rule. In *Proceedings of the Ninth International Conference on Uncertainty in Artificial Intelligence, UAI*, pages 527–534. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

- [57] P. Hajek. Deriving Dempster's rule. In : *Bouchon-Meunier, B., Valverde, L., Yager, R. R., (Eds.), Uncertainty in intelligent systems*, pages 75–84. Elsevier, 1993.
- [58] E. H. Ruspini. *The logical foundations of evidential reasoning*. Tech. note 408, SRI International Menlo Park, CA, USA, 1986.
- [59] L. Comtet. *Sperner Systems, sec.7.2 in Advanced combinatorics: The art of finite and infinite expansions*. D. Reidel Publ. Co., 1974.
- [60] R. Dedekind. Über zerlegungen von zahlen durch ihre grössten gemeinsamen theiler. In *Gesammelte Werke, Bd. 1*, pages 103–148. 1897.
- [61] A. Martin. Implementing general belief function framework with a practical codification for low complexity. In : *F. Smarandache and J. Dezert (Eds.), Advances and Applications of DSMT for Information Fusion*, volume 3, chapter 7, pages 217–274. American Research Press, Rehoboth, 2009.
- [62] P. Smets. Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35, 2008.
- [63] T. Denœux and L. M. Zouhal. An evidence-theoric k-NN rule with parameter optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 28(2):263–271, 1998.
- [64] J. Schubert. On nonspecific evidence. *International Journal of Intelligent Systems*, 8(6):711–725, 1993.
- [65] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE transactions on Systems, Man and Cybernetics A*, 30(2):131–150, 2000.
- [66] I. Borg and P. J. F. Groenen. *Modern multidimensional scaling*. Springer-Verlag, New-York, NY, USA, 1997.
- [67] A. Appriou. Discrimination multisignal par la théorie de l'évidence. In *R. Lengellé, (Ed.), Décision et Reconnaissance des formes en signal, (Hermès/Lavoisier, Collection Traité IC2 Série Traitement du Signal et de l'Image)*, chapter 7, pages 219–258. Hermes Science Publication, 2002.
- [68] A. Dromigny-Badin, S. Rossato, and Y. M. Zhu. Fusion de données radioscopiques et ultrasonores via la théorie de l'évidence. *Traitement du Signal*, 14(5):499–510, 1997.
- [69] A. Czigò, B. Solaiman, G. Cazuguel, C. Roux, and I. Loványi. Kohonen's self organizing feature maps with variable learning rate. Application to image compression. In *3rd international workshop on Image and Signal Processing, Santorini, Greece*, pages 11–14, 1997.
- [70] C.-I. Chang. An information theoretic-based measure for spectral similarity and discriminability. *IEEE Transactions on Information Theory*, 46(5):1927–1932, 2000.

-
- [71] T. Denceux. A k-Nearest Neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5):804–813, 1995.
- [72] H. Bostrm, S. F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. van_Laere, L. Niklasson, M. Nilsson, A. Persson, and T. Ziemke. On the definition of information fusion as a field of research. Technical report, University of Skovde, School of Humanities and Informatics, Skovde, Sweden, 2007.
- [73] I. Bloch and H. Maitre. Fusion de donnees en traitement d’images : modeles d’information et decisions. *Traitement du Signal*, 11(6):435–446, 1994.
- [74] M. Mangolini. *Apport de la fusion d’images satellitaires multicapteurs au niveau pixel en teledetection et photo-interpretation*. PhD thesis, Universite de Nice Sophia-Antipolis, France, 1994.
- [75] L. Wald. A European proposal for terms of reference in data fusion. In *International Archives of Photogrammetry and Remote Sensing*, volume XXXII, pages 651–654, 1998.
- [76] I. Bloch. *Fusion d’informations en traitement du signal et des images*. Hermes Science Publications, 2003.
- [77] C. Pohl and J. L. Van Genderen. Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, 19(5):823–854, 1998.
- [78] J. Dong, D. Zhuang, Y. Huang, and J. Fu. Survey of multispectral image fusion techniques in remote sensing applications. In *Image Fusion and Its Applications*, page 1–22. Intech., 2011.
- [79] A. Dromigny-Badin. *Image fusion using evidence theory: applications to medical and industrial images*. PhD thesis, INSA Lyon, France, 1998.
- [80] J. Zhang. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1(1):5–24, 2010.
- [81] N. Joshi, M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M. R. Jepsen, T. Kuemmerle, P. Meyfroidt, E. T. A. Mitchard, J. Reiche, C. M. Ryan, and B. Waske. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1), 2016.
- [82] M. Turker and M. Arikan. Sequential masking classification of multi-temporal Landsat7 ETM+ images for field-based crop mapping in Karacabey, Turkey. *International Journal of Remote Sensing*, 26(17):3813–3830, 2005.
- [83] T. Fisette, M. Maloley, R. Chenier, L. White, T. Huffman, R. Ogston, A. Pacheco, and P. Y. Gasser. Towards a national agricultural land cover classification-evaluating decision tree approach. In *Proceedings of the 26th Canadian Symposium on Remote Sensing, Wolfville, NS, Canada*, volume 14–16, 2005.

- [84] J. Inglada, M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, G. Dedieu, G. Sepulcre, S. Bontemps, P. Defourny, and B. Koetz. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9):12356–12379, 2015.
- [85] S. Foerster, K. Kaden, M. Foerster, and S. Itzerott. Crop type mapping using spectral-temporal profiles and phenological information. *Computers and Electronics in Agriculture*, 89:30–40, 2012.
- [86] M. Sandberg. Land cover mapping with multi-temporal SAR and optical satellite data. Master’s thesis, Aalto University, School of Engineering, 2016.
- [87] L. D. O. Pereira, C. D. C. Freitas, S. J. S. St Anna, D. Lu, and E. F. Moran. Optical and radar data integration for land use and land cover mapping in the brazilian amazon. *GIScience and Remote Sensing*, 50(3):301–321, 2013.
- [88] E. Lehmann, Z.-S. Zhou, P. Caccetta, A. Mitchell, A. Milne, K. Lowell, and S. McNeill. Combined analysis of optical and sar remote sensing data for forest mapping and monitoring. In *7th International Symposium on Digital Earth (ISDE7)*, Perth, Australia, 2011.
- [89] P. T. Wolter and P. A. Townsend. Multi-sensor data fusion for estimating forest species composition and abundance in northern minnesota. *Remote Sensing of Environment*, 115(2):671–691, 2011.
- [90] H. McNairn, C. Champagne, J. Shang, D. Holmstrom, and G. Reichert. Integration of optical and synthetic aperture radar (sar) imagery for delivering operational annual crop inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(5):434–449, 2009.
- [91] T. Idol, B. Haack, and R. Mahabir. Comparison and integration of spaceborne optical and radar data for mapping in sudan. *International Journal of Remote Sensing*, 36(6):1551–1569, 2015.
- [92] T. Kavzoglu and P. M. Mather. Pruning artificial neural networks: An example using land cover classification of multi-sensor images. *International Journal of Remote Sensing*, 20(14):2787–2803, 1999.
- [93] B. Waske and J. A. Benediktsson. Fusion of support vector machines for classification of multisensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3858–3866, 2007.
- [94] F. Roli, S. B. Serpico, and L. Bruzzone. Classification of multisensor remote-sensing images by multiple structured neural networks. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 4, pages 180–184, 1996.
- [95] J. Liu, M. Gong, K. Qin, and P. Zhang. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Transactions on Neural Networks and Learning Systems*, PP(99):1–15, 2017.

-
- [96] M. B. A. Gibril, S. A. Bakar, K. Yao, M. O. Idrees, and B. Pradhan. Fusion of radarsat-2 and multispectral optical remote sensing data for lulc extraction in a tropical agricultural area. *Geocarto International*, 32(7):735–748, 2017.
 - [97] J. Betbeder, M. Laslier, T. Corpetti, E. Pottier, S. Corgne, and L. Hubert-Moy. Multi-temporal optical and radar data fusion for crop monitoring: Application to an intensive agricultural area in brittany(france). In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 1493–1496, July 2014.
 - [98] H. Zhang, H. Lin, and Y. Li. Impacts of feature normalization on optical and sar data fusion for land use/land cover classification. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1061–1065, 2015.
 - [99] C. Oliver and S. Quegan. *Understanding synthetic aperture radar images*. Artech house, Norwood, MA, USA, 1998.
 - [100] A. Materka and M. Strzelecki. Texture analysis methods – a review. Technical report, Institue of Electronics, Technical University of Lodz, 1998.
 - [101] R. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.
 - [102] Z. Shao, H. Fu, P. Fu, and L. Yin. Mapping urban impervious surface by fusing optical and sar data at the decision level. *Remote Sensing*, 8(11), 2016.
 - [103] J. Inglada, A. Vincent, M. Arias, and C. Marais-Sicre. Improved early crop type identification by joint use of high temporal resolution sar and optical image time series. *Remote Sensing*, 8(5), 2016.
 - [104] Z. Elouedi, E. Lefevre, and D. Mercier. Discountings of a belief function using a confusion matrix. In *22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Los Alamitos, CA, USA*, pages 287–294, 2010.
 - [105] J. F. Lemmer. Confidence factors, empiricism and the dempster-shafer theory of evidence. In *Proceedings of the First Annual Conference on Uncertainty in Artificial Intelligence (UAI-85)*, pages 160–176, 1985.
 - [106] J. Pearl. Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning*, 4:363–389, 1990.
 - [107] P. Wang. A Defect in Dempster-Shafer Theory. In Ramon López de Mántaras and David Poole, editors, *UAI*, pages 560–566. Morgan Kaufmann, 1994.
 - [108] A. Gelman. The boxer, the wrestler, and the coin flip: A paradox of robust bayesian inference and belief functions. *The American Statistician*, 60(2):146–150, 2006.
 - [109] J. Dezert and A. Tchamova. On the validity of dempster’s fusion rule and its interpretation as a generalization of bayesian fusion rule. *International Journal of Intelligent Systems*, 29(3):223–252, 2014.

- [110] S. Destercke and D. Dubois. Idempotent conjunctive combination of belief functions: extending the minimum rule of possibility theory. *Information Sciences*, 181(18):3925–3945, 2011.
- [111] A. Chateauneuf. Combination of compatible belief functions and relations of specificity. *Papiers d’économie mathématique et applications*, Université Panthéon-Sorbonne (Paris 1), 1992.
- [112] M. Cattaneo. Combining belief functions issued from dependent sources. In *Proceedings of ISIPTA*, pages 133–147, 2003.
- [113] D. Dubois and R. R. Yager. Fuzzy set connectives as combinations of belief structures. *Information Sciences*, 66(3):245–276, 1992.
- [114] D. A. Alvarez. A monte carlo-based method for the estimation of lower and upper probabilities of events using infinite random sets of indexable type. *Fuzzy Sets and Systems*, 160(3):384 – 401, 2009.
- [115] R. R. Yager. Joint cumulative distribution functions for Dempster–Shafer belief structures using copulas. *Fuzzy Optimization and Decision Making*, 12(4):393–414, 2013.
- [116] B. Schmelzer. Joint distributions of random sets and their relation to copulas. *International Journal of Approximate Reasoning*, 65:59–69, 2015.
- [117] A. Sklar. *Fonctions de répartition à n dimensions et leurs marges*, volume 8. Publications de l’Institut de Statistique de Université Paris 8, 1959.
- [118] A. Sklar. Random variables, joint distributions, and copulas. *Kybernetika*, 9(6):449–460, 1973.
- [119] W. Hoeffding. Scale-invariant correlation theory. In N. I. Fisher and P. K. Sen, editors, *The Collected Works of Wassily Hoeffding*, pages 57–107. New York: Springer-Verlag, 1940.
- [120] W. Hoeffding. Scale-invariant correlation measures for discontinuous distributions. In N. I. Fisher and P. K. Sen, editors, *The Collected Works of Wassily Hoeffding*, pages 109–133. New York: Springer-Verlag, 1941.
- [121] G. Klir and Y. Bo. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, 1995.
- [122] D. Dubois and H. Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142(1):143–161, 2004.
- [123] H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1997.
- [124] H. T. Nguyen. On combining dependent evidence. Technical report, 2013.
- [125] B. Schmelzer. Sklar’s theorem for minitive belief functions. *International Journal of Approximate Reasoning*, 63:48–61, 2015.

-
- [126] G. Matheron. *Random sets and integral geometry*. Wiley, 1975.
- [127] O. Wolkenhauer. *Data Engineering*. John Wiley and Sons, New York, 2001.
- [128] C. Bertoluzza, M. A. Gil, and D. A. Ralescu editors. *Statistical modeling, analysis and management of fuzzy data*, volume 87 of Studies in fuzziness and soft computing. Physica Verlag, Heidelberg, New York, 2002.
- [129] T. Denœux. The cautious rule of combination for belief functions and some extensions. In *Proceedings of FUSION*, pages 1–8, Florence, Italy, 2006.
- [130] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer Publishing Company, Incorporated, 1st edition, 2008.
- [131] C. Alsina, B. Schweizer, and M. J. Frank. *Associative Functions: Triangular Norms and Copulas*. World Scientific, 2006.
- [132] E. J. Gumbel. Distributions à plusieurs variables dont les marges sont données. *Comptes Rendus de l'Académie des Sciences Paris*, (246):2717–2719, 1958.
- [133] M. J. Frank. On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes mathematicae*, 19:194–226, 1979.
- [134] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In M.M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. Elsevier/North-Holland, 1982.
- [135] G. Klir. A principle of uncertainty and information invariance. *International Journal of General Systems*, 17:249–275, 1990.

Avec l'avènement de nouvelles techniques d'acquisition d'image et l'émergence des systèmes satellitaires à haute résolution, les données de télédétection à exploiter sont devenues de plus en plus riches et variées. Leur combinaison est donc devenue essentielle pour améliorer le processus d'extraction des informations utiles liées à la nature physique des surfaces observées. Cependant, ces données sont généralement hétérogènes et imparfaites ce qui pose plusieurs problèmes au niveau de leur traitement conjoint et nécessite le développement de méthodes spécifiques. C'est dans ce contexte que s'inscrit cette thèse qui vise à élaborer une nouvelle méthode de fusion évidentielle dédiée au traitement des images de télédétection hétérogènes à haute résolution. Afin d'atteindre cet objectif, nous axons notre recherche, en premier lieu, sur le développement d'une nouvelle approche pour l'estimation des fonctions de croyance basée sur la carte de Kohonen pour simplifier l'opération d'affectation des masses des gros volumes de données occupées par ces images. La méthode proposée permet de modéliser non seulement l'ignorance et l'imprécision de nos sources d'information, mais aussi leur paradoxe. Ensuite, nous exploitons cette approche d'estimation pour proposer une technique de fusion originale qui permettra de remédier aux problèmes dus à la grande variété des connaissances apportées par ces capteurs hétérogènes. Finalement, nous étudions la manière dont la dépendance entre ces sources peut être considérée dans le processus de fusion moyennant la théorie des copules. Pour cette raison, une nouvelle technique pour choisir la copule la plus appropriée est introduite. La partie expérimentale de ce travail est dédiée à la cartographie de l'occupation des sols dans les zones agricoles en utilisant des images SPOT-5 et RADARSAT-2. L'étude expérimentale réalisée démontre la robustesse et l'efficacité des approches développées dans le cadre de cette thèse.

Mots-clés : Théorie des fonctions de croyance, Estimation, Carte de Kohonen, Fusion des données hétérogènes, Images optiques et radars, Dépendances, Théorie des copules

With the advent of new image acquisition techniques and the emergence of high-resolution satellite systems, remote sensing data to be exploited have become increasingly rich and varied. Their combination has thus become essential to improve the process of extracting useful information related to the physical nature of the observed surfaces. However, these data are generally heterogeneous and imperfect, which poses several problems in their joint treatment and requires the development of specific methods. It is in this context that falls this thesis that aimed at developing a new evidential fusion method dedicated to heterogeneous remote sensing images processing at high resolution. In order to achieve this objective, we first focus our research, firstly, on the development of a new approach for the belief functions estimation based on Kohonen's map in order to simplify the masses assignment operation of the large volumes of data occupied by these images. The proposed method allows to model not only the ignorance and the imprecision of our sources of information, but also their paradox. After that, we exploit this estimation approach to propose an original fusion technique that will solve problems due to the wide variety of knowledge provided by these heterogeneous sensors. Finally, we study the way in which the dependence between these sources can be considered in the fusion process using the copula theory. For this reason, a new technique for choosing the most appropriate copula is introduced. The experimental part of this work is devoted to land use mapping in case of agricultural areas using SPOT-5 and RADARSAT-2 images. The experimental study carried out demonstrates the robustness and effectiveness of the approaches developed in the framework of this thesis.

Keywords: Belief function theory, estimation, Kohonen's map, heterogeneous data fusion, optical and radar images, dependencies, copula theory



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

N° d'ordre : 2017IMTA0034

IMT Atlantique Bretagne-Pays de la Loire - www.imt-atlantique.fr

Campus de Brest
Technopôle Brest-Iroise
CS 83818
29238 Brest Cedex 03
T +33 (0)2 29 00 11 11
F +33 (0)2 29 00 10 00

Campus de Nantes
4, rue Alfred Kastler - La Chantrerie
CS 20722
44307 Nantes Cedex 3
T +33 (0)2 51 85 81 00
F +33 (0)2 51 85 81 99

Campus de Rennes
2, rue de la Châtaigneraie
CS 17607
35576 Cesson Sévigné Cedex
T +33 (0)2 99 12 70 00
F +33 (0)2 99 12 70 08