



HAL
open science

Semantic segmentation of forest stand by join analysis of very high resolution multispectral image and 3D airborne Lidar data

Clément Dechesne

► **To cite this version:**

Clément Dechesne. Semantic segmentation of forest stand by join analysis of very high resolution multispectral image and 3D airborne Lidar data. Image Processing [eess.IV]. Université Paris-Est, 2017. English. NNT: 2017PESC1224 . tel-01760483

HAL Id: tel-01760483

<https://theses.hal.science/tel-01760483>

Submitted on 6 Apr 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.



École doctorale n°532: Mathématiques et Sciences et Technologies de l'Information et de la Communication

THÈSE DE DOCTORAT

pour obtenir le grade de docteur délivré par

l'Université Paris-Est

Spécialité doctorale "Sciences et Technologies de l'Information Géographique"

présentée par

Clément DECHESNE

soutenue le 04/12/2017

à l'Institut National de l'Information Géographique et Forestière (IGN)

**Segmentation sémantique de peuplements forestiers par analyse conjointe d'imagerie
multispectrale très haute résolution et de données 3D Lidar aéroportées**

Jury

Jocelyn CHANUSSOT	Rapporteur
Felix MORSDORF	Rapporteur
Lorenzo BRUZZONE	Président
Laurence HUBERT-MOY	Examineur
Florence TUPIN	Examineur
Clément MALLET	Co-encadrant
Arnaud LE BRIS	Co-encadrant
Valérie GOUET-BRUNET	Directrice

Laboratoire LaSTIG

Institut National de l'Information Géographique et Forestière (IGN)

73 Avenue de Paris, 94165 Saint Mandé Cedex, France

Acknowledgments / Remerciements

This PhD thesis work was carried out in the Laboratoire LaSTIG of the Institut National de l'Information Géographique et Forestière (IGN).

First, I would like to acknowledge Valérie Gouet-Brunet, accepted to supervise this work during the three years I spend at IGN. I would like to thank my two thesis advisor Clément Mallet and Arnaud Le Bris for their support and for making them-self available whenever I needed. Working with them was a very pleasant and educational experience.

I thank Lorenzo Bruzzone for accepting to preside the committee, and for its interest to my work. I would like to address my acknowledgments to Jocelyn Chanussot and Felix Morsdorf, the reviewer of my thesis, for their thorough analysis of this work and their helpful comments. I would also like to thanks the other committee members, Florence Tupin and Laurence Hubert-Moy, for their relevant remarks.

Merci ensuite à toutes les personnes au MATIS qui ont participé et qui participent encore aux excellentes condition de recherche au laboratoire. Un grand merci également Marie-Claude Foubert et Marie-Jean Aboim, sans qui le laboratoire ne serait pas le même.

Je tiens également à remercier particulièrement mes collègues de RISOTO avec lesquels j'ai eu beaucoup de plaisir à travailler (en espérant que mes collaborations avec eux ne s'arrêtent pas là). Merci à Clément, pour ses traits d'esprit (pouvant s'apparenter à du harcèlement moral) apportant la légèreté dans le bureau et son expertise scientifique et bibliographique. Merci à Arnaud, pour ses précieux conseils et son expertise scientifique et technique qui m'ont permis de travailler dans d'excellentes conditions. Merci à Sebastien, pour ses anecdotes croustillantes et ses explications claires sur le Radar. Merci à Adrien, qui m'a beaucoup aidé au début de ma thèse. Merci aussi à Cécile qui a partagé avec moi les souffrances de la rédaction. Merci enfin à Tristan et Oussama, les courageux doctorant qui, je l'espère, finiront brillamment leur thèse. Je remercie également tous les autre membres de l'équipe: Bruno, David, Bahman, Mathieu, Lâman, Laurent, Loic, Bastien, Nathan, Stéphane et tous ceux que j'aurai oubliés.

Enfin, je tiens à remercier ma famille; PF (ou BIG) pour m'avoir supporté au quotidien ces tois dernières années, et mes deux sœurs Camille et Lucie. Un grand merci aussi à mes amis qui m'ont également soutenus au cours de ces trois années.

Clément

Contents

Contents	ii
Abstract	vii
Résumé	viii
Résumé étendu en français	ix
1.1 Introduction	ix
1.2 Méthode proposée	xi
1.2.1 Extraction des attributs	xi
Attributs lidar	xi
Attributs spectraux	xii
Extraction d'objets	xii
1.2.2 Classification	xiii
Sélection d'attributs	xiii
Classification	xiii
1.2.3 Régularisation	xiii
Formulation de l'énergie	xiv
Minimisation d'énergie	xv
1.3 Données	xv
1.4 Résultats	xvi
1.4.1 Sélection d'attributs	xvi
1.4.2 Classification	xvi
1.4.3 Régularisation	xvii
1.5 Conclusion	xviii
1 Introduction	1
1.1 Analysis of forested areas	2
1.2 Remote sensing for forested areas	4
1.3 Context of the thesis	8
1.4 Objectives	10
1.5 Strategy	11
1.6 Structure of the thesis	12
2 State of the art	15
2.1 Stand segmentation	16
2.1.1 Stand segmentation using VHR optical images	18
2.1.2 Stand segmentation using lidar data	18

2.1.3	Stand segmentation using VHR optical images and lidar	19
2.1.4	Challenges of stand segmentation	20
2.2	Segmentation	22
2.2.1	"Traditional" segmentation methods	22
2.2.2	Superpixels methods	24
2.2.3	Segmentation of point cloud	25
2.3	Classification	26
2.3.1	Supervised classification: common algorithms	26
2.3.2	Random Forest	26
2.4	Dimension reduction and feature selection	28
2.4.1	Dimension reduction: feature extraction	28
2.4.2	Feature selection	28
Existing methods	28	
Selection optimization	29	
2.5	Smoothing methods	30
2.5.1	Local methods	30
2.5.2	Global methods	31
2.6	Conclusion	32
3	Proposed framework	35
3.1	General flowchart	36
3.2	Feature extraction	36
3.2.1	Point-based lidar features.	38
3.2.2	Pixel-based lidar features.	39
3.2.3	Pixel-based multispectral images features.	40
3.2.4	Object-based feature map.	41
3.3	Over-segmentation	41
3.3.1	Segmentation of lidar data	43
3.3.2	Segmentation of optical images	45
3.4	Feature selection	46
3.5	Classification	47
3.5.1	Training set design	48
3.5.2	Training and prediction	49
3.6	Regularization	50
3.7	Conclusion and discussions	51
4	Flowchart assessment	55
4.1	Data	56
4.1.1	Remote sensing data	56
4.1.2	National Forest LC DB: "BD Forêt®"	57
4.1.3	Areas of interest	59
4.2	Framework experiments	64
4.2.1	Over-segmentation	64
4.2.2	Feature selection	65
4.2.3	Classification	74

4.2.4	Regularization	77
4.2.5	Computation times	80
4.3	Final results on multiple areas	80
4.3.1	Gironde	80
4.3.2	Ventoux	82
4.3.3	Vosges	86
4.3.4	Validity of the framework	93
4.4	Can forest stands be simply retrieved?	93
4.4.1	Segmentation of remote sensing data	95
4.4.2	Classification of the segments	96
4.5	Derivation of other outputs	99
4.5.1	Semi-automatic update process	99
4.5.2	Data enrichment for inventory	102
	Additional information extraction	102
	Segmentation using complementary criteria	103
4.6	Conclusion and perspectives	105
5	Regularization: How to obtain smooth relevant stands?	107
5.1	Local methods	108
5.1.1	Filtering	108
5.1.2	Probabilistic relaxation	109
5.2	Global smoothing	109
5.2.1	Unary/data term	112
5.2.2	Pairwise/prior term	112
5.3	Energy minimization	114
5.4	Constraints integration	115
5.4.1	Size constraint	115
5.4.2	Border constraint	115
5.5	How to efficiently smooth a classification?	116
5.5.1	Local methods	116
5.5.2	Global methods	117
5.5.3	Constraints integration	120
5.6	Conclusion	122
6	Data fusion	125
6.1	Levels of fusion	126
6.1.1	Object-level fusion	126
6.1.2	Classification-level fusion	127
6.1.3	Regularization-level fusion	128
6.2	Designing the best fusion scheme	128
6.3	Optimal fusion scheme	130
6.4	Conclusion	134

7	Conclusions and perspectives	137
7.1	Conclusion	138
7.1.1	Extraction of forest stands	138
7.1.2	Fusion schemes	140
7.1.3	Quantitative features	141
7.2	Perspectives	141
7.2.1	Relevance of other remote sensing data sources	142
7.2.2	Improvement	142
7.2.3	Application to other land-cover problems	143
7.3	Final outlook	143
A	Color code	145
B	Publications	149
B.1	Journal articles	150
B.2	Peer-reviewed conference papers	150
C	Results - Confusion matrices and accuracy metrics	153
C.1	Accuracy metrics	154
C.1.1	Metrics at the class level	154
C.1.2	Metrics at the global level	156
C.2	Flowchart assessment	157
C.2.1	Over-segmentation	157
C.2.2	Classification	163
C.2.3	Feature selection	166
C.2.4	Regularization	167
	Local methods	167
	Global methods	168
C.3	Can forest stands be simply retrieved?	173
C.3.1	VHR optical images	174
C.3.2	nDSM	175
C.4	Test on urban area	175
	Bibliography	179

Abstract

Forest stands the basic units for forest statistical inventory and mapping. Stands are defined as (large) forested areas (e.g., 2 ha) of homogeneous tree species composition and age. Their accurate delineation is usually performed by human operators through visual analysis of very high resolution (VHR) infra-red colored images. This task is tedious, highly time consuming, and should be automated for scalability, efficient updating purposes and deeper analysis.

Therefore, a framework based on the fusion of airborne lidar data and VHR multispectral images is proposed for the automatic delineation of forest stands containing one dominant species (i.e., which are more pure than to 75%). Indeed, this is a key preliminary task for forest land-cover database update. The framework is very adaptable and composed of four steps that are deeply analyzed draw the best from the remote sensing data sources (fusion of VHR optical images and airborne 3D lidar point cloud at different levels) as well as and the French Forest land-cover database. Multi-modal features are firstly extracted and their relevance is assessed. An over-segmentation is performed in order derive object-based features. They can be trees (obtained from the point cloud) or any other objects of similar size and/or shape. Because of the high number of features involved, a feature selection is carried out in order to reduce the computation times and to increase the discrimination as well as to assess the relevance of the extracted features and the complementarity of the remote sensing data. An object-based supervised classification is performed employing the supervised Random Forest classifier. A special design of the training set is proposed in order to cope with the potential errors of the Forest land-cover database. Finally, the classification result is further post processed in order to retrieve the forest stands, that is to say to obtain homogeneous areas with smooth borders. This smoothing is performed in an energy minimization framework where additional constraints are proposed to form the energy function. This problem is then formulated as a graphical one and solved by graph cut optimization algorithm.

The experimental results show that the proposed framework provides very satisfactory results both in terms of stand labeling and delineation, even for spatially distant regions exhibiting different landscapes. Multiple fusion schemes are proposed depending on the level of detail desired and operational constraints (computing times, data).

Keywords: *Lidar, multispectral imagery, fusion, feature selection, supervised classification, energy minimization, regularization, forest stand delineation, tree species.*

Résumé

Les peuplements forestiers constituent une entité de base pour l'inventaire forestier statistique et la cartographie. Ils sont définis comme de (grandes) zones forestières (par exemple, de plus de 2 ha) et de composition homogène en terme d'essence d'arbres et d'âge. Leur délimitation précise est généralement effectuée par des opérateurs humains par une analyse visuelle d'images contenant un canal infrarouges à très haute résolution (THR). Cette tâche est fastidieuse, nécessite beaucoup de temps et doit donc être automatisée pour un suivi de l'évolution et une mise à jour plus efficace des bases de données. Une méthode fondée sur la fusion de données lidar aéroportées et d'images multispectrales THR est proposée pour la délimitation automatique de peuplements forestiers contenant une essence dominante (c'est à dire, pure à plus de 75%). Il s'agit en effet d'une tâche préliminaire importante pour la mise à jour de la base de données de la couverture forestière.

Le méthode est adaptable à la donnée et au paysage étudié. Elle est composée de quatre étapes qui sont analysées en profondeur qui tirent le meilleur parti des différentes sources de données de télédétection, à l'aide de processus de fusion à plusieurs niveaux des images optiques VHR et du nuage de points lidar 3D aéroporté mais aussi de l'analyse de la base de données géographique (BD Forêt) décrivant la forêt Française. Des attributs multimodaux sont d'abord extraits et leur pertinence est évaluée. Ces attributs sont ensuite croisés avec une sursegmentation afin d'obtenir des attributs au niveau de l'objet. Il peut s'agir d'arbres (obtenus à partir du nuage de points) ou de tout autre objet de taille et/ou de forme similaire. En raison du nombre élevé d'attributs, une sélection d'attributs est ensuite effectuée. Elle permet de réduire les temps de calcul, d'améliorer la discrimination ainsi que d'évaluer la pertinence des attributs extraits et la complémentarité des données de télédétection. Une classification supervisée fondée objet est ensuite effectuée avec l'algorithme supervisé des Forêts Aléatoires. Une attention spéciale est apportée à la création du jeu d'apprentissage afin de faire face aux erreurs potentielles de la base de données Forêt. Enfin, le résultat de la classification est ensuite traité afin d'obtenir des zones homogènes avec des frontières lisses. Ce lissage est effectué de manière globale sur l'image en minimisant une énergie, dans laquelle contraintes supplémentaires sont proposées en plus des formulations classiques pour former la fonction d'énergie. Ce problème est reformulé de manière graphique et résolu par une approche de type coupe de graphe.

Une étude détaillée des différentes parties de la chaîne de traitement proposée a été réalisée. Les résultats expérimentaux montrent que la méthode proposée fournit des résultats très satisfaisants en termes d'étiquetage et de délimitation des peuplements, même pour des régions spatialement éloignées et présentant des paysages différents. La méthode proposée permet également d'évaluer la complémentarité des sources de données de télédétection (à savoir le lidar et les images optiques THR). Plusieurs schémas de fusion sont par ailleurs proposés en fonction du niveau de détail souhaité et des éventuelles contraintes opérationnelles (temps de calculs, données).

Mots clés: *Lidar, imagerie multispectrale, fusion, sélection d'attributs, classification supervisée, minimisation d'énergie, régularisation, délimitation de peuplement forestiers, essence d'arbre.*

Résumé étendu en français

1.1 Introduction

L'extraction de l'information dans les zones forestières, en particulier au niveau du peuplement, est motivée par deux objectifs principaux: l'inventaire statistique et la cartographie. Les peuplements forestiers sont les unités de base et peuvent être définis en termes d'espèces d'arbres ou de maturité des arbres. Du point de vue de la télédétection, la délimitation des peuplements est un problème de segmentation. Pour l'inventaire forestier statistique, la segmentation est utile pour extraire des points d'échantillonnage significatifs sur le plan statistique et des attributs fiables (surface terrière, hauteur dominante, etc.) (Means et al., 2000; Kangas et al., 2006). Pour la cartographie du couvert végétal, la segmentation est très utile pour la mise à jour des bases de données forestières (Kim et al., 2009). La plupart du temps, pour des raisons de fiabilité, chaque zone est interprétée manuellement par des opérateurs humains avec des images géospatiales de très haute résolution spatiale. Ce travail prend beaucoup de temps, de plus, dans de nombreux pays, la grande variété d'essences forestières (environ 20) rend le problème plus difficile.

L'utilisation de données de télédétection pour l'analyse automatique des forêts est de plus en plus répandue, en particulier avec l'utilisation combinée du lidar aéroporté et de l'imagerie optique (imagerie multispectrale à très haute résolution ou imagerie hyperspectrale) (Torabzadeh et al., 2014).

Quelques travaux proposant une délimitation automatique des peuplements forestiers avec des données de télédétection existent. Tout d'abord, la délimitation peut être effectuée avec une seule source de télédétection. Une technique de délimitation des peuplements utilisant des images hyperspectrales est proposée dans (Leckie et al., 2003). Les arbres sont extraits et classés selon 7 espèces d'arbres (5 conifères, 1 caduque et 1 non spécifié) à l'aide d'un classificateur à maximum de vraisemblance.

Une méthode de cartographie des peuplements utilisant des données lidar aéroportées à basse densité est proposée dans (Koch et al., 2009). Elle comporte plusieurs étapes; extraction d'attributs, rasterisation des attributs et classification à partir du raster. Les peuplements forestiers sont créés en regroupant les cellules voisines par classe. Ensuite, seuls les peuplements ayant une taille minimale prédéfinie sont acceptés. Les petites zones voisines d'espèce différentes qui n'atteignent pas la taille minimale sont fusionnées à un peuplement forestier proche.

La délimitation des peuplements forestiers proposée dans (Sullivan et al., 2009) utilise aussi du lidar aéroporté à basse densité pour une segmentation suivie d'une classification supervisée. Trois attributs (couverture de la canopée, densité de la tige et hauteur moyenne) sont calculés et rasterisés. La segmentation est réalisée par croissance de région. Les pixels spatialement adjacents sont regroupés en objets ou régions homogènes. Ensuite, une classification supervisée de l'image segmentée est réalisée à l'aide d'un classificateur de Bhattacharya, afin de définir la maturité des peuplements (les étiquettes correspondent à la maturité des peuplements).

Une délimitation de peuplements forestiers utilisant des données lidar aéroportées à haute densité est également proposée dans (Wu et al., 2014). Trois attributs sont d'abord extraits du nuage de points; l'indicateur de la taille de l'arbre, l'indice de densité forestière et l'indicateur d'espèces d'arbres. Une délimitation grossière du peuplement forestier est ensuite effectuée sur l'image des attributs en utilisant l'algorithme Mean-Shift, avec une valeur élevée des paramètres afin d'obtenir des peuplements forestiers grossiers sous-segmentés. Un masque forestier est ensuite appliqué à l'image segmentée afin de récupérer des peuplements forestiers et non-forestiers grossiers. Cette étape peut créer quelques petites zones isolées qui seront fusionnées à leur voisine la plus proche jusqu'à ce que leur taille soit supérieure à un seuil défini par l'utilisateur. Les peuplements forestiers sont ensuite raffinés, mais en utilisant des superpixels générés à partir des trois attributs au lieu d'utiliser les pixels d'origine de l'image 3 bandes. Le raffinement des peuplements forestiers est obtenu grâce à une croissance de région. Cette méthode fournit des peuplements relativement grands.

D'autres méthodes utilisant la fusion de différents types de données de télédétection ont également été développées. Deux méthodes de segmentation sont proposées dans (Leppänen et al., 2008) pour une forêt composée de pin sylvestre, d'Épicéa de Norvège et de feuillus. La première est une segmentation sur la hauteur de couronne suivi d'une croissance de région itérative sur une image composite de lidar et d'image IRC. La deuxième méthode propose une segmentation hiérarchique sur la hauteur de couronne. Chaque objet de l'image est connecté à la fois aux objets de l'image de niveaux supérieur et inférieur. Cela permet de considérer les segments finaux à partir des niveaux de segmentation les plus fins, comme les arbres individuels dans la zone.

L'analyse des données lidar et multispectrales est effectuée à trois niveaux dans (Tiede et al., 2004). Le premier niveau représente des petits objets (arbres individuels ou petit groupe d'arbres) qui peuvent être différenciés suivant des caractéristiques spectrales et structurelles en utilisant une classification fondée sur des règles. Le deuxième niveau correspond au peuplement. Il est construit en utilisant le même processus de classification que le niveau précédent, en se référant aux objets à petite échelle du niveau 1. Le troisième niveau est généré en fusionnant des objets du même développement forestier en unités spatiales plus grandes. Cette méthode produit une cartographie permettant d'évaluer la phase de développement forestier (les étiquettes ne correspondent pas aux espèces d'arbres).

Puisque les peuplements sont l'unité de base pour l'inventaire statistique, certaines méthodes de segmentation ont été développées à cette fins dans (Diedershagen et al., 2004) et (Hernando et al., 2012).

Au regard des méthodes existantes, il semble qu'il n'y ait pas de méthode de segmentation des peuplements forestiers, en termes d'espèce, capables de traiter de façon satisfaisante un grand nombre de classes (> 5). Il apparaît également que le fait de travailler au niveau de l'objet (habituellement le niveau de l'arbre), afin de discriminer les espèces d'arbres, produit de meilleurs résultats de segmentation des peuplements. Plusieurs méthodes de classification des espèces au niveau des arbres ont été étudiées dans (Heinzel et al., 2012), (Leckie et al., 2003) et (Dalponte et al., 2015). Toutefois, il est probable qu'elles soient imprécises, ce qui entraîne une classification inexacte des arbres. Cependant, le classement des essences forestières au niveau des arbres peut être utilisé pour la délimitation des peuplements forestiers.

Une méthode pour la segmentation des peuplements en termes d'espèce est proposée dans ce document. La méthode comprend trois étapes principales. Des attributs sont extraits au niveau du pixel et de l'objet. Les objets sont déterminés par une sur-segmentation. Une classification est effectuée au niveau de l'objet car elle améliore significativement les résultats de discrimination (environ 20% de mieux que la classification sur les pixels). Cette classification est ensuite régularisée par une minimisation d'énergie. La solution de cette régularisation, obtenue à l'aide d'une méthode de coupe de graphe, produit des zones homogènes d'espèces d'arbres avec des frontières lisses.

1.2 Méthode proposée

1.2.1 Extraction des attributs

L'extraction des attributs comporte trois étapes;

- Calcul et rastérisation des attributs lidar.
- Calcul des attributs spectraux.
- Extraction des objets (sur-segmentation) et création des image à l'objet.

Attributs lidar

Les attributs lidar nécessitent de prendre en compte un voisinage pour être cohérent. Pour chaque point lidar, 3 voisinages cylindrique, avec le même axe vertical, sont utilisés (rayons de 1 m , 3 m et 5m, hauteur infinie). Le cylindre est la forme la plus pertinente car elle permet de prendre en compte toute la variabilité de hauteur des points. Trois rayons sont utilisés afin de gérer les différentes tailles des arbres. Tout d'abord, deux indices de végétation, \mathcal{D}_1 et \mathcal{D}_2 , sont calculés: le premier est fondé sur le nombre de maxima locaux dans les voisinages et le second est lié au nombre de points hors-sol dans le voisinage (les points au sol ayant été déterminés précédemment par filtrage). \mathcal{D}_1 et \mathcal{D}_2 sont calculés comme suit:

$$\mathcal{D}_1 = \sum_{r_1 \in \{1,3,5\}} \sum_{r_2 \in \{1,3,5\}} Nt_{r_1, r_2}, \quad (1.1)$$

$$\mathcal{D}_2 = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{Ns_r}{Ntot_r}, \quad (1.2)$$

où Nt_{r_1, r_2} est le nombre de maxima locaux extraits d'un filtre maximal de rayon r_1 dans le voisinage cylindrique de rayon r_2 . Ns_r est le nombre de points classés comme "sol" dans le voisinage de rayon r et $Ntot_r$ est le nombre total de points dans le voisinage de rayon r .

La dispersion \mathcal{S} et la planarité \mathcal{P} sont aussi calculés en suivant la formulation de (Weinmann et al., 2015):

$$\mathcal{S} = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{\lambda_{3,r}}{\lambda_{1,r}}, \quad (1.3)$$

$$\mathcal{P} = \frac{1}{3} \sum_{r \in \{1,3,5\}} 2 \times (\lambda_{2,r} - \lambda_{3,r}), \quad (1.4)$$

où $\lambda_{1,r} \geq \lambda_{2,r} \geq \lambda_{3,r}$ sont les valeurs propres de la matrice de covariance des points dans le voisinage cylindrique de rayon r . Elles sont obtenues par une simple analyse en composante principale.

Des attributs statistiques, reconnus pour être pertinents pour la classification des espèces (Dalponte et al., 2014; Torabzadeh et al., 2015), sont aussi calculés. Pour chaque point lidar, les 3 mêmes voisinages cylindriques sont utilisés. Deux informations du lidar, la hauteur et l'intensité, sont utilisées afin de dériver des attributs statistiques. Un attribut statistique f_d , dérivé d'une information f_o , (hauteur ou intensité) est obtenu comme suit:

$$f_d = \frac{1}{3} \sum_{r \in \{1,3,5\}} f_s(\mathbf{P}_{r,f_o}), \quad (1.5)$$

où f_s est une fonction statistique (minimum; maximum; moyenne; médiane; écart-type; déviation absolue médiane d'une médiane (medADmed); déviation absolue moyenne d'une médiane (meanADmed); skewness; kurtosis; 10^{ème}, 20^{ème}, 30^{ème}, 40^{ème}, 50^{ème}, 60^{ème}, 70^{ème}, 80^{ème}, 90^{ème} et 95^{ème} centile), et \mathbf{p}_{r,f_o} un vecteur contenant les valeurs des points de l'information f_o dans le cylindre de rayon r . Toutes ces fonction statistiques sont utilisées pour la hauteur. Seule la moyenne est utilisée pour l'intensité: il est difficile de savoir si le capteur a été correctement calibré et une correction des valeurs d'intensité dans la canopée n'a pas encore été proposée.

24 attributs sont calculés au cours de cette étape; 2 liés à la densité de végétation, 2 liés à la distribution 3D locale du nuage de points et 20 attributs statistiques.

Les 24 attributs sont rastérisés à la même résolution spatiale que l'image multispectrale, en utilisant la méthode proposée dans (Khosravipour et al., 2014). Cette méthode est intéressante car elle produit des images lisses, qui permettent d'obtenir de meilleurs résultats pour la classification et la régularisation (Li et al., 2013a). Le modèle numérique de surface (MNS) est aussi obtenu en utilisant cette méthode, à la même résolution spatiale, en utilisant un modèle numérique de terrain (MNT) précédemment obtenu en filtrant le nuage de points. Le MNS est très important car il permet de déterminer la hauteur par rapport au sol et est un attribut très discriminant pour la classification (Mallet et al., 2011; Weinmann, 2016). Au total, 25 attributs lidar sont calculés.

Attributs spectraux

Les 4 bandes spectrales sont conservées et considérées comme des attributs spectraux. 3 indices de végétation pertinents; le NDVI, le DVI et le RVI sont calculés car ils peuvent fournir plus d'information que les bandes originales seules (Zargar et al., 2011). Comme pour les attributs lidar, des attributs statistiques sont calculés pour chaque bande et chaque indice de végétation en utilisant l'équation 1.5 (3 voisinages circulaires de rayon de 1m, 3m et 5m). D'autres fonctions statistiques sont utilisées (minimum; maximum; moyenne; médiane; écart-type; déviation absolue moyenne d'une médiane (meanADmed); déviation absolue moyenne d'une moyenne (meanADmean); déviation absolue médiane d'une médiane (medADmed); déviation absolue médiane d'une moyenne (medADmean)). Au total, 70 attributs spectraux sont calculés.

Extraction d'objets

L'extraction d'objets est une étape importante permettant d'améliorer les résultats de la classification. Cependant, les objets peuvent être extraits de différentes manières; il est possible d'effectuer une simple sur-segmentation sur l'un des 95 attributs disponibles (Watershed) ou bien d'utiliser des algorithmes plus avancés (SLIC, Quickshift, etc.) ou bien d'extraire les arbres directement du nuage

de points. Il apparaît que le choix de la méthode de sur-segmentation n'impacte que très peu les résultats finaux. La sur-segmentation est effectuée sur le MNS. Une fois la sur-segmentation effectuée, les attributs sont moyennés sur les objets: la valeur v_t d'un pixel appartenant à l'objet t est:

$$v_t = \frac{1}{N_t} \sum_{p \in t} v_p, \quad (1.6)$$

où N_t est le nombre de pixels dans l'objet t , et v_p est la valeur du pixel p .

1.2.2 Classification

La classification est composée de deux étapes; tout d'abord, le nombre d'attributs est réduit, en sélectionnant uniquement les plus pertinents, puis la classification est effectuée sur les attributs sélectionnés.

Sélection d'attributs

À cause du grand nombre d'attributs disponibles, une étape de sélection d'attributs doit être mise en place. La sélection d'attributs comporte deux étapes; l'une qui permet de déterminer le nombre d'attributs à sélectionner et l'autre de sélectionner les attributs.

L'algorithme SFFS (Sequential Forward Floating Search) (SFFS) (Pudil et al., 1994) est utilisé pour les deux étapes. Cet algorithme présente deux avantages: (i) il peut être utilisé avec plusieurs scores de classification (dans cette étude, le coefficient Kappa), (ii) il permet d'accéder à l'évolution du score de classification en fonction du nombre d'attributs sélectionnés. L'algorithme SFFS sélectionne p attributs qui maximisent le score de sélection d'attributs (le coefficient Kappa).

Classification

Un classificateur supervisé est utilisé afin de discriminer les essences forestières fournies par une base de données (BD) de couverture forestière existante. La classification est obtenue à l'aide de l'algorithme des forêts aléatoires. Cette méthode de classification a démontré être pertinente dans la littérature (Belgiu et al., 2016) et dans une étude précédente comparée aux Séparateurs à Vaste Marge (SVM) (Dechesne et al., 2016), car elle fournit des résultats similaires tout en étant plus rapide. Les résultats de la classification sont (i) une carte de label et (ii) une carte de probabilité (probabilités de chaque classe pour chaque pixel). Cette carte de probabilité est nécessaire pour l'étape de régularisation suivante.

1.2.3 Régularisation

La classification peut ne pas être suffisante pour obtenir des zones homogènes avec des frontières lisses. Par conséquent, afin de s'adapter au modèle de peuplement, la régularisation de la classification au niveau des pixels est nécessaire. Elle peut être effectuée de manière locale (par utilisation de filtre ou par relaxation probabiliste). Cependant, une régularisation globale par minimisation d'énergie reste la solution la plus optimale. De plus, une telle formulation permet d'insérer des contraintes supplémentaires pour une délimitation plus spécifique des peuplements forestiers.

Formulation de l'énergie

Le modèle énergétique proposé repose sur un modèle graphique, c'est-à-dire que le problème est modélisé par un graphe probabiliste prenant en compte les probabilités de classe P et la carte d'attributs normalisés (moyenne nulle et écart-type unitaire) A . Il se compose d'**un terme d'attache à la donnée** et d'**un terme de régularisation**. Pour une image I et une carte de classification C , l'énergie E est formulée comme suit:

$$E(I, C) = \sum_{u \in I} E_{\text{data}}(C(u)) + \gamma \sum_{u, v \in \mathcal{N}} E_{\text{pairwise}}(C(u), C(v)), \quad (1.7)$$

avec:

$$\begin{aligned} \gamma &\in [0, \infty[; \\ E_{\text{data}}(C(u)) &= f(P(C(u))); \\ E_{\text{pairwise}}(C(u) = C(v)) &= 0; \\ E_{\text{pairwise}}(C(u) \neq C(v)) &= V(u, v). \end{aligned}$$

\mathcal{N} désigne la 8-connexité, et $P(C(u))$ est la probabilité qu'un pixel u appartienne à la classe $C(u)$. γ est le paramètre de régularisation permettant de contrôler l'influence des deux termes et donc le niveau d'homogénéité des segments.

E_{data} est le terme d'attache aux données. E_{pairwise} est le terme de régularisation, permettant de mesurer la différence entre les attributs du pixel u et les attributs de ses 8 voisins. L'énergie exprime à quel point les pixels sont bien classés et à quel point les attributs sont similaires. D'autres modèles de champs aléatoires conditionnels pourraient être envisagés pour l'expression de l'énergie (Schindler, 2012; Volpi et al., 2015; Tuia et al., 2016). E_{pairwise} pourrait être exprimée relativement à plus de pixels.

La fonction f liée à E_{data} la plus efficace (voir Équation 1.7) est:

$$f(x) = 1 - x, \quad \text{avec } x \in [0, 1]. \quad (1.8)$$

Cette fonction permet de contrôler l'importance à donner au résultat de la classification. Une simple fonction linéaire permet de contrôler l'énergie: quand la probabilité est proche de zéro, l'énergie est maximale. Inversement, quand la probabilité est forte, l'énergie est minimale. Cette fonction permet de conserver ce terme dans $[0, 1]$ et ainsi simplifier le paramétrage de γ . D'autres formulations peuvent être envisagée mais il est apparu que cette formulation produit les meilleurs résultats pour notre problème.

Le terme de régularisation est une mesure qui contrôle la valeur de l'énergie en fonction de la valeur des attributs des 8 voisins. Deux pixels de classe différentes, mais avec des valeurs d'attributs proches sont plus susceptibles d'appartenir à la même classe que deux pixels de classes différentes avec des valeurs d'attributs différents. La valeur de l'énergie doit être proche de 1 lorsque les valeurs des attributs sont assez similaires et décroît à mesure de leur différence. Le terme de régularisation

V le plus efficace est exprimée comme suit:

$$V(u, v) = \frac{1}{n_{\text{opt}}} \sum_{i=1}^{n_{\text{opt}}} (-\exp(-\lambda_i |A_i(u) - A_i(v)|)). \quad (1.9)$$

$\forall i \in [1, n_{\text{opt}}], \lambda_i \in [0, \infty[$, où i est l'indice correspondant à l'attribut. $A_i(u)$ est la valeur du $i^{\text{ème}}$ attribut du pixel u .

$\lambda = [\lambda_1, \lambda_2, \dots, \lambda_{\text{opt}}]$ est un vecteur de longueur égale au nombre d'attributs (n_{opt}). Ce vecteur assigne des poids différents aux attributs. Si $\lambda_i = 0$ l'attribut i ne sera pas pris en compte dans le processus de régularisation. Quand λ_i est important, une petite différence entre les attributs impactera beaucoup l'énergie. Ainsi, l'attribut i aura un fort impact dans la régularisation. Les attributs étant de différents types (hauteur, réflectance, densité de végétation, etc.), il est donc important de garder les termes dans $[0, 1]$ pour chaque attribut, même si ils ne sont pas initialement dans la même gamme de valeurs. Dans notre étude, tous les poids λ_i sont fixés à 1 $\forall i$. D'autres formulations pour cette énergie peuvent être envisagées (un simple modèle de Potts par exemple).

Minimisation d'énergie

La minimisation de l'énergie est réalisée en utilisant des méthodes de coupe de graphe. L'algorithme de coupe de graphe utilisé est l'optimisation pseudo-boulienne quadratique (QPBO) avec α -expansion. Il s'agit d'une méthode de coupe de graphe très répandue qui résout efficacement les problèmes graphiques de minimisation d'énergie (Kolmogorov et al., 2007). L' α -expansion est une technique permettant de traiter les problèmes multi-classes (Kolmogorov et al., 2004).

Lorsque $\gamma = 0$, le terme de régularisation n'a aucun effet dans la formulation énergétique; la classe la plus probable est attribuée au pixel, conduisant au même résultat que la sortie de classification. Lorsque $\gamma \neq 0$, la carte d'étiquette résultante devient plus homogène, et les bords des segments sont plus lisses. Cependant, si γ est trop élevé, les petites zones sont liées pour être fusionnées dans des zones plus grandes, en supprimant une partie des informations utiles fournies par la classification.

1.3 Données

Les zones d'étude se situent sur des forêts de différentes régions de France présentant des paysages différents. Elles offrent une importante diversité des espèces d'arbres en présence. Chaque zone comprend un grand nombre d'espèces (4-5 espèces par zone), permettant de tester la robustesse de la méthode.

Les images multispectrales aéroportées ont été acquises par les caméras numériques IGN (Souchon et al., 2012). Elles sont composées de 4 bandes: 430-550 nm (bleu), 490-610 nm (vert), 600-720 nm (rouge) et 750-950 nm (proche infrarouge) avec une résolution spatiale de 0,5 m.

Les données lidar aéroportées ont été recueillies en utilisant un dispositif Optech 3100EA. L'empreinte était de 0,8 m afin d'augmenter la probabilité d'atteindre le sol. La densité de points pour tous les échos varie de 2 à 4 points/m². Nos données multispectrales et lidar sont en adéquation avec les normes utilisées dans de nombreux pays pour des fins de cartographie forestière à grande échelle (White et al., 2016). Les données ont été acquises en conditions foliaires, en mai et juin 2011 pour les images multispectrales et les données lidar, respectivement. Le recalage entre le nuage de points lidar et les images multispectrales VHR a été réalisé par l'IGN à partir de points de

contrôle au sol. Il s'agit d'une procédure standard de l'agence de cartographie française qui présente des résultats comparables aux solutions professionnelles standard.

La BD Forêt est composée de polygones 2D délimités par photo-interprétation. C'est une base de donnée géographique nationale française pour les forêts, librement accessible ¹. Elle est utilisée à la fois pour entraîner le classificateur et pour évaluer les résultats. Seuls les polygones contenant au moins 75% d'une espèce donnée sont utilisés pour la classification (c'est le seuil qui définit quand un peuplement peut être attribué à un type de végétation unique et considéré comme "pur"). Les polygones de végétation (*lande ligneuse* et *formation herbacée*) sont également pris en compte pour la classification. Comme cette étude ne s'intéresse qu'aux espèces, la vérité terrain utilisée ne couvrira donc pas l'ensemble de la zone (les zones contenant du mélange ne seront pas prise en compte).

1.4 Résultats

1.4.1 Sélection d'attributs

L'algorithme SFFS permet de déterminer le nombre optimal de d'attributs à sélectionner, et de les sélectionner. Le nombre optimal d'attributs à sélectionner est ici de 20. Ce nombre a été conservé pour les autres zones et les résultats montrent que le transfert n'a aucun impact sur la précision finale. Une fois le nombre optimal d'attributs déterminé, la sélection a été effectuée 40 fois sur toutes les zones de test afin de récupérer les attributs les plus pertinents. En moyenne, 61% des attributs sélectionnés proviennent de l'information spectrale et 39% de l'information lidar. Ceci montre la complémentarité des deux données de télédétection.

Pour les informations spectrales, les attributs dérivés des bandes originales sont plus pertinents que les indices de végétation. L'attribut statistique le plus pertinent pour l'information spectrale est le minimum (17% de la sélection spectrale). Le maximum (12%), la médiane (11%), la moyenne (11%) et l'écart-type (10%) sont également pertinents.

Pour l'information lidar, l'attribut le plus pertinent est étonnamment l'intensité, sélectionnée dans chacune des 40 sélections (12% de la sélection lidar, 5% de la sélection totale). L'écart-type de la hauteur(8% de la sélection lidar), le maximum de la hauteur(7%) et les densités (5% et 6% de la sélection lidar) sont également pertinents.

1.4.2 Classification

Les résultats de la classification et leur impact sur la segmentation finale sont illustrés par les Figures 1.1 et 1.2 pour une seule zone. La précision est calculée en comparant chaque pixel de la BD Forêt aux résultats de classification. Il apparaît clairement que l'approche de classification au pixels conduit à des cartes d'étiquettes bruitées, même si l'on obtient des résultats corrects (70.5% de précision globale). D'un autre côté, la classification sur les objet produit des étiquettes plus cohérentes sur le plan spatial (93.14% de précision globale). Cette différence est directement reflétée sur la segmentation finale. En effet, la régularisation basée sur la classification au pixel obtient 91.9% de correspondance correcte avec la BD Forêt, tandis que la régularisation basée sur la classification

¹<http://inventaire-forestier.ign.fr/spip/?rubrique67>

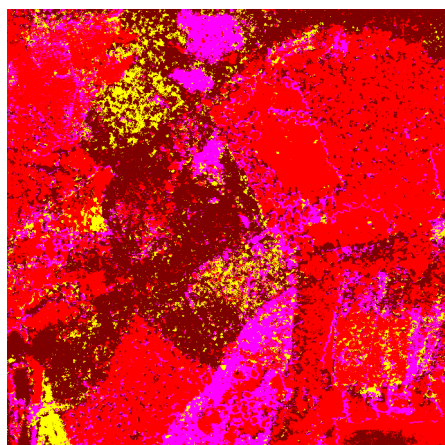
par objets 97.4%. Ces résultats confirment la pertinence du niveau objet pour la classification. Ces résultats ont également été observés pour les autres zones.



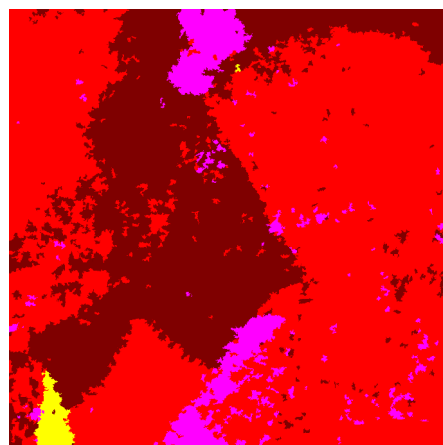
(A) Image optique IRC THR.



(B) MNS.



(C) Classification fondée pixel (précision globale: 70.48%, κ : 0.50).



(D) Classification fondée objet (PFF) (précision globale: 93.14%, κ : 0.86).

FIGURE 1.1: Résultats de classification; comparaison entre la méthode fondée pixel et la méthode fondée objet.

1.4.3 Régularisation

Les résultats finaux sont présentés par la figure 1.3. La précision globale montre que la méthode donne des résultats satisfaisants en termes de discrimination des espèces d'arbres, dans la gamme des documents existants de la littérature pour le même nombre d'espèces (Leckie et al., 2003; Lepänen et al., 2008). Deuxièmement, les résultats sont améliorés jusqu'à 15% grâce au lissage des résultats. Malgré des précisions très élevées, les résultats doivent être considérés avec précaution. Ils sont comparés à une BD qui comporte des défauts et ne couvre pas l'intégralité des zones d'étude. L'analyse visuelle des scores de segmentation peuvent en fait être pris avec une marge de $\pm 5\%$. Le réglage de γ est également une étape importante; quand il est trop bas (par exemple < 5), certains petits segments non pertinents peuvent rester dans la segmentation finale, ce qui entraîne un score de segmentation faible. Cependant, cette sur-segmentation peut être utile pour l'inventaire forestier

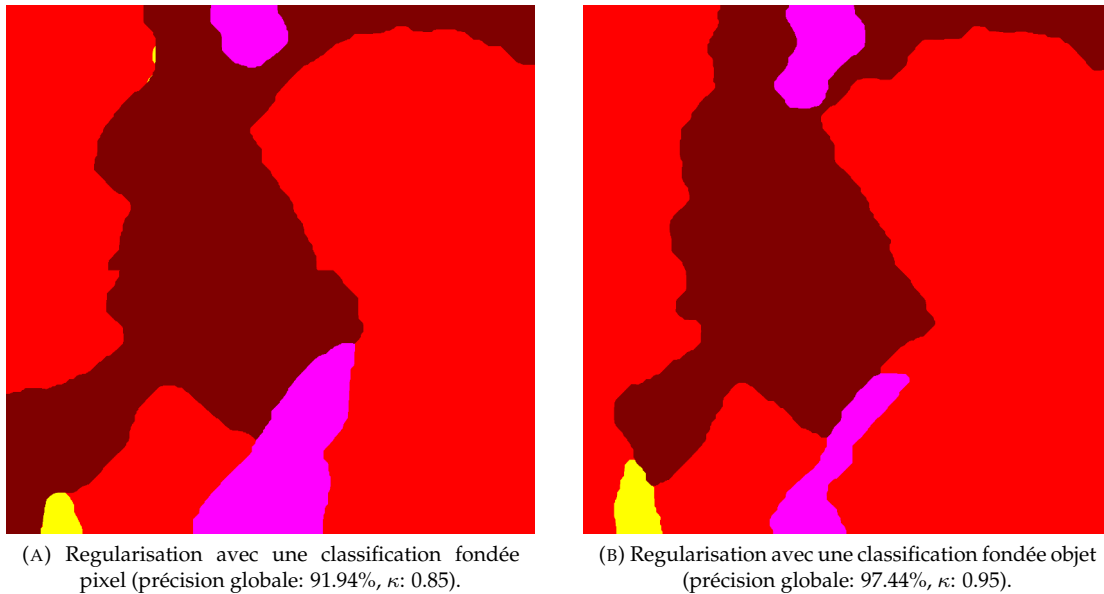


FIGURE 1.2: Résultats de régularisation; comparaison entre la méthode utilisant une classification fondée pixel et la méthode utilisant une classification fondée objet.

statistique et la cartographie, car elle fournit de petites zones homogènes au sein de grands peuplements. Lorsque γ se trouve dans $[10, 15]$, la précision de segmentation est élevée: chaque classe est correctement étiquetée à plus de 95%. Cependant, avoir un γ élevé (> 20) diminue légèrement la précision de la segmentation, mais réduit également considérablement la précision des classes sous-représentées (une classe ayant une petite superficie sera fusionnée à une classe ayant une plus grande superficie).

1.5 Conclusion

Une méthode en trois étapes pour la délimitation des peuplements forestiers, en termes d'espèce, a été proposée. La fusion des données lidar aéroporté et des images multispectrales produit des résultats très satisfaisants puisque les deux modalités de télédétection fournissent des observations complémentaires. Les segments finaux obtiennent un bon score de correspondance avec les peuplements délimités par les opérateurs humains. La méthode repose sur le calcul d'attributs lidar et spectraux à différents niveaux (pixel et objet) pour une classification supervisée des espèces d'arbres. De bons scores de classification sont obtenus, produisant une base solide pour une délimitation des peuplements. Cette délimitation est alors effectuée grâce à une régularisation globale de la classification. Elle repose sur un modèle énergétique formulé en fonction des résultats des probabilités de classification et des valeurs des attributs. Elle permet d'obtenir des zones homogènes en termes d'espèces avec des frontières lisses. De plus, il est possible de contrôler le niveau des détails requis pour la délimitation du peuplement, qui dépend de l'inventaire forestier ou des spécifications de la BD.

L'objectif étant de délimiter les peuplements forestiers en fonction du type de végétation (principalement des espèces d'arbres), l'utilisation d'images superspectrale ou hyperspectrales pourrait être intéressante pour obtenir plus d'informations sur les espèces et surtout surmonter le problème

de la variabilité des espèces. D'autres indices de végétation peuvent également être dérivés à partir des données hyperspectrales. De plus, l'utilisation de données lidar à haute densité (~ 10 pts/m²) pourrait également améliorer les résultats de la méthode.

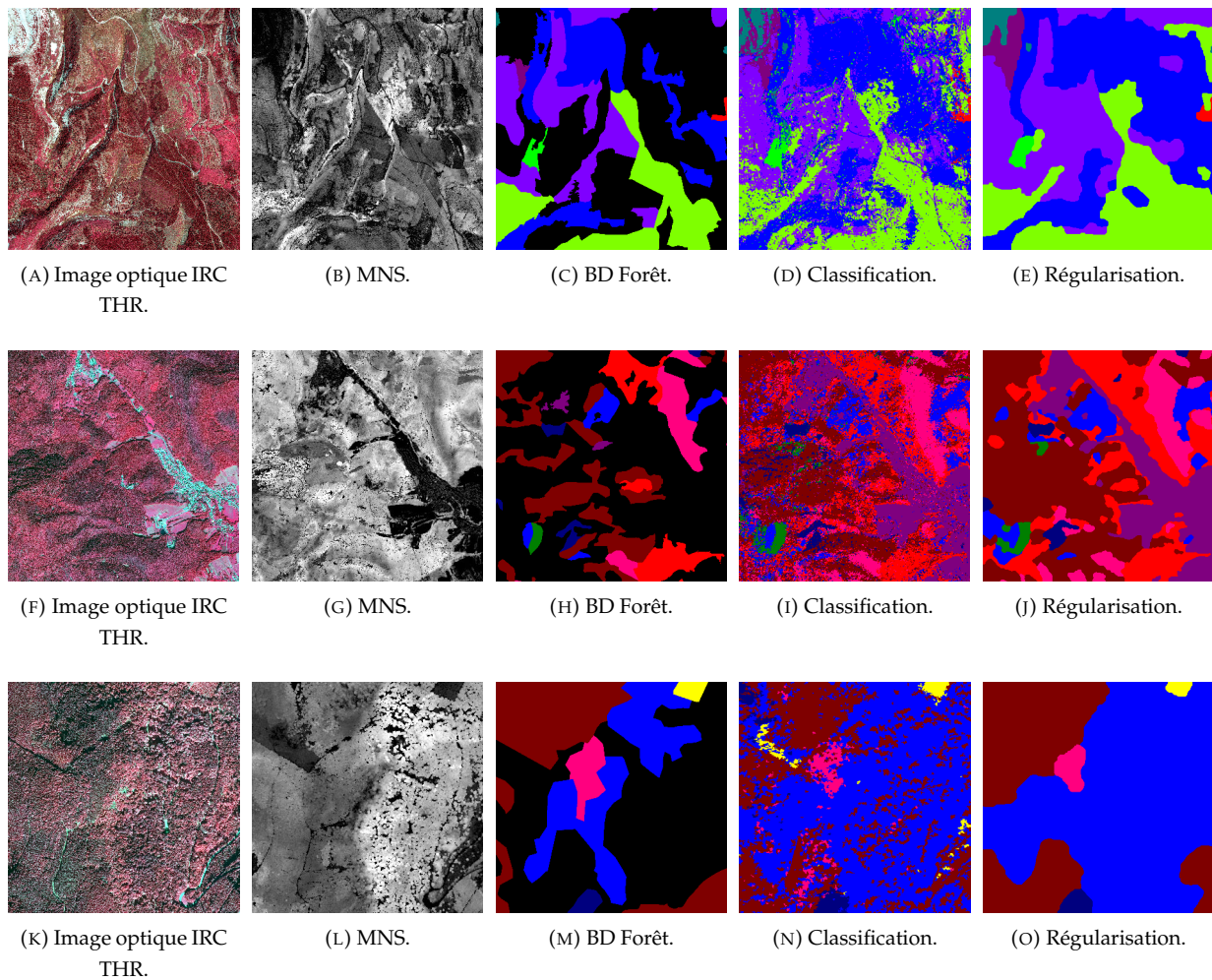


FIGURE 1.3: Résultats sur différentes zones.

Introduction

1.1	Analysis of forested areas	2
1.2	Remote sensing for forested areas	4
1.3	Context of the thesis	8
1.4	Objectives	10
1.5	Strategy	11
1.6	Structure of the thesis	12

1.1 Analysis of forested areas

Forests are a core component of planet's life. They are defined as large areas dominated by trees. Hundreds of other definitions of forest may be used all over the world, incorporating factors such as tree density, tree height, land use, legal standing and ecological function (Schuck et al., 2002; Achard, 2009).

Forest are commonly defined "**Land spanning more than 0.5 hectares (ha) with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds *in situ*. It does not include land that is predominantly under agricultural or urban land use**" according to *Food and Agriculture Organization (FAO)* ¹ (Keenan et al., 2015). They may consist either of closed forest formations where trees of various storeys and undergrowth cover a high proportion of the ground; or open forest formations with a continuous vegetation cover in which tree crown cover exceeds 10 %. Young natural stands and all plantations established for forestry purposes which have not yet reach a crown density of 10 percent or tree height of 5 m are included under forest, since they are normally forming part of the forest areas which are temporarily unstocked as a result of human intervention or natural causes, but which are expected to revert to forest.

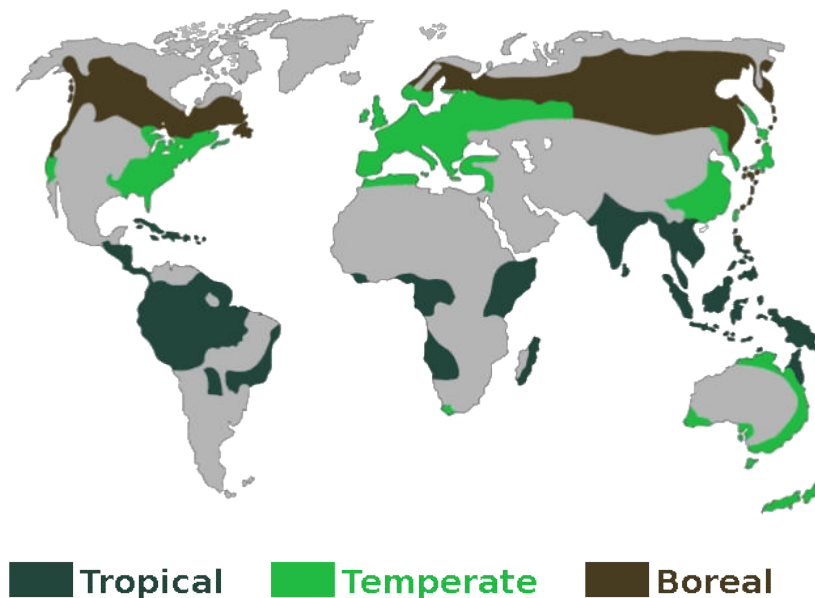


FIGURE 1.1: Forest repartition and categorization in the world. The three categories of forest are distributed according to their distance to the equator ².

Forests are the dominant terrestrial ecosystem of Earth, and are distributed across the globe (Pan et al., 2013). They cover about four billion hectares, or approximately 30% of the World's land area (see Figure 1.1). Forests at different latitudes and elevations form distinct ecozones: boreal forests near the poles, tropical forests near the Equator and temperate forests at mid-latitudes (see Figure 1.1). Higher elevation areas tend to support forests similar to those at higher latitudes, and the amount of precipitation also affects forest composition. Since these ecozones are very different,

¹<http://www.fao.org/docrep/017/ap862e/ap862e00.pdf>

²<http://dyntatrishana.blogspot.fr/2015/11/about-value-of-forests.html>

the study at a fine level (e.g. species composition) of forested areas must be restricted to a single ecozone at a time.

Human society and forests influence each other in both positive and negative ways (Vogt et al., 2006). On one hand, human activities, including harvesting forest resources, can affect forest ecosystems. On the other hand, forests have three main contributions to human: ecosystem services, tourist attraction and harvesting.

Ecosystem services.

Forests provide ecosystem services. They are involved in the provisioning of clean drinking water and the decomposition of wastes. Forests account for 75% of the gross primary productivity of the Earth's biosphere, and contain 80% of the Earth's plant biomass (Pan et al., 2013). They also hold about 90% of terrestrial biodiversity (Brooks et al., 2006; Wasiq et al., 2004). Forests are also beneficial for the environment; they capt and store the CO₂ (Fahey et al., 2010) (see Figure 1.2). About 45% of the total global carbon is held by forests. They also filter dust and microbial pollution of the air (Smith, 2012). Finally, they also play an important role in hydrological regulation and water purification (Lemprière et al., 2008) (see Figure 1.2).

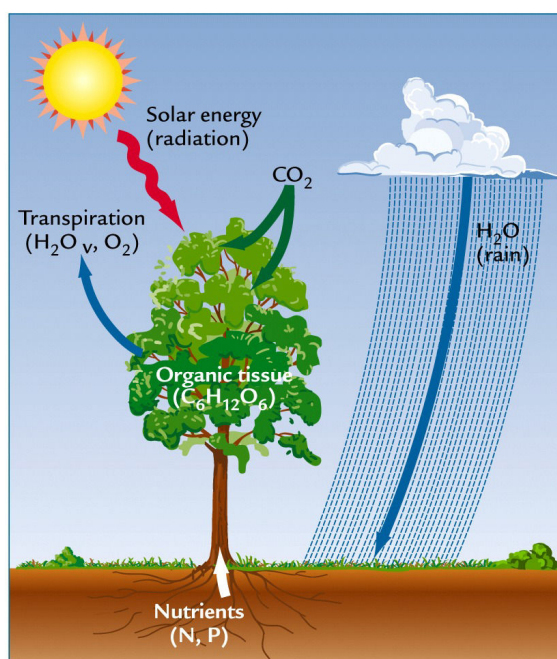


FIGURE 1.2: Carbon cycle: a process of CO₂ storage, water and air purification³.

Harvesting/wood resources.

Wood from trees displays many uses. It is still widely used for fuel (Sterrett, 1994). In this case, hardwood is preferred over softwood because it creates less smoke and burns longer. Wood is still an important construction material (Ramage et al., 2017): Elm was used for the construction of wood boats. In Europe, oak is still the preferred variety for all wood constructions, including beams, walls, doors, and floors (Thelandersson et al., 2003). A wider variety of woods is also used such as poplar, small-knotted pine, and Douglas fir. Wood is also needed in the paper industry since wood fibers

³<http://www.atmosedu.com/Geol390/Life/CarbonCycleShort.html>

are an important component of most papers. Eventually, wood is also extensively used for furniture or for making tools or music instruments.

Tourist attraction and recreative activities.

Forests serve as recreative attractions. In France, there are hundred of long distance footpaths (~60000km) through forests. Other activities such as rock climbing, mountain bike or adventure parks are mostly practiced in forests.

The evolution of forests need to be monitored in order to efficiently exploit the forest resources in a sustainable way (Paris Agreement 2015). For example, France is a significant wood importer (~25 millions of m³ per year), while the French forest is the third in Europe in term of volume ⁴. It is therefore needed to better manage and exploit the French wood stocks.

The assessments of these stocks are all the more difficult because a large part (about 75%) of the French forest is private, leading to a more complex management and exploitation. Furthermore, they can be assessed at different (i.e. forest, stand/plantation, tree) levels with more or less accuracy.

In order to evaluate the forest resources, a precise mapping combined with accurate statistics of forests therefore is needed. Such statistics are already operational at a national level and widely employed for the evaluation of forest resources. However, a precise mapping would allow to refine the evaluation of forest resources. Forests are complex structures (Pommerening, 2002), for which information is needed for management, exploitation and more generally for public and private policies. Such information can be the tree species or the tree maturity of the forest. There are two ways to extract such information from forest; **field inventory** or **remote sensing**. The field inventories are very expensive to set up and are adapted for statistics (considering a limited set of inventory sites) but only at a regional or national scale. Remote sensing is a more relevant way in order to obtain such information since it allows to extract them at larges scale.

In order to meet these needs, two synergistic products could be produced: **statistical inventory** or **forest mapping**.

1.2 Remote sensing for forested areas

Remote sensing through automatic Earth Observation image analysis has been widely recognized as the most economic and feasible approach to derive land-cover information over large areas and for a large range of needs. The obvious advantage is that remote sensing techniques can provide LC information on different levels of details in a homogeneous and reliable way over large scales. They can also provide bio-geophysical variables and change information (Hansen et al., 2013), in addition to the current cover of the Earth surface.

The analysis of forested areas from a remote sensing point of view can be performed at three different levels: pixel (straightforward analysis), object (mainly trees), plots or stands. In statistical national forest inventory (NFI), an automated and accurate tree segmentation would simplify the extraction of tree level features (basal area, dominant tree height, etc., (Means et al., 2000; Kangas et al., 2006)), since there is no straightforward way to obtain them. Two kinds of features can be

⁴<http://ec.europa.eu/eurostat/documents/3217494/5733109/KS-31-11-137-EN.PDF>

extracted, the ones estimated directly from remote sensing data and the ones interpolated using allometric equations.

The tree level is not the most suitable level of analysis for forest studies at a national scale but should be preferred for local studies. The plots correspond to the level of analysis in fields inventories; statistics are derived at this level and interpolated at a larger scale. When a joint mapping and statistical reasoning is required (e.g., land-cover (LC) mapping and forest inventory (Tomppo et al., 2008)), forest stands remain the prevailing scale of analysis (Means et al., 2000; White et al., 2016). A stand can be defined in many different ways in terms of homogeneity: tree species, age, height, maturity. Its definition varies according to the countries and agencies.

From a remote sensing point of view, the delineation of the stands is a segmentation problem. Forest stands are preferred, since they allow to extract reliable and statistically meaningful features and to provide an input for multi-source statistical inventory. For land-cover mapping, this is highly helpful for forest database updating (Kim et al., 2009), whether the labels of interest are *vegetated areas* (e.g., *deciduous/evergreen/mixed/non-forested*), or, even more precisely, the tree species.

- **Manual delineation.**

To obtain such information, most of the time in national forestry inventory institutes, for reliability purposes, each area is manually interpreted by human operators with very high spatial resolution (VHR) geospatial images focusing on the infra-red channel (Kangas et al., 2006). This work is extremely time consuming and subjective (Wulder et al., 2008b). Furthermore, in many countries, the wide variety of tree species (e.g., >20) significantly complicates the problem. This is all the more true than photo-interpretation may not always be sufficient (consensus may even not be reached between experts) and even in case of few species (3-5).

- **Automatic delineation.**

The design of an automatic procedure based on remote sensing data would fasten and ease such process. Additionally, the standard manual delineation procedure only takes into account the species, and few physical characteristics (alternatively height, age, stem density or crown closure). Instead, an automatic method could offer more flexibility not being limited to a visual analysis and using characteristics extracted from complementary data sources and not only Colored Infra-Red (CIR) ortho-images.

The use of remote sensing data for the automatic analysis of forests has been growing in the last 15 years, especially with the synergistic use of airborne laser scanning (ALS) and optical VHR imagery (multispectral imagery and hyperspectral imagery) (Torabzadeh et al., 2014; White et al., 2016). Several countries have already integrated such remote sensing data sources in their operational pipeline for forest management (Tokola, 2015; Wulder et al., 2008a; Patenaude et al., 2005) and characterization. Furthermore, they can be employed for forest management. ALS provides a joint direct access to the vertical distribution of the trees and to the ground underneath (Holmgren, 2004). Hyperspectral and multispectral optical images are particularly relevant for tree species classification: spectral and textural information from VHR images can allow a fine discrimination of many species (Clark et al., 2005; Franklin et al., 2000). Multispectral images are often preferred due to their higher availability, and higher spatial resolution (Belward et al., 2015). Multispectral and hyperspectral images can be acquired from airplanes or satellites. Spaceborne sensors allow to capture large areas with a higher temporal rate but generally suffer from a lower spatial resolution, even if the gap decreases every year (see Table 1.1). For a better spatial resolution,

airborne multispectral images are preferred since they also allow to extract more relevant texture features for tree species classification (Franklin et al., 2000). The airborne linear Lidar technology has been widely used for remote sensing tasks (Lim et al., 2003; Shan et al., 2008; Vosselman et al., 2010). Lidar has been successfully employed for many forest applications (Ferraz et al., 2016b). The new Geiger mode lidar (Ullrich et al., 2016) and single photon lidar (Viterbini et al., 1987) are also very promising, allowing a significantly higher point density with different angles at a higher altitude, enabling the coverage of larger areas at a better cost than classic Lidar systems. Employing such 3D data could have an important impact, especially on the studies of forested areas (Jakubowski et al., 2013; Strunk et al., 2012). Similarly to hyperspectral Lidar (Kaasalainen et al., 2007), additional research is required to accurately assess that relevance on our scope.

Synthetic Aperture Radar (SAR) is widely employed for the evaluation of biomass, especially in forested environment (Le Toan et al., 1992; Beaudoin et al., 1994). With its ability to penetrate the vegetation, SAR in P-band (0.3-1 GHz) allows to estimate efficiently the aboveground biomass. Thus, SAR can be employed in order to extract relevant information of forests but is not well adapted for the discrimination of the species.

	SPOT 6,7	Ikonos	Quickbird	Pléiades	RapidEye	Sentinel 2	Landsat 7	Worldview 3
Swath	60 km	11 km	16.5 km	20 km	77 km	290 km	185 km	13.1
Revisit time	2 d	2 d	1-3.5 d	1 d	5.5 d	5 d	16 d	<1 d
Resolution	6 m	4 m	2.44-2.88 m	2.8 m	6.5 m	10 m	30 m	1.24 m
Number of bands	4	4	4	4	5	13	8	28
Main spectral bands (nm)								
Blue	455-520	450-530	450-520	430-550	440-510	460-520	450-520	450-510
Green	530-600	520-610	520-600	500-620	520-590	545-575	520-600	510-580
Red	620-690	640-720	630-690	590-710	630-685	650-680	630-690	630-690
NIR	760-890	760-880	760-900	740-940	760-850	785-900	760-900	770-895

TABLE 1.1: Principal multispectral spatial optical sensors.

1.3 Context of the thesis

In France, the study of forests mainly consist in mapping and inventory. It can also be envisaged for the assessment of biodiversity, the impact of forests on human behaviors or on climate etc.

The forest inventory of IGN allows to obtain an estimation of the wood stock and the forestation rate at different scales (national, regional, departmental, see Figures 1.3 & 1.4). Statistics such as volume per hectare, deciduous volume or conifer volume can then be derived. The inventory is performed through extrapolation of field inventories.

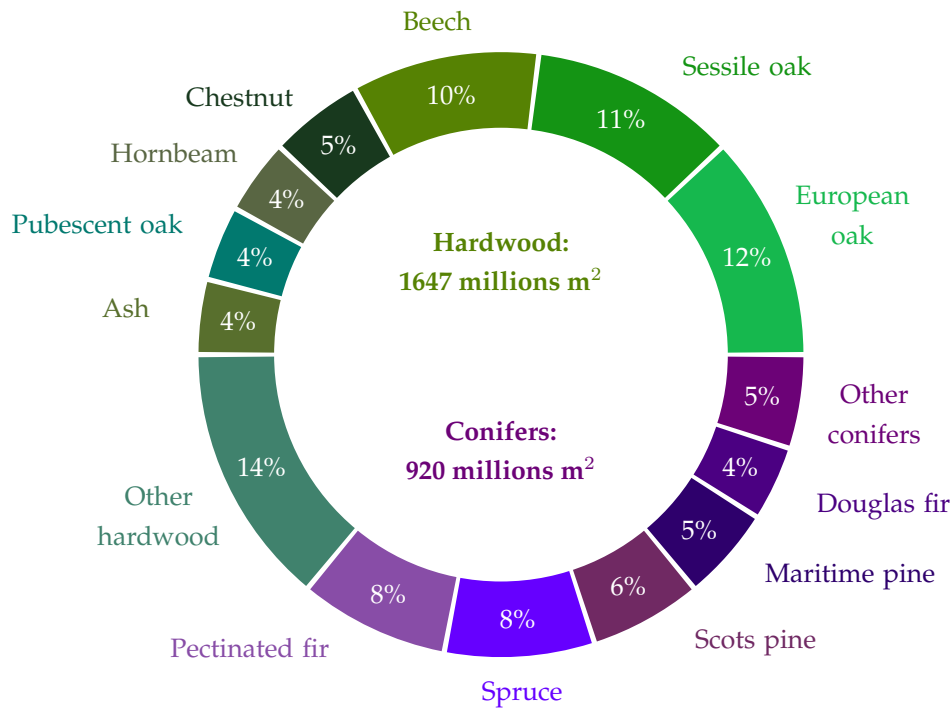


FIGURE 1.3: Distribution wood volume per species at the national scale⁵.

Forest mapping is also interesting for the understanding of forested areas. It is traditionally provided through a national forest land-cover (LC) database (DB) (see Figure 1.5). In France, it is manually interpreted by human operators using VHR CIR ortho-images. It assigns a vegetation type to each mapped beach of more than 5000 m². The nomenclature is composed of 32 classes based on hierarchical criteria such as pure stands of the main tree species of the French forest. The forest LC should be updated in a 10 years cycle.

⁵http://inventaire-forestier.ign.fr/spip/IMG/pdf/Int_memento_2013_BD.pdf

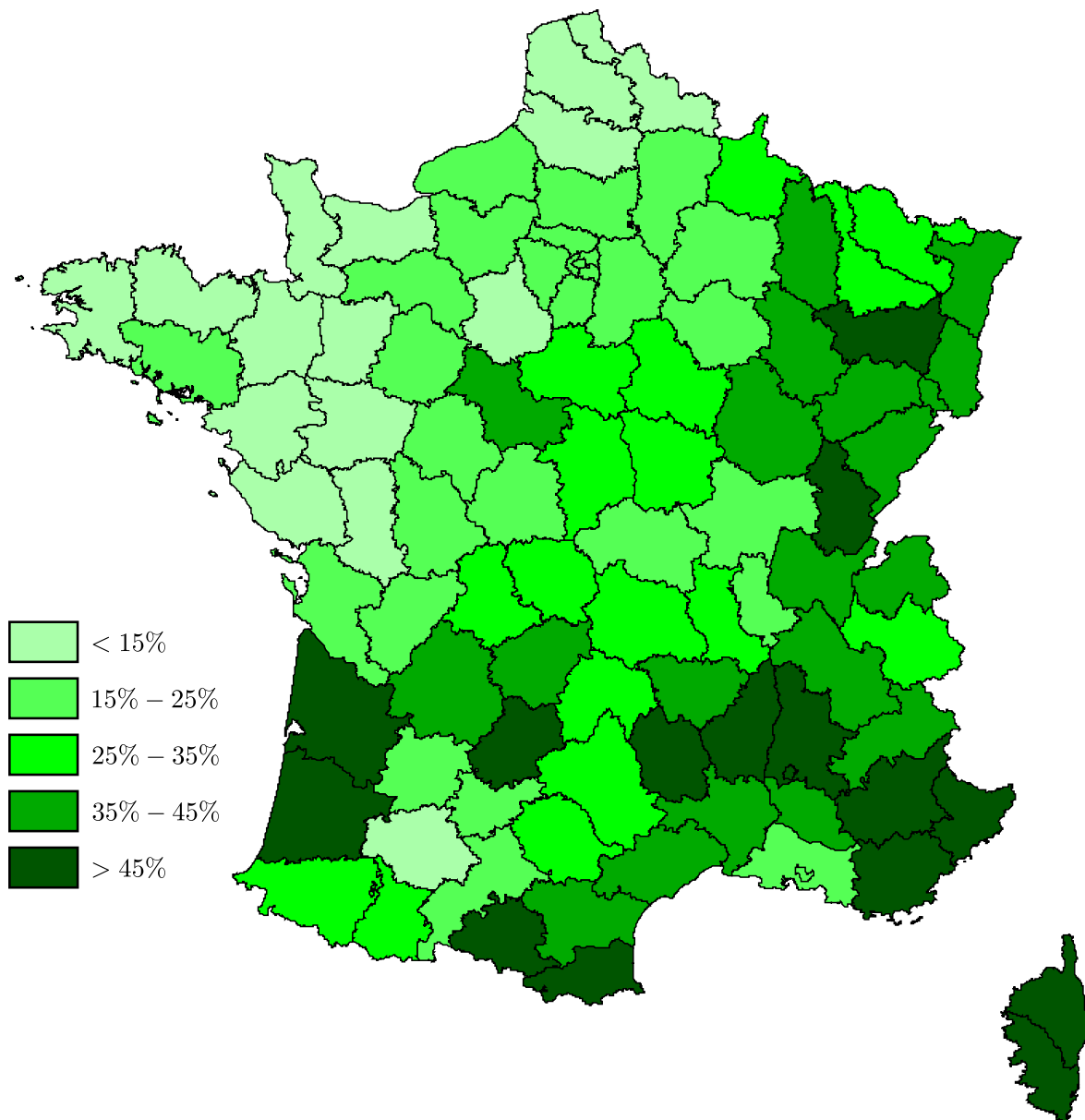


FIGURE 1.4: Forestation rate in France ⁶.

⁶http://inventaire-forestier.ign.fr/spip/IMG/pdf/Int_memento_2013_BD.pdf

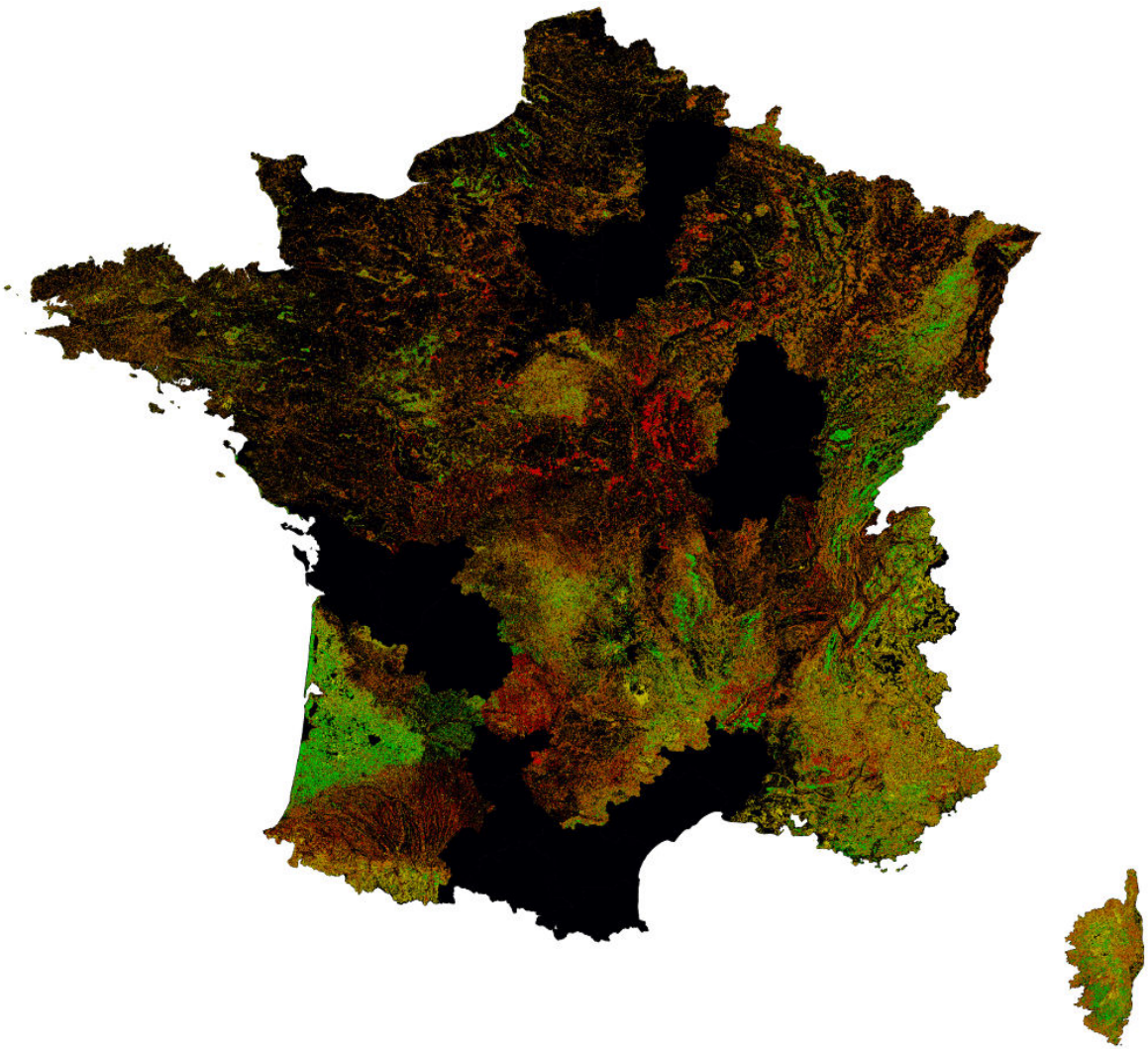


FIGURE 1.5: The French forest LC. Each color is associated to a single specie (~20 species in total) available on 76 department over 96, black corresponds to non-labeled zones (not operated or non forested).

1.4 Objectives

Currently, the forest LC DB is obtained through remote sensing (namely photo-interpretation), but it is a time consuming activity. A framework should be developed to update it automatically using remote sensing data processing. Since the forest LC is available, it can be used as an input for training a supervised classification (Gressin et al., 2013a). However, the learning process should be carefully addressed (Gressin et al., 2014). Indeed, some areas might have **changed** (e.g., forest cuts). Furthermore, the database is designed **generalized** (Smith et al., 1977). Indeed, forests are not perfectly homogeneous in term of species and there can be many gaps in the canopy, leading also to a noisy classification. Such classification would then not be sufficient in order to retrieve homogeneous patches similar to the forest LC. In order to retrieve homogeneous patches, the classification should be regularized using smoothing methods (Schindler, 2012). Furthermore, an automatic framework

considering more data sources than only CIR ortho-images would allow to enrich the LC, i.e. retrieve homogeneous tree species stands also homogeneous in terms of height.

1.5 Strategy

Two remote sensing modalities are available for the mapping of forested areas at IGN; VHR optical images and lidar points cloud. VHR optical images are acquired at the national level for various needs of IGN, while lidar is mostly employed order to derive an accurate Digital Surface Model especially in forested areas since it is the better solution to obtain it in such environment

VHR optical images

In this thesis, the VHR ortho-images employed have a spatial resolution of 50 cm. The ortho-images employed have 4 bands (red, green, blue near infra-red) captured by the IGN digital cameras (Souchon et al., 2012) that exhibit very high radiometric and geometric quality. Such VHR optical multispectral images are available over whole France. They are captured every 3 years and are one of the component of RGE (a public service mission of the IGN, that aims at describing the national land cover in a precise, complete and homogeneous way).

Airborne Laser Scanning

IGN also acquires 3D point clouds with laser scanning devices (Leica ALS 60 or Optech ALTM 3100). The point density for all echoes ranges from 2 to 4 points/m². Forested areas and areas subject to flooding are mainly flown. About 40000 km² are acquired each year for Digital Terrain Model generation as main purpose.

The registration between airborne lidar point clouds and VHR multispectral images was performed by IGN itself using ground control points, following a standard procedure production in the French mapping agency since IGN operates both sensors and has also a strong expertise in data georeferencing.

The combination of these two data sources is very relevant for the study of forests. Indeed, optical imagery provides the major information about the tree species (spectral and texture), while Lidar gives information about the vertical structure of the forests. Furthermore, Lidar allows to extract consistent objects such as trees, that could be used in the stand segmentation process, even if delineated coarsely.

The fusion of these two modalities is a way to extract the most information in order to retrieve forest stands. The fusion can be performed at different levels. 3 levels are frequently considered :

- Low level (or observation level): It corresponds to the fusion of the observations, in this case, the reflectance from the optical images and the coordinates of the lidar points. It is a straightforward fusion method that does not really extract information from the data. It is also simple way to validate the complementarity of the data.
- Medium level (or attribute level): It corresponds to the fusion of features, derived from both sources, and merged together. It also corresponds to the cooperative understanding of the

data; a feature is derived on a modality and applied to the other (e.g. segmentation of the point cloud applied to images). In this process, all the information from both data sources is directly exploited. However, attention should be paid to the choice of the employed features since it can lead to poor classification results.

- High level (or decision level or late fusion): It corresponds to decision fusion. Each data source has been processed independently (e.g. classified) and the final decision is an optimal combination of the classifications and the input data. This level of fusion is very important since it allows to refine the results and only keep the best from the intermediate results.

In this work, the fusion is mainly performed at the medium and high levels.

1.6 Structure of the thesis

This work is divided in 6 chapters.

- Chapter 2 presents and discusses the different existing methods for stand segmentation. Since stand segmentation is at the interplay between different kinds of image processing methods, they are also analyzed since they are also employed in this work for stand segmentation.
- Chapter 3 describes the proposed framework. It is composed of three steps. The first one is related to the extraction of features, it is composed of two core elements. Firstly an object extraction is carried out in order to derive features at the object level (medium level fusion for the cooperative understanding of the data). The desired objects aim to have a size similar to trees. Secondly, features (~ 100) are extracted from the two remote sensing modalities. The second part deals with the object-based classification. Here, a special attention is paid to the design of a training set. A feature selection is also carried out since it allows to validate the complementarity of data sources (medium level fusion) while reducing computation times. Finally, Random Forest classification is performed on the selected features (medium level fusion, since the features are spectral-based and lidar-based). This classification is then refined with regularization methods.
- Chapter 4 presents the results of the different experiments that have been carried out. Since many options can be envisaged for obtaining a relevant result (e.g. features, training set design, regularization), a large variety of experiments have been proposed for the contribution assessment of the different steps of the proposed framework.
- Chapter 5 emphasizes on the last step of the proposed framework. It aims at regularization of the classification. This step corresponds to the high level fusion. Indeed, the supervised classification does not allow to retrieve consistent forest stands (according to the forest LC DB). Thus, a regularization process allows to refine the results in order to obtain homogeneous segments with smooth borders and consistent with the forest LC DB. Such regularization can be performed using local or global methods both having their advantages and drawbacks.
- Chapter 6 summarizes and analyzes the different levels of fusion proposed in the framework. From the different fusion schemes possible, experiments are proposed in order to define what can be the optimal fusion schemes.

- Conclusions are drawn in Chapter 7. Eventually, perspectives are also proposed so as to alleviate remaining issues.

State of the art

2.1	Stand segmentation	16
2.1.1	Stand segmentation using VHR optical images	18
2.1.2	Stand segmentation using lidar data	18
2.1.3	Stand segmentation using VHR optical images and lidar	19
2.1.4	Challenges of stand segmentation	20
2.2	Segmentation	22
2.2.1	"Traditional" segmentation methods	22
2.2.2	Superpixels methods	24
2.2.3	Segmentation of point cloud	25
2.3	Classification	26
2.3.1	Supervised classification: common algorithms	26
2.3.2	Random Forest	26
2.4	Dimension reduction and feature selection	28
2.4.1	Dimension reduction: feature extraction	28
2.4.2	Feature selection	28
	Existing methods	28
	Selection optimization	29
2.5	Smoothing methods	30
2.5.1	Local methods	30
2.5.2	Global methods	31
2.6	Conclusion	32

Forest are complex areas, thus, the mapping of such environment needs the use of different image processing methods. Indeed, the extraction of "homogeneous" forest stands is at the interplay between different kinds of image processing methods (see Figure 2.1).

Several methods have already been proposed for forest stand segmentation depending on the definition of forest stand (Section 2.1). They involve different image processing algorithms which will be considered with details. Segmentation (Section 2.2) algorithms can be employed for a fine or coarse delineation of the principal components of the forests. Classification is also very useful to discriminate the different elements of the forest and especially detect tree species (Section 2.3). Furthermore, with the important number of features that can be derived from original data, feature selection algorithms are mandatory in order to improve the results while decreasing the computational load and times (section 2.4). Eventually, smoothing methods could be employed as a post processing in order to obtain a better labeling configuration. Especially global smoothing methods aims at the configuration corresponding to a minimum of an energy. Such energy minimization processes are used for a refinement of raw results (Section 2.5).

2.1 Stand segmentation

A forest stand is defined as a contiguous group of trees that is uniform in specie composition, structure, age and/or height, spatial arrangement, site quality or condition to distinguish it from adjacent groups of trees.

One should note that the literature remains heavily focused on individual tree extraction and tree species classification (Dalponte et al., 2014; Kandare et al., 2014; Véga et al., 2014; Dalponte et al., 2015), developing site-specific workflows with similar advantages, drawbacks, and classification performance. Some authors have focused on forest delineation (Eysn et al., 2012; Wang et al., 2012; Radoux et al., 2007), that most of the time do not convey information about the tree species and their spatial distribution. Forest stand delineation methods have been proposed but they generally remain very specific to the study area and commonly uniquely provide a binary mask as final output. Consequently, no operational framework embedding the automatic analysis of remote sensing data has been yet proposed in the literature for forest stand segmentation at large scale (Dechesne et al., 2017).

Hence, in the large amount of literature in the field, only few papers focus on the issue of stand segmentation or delineation. They can be categorized with regard to the type of data processed as presented in the following sections.

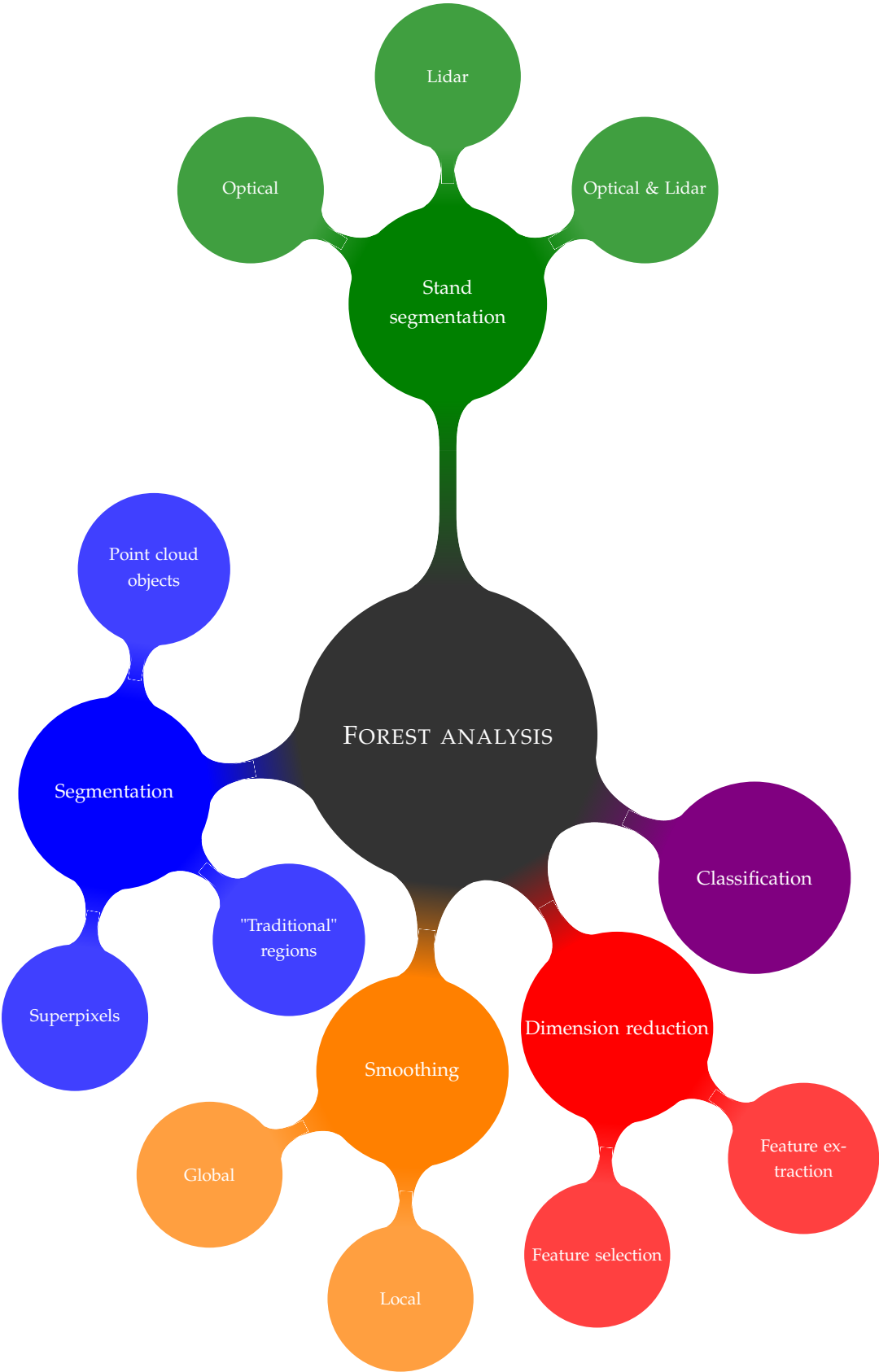


FIGURE 2.1: Processes involved in the analysis of forested areas.

2.1.1 Stand segmentation using VHR optical images

A stand delineation technique using VHR airborne superspectral imagery (0.6 m spatial resolution, 8 spectral bands ranging from 438.5 nm to 860.7 nm with an approximate bandwidth of 26 nm) is proposed in (Leckie et al., 2003). The trees are extracted using a valley following approach and classified into 7 tree species (5 coniferous, 1 deciduous, and 1 non-specified) with a maximum likelihood classifier. The classification is performed at the object (i.e. tree) level using statistical features (mean, maximum, standard deviation) and textural features. A semi-automatic iterative clustering procedure is then introduced to generate the forest polygons. The method produces relevant forest stands and consider many tree species. It shows the usefulness of an object-based classification using statistical and textural features. However, since experiments have been conducted on a small area (330 m × 800 m), no strong conclusion can be drawn.

A hierarchical and multi-scale approach for the identification of stands is adopted in (Hernando et al., 2012). The data inputs were the 4 bands of an airborne 0.5 m orthoimage (Red, Green, Blue, and Near Infra-Red) allowing to derive the Normalized Difference Vegetation Index (NDVI). The stand mapping solution is based on the Object-Based Image Analysis (OBIA) concept. It is composed of two main phases in a cyclic process: first, segmentation, then classification. The first level consists in over-segmenting (using the *multi-resolution segmentation algorithm* from *eCognition*) the area of interest and performing fine-grained land cover classification. The second level aims to transfer the vegetation type provided by a land cover geodatabase in the stand polygons, already retrieved from another segmentation procedure. The multi-scale analysis appears to have a significant benefit on the stand labeling but the process remains highly heuristic and requires a correct definition of the stand while we consider it is an interleaved problem.

Following the work of (Wulder et al., 2008b) with IKONOS images, Quickbird-2 panchromatic images are used in (Mora et al., 2010) to automatically delineate forest stands. A standard image segmentation technique from *eCognition* is used and the novelty mainly lies on the fact that its initial parameters are optimized with respect to NFI protocols. They show that meaningful stand heights can be derived, which are a critical input for various modeled inventory attributes.

The use of VHR optical images is very interesting since it is meaningful for tree species discrimination. Furthermore, statistical and textural features, that can be computed thanks to the high spatial resolution allows a better discrimination of tree species. Eventually, the Object-Based Image Analysis (OBIA) is also possible and preferred in order to obtain better results.

2.1.2 Stand segmentation using lidar data

A seminal stand mapping method using low density (1-5 point/m²) airborne lidar data is proposed in (Koch et al., 2009). It is composed of several steps of feature extraction, creation and raster-based classification of 15 forest types. Forest stands are created by grouping neighboring cells within each class. Then, only the stands with a pre-defined minimum size are accepted. Neighboring small areas of different forest types that do not reach the minimum size are merged together to an existing forest stand. The approach offers the advantage of detecting 15 forest types

(deciduous/coniferous and maturity) that match very well with the ground truth but to the detriment of simplicity: the flowchart has to be highly reconsidered to fit to other stand specifications. Additionally, the tree species discrimination is not addressed.

The forest stand delineation proposed in (Sullivan et al., 2009) also uses low density (3-5 point/m²) airborne lidar still coupling an object-oriented image segmentation and a supervised classification procedure implemented in *FUSION*. Three features are computed and rasterized also with the *FUSION* software. The segmentation is performed using a region growing approach. Spatially adjacent pixels are grouped into homogeneous objects or regions of the image. Then, a supervised discrimination of the segmented image is performed using a Battacharya classifier, in order to determine the maturity of the stands.

The method proposed in (Eysn et al., 2012) aims at generating a forest mask (*forested area* label only) using low density airborne lidar. A Canopy Height Model (CHM) with a spatial resolution of 1 m is derived. The positions and heights of single trees are determined from the CHM using a local maximum filter, based on a moving window approach. Only detected positions with a CHM height superior to 3 m are considered. The crown radii are estimated using an empirical function. The three neighboring trees are connected using a Delaunay triangulation applied to the previously-detected tree position. The crown cover is then calculated using the crown areas of three neighboring trees and the area of their convex hull for each tree triple. The forest mask is derived from the canopy cover values. While this is not a genuine stand delineation method, this approach could be easily extended to a multi-class problem and enlightens the necessity of individual tree extraction even with limited point densities as a basis for the stand-level analysis.

A forest stand delineation also based on airborne lidar data is proposed in (Wu et al., 2014). Three features are first directly extracted from the point cloud (related to tree height, density and shape). A coarse forest stand delineation is then performed on the feature image using the unsupervised Mean-Shift algorithm, in order to obtain under-segmented raw forest stands. A forest mask is then applied to the segmented image in order to distinguish forest and non-forest raw stands. It may create some small isolated areas, iteratively merged to their most similar neighbor until their size is larger than a user-defined threshold in order to generate larger and coarse forest stands. They are then refined into finer level using a seeded region growing based on superpixels. The idea is to select several different superpixels in a raw forest stand and merge them. This method provides a coarse-to-fine segmentation with relatively large stands. The process was only applied on a small area of a forest in Finland, thus, general conclusions can not be drawn.

2.1.3 Stand segmentation using VHR optical images and lidar

The analysis of the lidar and multispectral data is performed at three levels in (Tiede et al., 2004), following a given hierarchical nomenclature of classes standard for forested environments. The first level represents small objects (single tree scale, individual trees or small groups of trees) that can be differentiated by spectral and structural characteristics using here a rule-based classification. The second level corresponds to the stand level. It is built using the same classification process which summarizes forest development phases by referencing to small scale sub-objects at level 1. The third level is generated by merging objects of the same classified forest-development stage into

larger spatial units. The multi-scale analysis offers the advantage of alleviating the standard issue of individual tree crown detection and proposing development stage labels. Nevertheless, the pipeline is highly heuristic, under-exploits lidar data and significant confusions between classes are reported.

The automatic segmentation process of forests in (Diedershagen et al., 2004) is also supplied with Lidar and VHR multispectral images. The idea is to divide the forests into higher and lower strata with lidar. An unsupervised classification (with an algorithm similar to the *ISODATA*) process is applied to the two images (optical and rasterized lidar). The final stand delineation is achieved by segmenting the classification results with pre-defined thresholds. The segmentation results are improved using morphological operators such as opening and closing, which fill the gaps and holes at a specified extent. This method is efficient if the canopy structure is homogeneous and requires a strong knowledge on the area of interest. Since it is based on height information only, it cannot differentiate two stands of similar height but different species.

In (Leppänen et al., 2008) a stand segmentation technique for a forest composed of *Scots Pine*, *Norway Spruce* and *Hardwood* is defined. A hierarchical segmentation on the Crown Height Model followed by a restricted iterative region growing approach is performed on images composed of rasterized lidar data and Colored Infra-Red images. The process was only applied on a limited area of Finland (~ 70 ha) and prevents from drawing strong conclusions. However, the quantitative analysis carried out by the authors shows that lidar data can help to define statistically meaningful stands (here the criterion was the timber volume) and that multispectral images are inevitable inputs for tree species discrimination.

2.1.4 Challenges of stand segmentation

Table 2.1 summarizes the presented methods of forest stands segmentation. Firstly, it appears that the fusion of the two remote sensing modalities (optical images and lidar) improve the results for the problematic of forest stand delineation. However, the stands are not defined the same way in the different proposed methods, preventing from drawing general conclusion.

Regarding the existing state of the art on the forest stand segmentation, it appears that such task remains very complex to implement especially in an automatic way. Indeed, a simple segmentation of VHR optical image or lidar point cloud is not sufficient since it does not allow to retrieve consistent stands (in terms of species). However, segmentation algorithms are relevant for the extraction of small objects (ideally trees, or similar to trees). A classification is mandatory in order to obtain the tree species (i.e. semantic information). However, it is very difficult to discriminate species, since some have a very close looking (e.g. deciduous oak and beech), and the intra-class variability might be important (depending on age, maturity and other external features such as shape). Other issues related to the input data such as shadows in VHR optical images can also be reported. Eventually, the desired stands are not totally pure, a certain level of generalization is desired in order to have a consistent mapping at large scale. Thus, a regularization process can be employed for such purpose. It also appears that the type of data employed has an impact on the results.

- The VHR optical images permits to obtain information about the tree species, furthermore, textural features are very relevant (Franklin et al., 2000) if no hyperspectral data are available.

Reference	Data processed	Country	Segmentation criteria
Leckie et al. (2003)	Hyperspectral images	Canada	Tree species (7)
Hernando et al. (2012)	Multispectral images	Spain	Vegetation type
Mora et al. (2010)	Panchromatic images	Canada	Height
Koch et al. (2009)	Lidar	Germany	Forest types (15)
Sullivan et al. (2009)	Lidar	USA	Tree maturity
Eysn et al. (2012)	Lidar	Austria	Forest mask
Wu et al. (2014)	Lidar	Finland	Tree size, tree density
Tiede et al. (2004)	Lidar and multispectral images	Germany	Development phase
Diedershausen et al. (2004)	Lidar and multispectral images	Germany	Canopy structure, height
Leppänen et al. (2008)	Lidar and multispectral images	Finland	Tree species (3)

TABLE 2.1: Existing methods for forest stand segmentation, see text for more details.

- The lidar data provides information about the vertical structure of the forest that can also be useful for the discrimination of tree species (Brandtberg, 2007; Hovi et al., 2016; Li et al., 2013b). It also brings information about the height that allows to separate forest stands of different ages. Most of the time, lidar is deeply under-exploited since it is used only as a simple DSM/CHM.

The segmentation of forest stands must be envisaged as a region-based segmentation problem. Indeed, contour-based methods would be very difficult to employ since forested areas exhibit high variability and finding relevant borders is almost impossible especially in environment where no prior can be envisaged. Thus, for an optimal segmentation of forest stands, the strategies that are employed in this work are the following:

- **Extraction of small objects (similar to trees)**, in order to derive features at the object level, since it is very relevant for subsequent classification.
- **Extraction of multiple features from the two data sources.**
- **Object-based classification**, since it produces better results than a simple pixel-based classification.
- **Regularization of the classification** that leads to homogeneous forest stands with smooth borders.

Therefore, the fusion between VHR optical images and lidar is performed at the four levels since they allow to obtain relevant forest stands.

Main processing families adopted in this work are now presented and discussed with respect to our field of research.

2.2 Segmentation

The direct segmentation of optical image and/or lidar point clouds is not sufficient in order to retrieve forest stands. Indeed, such segmentation methods can not take into account the information needed to define the stand. However, with adapted parameters, segmentations algorithms might be useful to obtain objects that have an adapted size and shape for the desired study (Dechesne et al., 2017). They can be divided in two categories:

- The "traditional" segmentation methods; in these methods, a specific attention must be paid to the choice of the parameters in order to obtain relevant results. Such segmentation can be applied on an image or a point cloud. Specific methods have also been developed for the segmentation of lidar point cloud (Nguyen et al., 2013).
- The superpixels segmentation methods: they natively produce an over-segmentation of the image. The parameters control the size and the shape of the resulting segments (Stutz et al., 2017).

2.2.1 "Traditional" segmentation methods

The segmentation of an image can be performed using a large variety of techniques (Wilson et al., 1988; Nitzberg et al., 1993; Pal et al., 1993; Zhang, 2006).

The easiest way to segment an image is the thresholding of a gray level histogram of the image (Taxt et al., 1989). When the image is noisy or the background is uneven and illumination is poor, such thresholding is not sufficient. Thus, adaptive thresholding methods have been developed (Yanowitz et al., 1989).

The watershed transformation (Vincent et al., 1991) is also a simple segmentation method that considers the gradient magnitude of an image as a topographic surface. Pixels having the highest gradient magnitude intensities correspond to watershed lines, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common local intensity minimum. Pixels draining to a common minimum form a catch basin, which represents a segment.

The segmentation can also be considered as an unsupervised classification problem. Algorithms considering such classification problems adopt iterative process. The most popular algorithm is the *k-means* algorithm, or the *ISODATA* which is a variant of the *k-means*. Segmentation methods using the spatial interaction models like Markov Random Field (MRF) (Hansen et al., 1982) or Gibbs Random Field (GRF) (Derin et al., 1987) have also been proposed. Neural networks are also interesting for image segmentation (Ghosh et al., 1991) as they take into account the contextual information.

Conversely, the segmentation of an image can also be obtained from the detection of the edges of the image (Peli et al., 1982). The idea is to extract points of significant changes in depth values. Edges are local features and are determined based on local information and thus non suitable in our case.

Eventually, hierarchical or multi-scale segmentation algorithms can be employed. They analyze the image at several different scales. Their output is not a single partition, but a hierarchy of regions or a data structure that captures different partitions for different scales of analysis (Batz et al., 2004; Guigues et al., 2006; Trias-Sanz, 2006). These methods allow to control the complexity of the segmentation, which was not the case for the above-mentioned methods. The algorithm of (Guigues et al., 2006) is a bottom-up approach that starts with an initial over-segmentation (e.g. segmenting almost each pixel on a different own region) and uses this level as an initialization for the construction of subsequent significant levels. The segmentation process is guided by an energy E of the form:

$$E = D + \mu C \quad (2.1)$$

where, D is a fit-to-data measure (how well the segmentation fits to the original image, better fits give lower values of D); C is a measure of segmentation complexity (less complex solutions give lower values of C); and μ is a dimensional parameter, the scale parameter. The parameter μ balances between a perfect fit to the original data ($\mu = 0$), consisting of one segmentation region for each pixel in the original image, and the simplest segmentation, consisting of a single region containing the whole image (Guigues et al., 2006) (see Figure 2.2). The segmentation level can be adjusted gradually from the finest to the coarsest depending of the image complexity. The choice of a value of μ define a specific energy, leading to a unique segmentation.

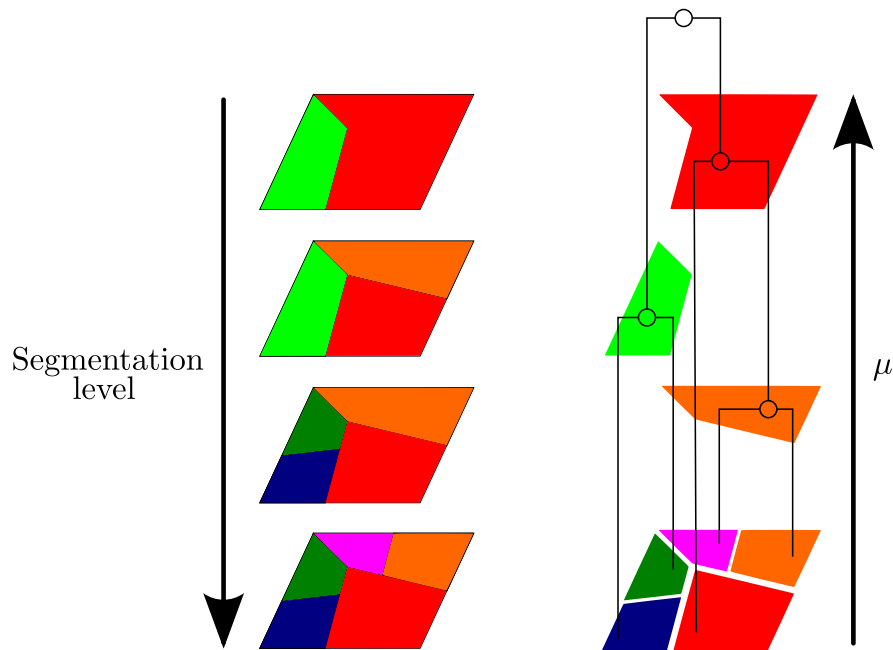


FIGURE 2.2: Graphical depiction of concepts related to hierarchical segmentation. The diagram on the left shows partitions of an image at four different scales μ . The partition at the top has the highest μ and is therefore the coarsest, the partition at the bottom is the finest.

Top-down approaches can also be employed for image segmentation. In (Landrieu et al., 2016) working-set/greedy algorithms to efficiently solve problems penalized respectively by the total variation on a general weighted graph are proposed. The algorithms exploit this structure by recursively splitting the level-sets of a piecewise-constant candidate solution using graph cuts.

Segmentation algorithms based on optical data have been also developed especially for forest analysis such as the approach proposed in (Tochon et al., 2015). It proposes a method for hyperspectral image segmentation, based on the binary partition tree algorithm, applied to tropical rainforests. Superpixel generation methods are applied to decrease spatial dimensionality and provide an initial segmentation map. Principal component analysis is performed to reduce the spectral dimensionality. A non-parametric region model based on histograms, combined with the diffusion distance to merge regions, is used to build the binary partition tree (Salembier et al., 2000). An adapted pruning strategy based on the size discontinuity of the merging regions is proposed. The resulting segmentation is coherent with the trees delineated manually, however, such method has been proposed for tropical rainforest and might not be adapted for temperate forests.

2.2.2 Superpixels methods

Dozen of superpixels algorithms have been developed (Achanta et al., 2012). They group pixels into perceptually meaningful atomic regions. Many traditional segmentation algorithms have been employed with more or less success to generate superpixels (Shi et al., 2000; Felzenszwalb et al., 2004; Comaniciu et al., 2002; Vedaldi et al., 2008; Vincent et al., 1991). These algorithms produce satisfactory results, however, they may be relatively slow and the number, size and shape of the

superpixels might not be specified, leading to a potential tedious parameter tuning step.

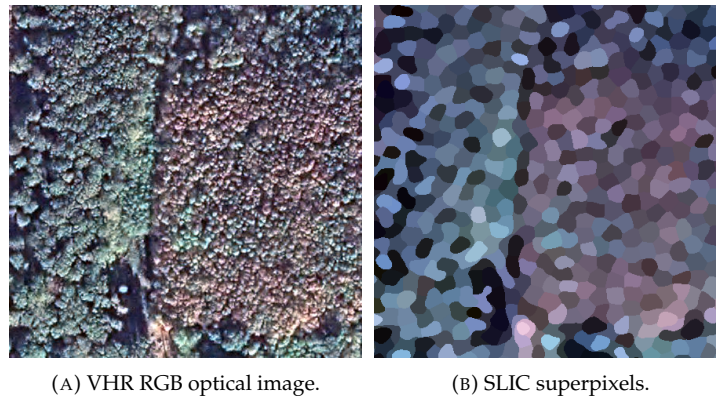


FIGURE 2.3: Comparison of the pixel (2.3a) and the superpixel (2.3b) approaches (125 m×125 m).

Superpixels algorithms have then been developed. One can control the number of superpixels, their size and their shape. Moore et al. (2008) create superpixels based on a grid. Optimal path are found using graph cut methods. Veksler et al. (2010) propose a generation of superpixels based on a global optimization. They are obtained by stitching together overlapping image patches such that each pixel belongs to only one of the overlapping regions. Levinshtein et al. (2009) generate superpixels by a dilatation of a set of seed locations using level-set geometric flow. Resulting superpixels are constrained to have uniform size, compactness, and boundary adherence. Finally, Achanta et al. (2012) propose a generation of superpixels based on the k-means algorithms. A weighted distance that combines color and spatial proximity is introduced in order to control the size and the compactness of the superpixels.

2.2.3 Segmentation of point cloud

Segmentation methods dedicated to 3D point cloud have been proposed (Nguyen et al., 2013). The aim is mainly to extract meaningful objects. Such extraction has two principal objectives:

- Objects are detected so as to ease or strengthen subsequent classification task. A precise extraction is not mandatory since the labels would be refined after.
- Objects are precisely delineated in order to derive features from these objects (e.g. surface, volume, etc.). A high spatial accuracy is therefore expected.

Several methods presented in previous sections can also be applied to 3D lidar point cloud.

In forested areas, the most reliable objects to extract are trees. The tree detection and extraction has been widely investigated (Wang et al., 2016; Kaartinen et al., 2012). The tree extraction from lidar point cloud can be envisaged in two ways:

- Rasterize the point cloud and use image-based segmentation techniques to obtain trees.
- Direct segmentation of the 3D point cloud.

A lot of methods have been developed for single tree delineation (Dalponte et al., 2014; Véga et al., 2014; Kandare et al., 2014; Reitberger et al., 2009). They all have their advantages and drawbacks, most of the time it is hard to assess the quality of the segmentation. None of them exhibit the ability to handle different kinds of forest.

2.3 Classification

A classification is a process that aims to categorize observations. The idea is to assign an observation to one or more classes. The classification can be unsupervised, in such cases the classes (i.e., the targeted labels) need to be discovered and the observation assigned. Such classification is similar to segmentation (see section 2.2) and is no further investigated here. The classification can be supervised, the target classes are known and observations with labels (employed for training and validation) are available. In our case, labels and training sets are given with the forest LC DB of interest.

2.3.1 Supervised classification: common algorithms

A great number of supervised classification algorithms have been developed and used for remote sensing issues (Landgrebe, 2005; Lu et al., 2007; Mather et al., 2016). There are two kinds of algorithms: the generative, often parametric, and the discriminative, often non-parametric.

The parametric methods assume that each class follows a specific distribution (mainly Gaussian). The parameters of the distribution are estimated using the learning set. This is the case for the maximum likelihood (Strahler, 1980) or maximum a posteriori (Fauvel et al., 2015).

The non parametric methods do not make any assumption on the classes distribution. In this category of algorithms, very popular ones are the Support Vector Machines (SVM) (Boser et al., 1992; Scholkopf et al., 2001) and the Random Forest (RF) (Breiman, 2001). The deep based-methods are also efficient algorithms (Hepner et al., 1990; Atkinson et al., 1997). However, despite their great performance in terms of accuracy, they have several drawbacks: firstly, the training process is time consuming and good GPU cards or specific architectures are required in order to reach decent training times (Dean et al., 2012; Moritz et al., 2015). Secondly, it requires an important amount of training data in order to correctly optimize the large number of parameters (e.g., hundred of millions). Simpler methods exist, such as the k-nearest neighbors (Indyk et al., 1998) or the decision trees (Breiman et al., 1984) but they produces quite low accuracy results. The non parametric methods are more efficient for the discrimination of complex classes (Paola et al., 1995; Foody, 2002), and are now considered as a basis for land cover classification (Camps-Valls et al., 2009).

We chose to use the RF, because of their widespread use, also offer the possibility of obtaining the probability of belonging of a pixel to a class. These posterior probabilities can be then integrated into a smoothing process. They also report good results, similar to SVM (see Chapter 4). The RF are described in section 2.3.2.

2.3.2 Random Forest

The RF have been introduced by (Breiman, 2001) and are defined by the aggregation of weak predictors (decision trees). Here, we refer to the RF with random inputs proposed in (Breiman, 2001).

The idea is to create an ensemble of sample sets $\mathcal{S}_n^{\Theta_1}, \dots, \mathcal{S}_n^{\Theta_k}$ randomly selected from an initial training set. A Classification and Regression Tree (CART) (Breiman et al., 1984) is built on each sample set $\mathcal{S}_n^{\Theta_i}$. Each tree is built using a random pool of m features among the M available features. The final classification is obtained by majority vote; each tree votes for a class and the class reaching the most votes wins (see Figure 2.4). This algorithm has two parameters: the number of trees k and the number of features m used to build a tree. The first parameter is arbitrary fixed to a high value. The second is generally fixed to the square root of the total number of feature (Gislason et al., 2006). Other parameters can be defined such as the maximal depth of the trees (or the purity of the leaves). In this thesis, the parameters of the Random Forest have not been fine tuned, since a standard tuning allows to obtain very good results. Other ameliorations to the Random Forest have been proposed such as the Rotation Forest (Rodriguez et al., 2006), the Random Ferns (Bosch et al., 2007) or the Extremely Randomized Trees (Geurts et al., 2006).

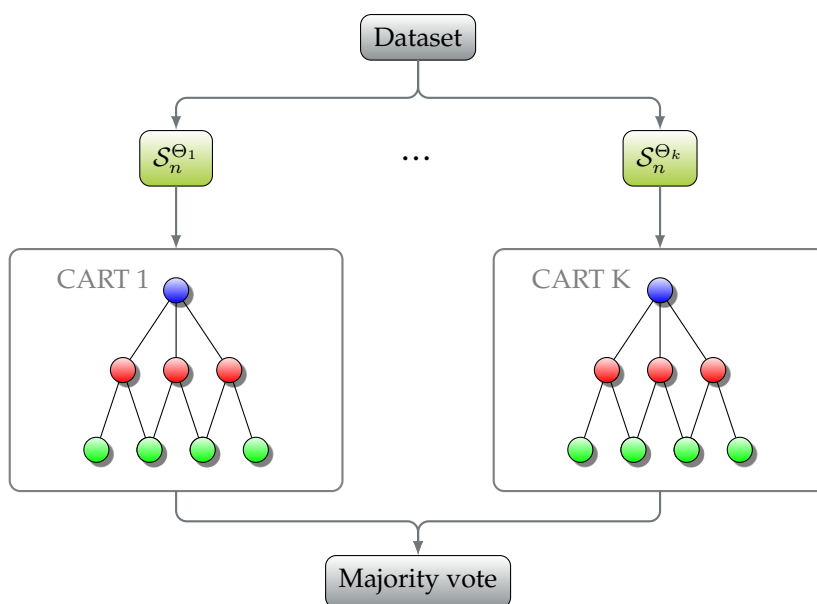


FIGURE 2.4: General flowchart of the Random Forest.

RF have shown classification performances comparable or better than traditional Boosting methods (Breiman, 2001) or SVM (Pal, 2005). They are also able to handle large datasets with high number of features. Furthermore, a measure of feature importance has been introduced in (Breiman, 2001). It allows to qualify the relevance of the features in the classification process (Strobl et al., 2007). Several other feature importance metrics have been proposed.

The importance of a feature $X_j, j \in \{1, \dots, q\}$ (with q the number of feature) is defined as follow. Let $\mathcal{S}_n^{\Theta_i}$ be a set of sample and OOB_i all the observations that do not belong to $\mathcal{S}_n^{\Theta_i}$. err_{OOB_i} , the error on OOB_i using $\mathcal{S}_n^{\Theta_i}$, is then computed. A random permutation on the value of the j^{th} feature of OOB_i is performed in order to obtain $\widetilde{err}_{OOB_i}^j$. $err_{OOB_i}^j$ is then computed. The importance of the feature j , $FI(\mathbf{X}_j)$ is the mean of the difference of the errors (see Equation 2.2).

$$FI(\mathbf{X}_j) = \frac{1}{k} \sum_{i=1}^k (err_{OOB_i}^j - err_{OOB_i}) \quad (2.2)$$

where k is the number of CART.

2.4 Dimension reduction and feature selection

Discriminative features are basis for classification without assuming which ones are the more salient, the standard strategy consists in computing a large set of features, especially in the multi-modal case (Tokarczyk et al., 2015). The feature selection methods try to overcome the curse of dimensionality (Bellman, 2015; Hughes, 1968). Indeed, the increasing number of features available tends to decrease the accuracy of the classifiers. Furthermore, the computation times increase with the number of features. Thus, reducing the feature dimension is beneficial for the classification task. Furthermore, in case of multi-modal features, feature selection allows to assess the contribution of the remote sensing modalities.

Two kinds of approaches exist: first the ones based on the extraction of new features summarizing the information by the transformation of the data, generally using a projection to an other space of lower dimensionality. Secondly, *feature selection* approaches that aim at identifying for an optimal subset of the features.

2.4.1 Dimension reduction: feature extraction

The most popular dimension reduction method is the Principal Component Analysis (PCA). It is an unsupervised method that aim to maximize the variance between data (Jolliffe, 2011). However, it has been demonstrated that PCA is not optimal for the purpose of classification (Cheriyadat et al., 2003). Other methods have been developed related to the PCA: the Independent Component Analysis (ICA) (Jutten et al., 1991) maximizes the statistical independence between data, and the Maximum Autocorrelation Factor (MAF) (Larsen, 2002) maximizes the spatial auto-correlation. When training samples are available, supervised methods exist, such as the linear discriminant analysis (LDA) that tries to maximize both the intra-class homogeneity and the inter-class variance (Fisher, 1936; Lebart et al., 1997).

2.4.2 Feature selection

Feature selection (FS) aims z defining an optimal subset of features without modifying them. Automatic methods have been proposed in order to obtain such subset. One can explore the subsets of features and need to define a criteria to evaluate the subsets. Furthermore, the selection can be supervised or unsupervised. The first aims at discriminating the better the classes while the second is looking for an optimal subset that contains the most informative and less redundant features. Many exploration methods for feature selection have been proposed in the literature. The naive exhaustive exploration of all the subsets can be envisaged only when the number of features is not important.

Existing methods

The feature selection methods can be separated into 3 categories: filters, wrapper and embedded.

• Filters

The filter methods are independent from any classifier. Within the filter methods, one can distinguish the supervised and unsupervised case depending on whether the notion of classes is taken into account or not. When supervised, they consider the features according to their capacity to bring together elements of the same class and separate the different elements (John, 1997). Separability

measures (e.g., Fisher (Fisher, 1936), Bhattacharya or Jeffries-Matusia) allow to determine whether a feature or a subset of feature is well adapted to discriminate the classes (Bruzzone et al., 2000; Herold et al., 2003; De Backer et al., 2005; Serpico et al., 2007). Among filters, ranking methods compute an individual importance score for each feature, classify the features according to this score and keep only the best. Such scores can be computed using training samples or not. Such methods are independent from a classifier and are used as preliminary step to classification. Statistical measures derived from information theory such as the divergence, the entropy or the mutual information have been proposed in the unsupervised case (Martínez-UsÓ et al., 2007; Le Moan et al., 2011) or supervised case (Battiti, 1994; Guo et al., 2008; Estévez et al., 2009; Sotoca et al., 2010; Cang et al., 2012).

To summarize, criteria and methods for filter selection methods are numerous and cover different approaches. The ranking filter methods, which sort features according to an individual importance score and retain only the n best remain limited since they do not take into account the dependencies between the selected features. Approaches that directly associate relevance scores with feature sets are more interesting. A distinction is made between supervised and unsupervised approaches. The unsupervised criteria are interesting, but present a risk of selecting attributes that would not all also be useful for classification. An optimization method must be then employed in order to select the best subset.

• Wrapper

The wrapper methods weight the feature subsets according to their pertinence for the prediction (Kohavi et al., 1997). This weighting is related to the performance of a classifier. Estévez et al. (2009), Li et al. (2011), Yang et al. (2007), and Zhuo et al. (2008) propose approaches with SVM classifiers. Zhang et al. (2007) and Fauvel et al. (2015) use maximum likelihood classifiers. The RF is also employed in (Díaz-Uriarte et al., 2006). Data are separated into two subsets. The first is used for the training, while the second for the evaluation. The use of a classifier is a big advantage as it fits more to the envisaged problem and produces better results with less features than the filters methods. However, the use of a classifier significantly increases the computation times. Furthermore, worse results could be obtained when using a feature subset with an other classifier.

• Embedded

Eventually, the embedded methods also involve a classifier but select the features during the training process (Tang et al., 2014) or using intermediate results from the training process. They have two advantages: since they take in consideration the data as training, they have the same advantages as wrappers (selecting features related to the classification problem). Furthermore, they are faster than the wrapper methods since they do not test feature sets on a test dataset. Many methods have been proposed. The RF allow to assess the feature importance (Breiman, 2001) and is also natively embedded since the irrelevant features are discarded while building the tree and will not be used in the classification process. Other methods are based on the SVM classifiers, the SVM-RFE (Recursive Feature Elimination) (Tuia et al., 2009) recursively removes the less pertinent features according to a weight estimated from a SVM model.

Selection optimization

Some methods involve a specific optimization method but when considering a generic feature relevance score, the set of possible solutions is generally too large to be visited entirely. Thus, using

heuristic rules allows to find a solution close enough to the optimal solution while visiting only a reasonable number of configurations. These optimization methods can generally be distinguished in sequential or incremental methods and stochastic methods.

• Sequential approaches

The first idea is to add features step by step (forward approaches), also called Sequential Forward Selection (SFS) (Marill et al., 1963). It could also be methods that start from the entire feature set and remove feature step by step (backward approaches), also called Sequential Backward Selection (SBS) (Whitney, 1971). A generalization of these methods have been proposed in (Kittler, 1978). Finally, the forward and backward methods could be combined in order to improve the process. The Sequential Floating Forward Selection (SFFS) and the Sequential Floating Backward Selection (SFBS) (Pudil et al., 1994) propose such improvement.

• Stochastic approaches

Stochastic algorithms will involve randomness in their exploration of the solution space. The random initialization and search for a solution can therefore propose different solutions of equivalent quality from a single dataset. The generation of the subset can be totally random (Liu et al., 1997). Genetic algorithms is a possible solution. They propose to weight the subsets according to their importance (Goldberg, 1989). They allow a faster convergence to a more stable solution. The Particle Swarm Optimization (PSO) algorithm (Yang et al., 2007) is faster and select relevant features. For finding an approximate optimal subset of features, simulated annealing (De Backer et al., 2005; Chang et al., 2011) is also possible.

2.5 Smoothing methods

Pixel-wise classification is not sufficient for both accurate and smooth land-cover mapping with VHR remote sensing data. This is particularly true in forested areas: the large intra-class and low inter-class variabilities of classes result in noisy label maps at pixel or tree levels. This is why various regularization solutions can be adopted from the literature (from simple smoothing to probabilistic graphical models).

According to (Schindler, 2012), both local and global methods can provide a regularization framework, with their own advantages and drawbacks.

2.5.1 Local methods

In local methods, the neighborhood of each element is analyzed by a filtering technique. The labels of the neighboring pixels (or the posterior class probabilities) are combined so as to derive a new label for the central pixel. Majority voting, Gaussian and bilateral filtering (Perona et al., 1990) can be employed if it is not targeted to smooth class edges. The majority vote can also be used especially when a segmentation is available: the majority class is assigned to the segment. The vote can be weighted by class probabilities of the different pixels.

The probabilistic relaxation is an other local smoothing method that aims at homogenizing probabilities of a pixel according to its neighboring pixels. The relaxation is an iterative algorithm in

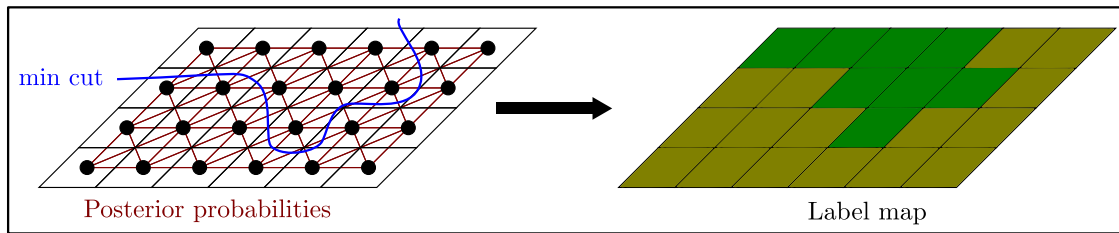
which the class probability at each pixel is updated at each iteration in order to have it closer to the probabilities of its neighbors (Gong et al., 1989; Smeeckaert et al., 2013). It reports good accuracies with decent computing time and offers an alternative to edge aware/gradient-based techniques that may not be adapted in semantically unstructured environments.

2.5.2 Global methods

Global regularization methods consider the whole image by connecting each pixels to its neighbors. They traditionally adopt the Markov Random Fields (MRF, see Figure 2.5), the labels at different locations are not considered to be independent and the global solution can be retrieved with the simple knowledge of the close neighborhood for each pixel. The optimal configuration of labels is retrieved when finding the Maximum A Posteriori over the entire field (Moser et al., 2013). The problem is therefore considered as the minimization procedure of an energy E over the whole image I . Despite a simple neighborhood encoding (pairwise relations are often preferred), the optimization procedure propagates over large distances. Depending on the formulation of the energy, the global minimum may be reachable. However, a large range of optimization techniques allow to reach local minima close to the real solution, in particular for random fields with pairwise terms (Kolmogorov et al., 2004). For genuine structured predictions, in the family of graphical probabilistic models, Conditional Random Fields (CRF, see Figure 2.5) have been massively adopted during the last decade. Interactions between neighboring objects, and subsequently the local context can be modeled and learned using an energy formulation. In particular, Discriminative Random Fields (DRF, (Kumar et al., 2006)) are CRF defined over 2D regular grids, and both unary/association and binary/interaction potentials are based on labeling procedure outputs. Many techniques extending this concept or focusing on the learning or inference steps have been proposed in the literature (Kohli et al., 2009; Ladický et al., 2012). A very recent trend even consists in jointly considering CRF and deep-learning techniques for the labeling task (Kirillov et al., 2015).

In standard land-cover classification tasks, global methods are known to provide significantly more accurate results (Schindler, 2012) since contextual knowledge is integrated. This is all the more true for VHR remote sensing data, especially in case of a large number of classes (e.g., 10, (Albert et al., 2016)), but presents two disadvantages. For large datasets, their learning and inference steps are computationally expensive. Furthermore, parameters should often be carefully chosen for optimal performance, and authors that managed to alleviate the latter problem still report a significant computation cost (Lucchi et al., 2011).

Markovian Random Field



Conditional Random Field

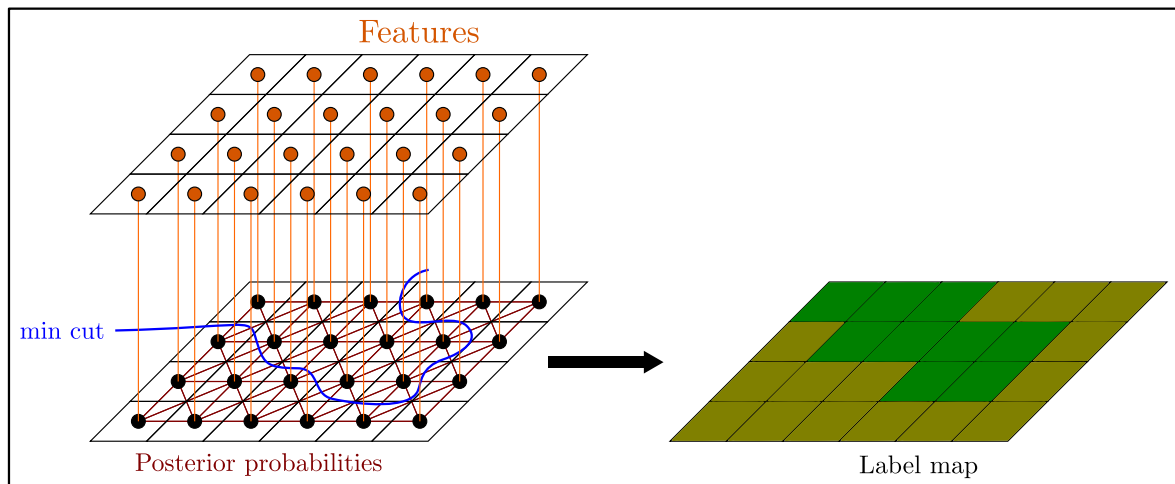


FIGURE 2.5: 8-connected MRF and CRF. The MRF only takes into account the posterior probabilities to compute the graph, while CRF also includes contextual information (e.g. the features).

2.6 Conclusion

In this chapter, several existing stand segmentation methods have been identified, leading to some important conclusions. Firstly, **it appears that the fusion of optical images and lidar has already been investigated and identified as useful for forest stand segmentation.** However, the stands are not defined the same way in the different proposed methods, preventing from drawing general conclusion for our problematic. Thus, proposed methods vary a lot depending on the definition of stands (species, age, height, etc.). Besides, several existing methods are very dedicated to human/operator interaction.

The segmentation of forest stands is often envisaged as a region-based segmentation problem, and specific strategies are employed in this work in order to retrieve forest stands. First, it appears that the extraction of features is relevant for the tree species discrimination. Such features can be derived from both data sources. From the literature, it came out that working at the object-level improves the tree species discrimination results. The small objects are then desired to have a size and shape similar to trees. Many algorithms can be envisaged for the extraction of such objects. A supervised classification is then mandatory in order to discriminate the tree species. However, since many features have been derived, it is also interesting to select a limited number of features to train

the classifier. Such selection is also interesting in order to validate the complementarity of the data sources. Eventually, the classification might be noisy and can be smoothed through regularization methods. Such smoothing allows to obtain homogeneous segments with regular borders that fit with the specifications of the forest stands desired.

All these problems are addressed in the following chapters.

Proposed framework

3.1	General flowchart	36
3.2	Feature extraction	36
3.2.1	Point-based lidar features.	38
3.2.2	Pixel-based lidar features.	39
3.2.3	Pixel-based multispectral images features.	40
3.2.4	Object-based feature map.	41
3.3	Over-segmentation	41
3.3.1	Segmentation of lidar data	43
3.3.2	Segmentation of optical images	45
3.4	Feature selection	46
3.5	Classification	47
3.5.1	Training set design	48
3.5.2	Training and prediction	49
3.6	Regularization	50
3.7	Conclusion and discussions	51

3.1 General flowchart

With respect to the methods mentioned in Chapter 2, it appears that there is no operational automatic forest stand segmentation method, where the target labels are the tree species, that can satisfactorily handle a large number of classes (>5). This problem has been barely investigated and only *ad hoc* methods have been developed. The proposed framework is fully automatic, modular and versatile for species-based forest stand segmentation. It strongly relies on the exploitation of an existing Forest LC DB (in order to update and improve it), taking into account potential errors that it may contain, subsequently providing an adapted model for the studied area. The proposed framework propose the following characteristics:

- No heavy and sensitive parameter tuning;
- Several steps can be substituted with almost equivalent solution, depending on the requirements: accuracy, speed, memory usage, etc.
- Depending on the input and the prior knowledge, one can select the shape and the level of details of the output maps.

The framework is composed of four main steps. Features are then computed at the pixel level for the optical images and at the point level for the lidar data. An over-segmentation is then performed in order to retrieve small object that will be employed for subsequent classification. The object extracted from the over-segmentation and the computed features allow to derive features at the object level. Such features are then used by the classification of the vegetation types (mainly tree species). Indeed, performing classification at the object level significantly improves the discrimination results compared to a pixel-based classification. Here, the training set is automatically derived from an existing forest LC database. Specific attention is paid to the extraction of the most relevant training pixels, which is highly challenging with partly outdated and generalized vector databases. Because of the high number of features, a feature selection is also carried out in order to have a more efficient classification and reduce the computational load and time, but also in order to assess the complementarity of the multi-source features (namely multispectral optical images/lidar point cloud). Finally, a regularization of the label map is performed in order to remove the noise and to retrieve homogeneous forest stands according to a given criteria (here tree species). Each step is presented in the next sections. A particular focus is made on the regularization methods in Chapter 5 and the discussion of the most relevant fusion scheme in Chapter 6. The flowchart of the framework is presented in Figure 3.1.

3.2 Feature extraction

The extraction of discriminative features is an important preliminary step in order to obtain an accurate classification. The features can be handcrafted. Such strategy has been extensively employed for remote sensing applications. The features could also be learned for a specific classification task using convolutional neural networks (Demuth et al., 2014). Here, the proposed features have been derived manually, since they are physically interpretable. Most are standards in the literature. The lidar features are derived at the 3D level and the spectral features at the 2D (image) level. Two strategies can be envisaged:

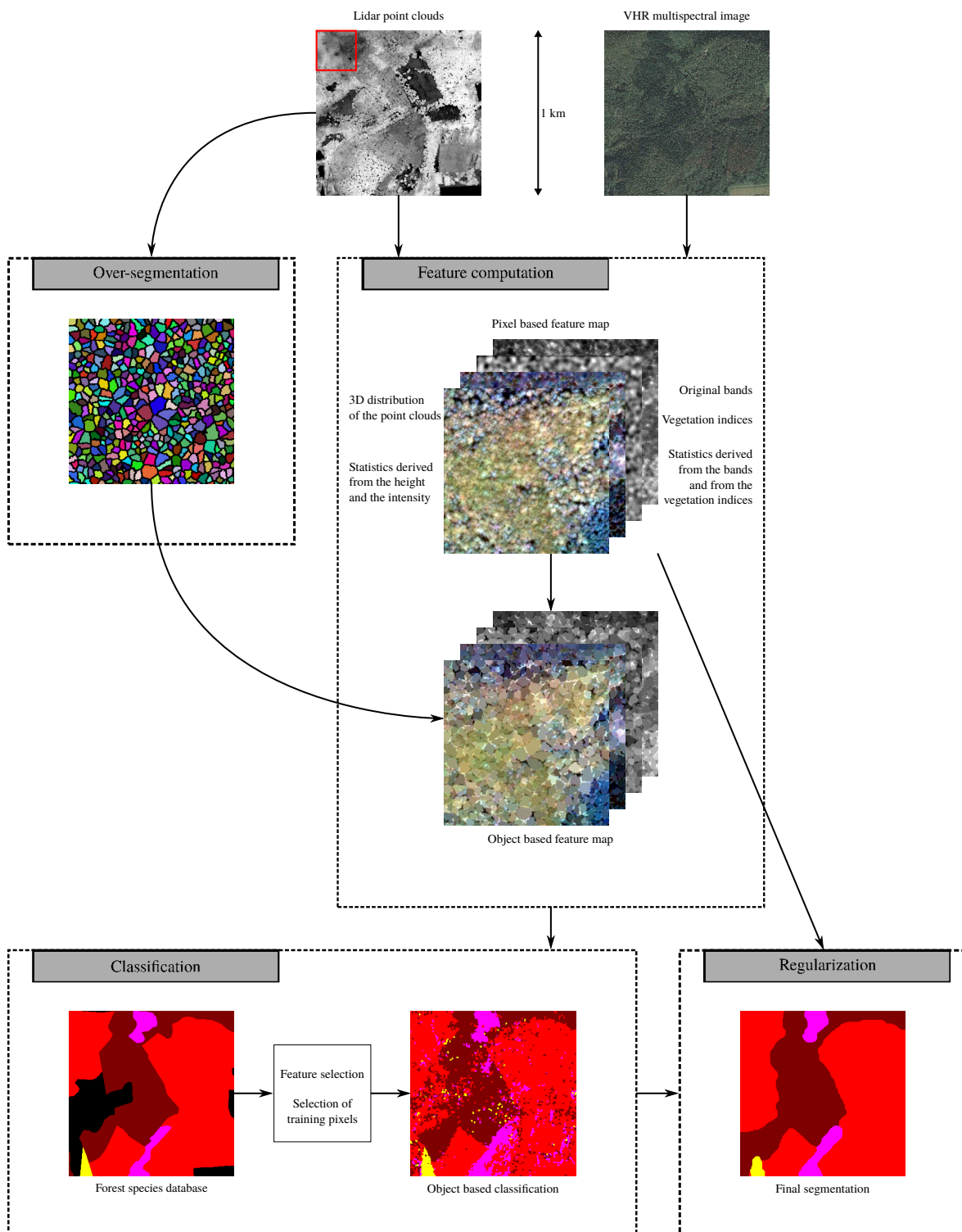


FIGURE 3.1: Flowchart of the proposed framework, see text for more details.

- The rasterization of the 3D lidar features at the same spatial resolution as the spectral features.
- The projection of the 2D spectral features in the 3D point cloud.

The projection of the 2D spectral features raises three main issues:

- In the 3D space, it is impossible to attribute a relevant spectral value to a point below the canopy (i.e. not visible in the optical image).
- A single pixel can be attributed to many 3D points, the information will be then duplicated. As an example for an image of 1 km² at a spatial resolution of 0.5 m there are 4 millions of pixels, while the number of 3D points is about 16 millions.
- The processing of 3D points is tedious (especially for classification and regularization) while pixels are more easy to handle.

On the opposite, the rasterization of the 3D lidar features has been widely employed, many efficient rasterization algorithms have been proposed. Thus this strategy was adopted.

3.2.1 Point-based lidar features.

24 features are extracted during this step; 2 related to vegetation density, 2 related to the 3D local distribution of the point cloud (planarity and scatter), and 20 statistical features.

Lidar-derived features require a consistent neighborhood for their computation (Demantké et al., 2011; Filin et al., 2005). For each lidar point, 3 cylindrical neighborhoods, aligned with the vertical axis, are used (1 m, 3 m and 5 m radii, infinite height). A cylinder appears to be the most relevant environment in forested areas so as to take into account the variance in altitude of the lidar points. Three radius values are considered so as to handle the various sizes of the trees, assuming a feature selection step will prune the initial set of attributes.

Density features.

Two vegetation density features, \mathcal{D}_1 and \mathcal{D}_2 , are computed: the first one based on the number of local height maxima within the neighborhoods, and the second one related to the number of non-ground points within the neighborhoods (ground points were previously determined by a filtering step). \mathcal{D}_1 and \mathcal{D}_2 are calculated as follows:

$$\mathcal{D}_1 = \sum_{r_1 \in \{1,3,5\}} \sum_{r_2 \in \{1,3,5\}} Nt_{r_1, r_2}, \quad (3.1)$$

$$\mathcal{D}_2 = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{Ns_r}{Ntot_r}, \quad (3.2)$$

where Nt_{r_1, r_2} is the number of local maxima retrieved from a r_1 maximum filter within the cylindrical neighborhood of radius r_2 . Ns_r is the number of points classified as ground points within the cylindrical neighborhood of radius r and $Ntot_r$ is the total number of points within the cylindrical neighborhood of radius r . \mathcal{D}_1 describes how trees are close to each other and gives information about the tree crown width. Such information is very discriminative for tree species classification (Korpela et al., 2010). \mathcal{D}_2 provides information on the penetration rate of the lidar beam. It has been proven to be relevant for tree species classification (Vauhkonen et al., 2013).

Shape features.

Additionally, the scatter \mathcal{S} and the planarity \mathcal{P} features are computed following (Weinmann et al., 2015):

$$\mathcal{S} = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{\lambda_{3,r}}{\lambda_{1,r}}, \quad (3.3)$$

$$\mathcal{P} = \frac{1}{3} \sum_{r \in \{1,3,5\}} 2 \times (\lambda_{2,r} - \lambda_{3,r}), \quad (3.4)$$

where $\lambda_{1,r} \geq \lambda_{2,r} \geq \lambda_{3,r}$ are the eigenvalues of the covariance matrix within the cylindrical neighborhood of radius r . They are retrieved with a standard Principal Component Analysis. They have been shown to be relevant for classification tasks (Weinmann et al., 2015).

Statistical features.

Statistical features, known to be relevant for vegetation type (mainly tree species) classification (Dalponte et al., 2014; Torabzadeh et al., 2015), are also derived. For each lidar point, the same 3 cylindrical neighborhoods are used. Two basic information from the lidar data, namely height and intensity (Kim et al., 2011), are used to derive statistical features. A statistical feature f_d , derived from an original feature f_o , (normalized height or intensity) is computed as follows:

$$f_d = \frac{1}{3} \sum_{r \in \{1,3,5\}} f_s(\mathbf{p}_{r,f_o}), \quad (3.5)$$

where f_s is a statistical function, and \mathbf{p}_{r,f_o} a vector containing the sorted values of the original feature f_o within the cylindrical neighborhoods of radius r . The statistical functions f_s employed are standard ones (minimum; maximum; mean; median; standard deviation; median absolute deviation from median (*medADmed*); mean absolute deviation from median (*meanADmed*); skewness; kurtosis; 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th and 95th percentiles).

All the statistical functions are used for the height. Only the mean is used for the intensity: indeed it is hard to know how well the sensor is calibrated and a suitable correction of intensity values within tree canopies has not yet been proposed.

3.2.2 Pixel-based lidar features.

The lidar features are rasterized at the resolution of the multispectral image using a pit-free method proposed in (Khosravipour et al., 2014). Indeed, the main problem for the rasterization of lidar features is that the point density is not homogeneous, thus applying a regular grid can lead to a pixel-based feature map with pits. Therefore, a special pit-free rasterization methods need to be employed. The idea is to retain points within a cylindrical neighborhood. The axis of the cylindrical neighborhood is the center of the pixels, the radius is chosen in order to have enough retained points (at least 10). The features values of the points are weighted according to their distance to the center of the pixel. The value of the pixel is the mean of the weighted feature values of the points. The rasterization method used here is interesting because it produces smooth images, compared to rough rasterization and will lead to better results for classification and regularization (Li et al., 2013a). This rasterization process at the feature level is valid since both datasets have approximately

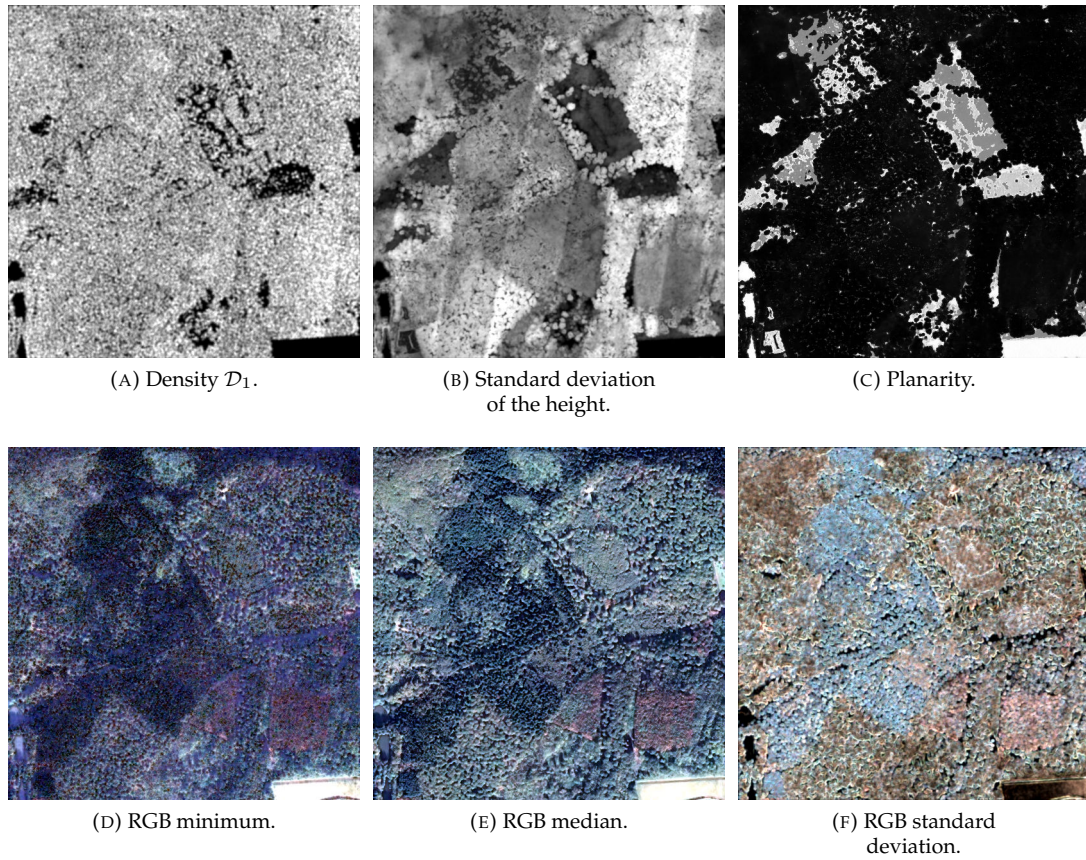


FIGURE 3.2: Some features at the pixel level (1 km²).

the same initial spatial resolution.

A nDSM is also computed using this method, at the same spatial resolution using an existing 1 m Digital Terrain Model computed from the initial point cloud (Ferraz et al., 2016a). The nDSM is very important as it allows to derive the height above the ground and is known as a very discriminative feature for classification (Mallet et al., 2011; Weinmann, 2016). Some lidar features are presented in Figure 3.2.

3.2.3 Pixel-based multispectral images features.

The original 4 spectral bands of the VHR airborne optical image are kept and considered as multispectral features. The Normalized Difference Vegetation Index (NDVI), (Tucker, 1979), the Difference Vegetation Index (DVI), (Bacour et al., 2006) and the Ratio Vegetation Index (RVI) (Jordan, 1969) are also computed as they are standard relevant vegetation indices (Anderson et al., 1992; Lee et al., 1997). Many other vegetation indices have been proposed (Bannari et al., 1995) that have shown relevance for vegetation classification tasks. Indeed, they mainly provide information about the chlorophyll contents and are therefore more relevant for vegetation discrimination more than the original bands alone (Zargar et al., 2011) (emphasizing specific spectral behaviors).

As for the point-based lidar features, statistical features are also derived from each band and each vegetation index according to Equation 3.5 (3 circular neighborhoods of 1 m, 3 m and 5 m radii). Other statistical functions are used (minimum; maximum; mean; median; standard deviation; mean absolute deviation from median (*meanADmed*); mean absolute deviation from mean (*meanADmean*); median absolute deviation from median (*medADmed*); median absolute deviation from mean (*medADmean*). Such features are related to texture features that are relevant for classification tasks (Haralick et al., 1973).

Finally, the pixel-based multispectral feature set is composed of 70 attributes. Some of these spectral features are presented in Figure 3.2.

The pixel-based features (lidar and spectral) that have been derived are summarized in Figure 3.3.

3.2.4 Object-based feature map.

The pixel-based multispectral and lidar maps are merged so as to obtain a pixel-based feature map. Then, an object-based feature map is created using the over-segmentation and the pixel-based feature map. The value v_t associated to an object t in the object-based feature map is computed as follows:

$$v_t = \frac{1}{N_t} \sum_{p \in t} v_p, \quad (3.6)$$

where N_t is the number of pixels in object t , and v_p is the value of the pixel p . If a pixel does not belong to an object (e.g. when the extracted objects are trees), it keeps the value of the pixel-based feature map. Here, only the mean value of the pixels within the object is envisaged but one can also consider other statistical values (minimum, maximum, percentiles etc.). It has not been considered here since it would drastically increase the total number of features and there is no warranty that such statistical values are relevant or are not redundant with the features already produced here.

Other morphological features could also be directly derived from the lidar points cloud at the object-level. For instance, an alpha-shape could be performed on the individual trees (Vauhkonen et al., 2010) and penetration features could be derived as it can help to classify vegetation type (mainly tree species) (Ko et al., 2013). However, low point densities (1-5 points/m²) compatible with large-scale lidar surveys are not sufficient in order to derive such discriminative features, they are therefore not considered here.

An illustration of the pixel-based and object-based feature map is presented in Figure 3.4.

3.3 Over-segmentation

Over-segmentation proposes a full coarse partition of the area of interest. "Objects" are detected so as to ease and strengthen subsequent classification task. An accurate object extraction is not mandatory since the labels are refined after. Both 3D and 2D mono-modal solutions are investigated, depending on the input data and the desired level of detail for the objects. Indeed, multi-modal segmentation has also been proposed (Tochon, 2015) and can be adopted for over-segmentation. However, they are not employed here since they mostly aim at producing very accurate segmentation results which are not needed in this framework. In this section, the over-segmentation aims at

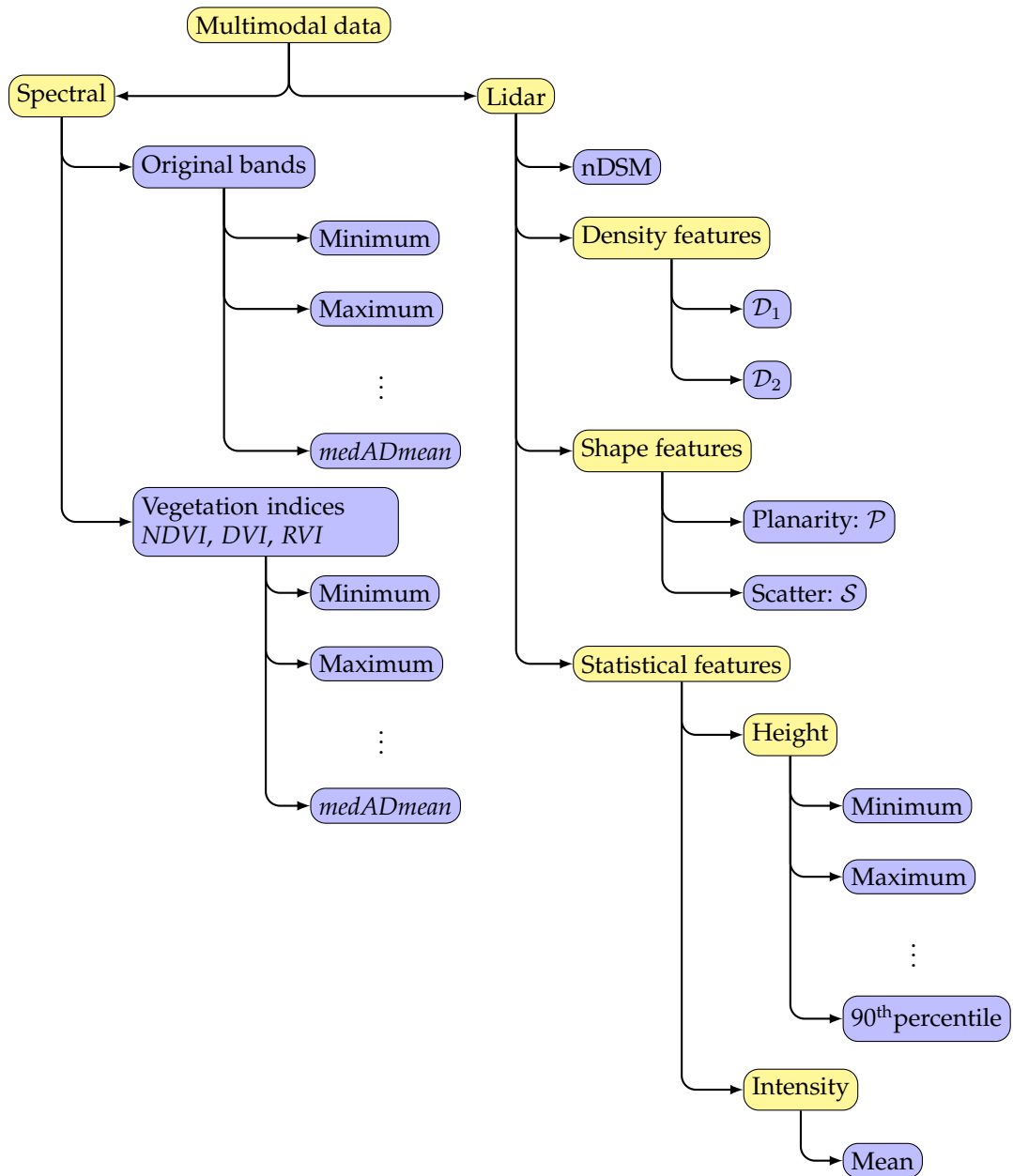


FIGURE 3.3: Features derived at the pixel level, yellow corresponds to the categories of the features, blue corresponds to the computed features.

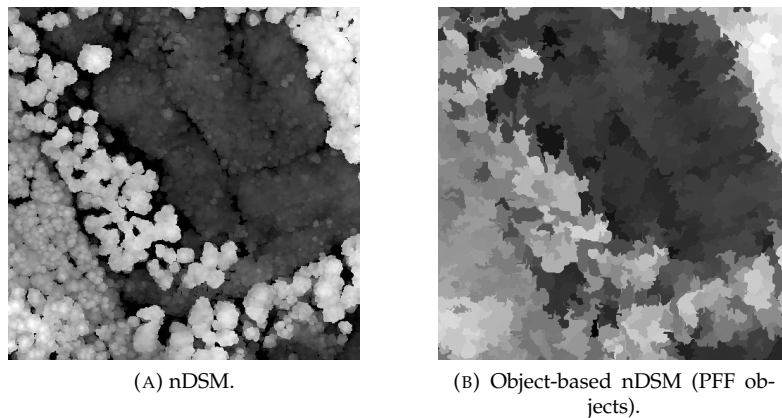


FIGURE 3.4: Illustration of the pixel-based and object-based features (250 m×250 m).

extracting objects at the tree scale (i.e. extract trees or objects of size that are similar to trees). Individual Tree Crown (ITC) delineation is a complex task and no universal automatic solution has been proposed so far. Thus different methods have been tested in order to extract consistent objects (i.e., object similar to trees). Five existing image-based segmentation methods have been tested either on VHR optical images or rasterized lidar features. In addition a coarse 3D tree extraction method has been developed.

3.3.1 Segmentation of lidar data

Two approaches can be envisaged: the direct segmentation in 3D of the point cloud or the segmentation of a given rasterized lidar feature using standard image-based segmentation algorithms.

The tree extraction directly from the 3D point cloud is a complex task that has been widely tackled and discussed (Dalponte et al., 2014; Véga et al., 2014; Morsdorf et al., 2003; Kandare et al., 2014; Wang et al., 2016). Variations exist according to the tree species, the number of vegetation strata, forest complexity, location and data specification. However, a precise tree extraction is not needed here, since the extracted objects are only needed to improve the classification task. A coarse and standard method is therefore adopted: the tree tops are first extracted from the lidar points cloud using a local maximum filter (Figure 3.5b). A point is considered as a tree top when it has the highest height value within a 5 meter radius neighborhood. Only the points above 3 meters are retained as it is a common threshold of the literature (Eysn et al., 2012), and appears to be highly discriminative in non-urban areas. Points belonging to a tree are obtained through a two step procedure.

1. If the height of a point within a 5 m radius is greater or equal than 80% the height of the closest tree top, it is aggregated to the tree top (Figure 3.5c).
2. If the distance in the (x, y) plane between an unlabeled point and the closest tree point is smaller than 3 m they are also aggregated (Figure 3.5d).

This delineation method allows to discard low vegetation, but buildings might be extracted and considered as trees. However, it is not a big issue since the purpose of this segmentation is not to precisely extract trees but only provide relevant object for the subsequent object-based classification.

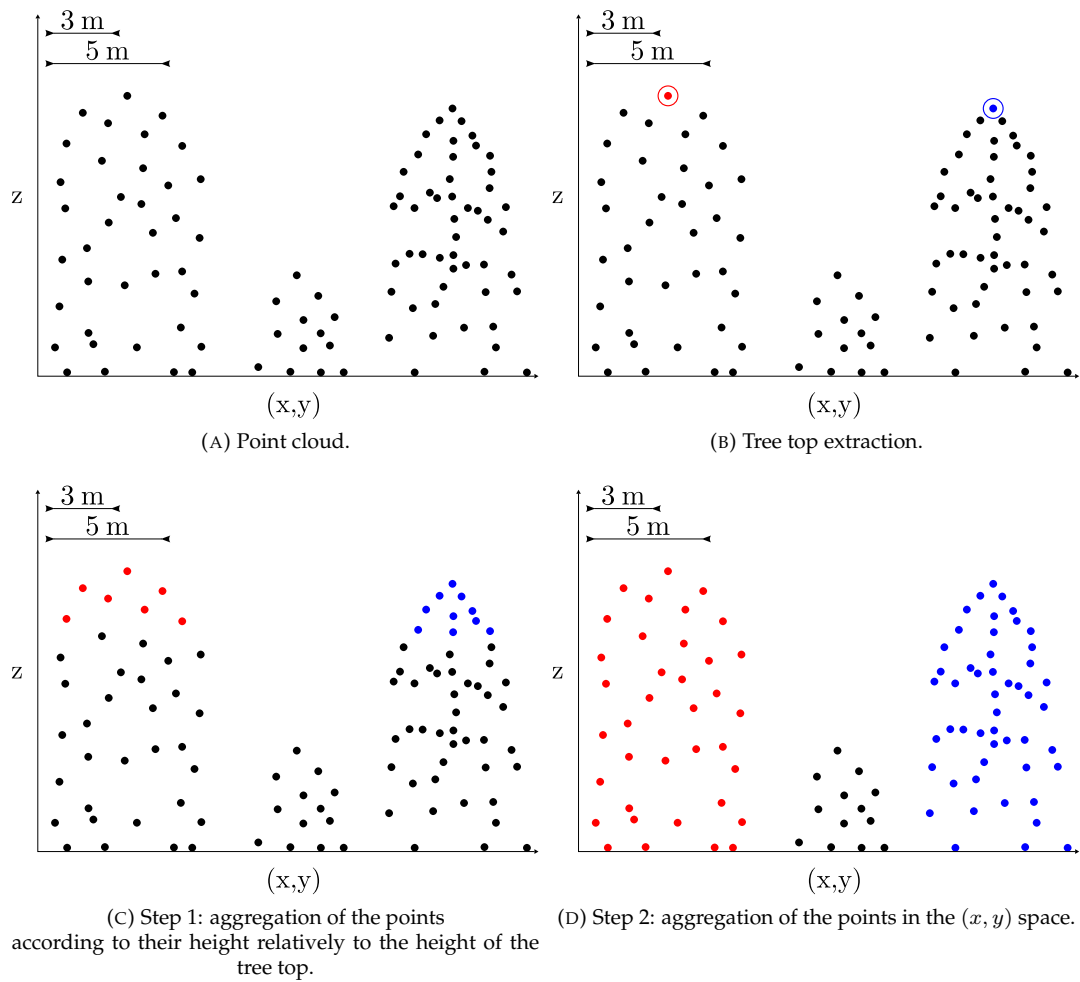


FIGURE 3.5: The proposed tree extraction procedure. Red and blue points are the points assigned to different trees, black points correspond to unlabeled points.

The segmentation of lidar data can be performed using image-based segmentation algorithm on a given rasterized lidar feature. They are mainly applied on the normalized Digital Surface Model (i.e. true height of topographic objects on the ground). Thus a method using a single feature must/can be employed. The watershed algorithm (Vincent et al., 1991) with specific parameters allows to obtain quickly a consistent over-segmentation of the image. A hierarchical segmentation (Guigues et al., 2006) is also relevant with the advantages that only one parameter that controls the segmentation level needs to be provided and different suitable segmentations can be obtained. These two algorithms are detailed below.

Watershed.

Watershed has been proposed after the natural observation of water raining onto a landscape topology and flowing with gravity to collect in low basins (Beucher et al., 1979). The size of those basins will grow with increasing amounts of precipitation until they spill into one another, causing small basins to merge together into larger basins. Regions (catchment basins) are formed by using local geometric structure to associate points in the image domain with local extrema in some

feature measurement such as curvature or gradient magnitude. This technique is less sensitive to user-defined thresholds than classic region-growing methods, and may be better suited for fusing different types of features from different data sets. The watersheds technique is also more flexible in that it does not produce a single image segmentation, but rather a hierarchy of segmentations from which a single region or set of regions can be extracted a-priori (Vincent et al., 1991).

Hierarchical segmentation.

This segmentation method (Guigues et al., 2006) introduces a multi-scale theory of piecewise image modeling, called the scale-sets theory, and which can be regarded as a region-oriented scale-space theory. A general formulation of the partitioning problem which involves minimizing a two-term-based energy, of the form $D + \mu C$, where D is a goodness-of-fit term and C is a regularization term. Such energies arise from basic principles of approximate modeling and relate them to operational rate/distortion problems involved in lossy compression problems. An important subset of these energies constitutes a class of multi-scale energies in that the minimal cut of a hierarchy gets coarser and coarser as parameter μ increases. This allows to define a procedure to find the complete scale-sets representation of this family of minimal cuts. Considering then the construction of the hierarchy from which the minimal cuts are extracted, ending up with an exact and parameter-free algorithm to build scale-sets image descriptions whose sections constitute a monotone sequence of upward global minima of a multi-scale energy, which is called the "scale climbing" algorithm. This algorithm can be viewed as a continuation method along the scale dimension or as a minimum pursuit along the operational rate/distortion curve. Furthermore, the solution verifies a linear scale invariance property which allows to completely postpone the tuning of the scale parameter to a subsequent stage.

3.3.2 Segmentation of optical images

Many algorithms have been developed for the over-segmentation of optical RGB (Red-Green-Blue) images. Superpixels methods (Achanta et al., 2012) are specific over-segmentation methods that put a special effort on the size and shape of the extracted objects. Pseudo-superpixels can be generated using segmentation algorithms (Shi et al., 2000; Felzenszwalb et al., 2004; Comaniciu et al., 2002; Vedaldi et al., 2008; Vincent et al., 1991). A special attention must be paid to the choice of the parameters. These methods produce superpixels that might not be homogeneous in terms of size and shape but with a good visual delineation. Superpixels algorithms have then been developed. They allow to control one or many parameters of the superpixels; their number, their size and their shape (Moore et al., 2008; Veksler et al., 2010; Levinshtein et al., 2009; Achanta et al., 2012).

For the over-segmentation of the optical images, we will use both "traditional" (i.e. non superpixels methods) and superpixels methods.

The three methods have been employed for the segmentation of RGB VHR optical images are detailed below:

- "PFF": A segmentation algorithm based on graph-cut (Felzenszwalb et al., 2004),
- Quickshift: A segmentation algorithm based on the mean shift algorithm (Vedaldi et al., 2008),
- SLIC: A genuine superpixel algorithm working in the CIELab space (Achanta et al., 2012).

PFF.

This algorithm defines a predicate for measuring the evidence for a boundary between two regions using a graph-based representation of the image (Felzenszwalb et al., 2004). An efficient segmentation algorithm based on this predicate is employed, and shows that although this algorithm makes greedy decisions, it produces a segmentation that satisfies global properties. The algorithm runs in near linear time with the number of graph edges. An important characteristic of the method is its ability to preserve details in low-variability image regions while ignoring details in high-variability regions.

Quickshift

Quickshift (Vedaldi et al., 2008) is a kernelized version of a mode seeking algorithm similar in concept to mean shift (Comaniciu et al., 2002; Fukunaga et al., 1975) or metroid shift (Sheikh et al., 2007). Quickshift is faster and reports better results than traditional mean shift or metroid shift for standard segmentation tasks. Given N data points x_1, \dots, x_N , it computes a Parzen density estimate around each point using, for example, an isotropic Gaussian window of standard deviation σ :

$$P(x) = \frac{1}{2\pi\sigma^2N} \sum_{i=1}^N e^{-\frac{\|x-x_i\|^2}{2\sigma^2}} \quad (3.7)$$

Once the density estimate $P(x)$ has been computed, Quickshift connects each point to the nearest point in the feature space which has a higher density estimate. Each connection has a distance d_x associated with it, and the set of connections for all pixels forms a tree, where the root of the tree is the point with the highest density estimate. To obtain a segmentation from a tree of links formed by quick shift, a threshold τ is chosen and break all links in the tree with $d_x > \tau$. The pixels which are a member of each resulting disconnected tree form each segment.

SLIC superpixels.

The SLIC superpixel algorithm (Achanta et al., 2012) clusters pixels in the combined five-dimensional color (CIE Lab color space) and image plane (x,y) space to efficiently generate compact, nearly uniform superpixels. It is basically based on the k-means algorithm. The number of desired clusters corresponds to the number of desired superpixels. The employed distance is based on a weighted sum of a color based distance and a plane space distance. This method produces superpixels achieving a good segmentation quality measured by boundary recall and under-segmentation error. The benefits of superpixels approaches have already been shown, as they increase classification performance over pixel-based methods. A major advantage of this superpixels method is that the produced segments are compact and regularly distributed.

3.4 Feature selection

Due to the high number of possible features involved (95), an automatic Feature Selection (FS) step has been also integrated. This selection is composed of two steps: the choice of the number of features to select and the selection of the feature subset itself. Indeed, the choice of the number of features is very important because it enables to greatly decrease the classification complexity and computation times without limiting the classification quality.

An incremental optimization heuristic called the Sequential Forward Floating Search (SFFS) (Pudil et al., 1994) algorithm is used for both steps. The SFFS algorithm has two main advantages:

- it can be used with many FS scores (in this study, the Kappa coefficient of the classification),
- it enables to access to the evolution of the classification score/accuracy according to the number of selected features.

Here, the classification accuracy is assessed through the Kappa coefficient of the RF classifier and the SFFS algorithm selects p features by maximizing this FS score criterion. In order to retrieve the optimal number of features, the SFFS algorithm was performed n times on different sample sets (i.e., training pixels from different areas of France that exhibit different tree species) with p equal to the total number of features (95, but in order to reduce the computation times of this step, one can choose a lower value). The classification accuracy a_s is conserved for each selection of s features ($s \in \llbracket 1, p \rrbracket$) and averaged over the n iterations. The number of optimal features n_{opt} is obtained as follow:

$$n_{\text{opt}} = \underset{s \in \llbracket 1, p \rrbracket}{\operatorname{argmax}} \left(\frac{1}{n} \sum_{i=1}^n a_s \right). \quad (3.8)$$

It corresponds to the size of the selection of s features having the maximal mean accuracy.

The feature selection was carried out for different areas of interest (one selection for each area) with $p = n_{\text{opt}}$. The selected features then are used for both the classification (at the object level) and the energy minimization (at the pixel level) steps.

The feature selection has been carried out only once on different training sets (i.e., training pixels from 3 different regions of France that exhibit different tree species, cf Chapter 4, Section 4.1). The selected features are then retained as the single relevant features. It is therefore not necessary to compute the others features. Such process greatly decreases the computation times but also might reduce the performance of the classification since the selected features are relevant for many regions but not necessarily for a given area of interest.

3.5 Classification

The classification is performed using a supervised classifier, in order to discriminate the vegetation types (mainly tree species) provided by the existing forest LC DB. Only the "pure" polygons (i.e., polygons containing at least 75% of a single tree specie) are employed. The classifier used in this study is the Random Forest (RF), implemented in OpenCV (Bradski et al., 2008), as it has been extensively shown relevant in the remote sensing literature (Belgiu et al., 2016; Fernández-Delgado et al., 2014). It was compared to Support Vector Machine (SVM) (Dechesne et al., 2016), and provided similar results while being faster. The RF has many advantages:

- It can handle a large number of classes. In our problem more than 20 classes can be envisaged but in most cases the studied areas exhibit between 4 and 7 different classes.
- It is an ensemble learning method, which brings a high level of generalization.
- It can handle a large number of features, even if they are derived from different remote sensing modalities (which is the case in this study).
- The feature importance can be easily assessed.

- The posterior probabilities/uncertainties are natively obtained.
- Only few parameters are needed, the tuning of the parameters is not investigated in this study since a standard parametrization produces satisfactory results.
- It is robust to noise and mislabels (Mellor et al., 2013; Mellor et al., 2014; Mellor et al., 2015)

The SVM classifier is very efficient (Vapnik, 2013), however, the training of such classifier is time consuming, especially when the number of training samples increases (which is the case when the learning is based on a database). Furthermore, when using different types of features (here spectral and lidar features) a special attention should be paid to the employed kernel. The RF classifier is therefore preferred because it natively handles features of different types and works better when the number of samples increases.

The outputs of the classification are:

- a label map that allow to evaluate the accuracy of the classification,
- a probability map (posterior class probabilities for each pixel/object). This probability map is the main input for the subsequent regularization step.

In order to overcome the issues of the generalized and partly outdated forest LC DB, a strategy is proposed in order to automatically select the most suitable training set out of an existing land-cover forest maps, subsequently improving the classification accuracy. Additionally, in order to reduce the classification complexity and computation times, a feature selection has previously carried out in order to identify an "optimal" feature subset.

3.5.1 Training set design

Using an existing LC DB to train a model is not straightforward (Pelletier et al., 2016; Gressin et al., 2013b; Radoux et al., 2014; Maas et al., 2016). First, locally it can suffer from a lack of semantic information (not all the classes of interest are present). Secondly, this database may also be semantically mislabeled and, more frequently, geometrically incorrect: changes may have happened (forest cut or grow, see Figure 3.6a) and the geodatabase may have been generalized (see Figure 3.6b), resulting in sharp polygon vertices that do not exactly correspond to the class borders. Thirdly, in many forest LC databases, polygons of a given vegetation type (mainly tree species) may contain other vegetation types in a small proportion. This is the case in the French forest LC DB where a vegetation type is assigned to a polygon if the latter is composed at least of 75% of this type. Such errors might highly penalize the classification (Carlotto, 2009), especially if random sampling is performed for constituting the training set. Deep neural networks can handle noisy labels without specific approach but it requires larger amount of training samples.

In order to correct the potential errors of the LC database or discard pixels that do not correspond to the class of interest, a k-means clustering has been therefore performed for each label in the training area (which is the area covered by the polygons of this label). It is assumed that erroneous pixels are present but in a small proportion and that therefore the main clusters corresponds to the class of interest. Let $p_{i-c,t}$ be the i^{th} pixel of the vegetation type (mainly tree species) t in the cluster c of the k-means. The pixels P_t used to train the model for the vegetation type (mainly tree species) t

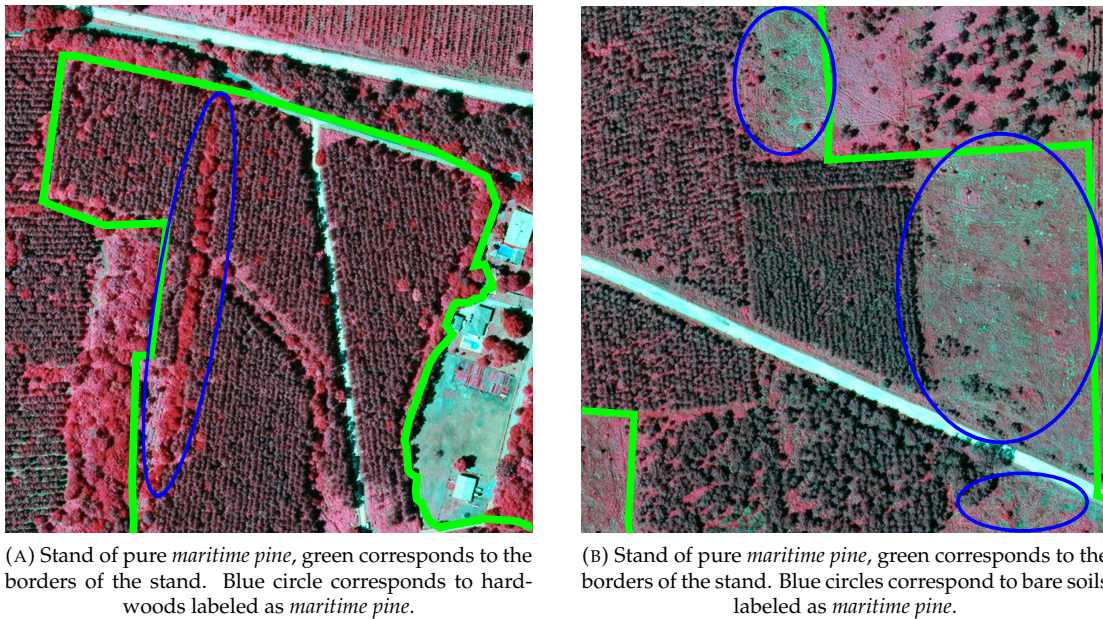


FIGURE 3.6: Two main errors in the forest LC DB; generalization (3.6a) and miss-labeling/change (3.6b).

correspond to the set:

$$P_t = \{p_{i-c,t} \mid c = \underset{c \in [1,k]}{\operatorname{argmax}} \operatorname{Card}(\cup_i p_{i-c,t})\}. \quad (3.9)$$

That is to say, only samples belonging to the main k-means cluster among training pixels for one class are kept in the training dataset.

Such selection is a bit exclusive and does not allow a lot of variability of the selected training pixels. Thus, a more reasonable way to select training pixels is to keep the pixels such as:

$$P_t = \{p_{i-c,t} \mid \frac{\operatorname{Card}\left(\cup_{\forall i} p_{i-c,t}\right)}{\operatorname{Card}\left(\cup_{\forall i, \forall c_k} p_{i-c_k,t}\right)} \geq u_p\}, \quad (3.10)$$

where u_p is a proportion defined by the user. Such selection is equivalent to select the pixels that are in a cluster which size represents a significant proportion of the total pixels labeled as t in the Forest LC. in this case, the number of clusters for the k-means can be increased. In the experiments, u_p was set to 0.25 (i.e. 25%) and k to 4. The second formulation is preferred since it allows to keep some inter-class variability.

3.5.2 Training and prediction

For the prediction, the RF classifier is employed. It is performed at the object level. On each area, the RF model is trained using 1000 samples per class. The samples are randomly drawn from the pixels selected by the training set design. The selected features are employed for the construction of the model and for the prediction.

Four main parameters are needed for the construction of the RF model. Firstly, the maximum number of trees in the forest needs to be set. Typically the more trees are employed, the better the accuracy is. However, the improvement in accuracy generally diminishes and asymptotes pass a certain number of trees. The number of trees increases the prediction time linearly. Here, 100 trees are computed in order to reach sufficient accuracy without huge computation times. The maximal depth of the tree needs then to be set. A low value will likely underfit and conversely a high value will likely overfit. The optimal value can be obtained using cross validation or other suitable methods but a depth of 25 is commonly accepted. The number of features employed at each tree node and that are used to find the best split(s) is another parameter. It is commonly set to the square root of the total number of features. Since 20 features are employed for the classification, this number is here set to 4. Besides, the tree computation can be stopped when each leaf of the tree has reached a sufficient purity or maximum population. Here, when the OOB error is below 0.01, the tree computation stops.

To summarize, the parameters of the RF employed in this study are the following:

- Number of trees: 100.
- Maximum depth of the trees: 25.
- Sufficient accuracy (OOB error): 0.01.
- Number of features at each tree node: 4

3.6 Regularization

The obtained classification might remain very noisy due to the complexity of the tree species discrimination task. The forest stand could not be clearly defined and a refinement should be employed in order to smooth the classification results. Many smoothing methods have been proposed and are evaluated in Chapter 5. They can be based on the classification label map results or on a class membership probability map. More advanced framework can include object contours (Ronfard, 1994; Chan et al., 2001) or output of object detection for higher level regularization. Here for such unstructured environments, it does not appear to be relevant.

The smoothing is here performed at the pixel level. Both local and global methods have been investigated. The local methods only consider a limited number of pixels while the global methods consider all the pixels of the image.

The global smoothing method uses only a small number of relation between neighboring pixels (8-connectivity) to describe the smoothness. An energy is computed and its minimum leads to a smoothed result. The energy $E(I, C, A)$ for an image I is expressed as follows:

$$E(I, C, A) = \sum_{\mathbf{u} \in I} E_{\text{data}}(\mathbf{u}, P(C(\mathbf{u}))) + \gamma \sum_{\mathbf{u} \in I, \mathbf{v} \in \mathcal{N}_{\mathbf{u}}} E_{\text{pairwise}}(\mathbf{u}, \mathbf{v}, C(\mathbf{u}), C(\mathbf{v}), A(\mathbf{u}), A(\mathbf{v})), \quad (3.11)$$

where $A(\mathbf{u})$ is a vector of the values of the features at pixel \mathbf{u} to be selected so as to constrain the problem according to a given criterion (height for instance) and $\mathcal{N}_{\mathbf{u}}$ is the 8-connectivity neighborhood of the pixel \mathbf{u} . When $\gamma = 0$, the pairwise prior term has no effect in the energy formulation; the most

probable class is attributed to the pixel, leading to the same result as the classification output. When $\gamma > 0$, the resulting label map becomes more homogeneous.

In spite of having only connections between local neighbors, the optimization propagates information over large distances (Schindler, 2012). The problem is NP-hard to solve but an efficient algorithm called Quadratic Pseudo-Boolean Optimization (QPBO) allows to efficiently solve it.

The principal difficulty lies in the formulation of E_{data} and E_{pairwise} , several formulations are investigated in Chapter 5. The ones that produce the best results and were finally retained are defined as follows:

$$E_{\text{data}} = 1 - P(C(\mathbf{u})). \quad (3.12)$$

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= \frac{1}{n} \sum_{i=1}^n \exp(-\lambda_i |A_i(\mathbf{u}) - A_i(\mathbf{v})|), \end{aligned} \quad (3.13)$$

where $A_i(\mathbf{u})$ is the value of the i^{th} feature of the pixel \mathbf{u} and $\lambda_i \in \mathbb{R}^{+*}$ the importance given to feature i in the regularization (λ_i are set to 1 $\forall i$). To compute such energy, the features need to be first normalized (i.e., zero mean, unit standard deviation) in order to ensure that they all have the same range.

3.7 Conclusion and discussions

In this chapter, an automatic framework for the extraction of forest stand has been proposed. It is composed of four main steps. An over-segmentation is firstly performed in order to retrieve small object that will be employed for subsequent classification. Multi-modal features are computed at the pixel level and object level. Because of the high number of features, a feature selection is also carried out in order to have a more efficient classification and reduce the computational load and time, but also in order to assess the complementarity of the multi-modal features. Classification is then performed at the object level since it improves the discrimination of tree species. When training this classifier, a specific attention is paid to the design of the training set, in order to cope with the errors of the Forest LC DB. Finally, a regularization of the label map is performed in order to remove the noise and to retrieve homogeneous forest stands according to a given criteria (here tree species). Next chapters will come into details about the different steps of the proposed framework, and will investigate several variants, so as to justify the choices and to define at the end the best joint use of VHR optical images and lidar point cloud for forest stand delineation.

The contribution of our framework is multiple:

- The extraction of features from VHR optical images and lidar 3D point cloud. From the VHR optical images, vegetation indices are derived. From the original bands and the vegetation indices, statistical features are derived. From the lidar point cloud, 3 types of features are extracted. The first ones are related to vegetation density. The seconds are related to the local shape of the lidar point cloud. Finally, statistical features (related to height or intensity of the 3D points) are extracted.
- The feature selection that allow to reduce computation times. It also allow to assess the relevance of the extracted feature and the complementarity of data sources. It is perform using an

incremental optimization heuristic called the Sequential Forward Floating Search based on the κ of the Random Forest.

- The classification is performed using the Random Forest classifier. A special attention is paid to the selection of training pixel using unsupervised classification algorithm (namely *k-means*). The training of the classifier is performed with standard parameters.
- The regularization can be performed using local or global methods (see Chapter 5). It is possible to integrate informations from the derived features. The regularization allows to smooth the classification that might be noisy. It is a necessary step in order to obtain relevant forest stands.

The use of deep-based features could be interesting since they produce good discrimination results for remote sensing application (Kontschieder et al., 2015). Furthermore, such methods can also be employed for classification, reporting good results (Paisitkriangkrai et al., 2016; Workman et al., 2017). However, since we wanted to draw the best from the data sources (especially lidar), hand-crafted features have been preferred. Indeed, the integration of lidar is generally limited to a simple nDSM (Audebert et al., 2016).

Flowchart assessment

4.1	Data	56
4.1.1	Remote sensing data	56
4.1.2	National Forest LC DB: "BD Forêt®"	57
4.1.3	Areas of interest	59
4.2	Framework experiments	64
4.2.1	Over-segmentation	64
4.2.2	Feature selection	65
4.2.3	Classification	74
4.2.4	Regularization	77
4.2.5	Computation times	80
4.3	Final results on multiple areas	80
4.3.1	Gironde	80
4.3.2	Ventoux	82
4.3.3	Vosges	86
4.3.4	Validity of the framework	93
4.4	Can forest stands be simply retrieved?	93
4.4.1	Segmentation of remote sensing data	95
4.4.2	Classification of the segments	96
4.5	Derivation of other outputs	99
4.5.1	Semi-automatic update process	99
4.5.2	Data enrichment for inventory	102
	Additional information extraction	102
	Segmentation using complementary criteria	103
4.6	Conclusion and perspectives	105

In this chapter, the reliability of the proposed framework for the segmentation of forest stands is assessed.

First the data employed in this framework are introduced. The specifications the the remote sensing modalities (VHR optical images and airborne lidar point cloud) are presented, then, the nomenclature of the French Forest land cover (LC) database (DB) is detailed. The test areas where the framework has been employed are also presented.

Experiments are then conducted in order to optimize the first steps of the method (i.e., features reliability, over-segmentation, feature selection and classification). The last step (regularization) will be experimented with detail in Chapter 5.

After validation of the framework, it is also applied several areas from different regions of France, in order to validate its adaptability.

A simple and naive segmentation method is then proposed. It aims at justifying the relevance of our framework, showing that the problem of stand segmentation can not be envisaged as a simple segmentation problem.

Eventually, examples other interesting outputs of the framework, such as a semi-automatic update process and additional features for the enrichment of the Forest LC DB are proposed.

4.1 Data

In this section, the specifications of the different input data are presented. Firstly, the two remote sensing modalities are presented. The forest LC DB is then detailed. Finally, the areas where the framework has been tested are presented.

4.1.1 Remote sensing data

VHR optical images.

IGN actually acquires airborne images over the whole territory within a 3 years update rate. Such images are part of the French official state requirement. In this thesis, the images have been ortho-rectified at the spatial resolution of 50 cm. Such resolution allows to derive consistent statistical features from optical images (see Chapter 3). The ortho-images are composed of 4 bands (red: 600-720 nm, green: 490-610 nm, blue: 430-550 nm and near infra-red: 750-950 nm) captured by the IGN digital cameras (Souchon et al., 2012) and with its own aircraft granting high radiometric and geometric quality.

Airborne Laser Scanning.

IGN also acquires 3D lidar point clouds over various areas of interest (forest, shorelines, rivers). Here, forested areas have been acquired by an Optech 3100EA device. Such data are employed in order to derive an accurate Digital Surface Model since it is the better solution to obtain it in such environment. The footprint was 0.8 m in order to increase the probability of the laser beam to reach the ground and subsequently acquire ground points, below tree canopy. The point density for all echoes ranges from 2 to 4 points/m². The points are given with an intensity value that correspond to the quantity of energy that came back to the sensor. Such information is hard to interpret since the beam might have been reflected multiple times and it is therefore difficult to calibrate the sensor. Data are usually acquired under leaf-on conditions and fit with the standards used in many

countries for large-scale operational forest mapping purposes (Kangas et al., 2006; Næsset, 2002). One may have note that superior densities are now recommended to boost metric extraction.

Data registration.

A prerequisite for data fusion is the most accurate alignment of the two remote sensing data sources (Torabzadeh et al., 2014). A frequently used technique is to geo-rectify images using ground controls points (GCPs). A geometric transformation is established between the coordinates of GCPs and their corresponding pixels in the image or 3D point in the point cloud. It is then applied to each pixel/point, so that coordinate differences on those points are reduced to the lowest possible level. This method can be easily applied and is relatively fast in terms of computation time.

The registration between airborne lidar point clouds and VHR multispectral images was performed by IGN itself using GCPs. This is a standard procedure, since IGN operates both sensors and has also a strong expertise in data georeferencing. IGN is the national institute responsible for geo-referencing in France for both airborne and spaceborne sensors. No spatial discrepancies are notice in the processed areas.

4.1.2 National Forest LC DB: "BD Forêt®"

The IGN forest geodatabase is a reference tool for professionals in the wood industry and for environmental and spatial planning stakeholders.

The forest LC database is a reference vector database for forest and semi-natural environments. Produced by photo-interpretation of VHR CIR optical images completed with extensive field surveys, the forest LC database is realized following the departmental division in the metropolitan territory.

• Forest LC DB, version 1

The version 1 of the forest LC DB, was developed by photo-interpretation of aerial images in infrared colors. Its minimum mapped surface area is 2.25 ha. It describes the soil cover (by description of the structure and the dominant composition of wooded or natural formations), based on a departmental nomenclature ranging from 15 to 60 classes according to vegetation's diversity of the mapped department. Constituted, until 2006, at the departmental level, it is available throughout the metropolitan territory. For more than half of the departments, several versions of the version 1 of the forest LC database are available.

This version of the forest LC DB is not employed in this work since it highly depends on the area investigated.

• Forest LC DB, version 2

The forest LC DB version 2 has been produced since 2007 by photo-interpretation of VHR CIR optical images. It describes the forest at the national level. Indeed, the classes are the same for the whole metropolitan part of the country which was not the case for the version 1 of the Forest LC DB. It assigns to each mapped range of more than 5000 m² a vegetation formation type. Its main characteristics are the following:

- A national nomenclature of 32 items based on a hierarchical breakdown of the criteria, distinguishing, for example, pure stands from the main forest tree species in the French forest (see Figure 4.1). Within the 32 classes, there are mixed forest (not specific information on tree species) that are not employed in our analysis and 2 classes of low vegetation formation that do not correspond to forest but that are employed during the classification.
- A type of vegetation formation assigned to each mapped range greater than or equal to 50 ares (5000 m²).
- A layer geometrically compatible with the other geodatabases layers produced by the IGN and remote sensing data sources.

Produced by department in metropolitan France, the version 2 of the forest LC database covers now 75 (out of 95) departments.

The quality of this version of the Forest LC DB has not been assessed for all departments. Evaluations have been conducted by IGN on two departments (namely Corrèze (19) and Haute-Saône (70)), enlightening that the quality of the Forest LC DB is unequal. The quality on the departments 70 was good while many confusions are reported on the departments 19 (up to 20%, showing the limits of photo-interpretation).

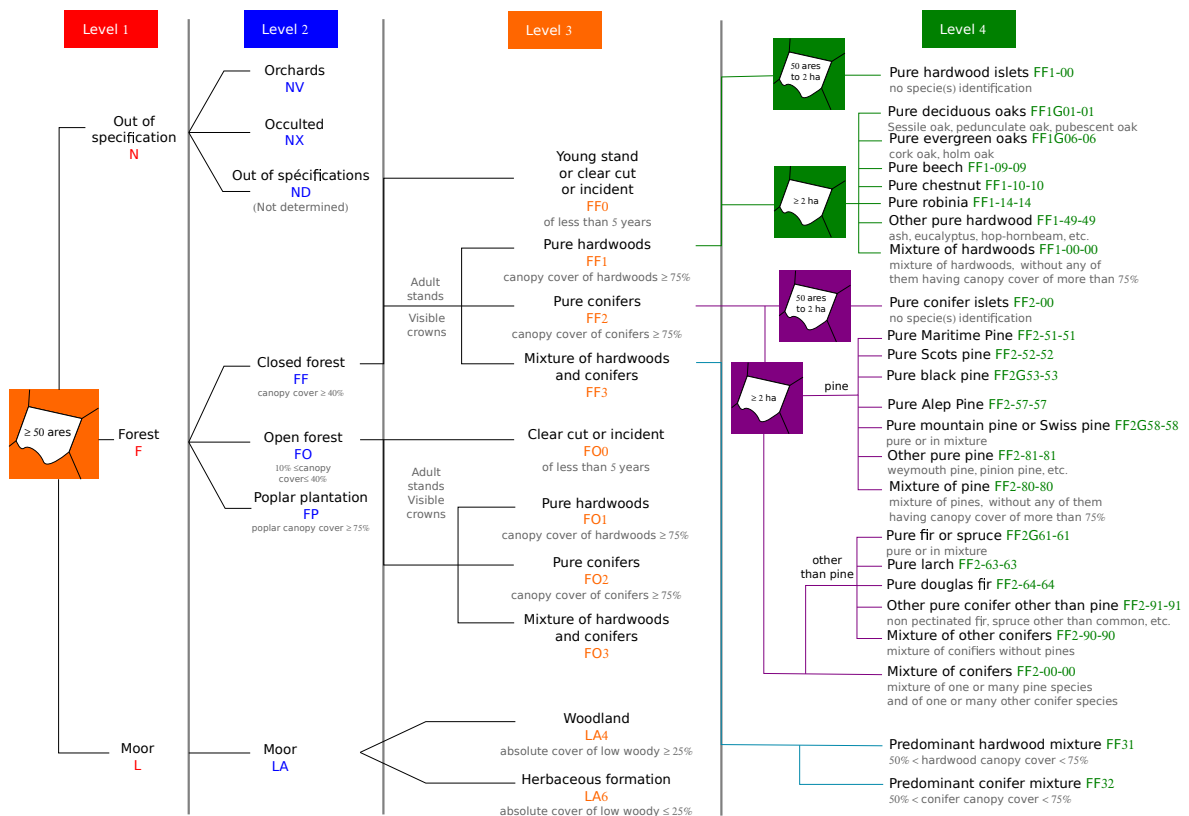


FIGURE 4.1: Organizational chart of the version 2 of the forest LC database ¹.

¹http://inventaire-forestier.ign.fr/spip/IMG/pdf/IF_25_proche_infrarouge_2.pdf

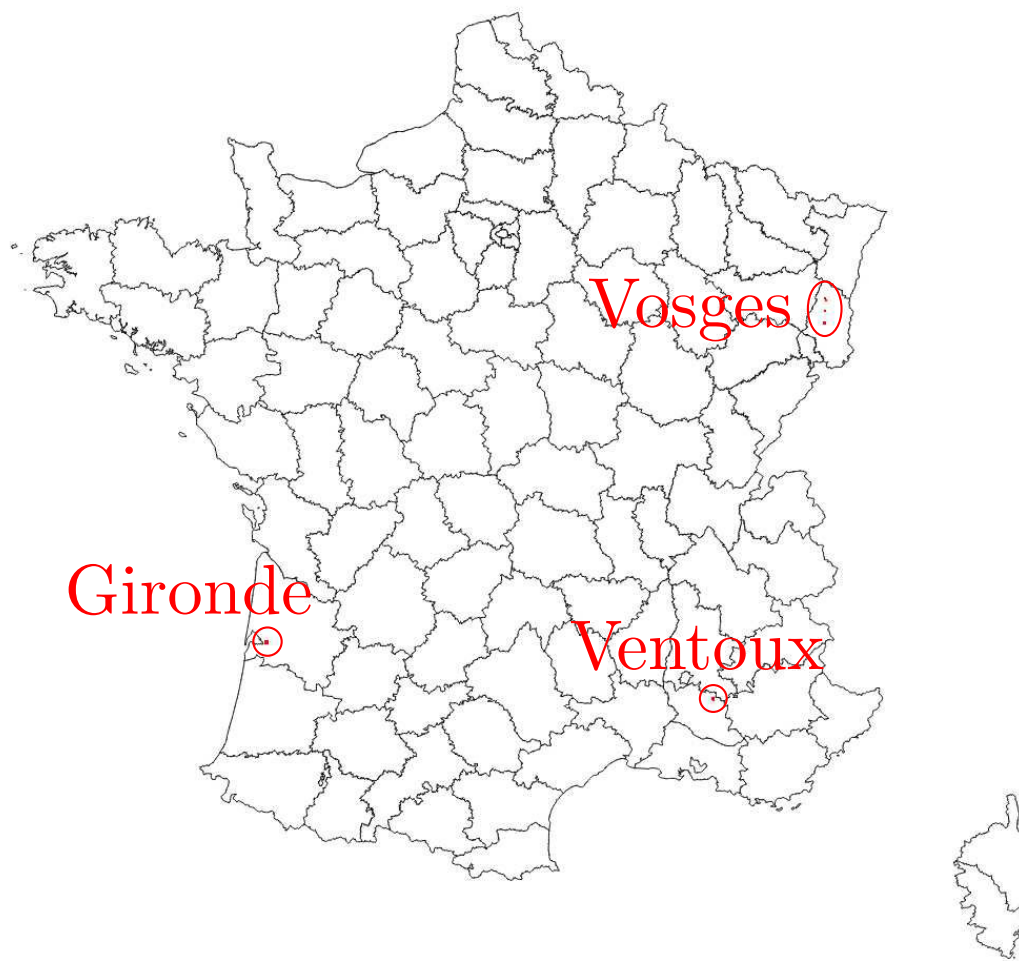


FIGURE 4.2: The 3 geographical areas of interest. 8 areas being processed in total.

4.1.3 Areas of interest

In order to assess the relevance of the proposed framework, we have selected 3 different regions of France (see Figure 4.2) which exhibit specific characteristics. The color code of the Forest LC DB is presented in Table A.1.

Gironde.

In the Gironde department, and more generally in the **Landes** forests (South-Western Atlantic coast in France), the dominant tree species is the maritime pine. A second species, oaks, cohabits with maritime pines. In the XIXth century, maritime pines were planted in the **Landes** (called **Landes de Gascogne**). The goal was multiple: clean up the marshland, retain the dunes, and provide an interesting tree exploitation to a population having at the time few sources of income.

In the studied area, two main species (namely *maritime pine* (7, ●) and *deciduous oaks* (1, ●)) are reported, but also a less common species: *elm* (labeled as *other pure hardwood* (6, ●)). A low vegetation class (namely *woody heathland* (17, ●)) is also represented. After a visual inspection of the Forest LC DB and remote sensing data, it appears that stands of *maritime pine* (7, ●) exhibit two interesting aspects:

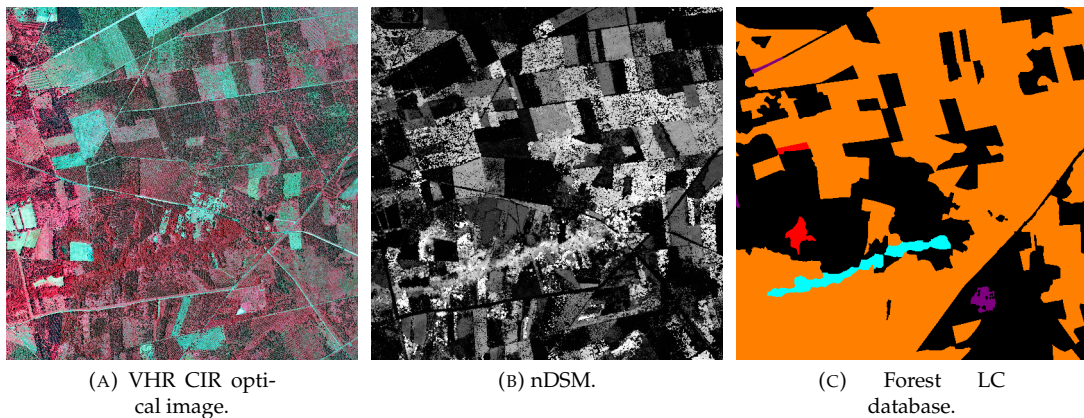


FIGURE 4.3: VHR CIR optical image, rasterized nDSM and forest LC on the selected area from the Gironde (9 km²).

- clear cuts, corresponding to areas labeled as maritime pine but that are not planted (or replanted yet),
- variation in height of trees, that is to say, a single polygon can contain stands of different height.

Ventoux.

The Ventoux is a mountainous area in South Eastern of France. The vegetation is diverse because of its location and weather. Indeed, it is a mountainous formation in a Mediterranean environment. Thus many tree species are represented, making it an interesting area for forest stand segmentation. Two areas were selected, since they exhibit various species. The first area (called *Ventoux1*) is large (2.4×2.5 km) and exhibits a large number of vegetation types (5 tree species and 2 herbaceous formation) and is therefore interesting in order to validate the framework. The second area (called *Ventoux2*) also exhibits a large number of vegetation types (4 tree species and 1 herbaceous formation) and is interesting in order to validate that under-represented classes can be well retrieved with the proposed framework. In the two proposed areas, the represented vegetation types are: *Deciduous oaks* (1, ●), *Evergreen oaks* (2, ●), *Beech* (3, ●), *Black pine* (9, ●), *Mountain pine or Swiss pine* (11, ●), *Larch* (14, ●), *Woody heathland* (17, ●) and *Herbaceous formation* (18, ●) (6 species, 2 low vegetation formations).

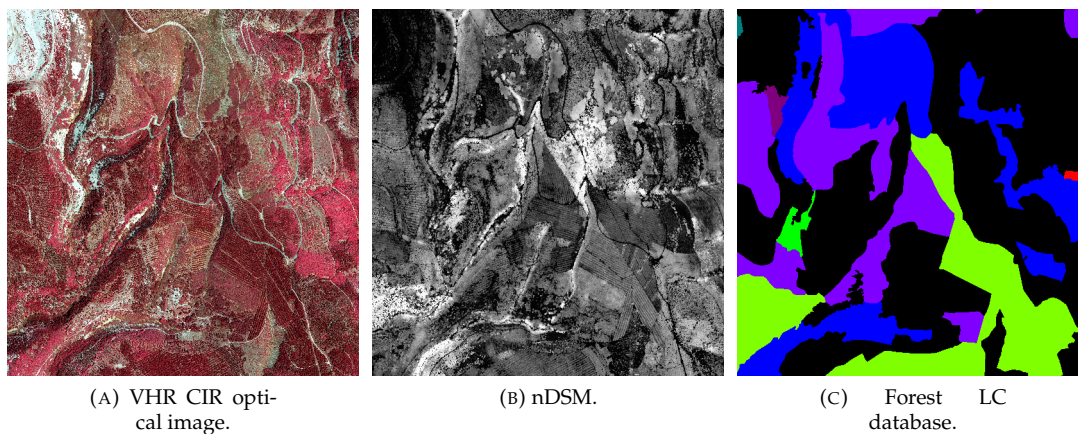


FIGURE 4.4: VHR CIR optical image, rasterized nDSM and forest LC on the selected area from the Ventoux (*Ventoux1*, 2.4×2.5 km).

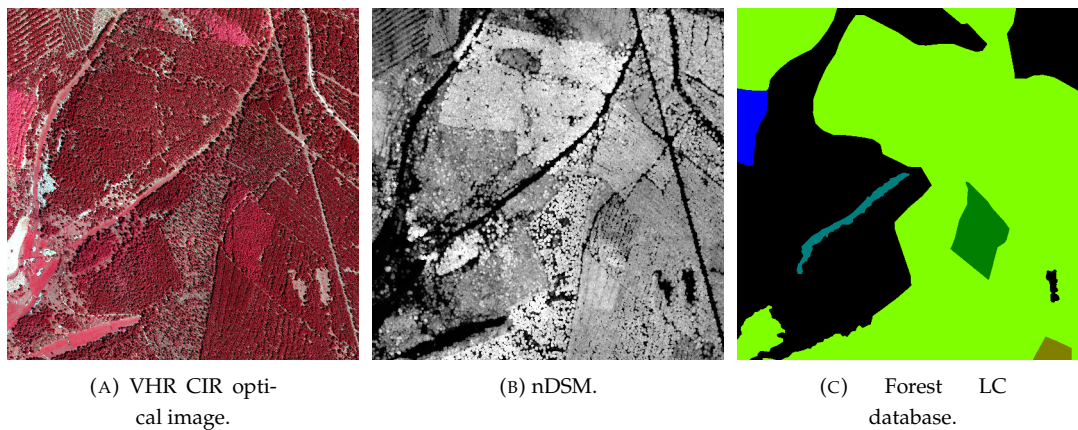


FIGURE 4.5: VHR CIR optical image, rasterized nDSM and forest LC on the selected area from the Ventoux (*Ventoux2*, 1 km²).

Vosges.

The Vosges are a mountainous area in Eastern France (up to 1,500 m). The vegetation is very varied because of the environmental conditions (altitude, climate, topography, soils type, etc.). Such forested areas are therefore very interesting for forest stand extraction since many species cohabit there. Five areas have been processed.

Vosges1 is the main area of interest of 1 km². Most of the framework has been validated on this area. Indeed, it is a very interesting area for 3 reasons:

- It contains four tree species, which is quite important for a 1 km².
- **The stands are adjacent to each other, thus, it is possible to assess if the borders are well retrieved by the proposed framework.**
- The stands exhibit height variation. Thus, stands of same specie but with different height might be extracted.

The second area (*Vosges2*) is a large area (km^2) that exhibit a large number of species (6 species, 1 low vegetation formation) and is therefore interesting in order to validate the framework. In the other areas (*Vosges3*, *Vosges4* and *Vosges5*) a relevant number of species are represented (4 or 5) but the stands are most of the time not adjacent. However, such areas allow to assess how the proposed framework operates in the mixed stands (i.e., the unlabeled areas). In the five proposed areas, the vegetation types represented are: *Deciduous oaks* (1, ●), *Beech* (3, ●), *Chestnut* (4, ●), *Robinia* (5, ●), *Scots pine* (8, ●), *Fir or Spruce* (13, ●), *Larch* (14, ●), *Douglas fir* (15, ●), *Woody heathland* (17, ●) (8 species, 1 low vegetation formation).

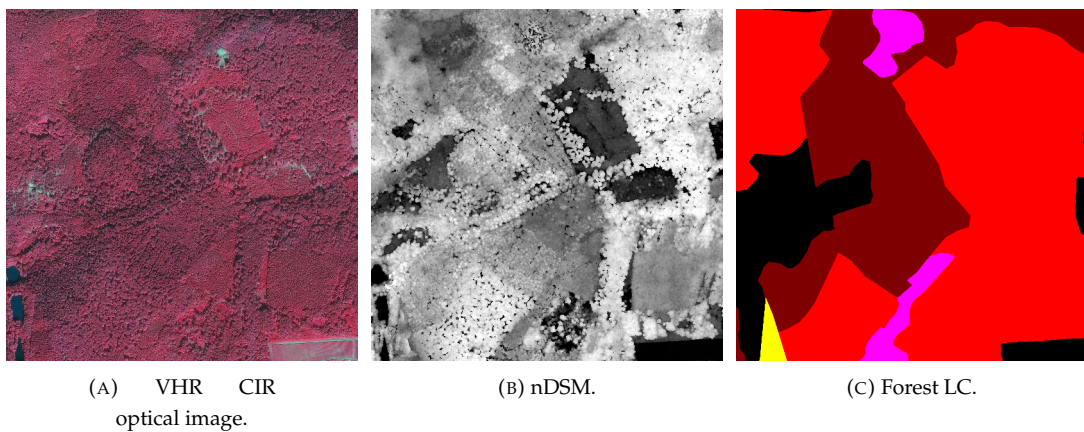


FIGURE 4.6: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges1*, 1 km^2).

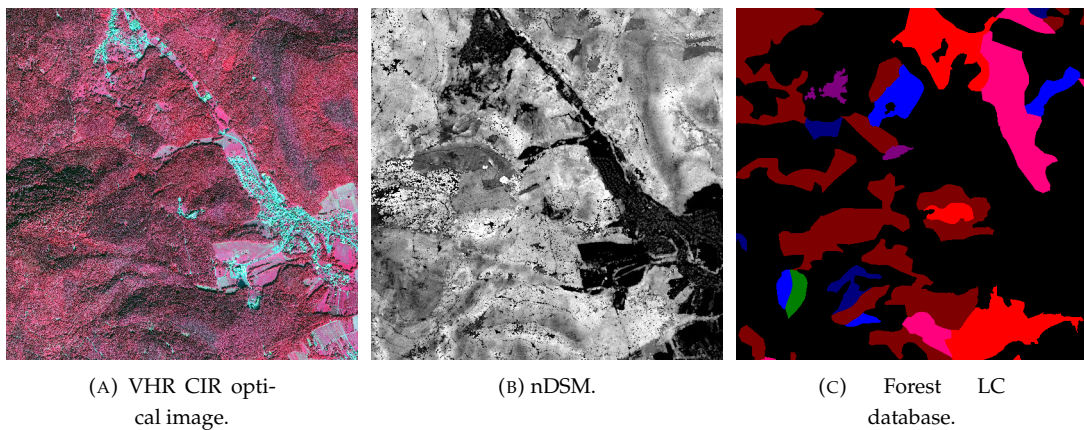


FIGURE 4.7: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges2*, 9 km^2).

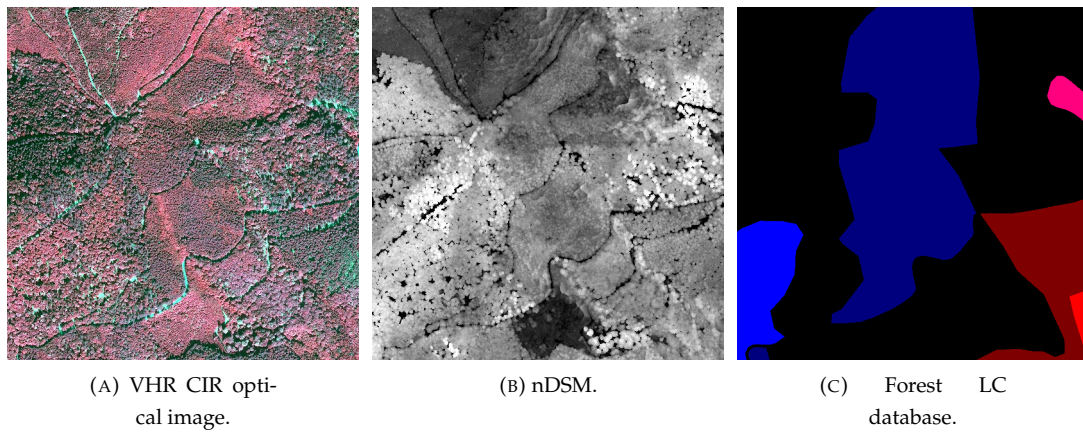


FIGURE 4.8: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges3*, 1 km²).

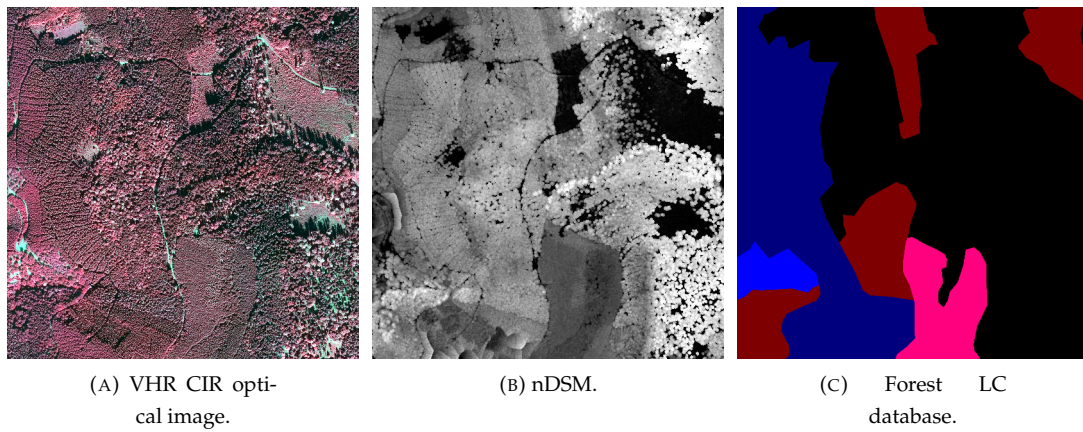


FIGURE 4.9: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges4*, 1 km²).

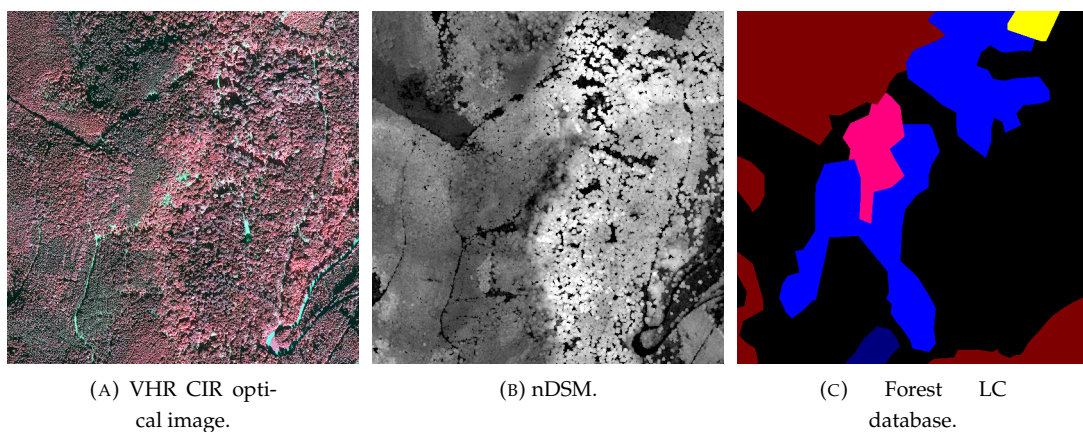


FIGURE 4.10: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges5*, 1 km²).

4.2 Framework experiments

The proposed framework is composed of different steps (defined above) they have been evaluated on a single area (*Vosges1*). Indeed, this area is very interesting since it exhibit forest stands of pure species that are adjacent to each other. Thus, the retrieval of the borders can be assessed. In this section, each step is evaluated at 2 levels; the direct output will be first considered and the impact on the final segmentation will also be investigated. In order to simplify the reading of captions, when no specific information is provided the framework is employed using the following parameters, defining a standard pipeline:

- Feature selection:
 - Selection of training samples, to cope with the Forest LC DB errors,
 - Selection of 20 features,
 - Object-based features (lidar and spectral) obtained from the PFF over-segmentation,
 - Pixel-based features (lidar and spectral) for the regularization.
- Classification performed with:
 - Selection of 500 per class training samples ($k = 4$ and clusters that amount for more than 25% of the total labeled pixel of the processed class are kept), to cope with the Forest LC DB errors,
 - Object-based features (lidar and spectral) obtained from the feature selection.
- Regularization using global method with:
 - $\gamma = 10$,
 - Linear data formulation for the unary term,
 - *Exponential-feature model* with pixel-based features obtained from the feature selection.

The validity of the framework is assessed by a visual analysis and 4 standard accuracy metrics that are derived from the confusion matrices. The employed metrics are the Intersection over Union (IoU), the mean F-score, the overall accuracy and the κ coefficient. Since they are standard metrics for the evaluation of a classification/object detection, they are not detailed here but in Section C.1 of Appendix C.

4.2.1 Over-segmentation

The over-segmentation aims at extracting "objects" so as to ease and strengthen subsequent classification task. An accurate object extraction is not mandatory since the labels are refined after. Both 3D and 2D mono-modal solutions are investigated, depending on the input data and the desired level of detail for the objects.

At the beginning, the idea was to extract trees from the lidar point cloud. It seemed meaningful since trees are the basic units of forest stands. Thus, a simple bottom-up method for tree extraction has firstly been proposed. However, a precise extraction is hard to obtain whatever the adopted technique for a large range of forested environment (Karttinen et al., 2012; Wang et al., 2016). That is why a comparison with other segmentation algorithms has then been investigated to identify whether a

tree detection was necessary or whether only homogeneous objects were sufficient. The idea here is to compare different over-segmentation methods. The advantage of using an object-based analysis instead of a pixel-based analysis is discussed in Section 4.2.3. The results of the over-segmentation provided by the several segmentation algorithms are presented in Figure 4.11.

In all the tested methods, the resulting segments are relevant since they all represent small homogeneous objects. In a given over-segmentation, objects are mostly not of the same size and shape, except for the SLIC superpixels (indeed, the aim of the methods being to obtain such uniform segments). The PFF algorithm produces objects with rough borders that follow the more precisely the borders observed in the image.

From a qualitative analysis, as expected, it appears that no segmentation method performs better than the others. Furthermore, it is impossible to evaluate the segmentation quantitatively since no over-segmentation ground truth is available. Thus, the different segmentation methods are compared through the results they produce after the object-based classification (which indicates how the objects are relevant for classification) and after the regularization (how the objects impacts the final results). These results are presented in Figure 4.12, 4.13 and 4.14.

From the results, it appears that the choice of the over-segmentation methods does not significantly impact the final results (after regularization). The watershed applied to the nDDSM tends to under-perform the other algorithms (94% of overall accuracy after regularization compared to 97% for other methods, with similar results observed using other metrics). The most visible contribution of the choice of the over-segmentation method stands in the classification results (before smoothing). The hierarchical segmentation and the watershed have poorer classification results compared to the other methods (respectively 88% and 83% in terms of overall accuracy compared to values $\geq 91\%$ for the 4 other methods, with similar results observed using other metrics). From the experiments of the over-segmentation algorithms several conclusions can be drawn:

- The watershed algorithm applied to the nDSM is not relevant for this framework.
- A precise tree delineation is not mandatory, a coarse extraction is sufficient in order to obtain relevant objects for a subsequent classification that will be regularized
- There is no preferential data type to operate the segmentation on, since both lidar (especially nDSM) and VHR optical image produce relevant over-segmentation that can be employed for classification.

It leads to the conclusion that the choice of the over-segmentation algorithm should not be guided by performance but on how complicated the tuning step is and how long it takes to operate it.

4.2.2 Feature selection

In the previous step, a high number of features has been derived (95). Thus, an automatic feature selection step is carried out for 4 main reasons.

- It allows to determine how many features are needed for an optimal classification.
- It shows the complementarity of the data sources (optical images and lidar).
- It permits to understand which features are interesting for tree species classification.
- It reduces the computational loads and times.

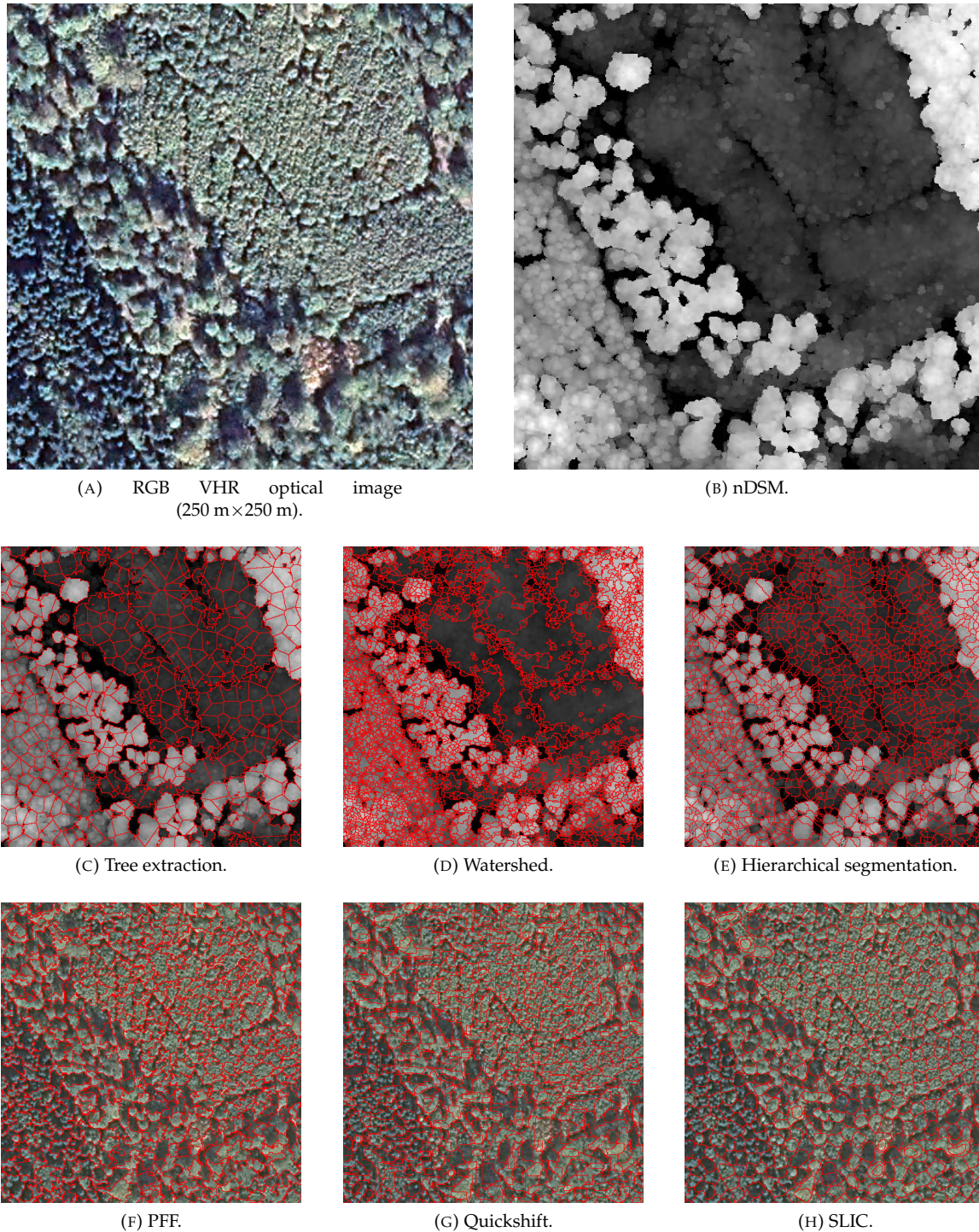


FIGURE 4.11: Illustration of over segmentation results superposed to the input remote sensing source. red corresponds to the borders found by the segmentation algorithm.

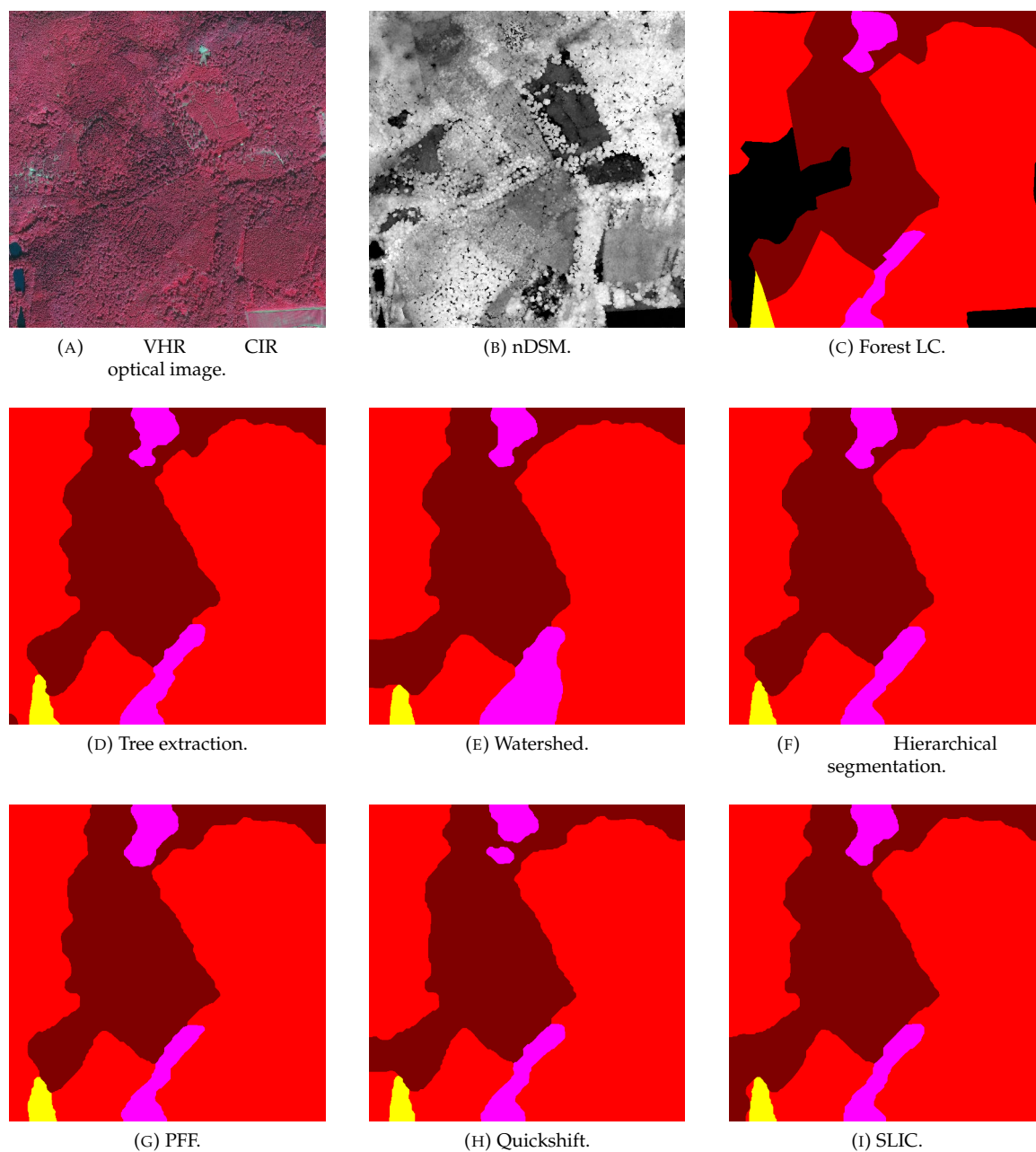


FIGURE 4.12: Final regularization results for different over-segmentation methods.

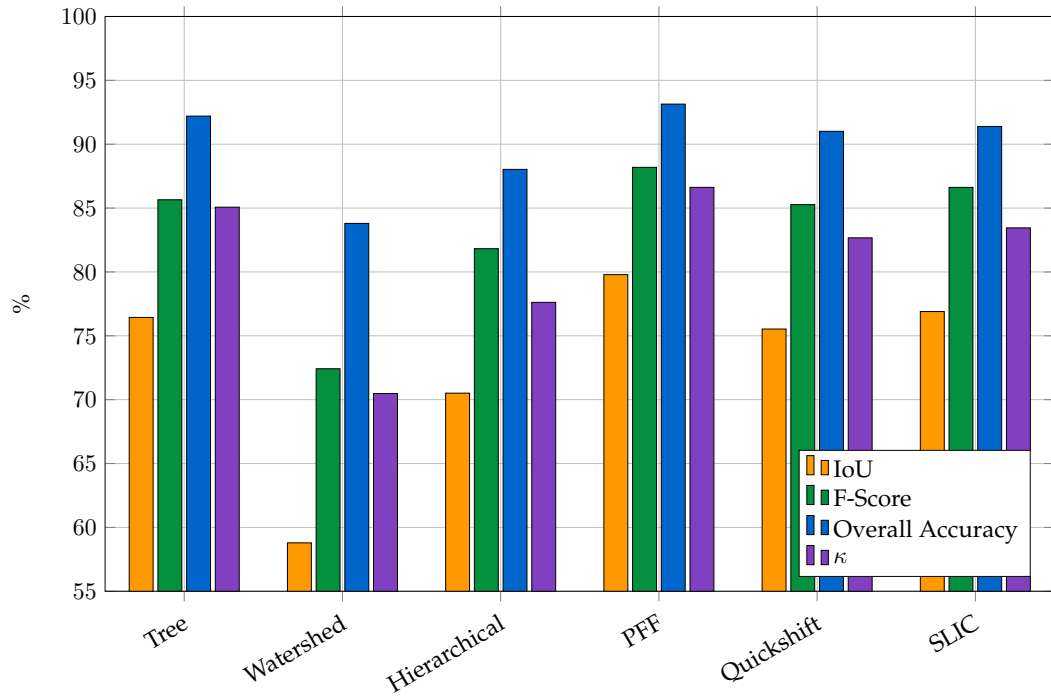


FIGURE 4.13: Quantitative evaluation of classification accuracy (according to the Forest LC DB) using different over-segmentation methods.

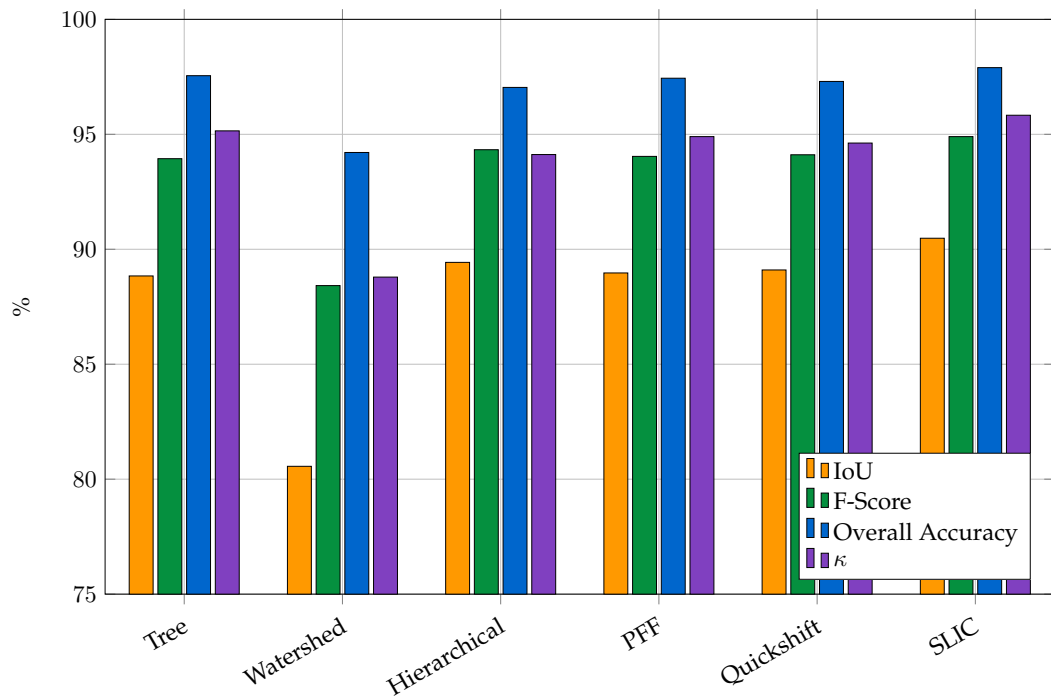


FIGURE 4.14: Quantitative evaluation of regularization accuracy (according to the Forest LC DB) using different over-segmentation methods.

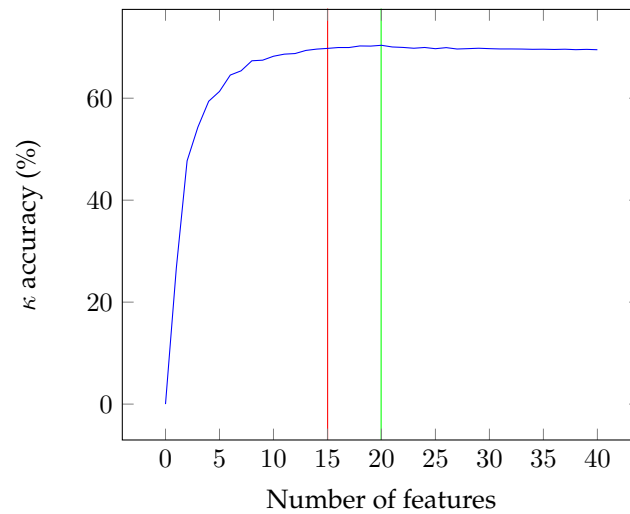


FIGURE 4.15: κ accuracy as a function of the number of selected features. The red bar corresponds to the minimal number of features needed in order to obtain sufficient classification accuracy (99% of the maximum accuracy). The green bar corresponds to the number of features needed in order to obtain the best classification accuracy

How many features for the classification?

It is aimed here at obtaining a number of features for an optimal classification. The idea is to run the feature selection N times using all features. Let $a_{n,k}$ be an accuracy metric for the classification of the n^{th} iteration using the k ($k \in \llbracket 1, K \rrbracket$, where K is the total number of available features) best features. The optimal number of feature k_{opt} is defined as follow:

$$\forall k \in \llbracket 1, K \rrbracket, \quad \sum_{n=1}^N a_{n,k} \leq \sum_{n=1}^N a_{n,k_{opt}} \quad (4.1)$$

Here, $K = 95$ and $N = 50$. **This experiment has been carried out on different areas (i.e. using training and validation samples from different geographical areas)**, and the optimal number of features found is $k_{opt} = 20$ (see Figure 4.15). When k is too low, poor classification results are obtained. When the number of features selected increases, the classification performs better, but when employing too much feature (here, more than 20), the accuracy decreases. It is a well know phenomena (Bellman, 2015; Hughes, 1968) Thus, in all the following experiments, the classification is performed using only 20 features.

Complementarity of data sources.

Once the optimal number of features was determined, the feature selection was performed 40 times over all the test areas in order to retrieve the most relevant features, in order to obtain statistically relevant results. The retained attributes are presented in Figures 4.16, 4.17 and 4.18. On average, 61% of the selected features are derived from the spectral information and 39% from the lidar information for a single selection (i.e., in a random selection of 20 features, 12 are derived from VHR optical images and 8 from lidar). This shows the complementarity of both remote sensing data.

Features for tree species classification.

For the spectral information, over the 40 selections, the features derived from the original band

set are more relevant than the ones generated from the vegetation indices: the near-infrared derived features represent 18% of the spectral selected features, 16% for the red and the green, 15% for the blue and the DVI, only 11% for the NDVI and 10% for the RVI.

The most relevant statistical features for the spectral information is surprisingly the minimum (17% of the spectral selection). The maximum (12%), the median (11%), the mean (11%) and the standard deviation (10%) are also particularly relevant. The other statistics are selected less than 9% each. For more details see Figures 4.17 and 4.18.

For the Lidar information, the most relevant feature is surprisingly the intensity, selected in each of the 40 selections (12% of the lidar selection, 5% of the total selection). This feature highly depends on the calibration of the sensor. Thus when using an other sensor, the intensity might return different results and would be not adapted for an accurate classification. The standard deviation (8% of the lidar selection), the maximum (7%) and the densities (5% and 6%) are also relevant. The other lidar derived features count for less than 4% each. For more details, see Figure 4.16.

Reduce computational loads and times.

It is obvious that processing a reduced number of features (20 instead of 95) reduces the computational loads and times. Furthermore, if an optimal feature subset is found, only the concerned features need to be computed. Thus, the feature computation step would be reduced to a minimum step by computing only the relevant features.

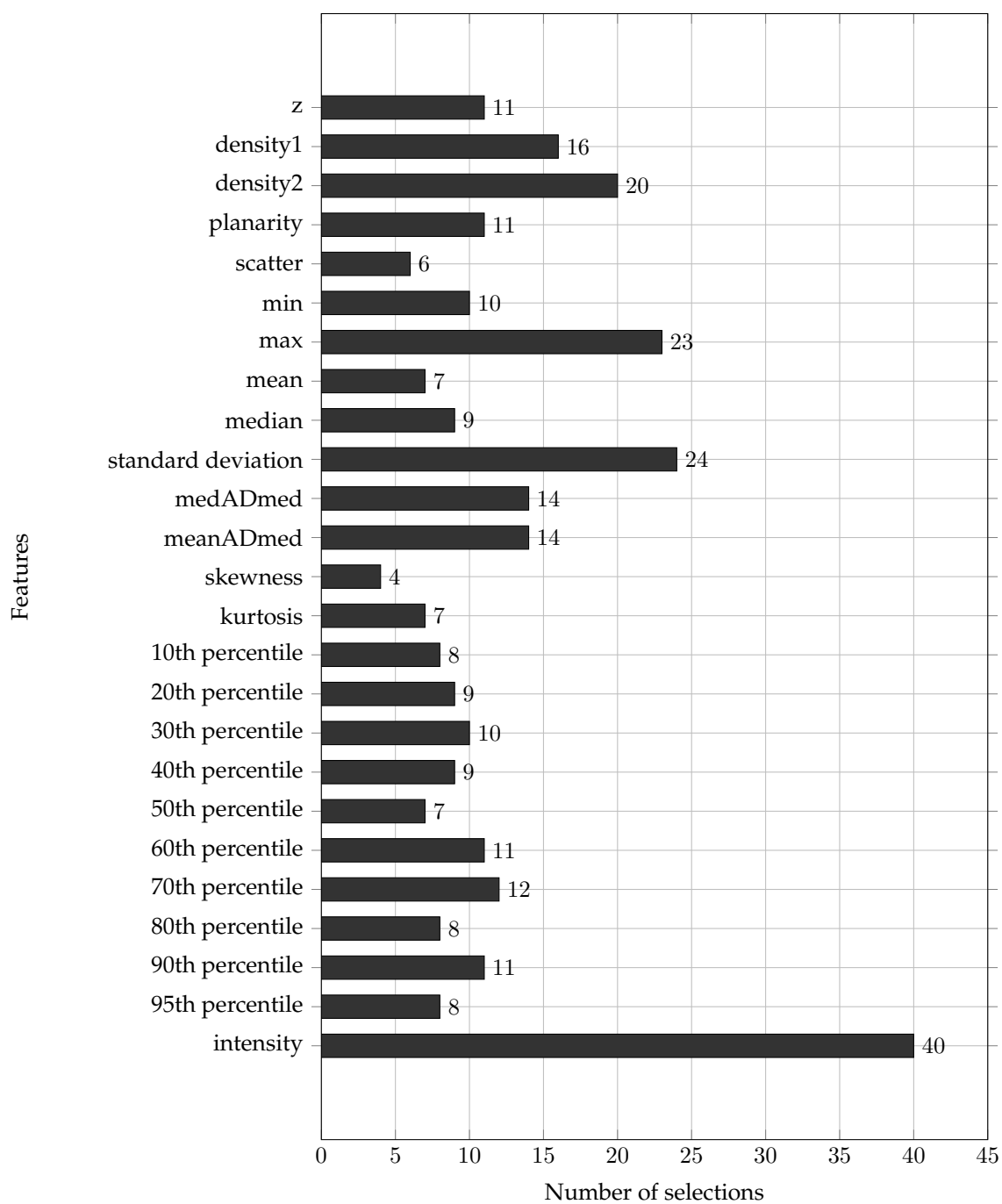


FIGURE 4.16: Result of the feature selection for the Lidar features over 40 trials of 20 optimal features.

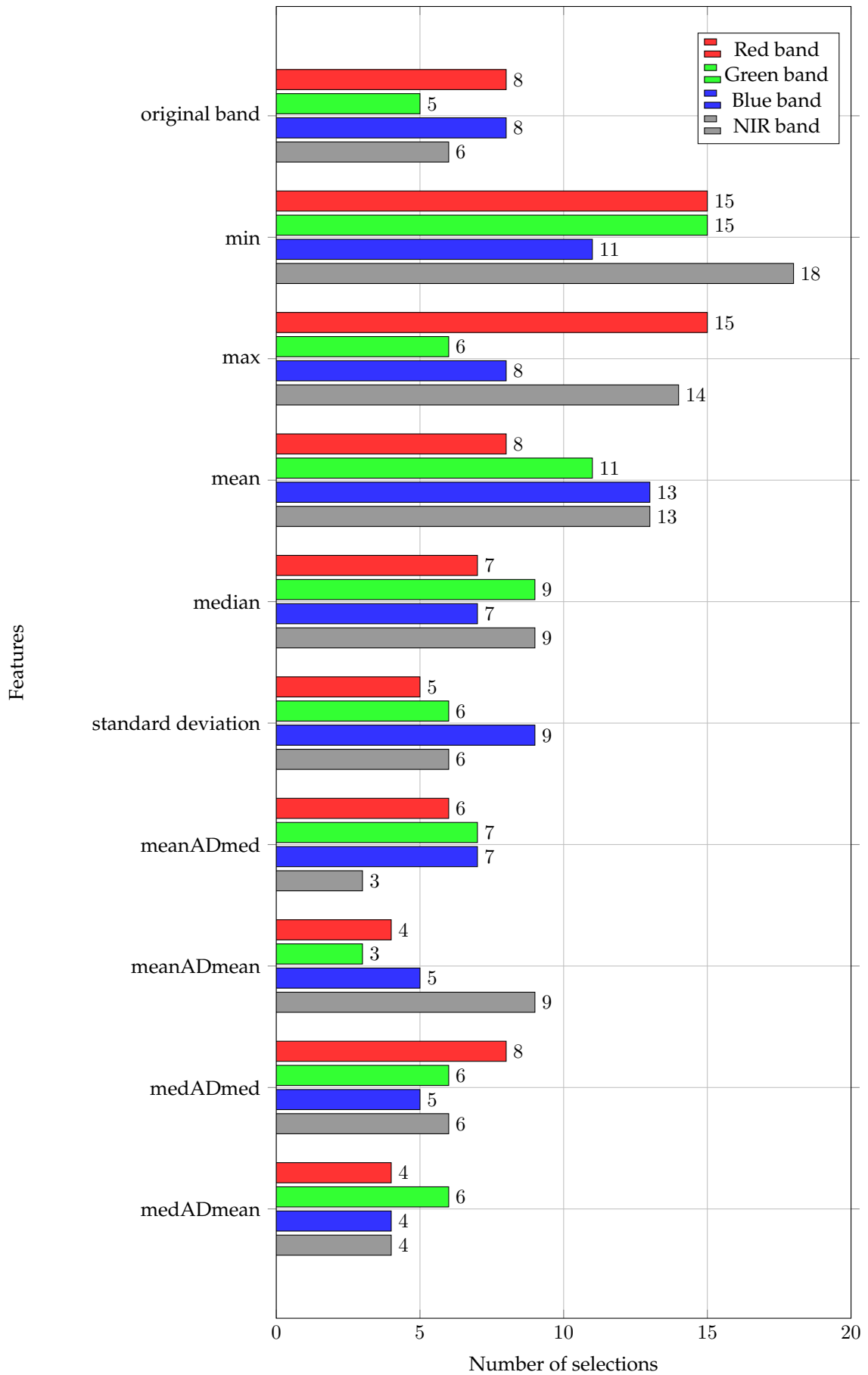


FIGURE 4.17: Result of the feature selection for the spectral features over 40 trials of 20 optimal features.

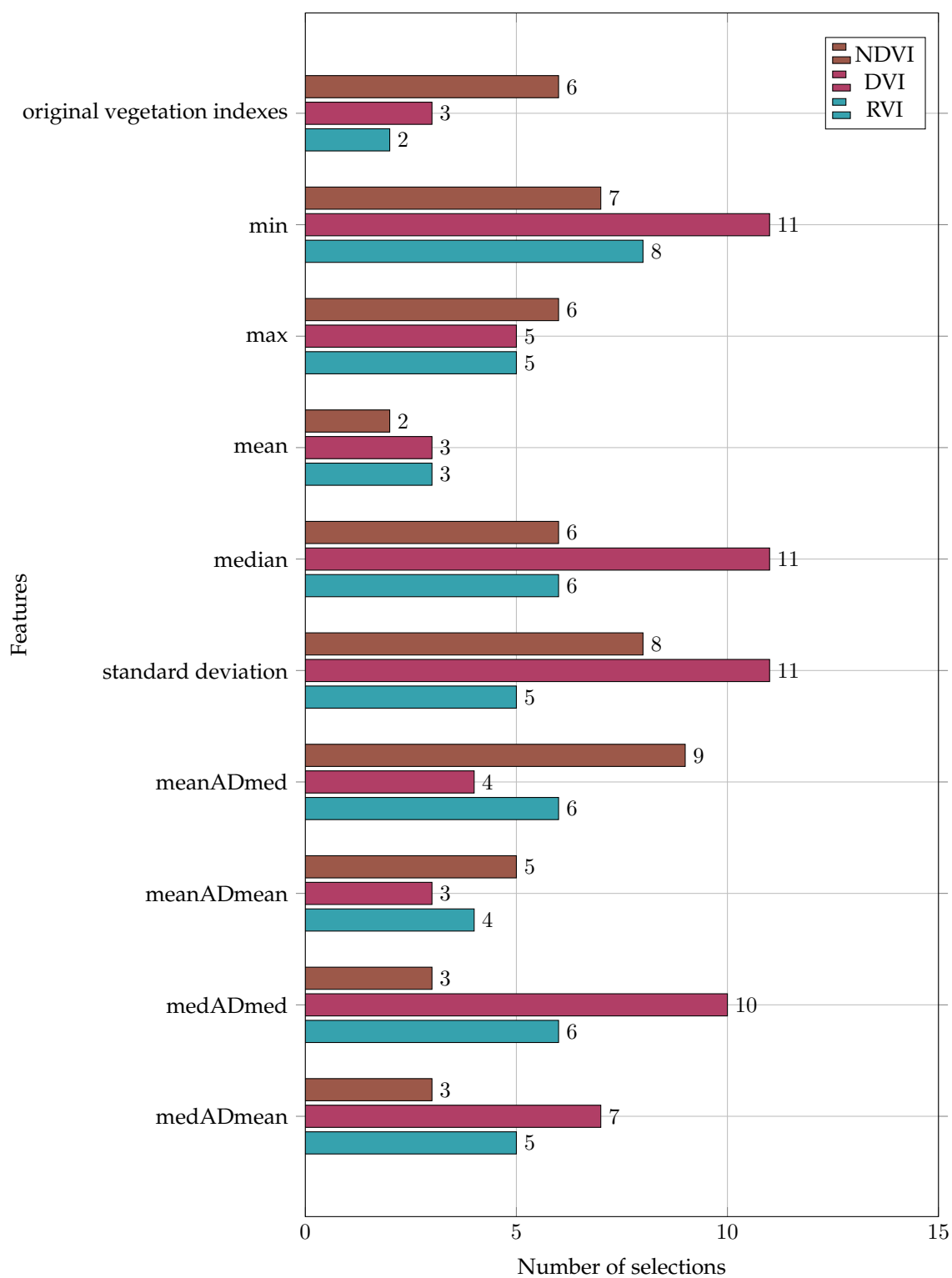


FIGURE 4.18: Result of the feature selection for the vegetation indices features over 40 trials of 20 optimal features.

4.2.3 Classification

In order to discriminate the vegetation types (mainly tree species) provided by the existing forest LC DB, a supervised classification is carried out, since such information about tree species is not straightforward to extract. Here, the classification is composed of two steps, the selection of training samples to cope with database errors and generalization and the classification using the Random Forest classifier (i.e., model training). The classification is mainly performed using standard RF classifier (Bradski et al., 2008). Tests have also been conducted using a SVM classifier with RBF kernel (Vapnik, 2013), leading to similar results (Dechesne et al., 2016). The classification can be impacted by two factors:

- The use of object-based or pixel-based features.
- The selection or non-selection of training pixels to take into account the potential errors of the Forest LC DB.

Pixel-based classification versus object-based classification.

The results of pixel-based and object-based classifications are shown by Figure 4.19. The corresponding confusion matrices and accuracy metrics are presented in Tables C.15 and C.16.

The pixel-based classification (Figure 4.19c) appears more noisy than the object-based classification (Figure 4.19d), as expected. However, the pixel-based classification already provides bad discrimination results (overall accuracy: 70.48%, κ : 0.5, mean F-score: 50%, IoU: 38.1%) since many confusions are reported. Conversely, even if the objects are roughly extracted (no specific attention is paid to the relevance of the extracted objects), the object-based classification produces more spatially consistent labels (overall accuracy: 93.14%, κ : 0.86, mean F-score: 88.19%, IoU: 79.79%). Such results impacts the final output (see Figures 4.19e and 4.19f and Tables C.17 and C.18). The regularization allows to greatly improve the bad results of the pixel-based classification (overall accuracy: 91.94%, κ : 0.84, mean F-score: 84.74%, IoU: 71.13%). However, when regularizing an object-based classification, the results are still better (overall accuracy: 97.44%, κ : 0.95, mean F-score: 94.04%, IoU: 88.97%).

Training set design.

The selection of the training pixels is an important step for the classification. It is a two sided problem:

- If the selected pixels are randomly taken from the polygons of the forest LC, some might not correspond to the target class, leading to confusions in the final classification.
- If the pixels are selected using a too discriminative criteria, the variability of the target class will not be taken into account, also leading to confusions in the final classification (over-fitting).

The Figure 4.20 shows the training pixels that have been selected by the k-means algorithm. Here, k was set to 4, in order to conserve some variability in the selected set of pixels, the cluster that account for more than 25% size of the processed class are kept. Most pixels from the forest LC are retained. However, the pixels that are excluded from the training set are visually relevant as erroneous pixels. Indeed, they mostly correspond to:

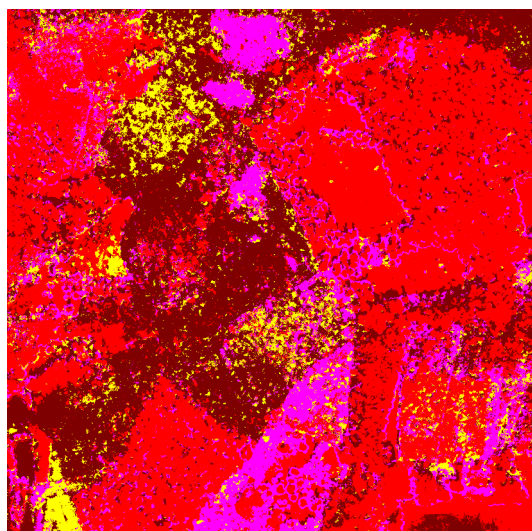
- shadows in the optical images,
- gaps in the canopy (retrieved thanks to the lidar data),



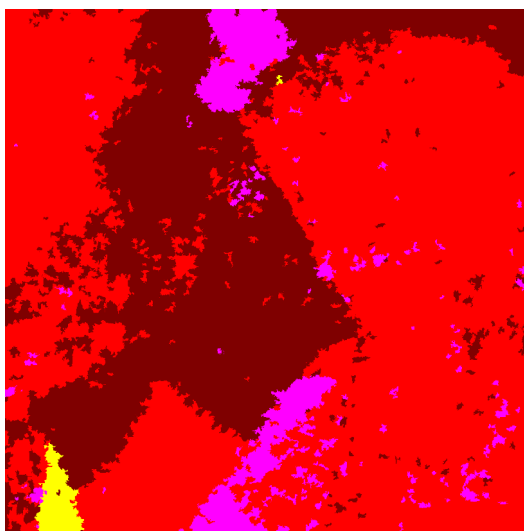
(A) VHR CIR optical image.



(B) Forest LC.



(C) Pixel-based classification (overall accuracy: 70.48%, κ : 0.50).



(D) Object-based classification (PFF) (overall accuracy: 93.14%, κ : 0.86).



(E) Regularization using pixel-based classification (overall accuracy: 91.94%, κ : 0.85).



(F) Regularization using object-based classification (PFF) (overall accuracy: 97.44%, κ : 0.95).

FIGURE 4.19: Results of the regularization; pixel-based versus object-based (*Vosges1* 1 km²).

- pixels that are visually different from the other pixels of the considered class (i.e. other minor species).

The selection of training pixels is beneficial since it allows to remove the obviously irrelevant pixels from the training set while maintaining a certain variability within classes.

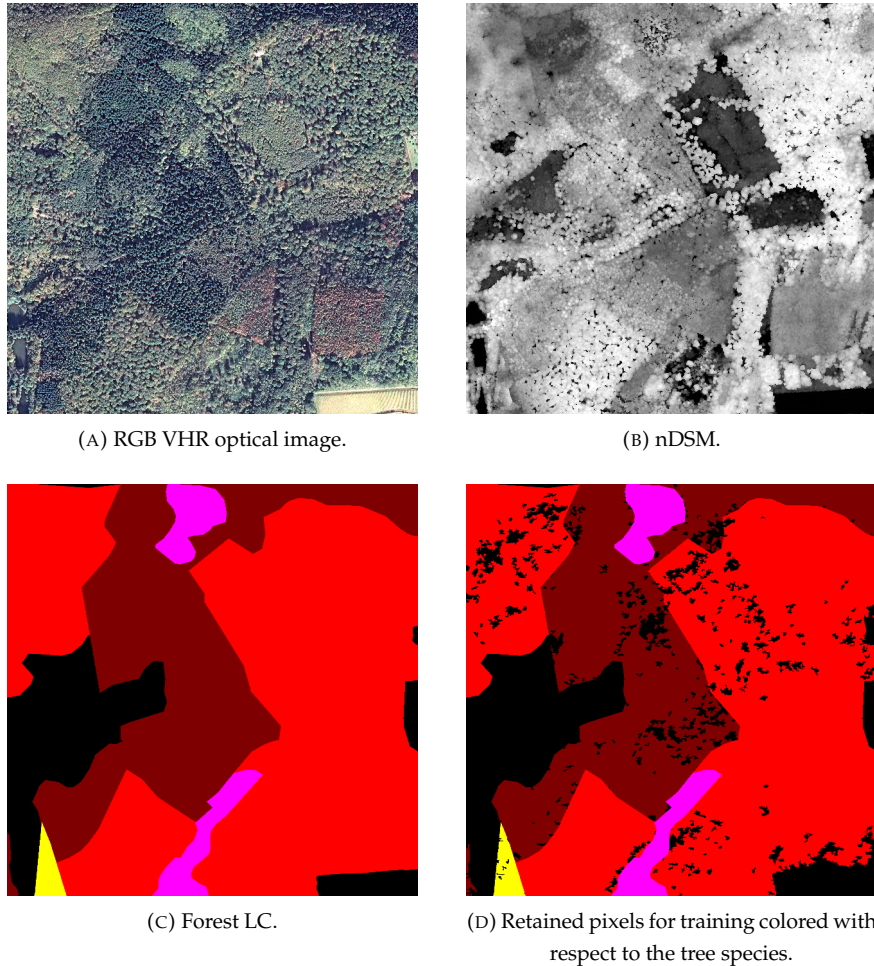


FIGURE 4.20: Selection of the training pixels.

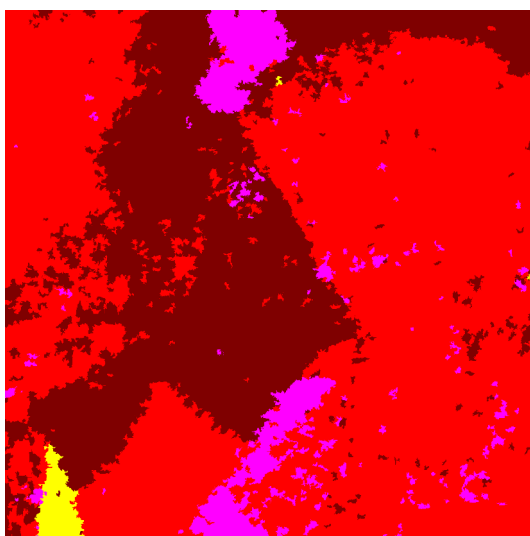
The results of the classification and regularization using different training sets is presented in Figure 4.21. The confusion matrices and accuracy metrics when the training pixels are not selected are presented in Tables C.19 and C.20.

The selection of training pixels greatly increases the classification results. Indeed, without any selection, the classification is very noisy (even when working at the object level, overall accuracy: 78.98%, κ : 0.62, mean F-score: 60.44% and IoU: 47.14%) and many confusions are reported (especially for *Chestnut* (4, ●) and *Robinia* (5, ●)). The regularization attenuates the errors, but the result remains worse than the one obtained with the selection of training pixels (overall accuracy: 94.67%, κ : 0.89, mean F-score: 90.17% and IoU: 82.61%), confusions are still reported for *Robinia* (5, ●).

4.2.4 Regularization

The obtained classification might remain very noisy due to the complexity of the tree species discrimination task. The forest stand could not be clearly defined and a refinement should be employed in order to smooth the classification results. Regularization is the final step: it smooths the former classification. Further discussions and analysis are proposed in Chapter 5 to justify the used regularization framework and its parametrization. The smoothing of the classification is performed using local or global methods. Here, we only aim to validate the relevance of this step showing examples of results and illustrating the influence of the smoothing parameter γ . Thus, only the best results are presented and discussed.

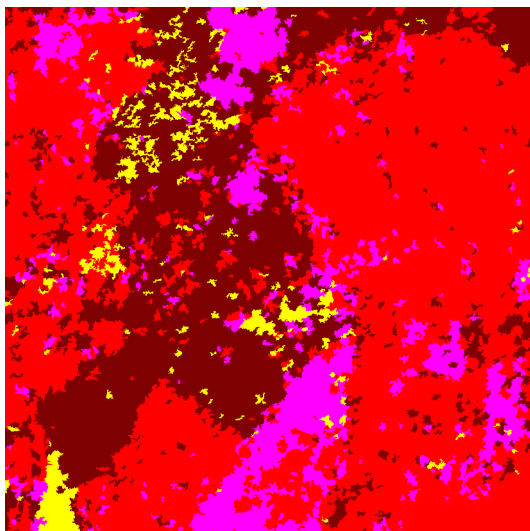
It appears that the global method produces the best results (+4% and +2% in terms of overall



(A) Classification with selection of training pixels (overall accuracy: 93.14%, κ : 0.86).



(B) Regularization with selection of training pixels (overall accuracy: 97.44%, κ : 0.95).



(C) Classification without selection of training pixels (overall accuracy: 78.98%, κ : 0.62).



(D) Regularization without selection of training pixels (overall accuracy: 94.67%, κ : 0.89).

FIGURE 4.21: Selection of the training pixels.

accuracy compared to filtering and probabilistic relaxation respectively). At this step, only the choice of the parameter γ is needed. It influences the final results and more precisely how smooth the borders of the resulting segments are.

Influence of γ

The effect of the parameter γ is presented in Figures 4.23 and 4.22. When γ is low, the borders are rough and small regions might appear (Figure 4.23c). The match with the forest LC DB is very good (overall accuracy: 97.4%, κ : 0.95, mean F-score: 94.1% and IoU: 89.16%) Increasing γ smooths the borders with, again, very good results (overall accuracy: 97.44%, κ : 0.95, mean F-score: 94.04% and IoU: 88.97%). However, a too high value has a negative impact on the results, reducing the size of meaningful segments (Figure 4.23e) or even removing them (Figure 4.23f), drastically reducing the match with the forest LC DB (overall accuracy: 96.06%, κ : 0.92, mean F-score: 89.02% and IoU: 81.32%). The tuning of the parameter γ is an important issue, since different values of γ might be acceptable depending on the level of detail expected for the segmentation. In forest inventory, having small regions of pure species is interesting for the understanding of the behavior of the forest. For generalization purposes (such as forest LC), the segments must have a decent size and may exhibit variability.

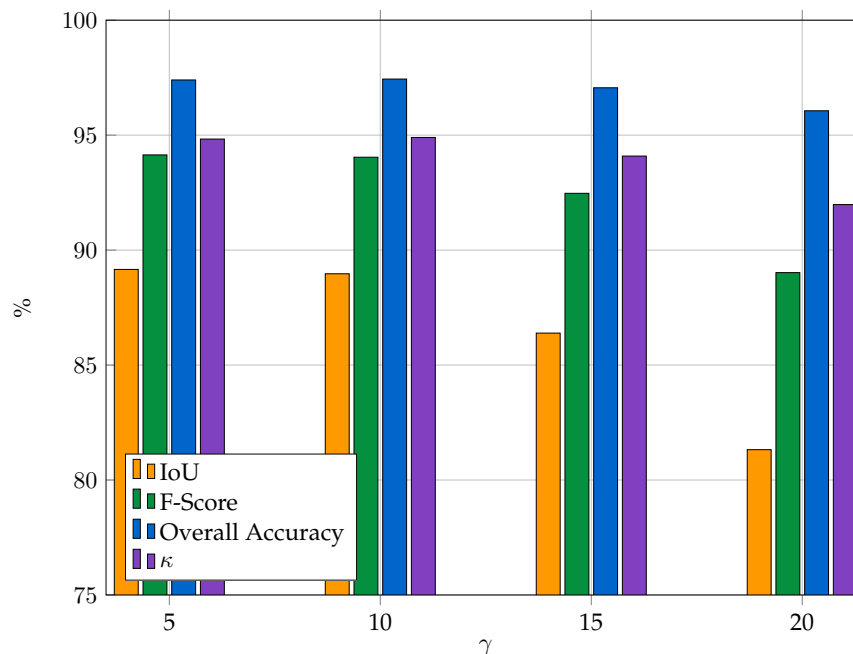


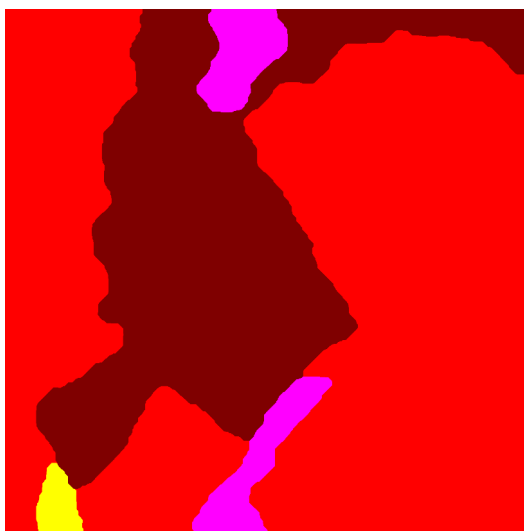
FIGURE 4.22: Regularization accuracy metrics using different values of γ .



(A) VHR CIR optical image.



(B) Forest LC.

(C) Regularization ($\gamma = 5$, overall accuracy: 97.4, κ : 0.95).(D) Regularization ($\gamma = 10$, overall accuracy: 97.44, κ : 0.95).(E) Regularization ($\gamma = 15$, overall accuracy: 97.06, κ : 0.94).(F) Regularization ($\gamma = 20$, overall accuracy: 96.06, κ : 0.92).FIGURE 4.23: Regularization results, effect of the parameter γ .

4.2.5 Computation times

In the proposed framework, the object-extraction step is the only step that exhibits different computation times depending on the over-segmentation method employed. All these methods are fast (less than 5 minutes for a 1 km² area), except for the tree extraction from the Lidar point cloud. Indeed, the tree extraction needs to iterate several times on all the points from the lidar cloud, leading to high computation times (about 1 h30 for a 1 km² area). These computation times, with regard to results of the over-segmentation methods confirm that the tree extraction is not necessary. It produces similar results compared to other methods, but with higher computation times.

For the other steps of the framework, the computation times are presented in Table 4.1. The computation of the features is the most time consuming step (2h). The feature selection (1h) here appears also quite long but can easily be reduced if necessary (e.g., by employing an other method). The implementation of the different steps of the framework can be improved). This times are provided for the retrieval of the best features within the whole feature set. However, once they are identified, only a limited number of optimal features needs to be computed. It results in a decrease of the computation times (less features are computed and no feature selection is carried out). Initial, the proposed framework takes 4 h for 1 km², with optimal features identified, it takes 1 h 30.

	Computation time
<i>Lidar features (all)</i>	~ 1 h
<i>Optical features (all)</i>	~ 1 h
<i>Object-based feature map</i>	~ 10 min
<i>Feature selection</i>	~ 1 h
<i>Classification</i>	~ 10 min
<i>Regularization</i>	~ 30 min
<i>Full algorithm</i>	~ 4 h

TABLE 4.1: Average computation times of the different steps of the framework for a 1 km² area.

4.3 Final results on multiple areas

In the previous section, results have been presented on a single zone in order to evaluate the different possibilities of the framework and find out optimal schemes for forest stands extraction. In this section, results are presented over different regions of France using only the best configuration (see Section 4.2).

4.3.1 Gironde

The area and the results of the framework are presented in Figure 4.24. The confusion matrices and accuracy metrics for the classification and regularization are presented in Table 4.2 and 4.3 respectively.

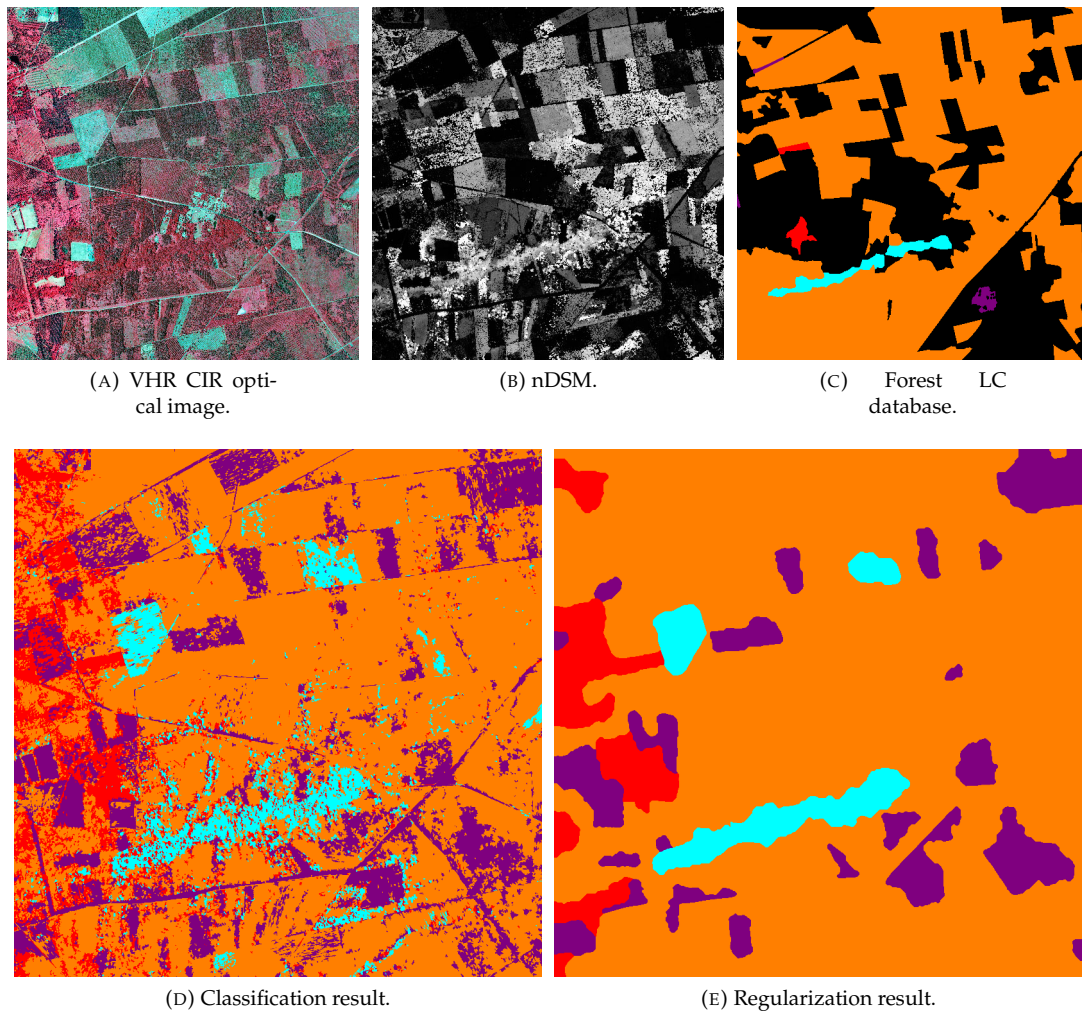


FIGURE 4.24: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Gironde (9 km²).

Confusion matrix					
Label	● 1	● 6	● 7	● 17	Precision
● 1	175555	714	5360	2817	95.18
● 6	93503	340999	70548	4865	66.87
● 7	1525795	978053	15704988	3466084	72.46
● 17	11467	1489	20525	114102	77.31
Recall	9.719	25.81	99.39	3.18	

Accuracy metrics					
Label	● 1	● 6	● 7	● 17	Overall
IU	9.671	22.88	72.14	3.151	26.96
F-score	17.64	37.24	83.81	6.109	36.2
Accuracy	92.72	94.9	73.06	84.42	92.72
P0	0.93	0.95	0.73	0.84	0.73
Pe	0.91	0.92	0.69	0.84	0.68
κ	0.1639	0.3512	0.1402	0.04912	0.146

TABLE 4.2: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	● 1	● 6	● 7	● 17	Precision
● 1	179476	2	4968	0	97.31
● 6	3417	419132	87366	0	82.2
● 7	615727	534859	16777216	1943289	84.43
● 17	0	0	36207	111376	75.47
Recall	22.47	43.93	99.24	5.421	

Accuracy metrics					
Label	● 1	● 6	● 7	● 17	Overall
IoU	22.33	40.12	83.89	5.327	37.92
F-score	36.51	57.26	91.24	10.11	48.78
Accuracy	96.99	96.98	84.44	90.44	99.25
P0	0.97	0.97	0.84	0.9	0.84
Pe	0.95	0.93	0.79	0.9	0.79
κ	0.3558	0.5585	0.2575	0.08904	0.275

TABLE 4.3: Confusion Matrix and accuracy metrics of the regularization.

In this area, a lot of confusions are reported. The main class (namely *maritime pine* (7, ●)) is well retrieved. The confusions are due to the over-representation of the *maritime pines* (7, ●) which exhibits high variability (different heights). Furthermore, some areas are labeled as *maritime pine* (7, ●) in the forest LC have probably been harvested. They clearly appear as bare soils that can be easily confused with *Woody heathland* (17, ●). However, from a visual point of view the results are coherent. Such errors are easily automatically detected and can be employed for the update of the Forest LC DB or the detection of clearcut areas. Furthermore, it is interesting to note that the stand of *elm* labeled as *other hardwood* (6, ●), is well retrieved.

The poor results observed can be explained:

- The intense harvesting in this area, leading to many clearcut that can be confused with bare soil by the classifier.
- The major forest damage caused by Cyclone Klaus in January 2009.

4.3.2 Ventoux

The results of the framework over two areas presented in Figures 4.25 and 4.26. The confusion matrices and accuracy metrics of the classification and regularization are respectively presented in Tables 4.4 and 4.5 for *Ventoux1* and Tables 4.6 and 4.7 for *Ventoux2*.

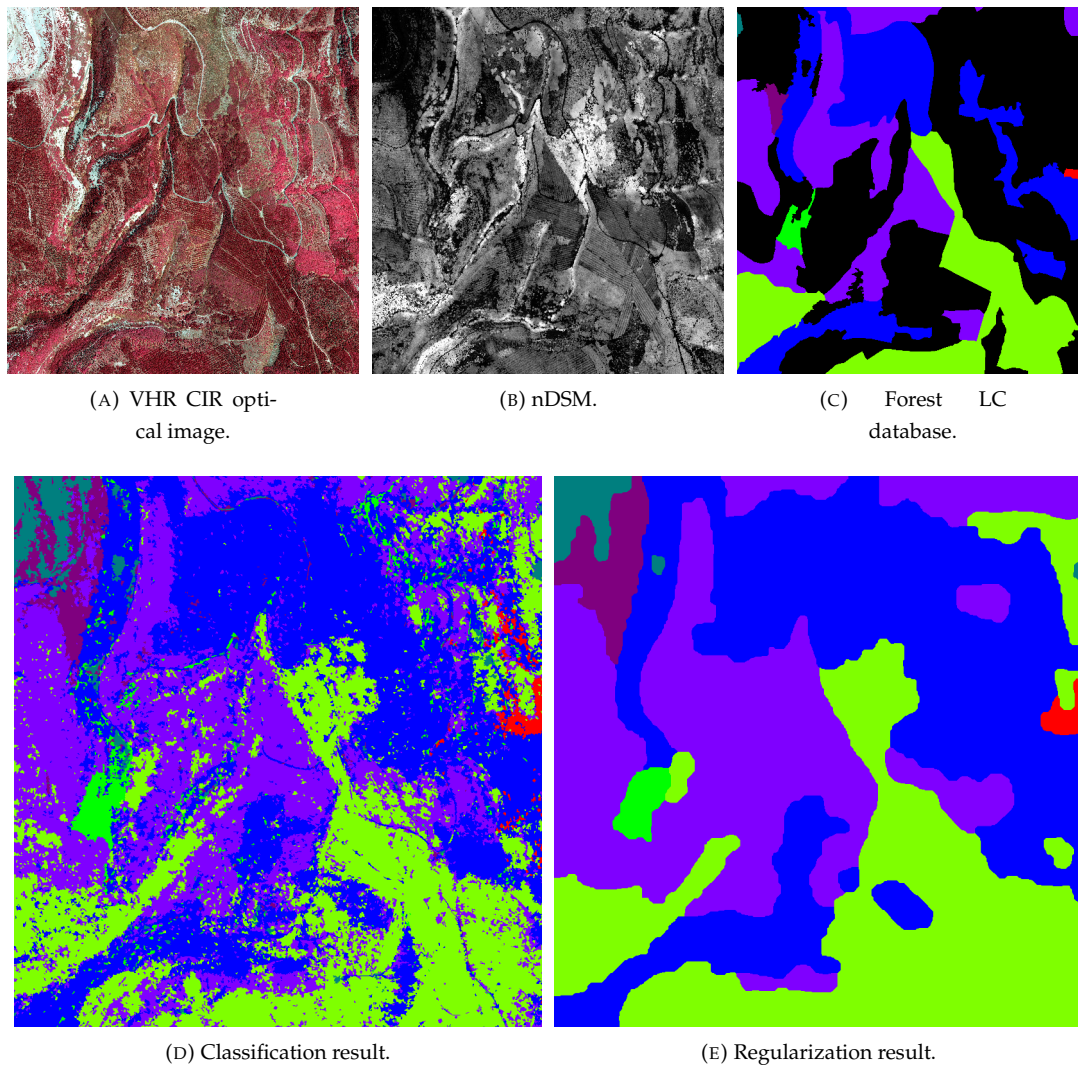


FIGURE 4.25: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Ventoux (*Ventoux1*, 2.4×2.5 km).

On this area, very good results are reported. In the classification process, some confusions are reported (between *Evergreen oaks* (2, ●) and *Beech* (3, ●) and also between *Beech* (3, ●) and *Herbaceous formation* (18, ●)), even if the results are already good. The classification precisions for all classes are greater than 80%, and most of the recall rates are also satisfactory ($> 70\%$, except for *Evergreen oaks* (2, ●), *Woody heathland* (17, ●) and *Herbaceous formation* (18, ●)). After regularization, the results are improved a lot (precisions greater than 90% for all classes). Regarding the recalls, only the *Woody heathland* (18, ●) class suffers from limited confusion with *Deciduous oaks* (1, ●) and *Mountain pine or Swiss pine* (11, ●). The global results are very satisfactory with an IoU of 89.45%, showing that stands borders are well retrieved. The κ of 0.94 shows an agreement nearly perfect. Finally, the mean F-score and overall accuracy confirm the relevance of the results. Furthermore, in the mixed areas (i.e., areas not labeled in the Forest LC DB), we retrieve small stands that are visually coherent, showing that the proposed framework allows to obtain finest results.

Confusion matrix								
Label	● 1	● 2	● 3	● 9	● 11	● 17	● 18	Precision
● 1	58218	0	3247	2759	331	0	0	90.18
● 2	0	154297	772	79	186	439	2404	97.55
● 3	21038	85103	3827279	66812	188176	41743	32783	89.78
● 9	0	11572	175368	3043399	292148	4760	4766	86.17
● 11	139	45061	384619	294910	3083873	43374	9964	79.85
● 17	0	0	1045	0	5905	104020	13	93.73
● 18	0	0	31	829	154	126	29367	96.26
Recall	73.33	52.12	87.13	89.28	86.36	53.49	37.03	

Accuracy metrics								
Label	● 1	● 2	● 3	● 9	● 11	● 17	● 18	Overall
IoU	67.91	51.45	79.27	78.09	70.91	51.64	36.51	62.25
F-score	80.89	67.94	88.44	87.7	82.98	68.11	53.49	75.65
Accuracy	99.77	98.79	91.68	92.9	89.48	99.19	99.58	85.69
P0	1	0.99	0.92	0.93	0.89	0.99	1	0.86
Pe	0.99	0.96	0.54	0.59	0.57	0.97	0.99	0.31
κ	0.8077	0.6738	0.8194	0.827	0.7538	0.6773	0.5332	0.7929

TABLE 4.4: Confusion Matrix and accuracy metrics of the classification (*Ventoux1*).

Confusion matrix								
Label	● 1	● 2	● 3	● 9	● 11	● 17	● 18	Precision
● 1	62834	0	6	1715	0	0	0	97.33
● 2	0	150450	5533	0	2194	0	0	95.11
● 3	6556	0	4185074	24676	30802	15531	295	98.17
● 9	0	0	28668	3463198	40147	0	0	98.05
● 11	0	5123	145885	193255	3500313	17364	0	90.64
● 17	0	0	236	0	5770	104977	0	94.59
● 18	0	0	0	517	0	0	29990	98.31
Recall	90.55	96.71	95.87	94.02	97.8	76.14	99.03	

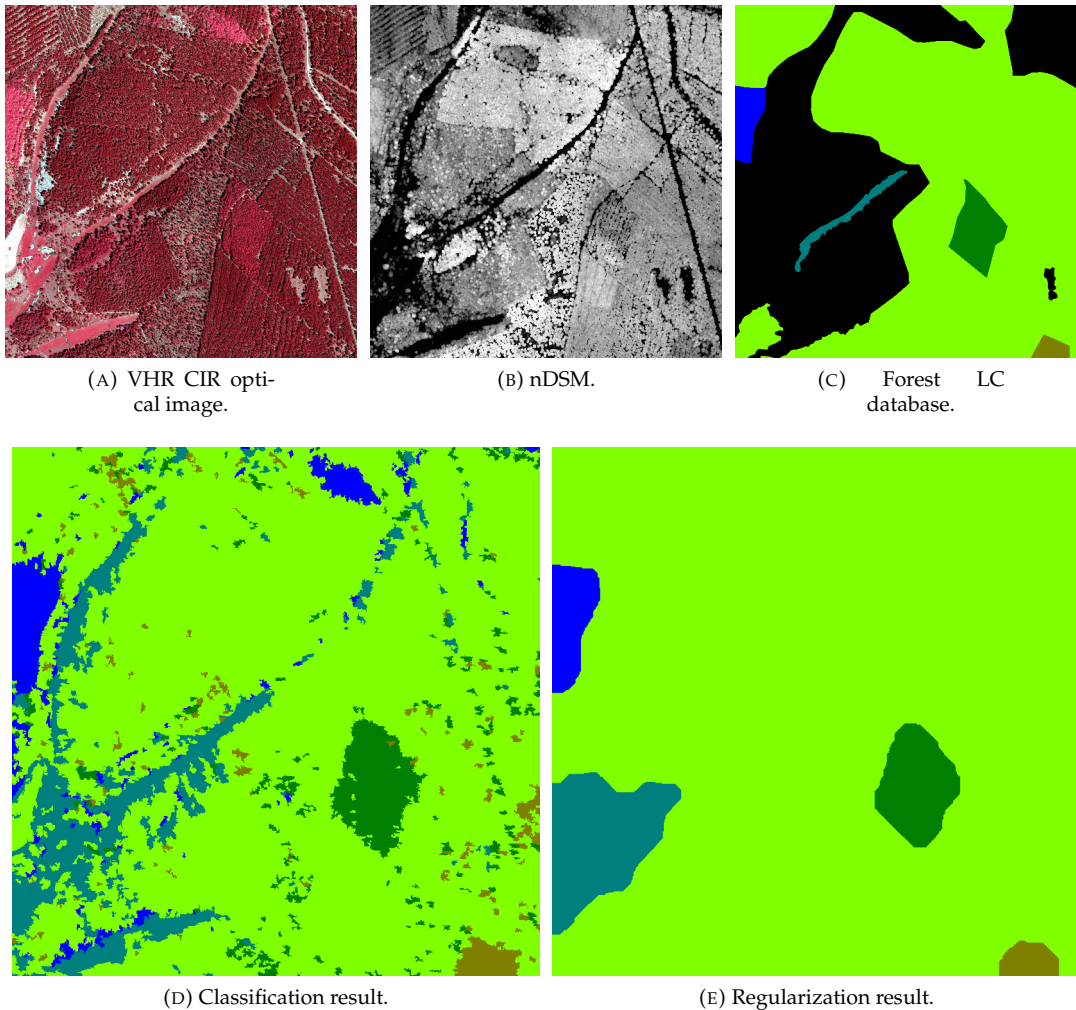
Accuracy metrics								
Label	● 1	● 2	● 3	● 9	● 11	● 17	● 18	Overall
IoU	88.36	92.13	94.19	92.3	88.82	72.96	97.36	89.45
F-score	93.82	95.9	97.01	95.99	94.08	84.37	98.66	94.26
Accuracy	99.93	99.89	97.85	97.6	96.34	99.68	99.99	95.64
P0	1	1	0.98	0.98	0.96	1	1	0.96
Pe	0.99	0.97	0.54	0.58	0.57	0.98	0.99	0.31
κ	0.9379	0.9585	0.9533	0.9428	0.9143	0.8421	0.9866	0.9364

TABLE 4.5: Confusion Matrix and accuracy metrics of the regularization (*Ventoux1*).

Similar results are observed on the small area. The classification reports great precisions (overall accuracy: 92.06%, κ : 0.64, mean F-score: 71.06%, IoU: 58.08%), however, recall rates are not sufficient for *Larch* (14, ●), *non-pectinated fir* (labeled as *Other conifer other than pine* (16, ●)) and *Herbaceous formation* (18, ●). The global metrics also report good results. After regularization, most of the precision and recall rates are improved, except for the *Herbaceous formation* (18, ●). Indeed, this class is represented as a thin strip that is totally merged with *Black pine* (9, ●). This is a quite specific situation in the Forest LC DB (usually, object are less elongated / thin). The global metrics confirm the good results (overall accuracy: 98.09%, κ : 0.87, mean F-score: 84.58%, IoU: 77.91%), the mean F-score reflects the discussed confusion in the final result.

Confusion matrix						
Label	● 3	● 9	● 14	● 16	● 18	Precision
● 3	55299	316	0	0	0	99.43
● 9	19303	2148084	89687	49739	38944	91.57
● 14	0	3770	92067	260	0	95.81
● 16	0	8	0	20508	0	99.96
● 18	35	398	2	165	34587	98.29
Recall	74.09	99.79	50.65	29.02	47.04	

Accuracy metrics						
Label	● 3	● 9	● 14	● 16	● 18	Overall
IoU	73.78	91.4	49.56	29.02	46.66	58.08
F-score	84.91	95.51	66.27	44.98	63.63	71.06
Accuracy	99.23	92.08	96.33	98.03	98.45	92.06
P0	0.99	0.92	0.96	0.98	0.98	0.92
Pe	0.95	0.79	0.9	0.96	0.96	0.78
κ	0.8452	0.6276	0.6452	0.4429	0.6294	0.6416

TABLE 4.6: Confusion Matrix and accuracy metrics of the classification (*Ventoux2*).FIGURE 4.26: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the *Ventoux* (*Ventoux2*, 1 km²).

Confusion matrix						
Label	● 3	● 9	● 14	● 16	● 18	Precision
● 3	55365	250	0	0	0	99.55
● 9	2055	2325662	8149	5015	4876	99.14
● 14	0	4198	91899	0	0	95.63
● 16	0	208	0	20308	0	98.99
● 18	0	24004	0	0	11183	31.78
Recall	96.42	98.78	91.85	80.2	69.64	

Accuracy metrics						
Label	● 3	● 9	● 14	● 16	● 18	Overall
IoU	96	97.95	88.16	79.54	27.91	77.91
F-score	97.96	98.96	93.71	88.61	43.64	84.58
Accuracy	99.91	98.09	99.52	99.8	98.87	98.09
P0	1	0.98	1	1	0.99	0.98
Pe	0.96	0.85	0.93	0.98	0.98	0.85
κ	0.9791	0.8696	0.9345	0.885	0.4315	0.8733

TABLE 4.7: Confusion Matrix and accuracy metrics of the regularization (*Ventoux2*).

4.3.3 Vosges

Similar results are also observed in this area (see Figures 4.27, 4.28, 4.29 and 4.30, and Tables 4.8, 4.9, 4.10, 4.11, 4.12, 4.13, 4.14 and 4.15).

The results on the large area *Vosges2* allow to validate the proposed framework. The classification is noisy, leading to many confusions (mean F-score: 76.92%, IoU: 63.03%). After regularization, the results are greatly improved with a nearly perfect match with the forest LC DB (overall accuracy: 96.26%, κ : 0.95, mean F-score: 94.62%, IoU: 90.12%). Precision and recall rates are also improved for each class. Furthermore, small segments that do not exist in the Forest LC DB are obtained and a visually coherent. **The proposed framework is therefore relevant for the retrieval of forest stands but also allows to obtain further information about small isolated stands within a larger one.**

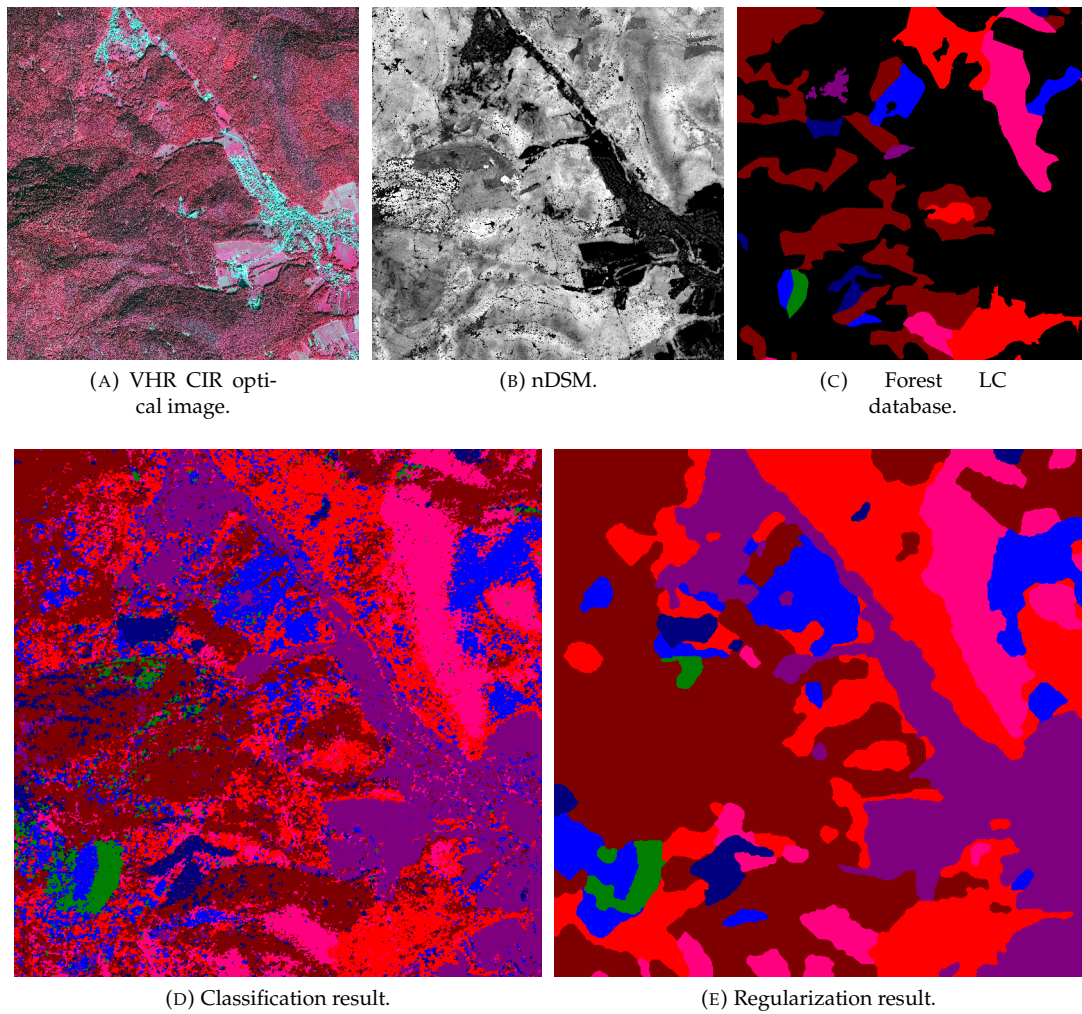


FIGURE 4.27: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges2*, 9 km²).

Confusion matrix								
Label	● 1	● 3	● 8	● 13	● 14	● 15	● 17	Precision
● 1	1984718	134687	98420	267719	8668	37730	50570	76.85
● 3	39022	765057	14290	70204	14742	5076	18861	82.51
● 8	110055	42759	1783163	81597	5141	13510	3845	87.41
● 13	345941	194221	276821	4914574	41583	373732	121745	78.4
● 14	78	1718	128	190	178704	282	270	98.53
● 15	4739	2582	7455	25017	1940	434978	1175	91.02
● 17	586	1700	514	804	15	439	230820	98.27
Recall	79.86	66.95	81.77	91.69	71.26	50.24	54.02	

Accuracy metrics								
Label	● 1	● 3	● 8	● 13	● 14	● 15	● 17	Overall
IoU	64.38	58.63	73.15	73.2	70.51	47.87	53.51	63.03
F-score	78.33	73.92	84.49	84.52	82.7	64.75	69.72	76.92
Accuracy	91.36	95.75	94.85	85.84	99.41	96.27	98.42	80.96
P0	0.91	0.96	0.95	0.86	0.99	0.96	0.98	0.81
Pe	0.68	0.85	0.72	0.5	0.97	0.9	0.95	0.29
κ	0.7294	0.7164	0.8141	0.7163	0.8241	0.6295	0.6898	0.7336

TABLE 4.8: Confusion Matrix and accuracy metrics of the classification (*Vosges2*).

Confusion matrix								
Label	● 1	● 3	● 8	● 13	● 14	● 15	● 17	Precision
● 1	2428277	0	16432	102640	0	66	35097	94.03
● 3	2	898716	239	15216	7345	0	5734	96.92
● 8	26741	36035	1937428	37313	0	2553	0	94.97
● 13	60069	27193	32422	6088861	5109	4358	50605	97.13
● 14	0	370	0	0	181000	0	0	99.8
● 15	1066	0	1237	7285	0	468298	0	97.99
● 17	281	0	0	0	0	0	234597	99.88
Recall	96.5	93.39	97.47	97.4	93.56	98.53	71.95	

Accuracy metrics								
Label	● 1	● 3	● 8	● 13	● 14	● 15	● 17	Overall
IoU	90.92	90.7	92.68	94.68	93.38	96.58	71.89	90.12
F-score	95.25	95.12	96.2	97.27	96.58	98.26	83.65	94.62
Accuracy	98.09	99.28	98.8	97.31	99.9	99.87	99.28	96.26
P0	0.98	0.99	0.99	0.97	1	1	0.99	0.96
Pe	0.68	0.86	0.73	0.5	0.97	0.93	0.96	0.32
κ	0.9405	0.9473	0.9549	0.9461	0.9653	0.9819	0.8329	0.9454

TABLE 4.9: Confusion Matrix and accuracy metrics of the regularization (Vosges2).

On *Vosges3*, similar results are observed, the classification is not relevant (mean F-score: 80.59%, IoU: 68.14%). Indeed, confusions are reported for *Douglas fir* (15, ●) (precision:81.01%). Indeed, young stands of *Douglas fir* (15, ●) generally contains a significant amount of broadleaved (such as *Deciduous oaks* (1, ●)) and have similar aspect with other coniferous (such as *Fir or Spruce* (13, ●)), explaining the observed confusions. After regularization, the results are again greatly improved (overall accuracy: 98.76%, κ : 0.98, mean F-score: 97.29%, IoU: 94.89%). The precisions and recall rates have globally been improved. Only the precision for *Deciduous oaks* (1, ●) has decreased because the regularization process has eroded the border of this class. In order to retrieve this class more precisely, the parameter γ should be decreased. As it has been explained in Section 4.2.4, the borders will be less smooth in this case and some new small segments can appear.

Confusion matrix						
Label	● 1	● 3	● 8	● 13	● 15	Precision
● 1	25221	0	0	248	66	98.77
● 3	806	195007	330	1691	3547	96.83
● 8	0	231	38520	382	0	98.43
● 13	8148	1844	3356	266015	13680	90.78
● 15	11065	35230	19069	136006	859279	81.01
Recall	55.75	83.94	62.86	65.79	98.03	

Accuracy metrics						
Label	● 1	● 3	● 8	● 13	● 15	Overall
IoU	55.37	81.7	62.24	61.67	79.71	68.14
F-score	71.27	89.93	76.73	76.29	88.71	80.59
Accuracy	98.74	97.3	98.56	89.79	86.5	85.45
P0	0.99	0.97	0.99	0.9	0.87	0.85
Pe	0.96	0.77	0.94	0.66	0.51	0.42
κ	0.7068	0.8838	0.7602	0.6999	0.7229	0.7497

TABLE 4.10: Confusion Matrix and accuracy metrics of the classification (Vosges3).

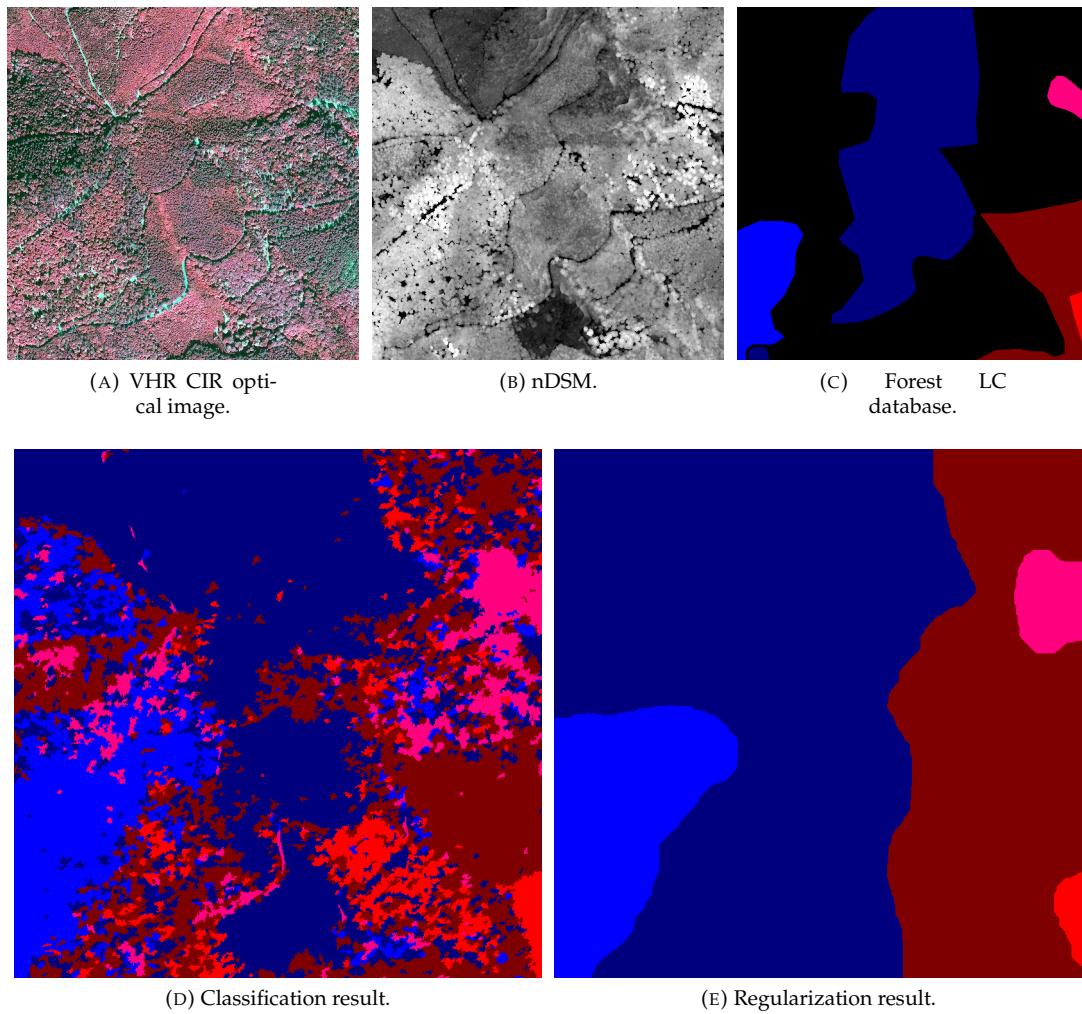


FIGURE 4.28: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges3*, 1 km²).

Confusion matrix						
Label	● 1	● 3	● 8	● 13	● 15	Precision
● 1	24016	0	0	1519	0	94.05
● 3	0	201037	0	0	344	99.83
● 8	0	0	39133	0	0	100
● 13	2881	0	0	290162	0	99.02
● 15	0	14349	0	922	1045378	98.56
Recall	89.29	93.34	100	99.17	99.97	

Accuracy metrics						
Label	● 1	● 3	● 8	● 13	● 15	Overall
IoU	84.52	93.19	100	98.2	98.53	94.89
F-score	91.61	96.47	100	99.09	99.26	97.29
Accuracy	99.73	99.09	100	99.67	99.04	98.76
P0	1	0.99	1	1	0.99	0.99
Pe	0.97	0.78	0.95	0.7	0.55	0.47
κ	0.9147	0.9595	1	0.9889	0.9788	0.9766

TABLE 4.11: Confusion Matrix and accuracy metrics of the regularization (*Vosges3*).

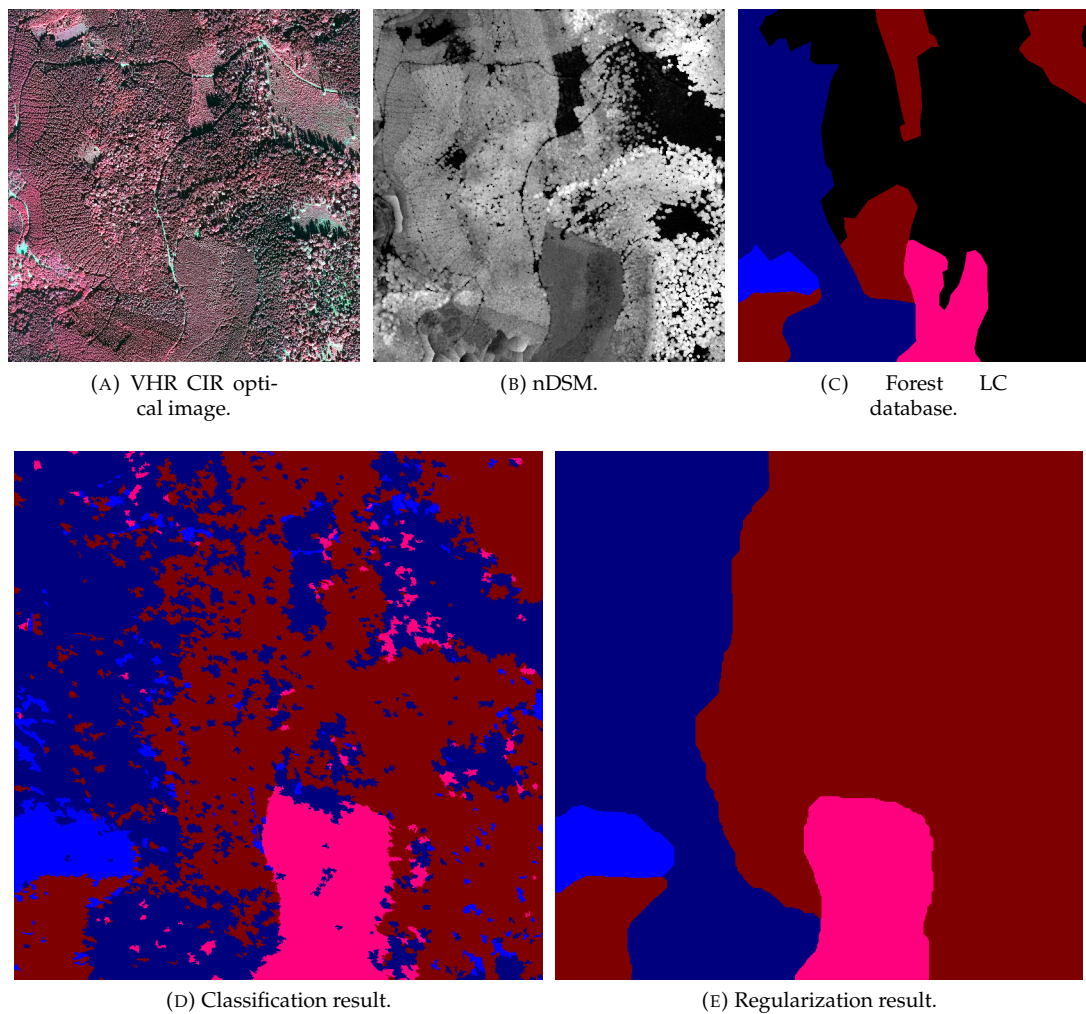


FIGURE 4.29: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges4*, 1 km²).

On *Vosges4*, the classification results are very good (overall accuracy: 88.31%, κ : 0.82, mean F-score: 87%, IoU: 77.68%). It is still noisy (the recall rate for the *Beech* (3, ●) is 62.83%). The precision rates are also good (> 80%). After regularization, all scores are improved (overall accuracy: 97.27%, κ : 0.96, mean F-score: 96.45%, IoU: 93.21%) showing a nearly perfect fit with the Forest LC DB. The recall is > 85% and the precision is > 95% for each class.

Confusion matrix					
Label	● 3	● 8	● 13	● 15	Precision
● 3	89744	0	1135	1189	97.48
● 8	79	234733	376	3432	98.37
● 13	16287	3192	533305	47949	88.78
● 15	36733	12171	95749	791577	84.55
Recall	62.83	93.86	84.58	93.77	

Accuracy metrics					
Label	● 3	● 8	● 13	● 15	Overall
IoU	61.82	92.42	76.41	80.05	77.68
F-score	76.41	96.06	86.62	88.92	87
Accuracy	97.03	98.97	91.18	89.44	88.31
P0	0.97	0.99	0.91	0.89	0.88
Pe	0.88	0.77	0.56	0.5	0.36
κ	0.749	0.9547	0.8005	0.7889	0.8185

TABLE 4.12: Confusion Matrix and accuracy metrics of the classification (*Vosges4*).

Confusion matrix					
Label	● 3	● 8	● 13	● 15	Precision
● 3	90435	0	578	1055	98.23
● 8	0	237335	1075	210	99.46
● 13	2140	966	587349	10278	97.77
● 15	11513	2137	21032	901548	96.3
Recall	86.88	98.71	96.28	98.74	

Accuracy metrics					
Label	● 3	● 8	● 13	● 15	Overall
IoU	85.54	98.18	94.21	95.12	93.27
F-score	92.21	99.08	97.02	97.5	96.45
Accuracy	99.18	99.77	98.07	97.52	97.27
P0	0.99	1	0.98	0.98	0.97
Pe	0.9	0.78	0.56	0.5	0.37
κ	0.9178	0.9895	0.9559	0.9505	0.9567

TABLE 4.13: Confusion Matrix and accuracy metrics of the regularization (*Vosges4*).

On the last areas (*Vosges5*), the classification results are poor (overall accuracy: 74.7%, κ : 0.58, mean F-score: 66.57%, IoU: 50.52%) probably because of the difference in illumination and height of a stand of the same specie. Furthermore, the *Scots pine* (8, ●) is composed of a thin strip, leading to the same problems discussed for *Ventoux2*. After regularization, the results are improved (overall accuracy: 89.76%, κ : 0.83, mean F-score: 85.33%, IoU: 75.84%) but are similar to results we can obtain after classification. Here, two annoying effects are combined:

- a class (namely *Scots pine* (8, ●)) has a significant part represented as a thin strip, that is totally merged into an other class (namely *Beech* (3, ●)),
- the regularization process has eroded the border of a class (namely *Chestnut* (4, ●))

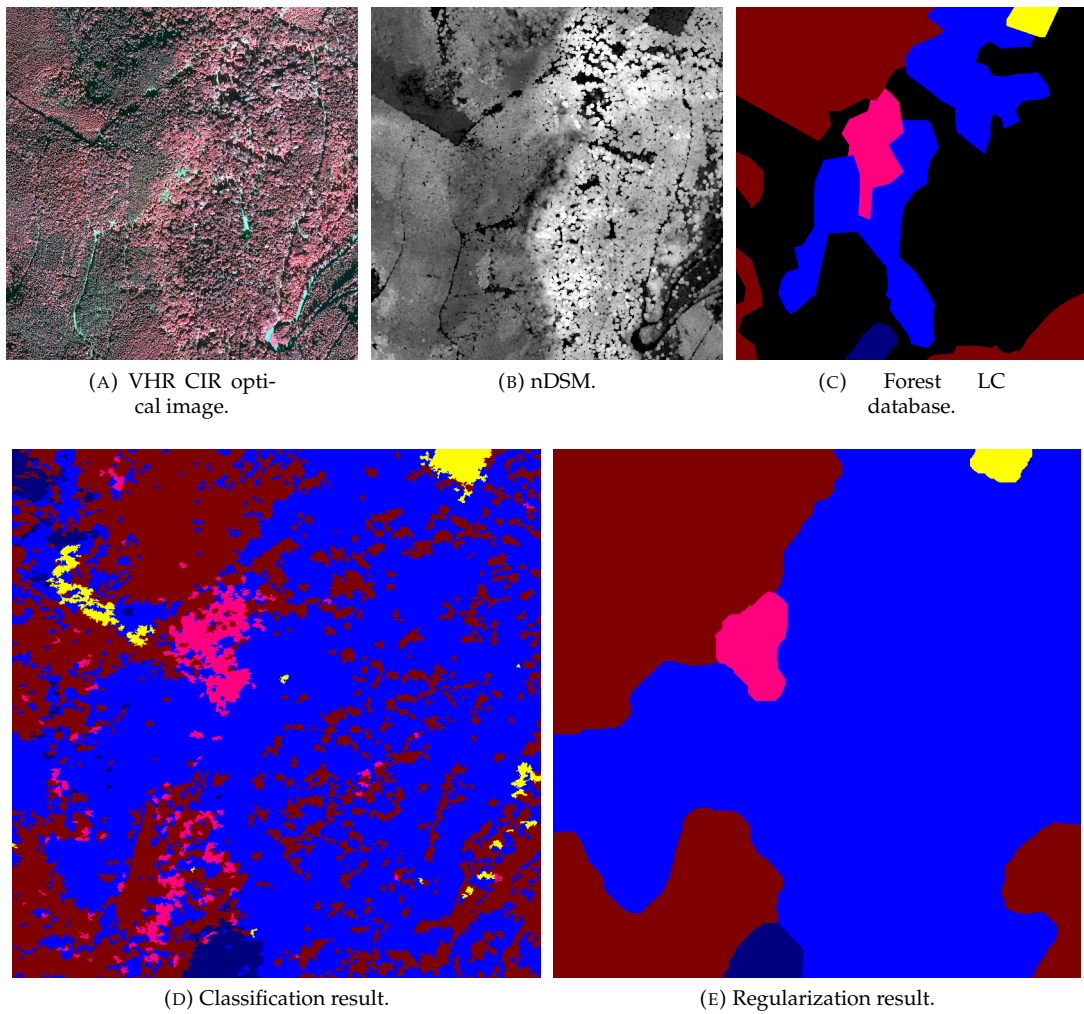


FIGURE 4.30: VHR CIR optical image, rasterized nDSM and forest LC of the selected area from the Vosges (*Vosges5*, 1 km²).

Confusion matrix						
Label	● 3	● 4	● 8	● 13	● 15	Precision
● 3	698791	945	3118	70685	343	90.3
● 4	7874	23049	0	5575	0	63.15
● 8	56145	0	65952	14528	0	48.27
● 13	224194	23458	5184	515485	37883	63.94
● 15	954	0	752	1999	36299	90.74
Recall	70.73	48.57	87.93	84.75	48.71	

Accuracy metrics						
Label	● 3	● 4	● 8	● 13	● 15	Overall
IU	65.73	37.85	45.27	57.34	46.4	50.52
F-score	79.33	54.91	62.33	72.89	63.39	66.57
Accuracy	79.69	97.89	95.55	78.61	97.66	74.7
P0	0.8	0.98	0.96	0.79	0.98	0.75
Pe	0.49	0.95	0.89	0.52	0.94	0.39
κ	0.5993	0.5385	0.6018	0.5579	0.6229	0.5819

TABLE 4.14: Confusion Matrix and accuracy metrics of the classification (*Vosges5*).

Confusion matrix						
Label	● 3	● 4	● 8	● 13	● 15	Precision
● 3	771298	0	0	2584	0	99.67
● 4	13611	22887	0	0	0	62.71
● 8	54417	0	74760	7448	0	54.72
● 13	103070	1021	0	702113	0	87.09
● 15	0	224	0	1234	38546	96.36
Recall	81.84	94.84	100	98.42	100	

Accuracy metrics						
Label	● 3	● 4	● 8	● 13	● 15	Overall
IU	81.62	60.64	54.72	85.89	96.36	75.84
F-score	89.88	75.5	70.73	92.41	98.14	85.33
Accuracy	90.31	99.17	96.55	93.57	99.92	89.76
P0	0.9	0.99	0.97	0.94	1	0.9
Pe	0.5	0.97	0.89	0.51	0.96	0.41
κ	0.8076	0.7509	0.6907	0.8686	0.981	0.8266

TABLE 4.15: Confusion Matrix and accuracy metrics of the regularization (Vosges5).

4.3.4 Validity of the framework

The experiments on multiple areas allow to draw general conclusions on the proposed framework. The most important point is that **the framework can be applied to any forested area**. It will generally produce relevant results from a visual point of view. Three aspects have to be taken into account when quantitatively evaluating the results.

- The employed Forest LC DB is generalized and may contain errors (e.g., clear cuts). The proposed framework proposes **the creation of an optimal training set** to cope with such errors. However, such errors remain in the forest LC DB, which is employed for evaluation. **Thus, the obtained results are biased by the employment of a potentially incorrect Forest LC DB for evaluation.**
- Some stands might be or contain parts represented as a thin strip. **The proposed regularization method is likely to merge such strip with the adjacent class, leading to poor results.** In order to overcome this problem, the value of γ could be decreased. The strip will be kept but some small segment may also appear, giving a precise information of small forest stands within larger stands, but also probably decreasing the global results.
- Sometimes, **the proposed regularization method erodes classes edges, leading to poorer results.** Again, decreasing the γ parameter would allow to retrieve the borders more precisely while have the same effects mentioned above.

4.4 Can forest stands be simply retrieved?

A framework involving classification and global regularization was proposed and successfully employed to retrieve forest stands. This section aims at demonstrating this complex framework was necessary compared to simpler traditional segmentation tools.

A straightforward method to retrieve forest stands could be to simply segment one of the 2 input remote sensing data source. Such segmentation algorithms do not take into account information about species. However, they could allow to retrieve stand borders easily. Furthermore, once the segmentation is performed, one can add semantic information using imperfect classification results. Two algorithms were employed in order to obtain relevant stands only through the segmentation of the data. It helps performing a baseline comparison with the framework presented in Chapter 3.

The first segmentation algorithm is the one proposed in (Guigues et al., 2006). It is a multi-scale hierarchical segmentation algorithm that allows to control the segmentation level through a unique scale parameter μ .

The second segmentation algorithm (called here PFF) employed (Felzenszwalb et al., 2004) is a method for image segmentation based on pairwise region comparison considering the minimum weight edge between two regions in measuring the intensity difference between each pixel of them. 3 parameters need to be tuned in order to obtain relevant segmentation.

- σ is the standard deviation of the Gaussian filter employed to smooth the image as a pre-processing (we followed the authors' recommendation employing $\sigma = 0.8$).
- k is a second parameter that sets an observation scale (the larger k , the larger segments).
- m permits to define the minimum size of a segment.

These experiments have been performed only on area *Vosges1* (presented in Figure 4.31). Since it did not produce relevant results, no further experiment have been carried out on other areas. It is a 1 km² area, the spatial resolution of the VHR optical image is 0.5 m, and the nDSM has been rasterized at the same resolution. From these data, it can be seen that a stand is composed of areas that are not homogeneous in term of reflectance and/or height. Furthermore, one can also note that the difference between two stands in terms of reflectance and/or height might not be so important (two distinct stands can be similar).

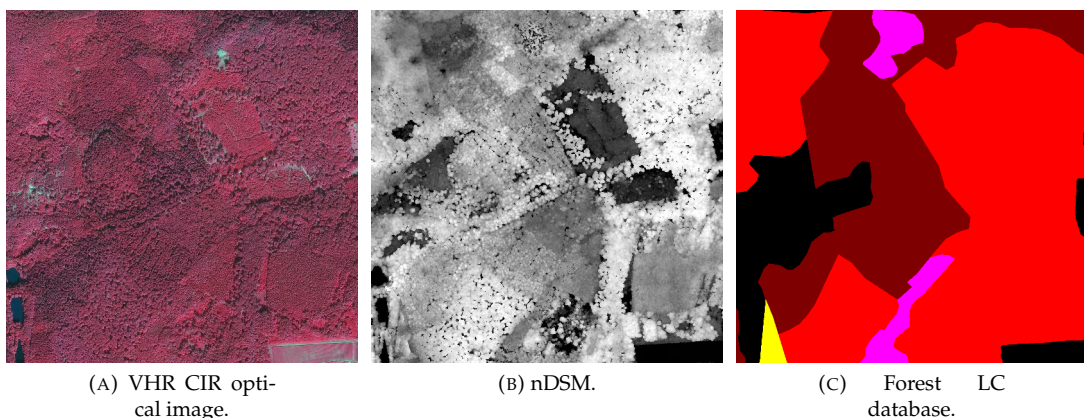


FIGURE 4.31: VHR CIR optical image, rasterized nDSM and forest LC of the proposed area for the direct segmentation tests (1 km²).

4.4.1 Segmentation of remote sensing data

Two segmentation strategies are employed to segment to retrieve the forest stands borders of the forest LC DB:

- The segmentation is applied to the VHR optical images. Thus, the resulting segments will correspond to "stands" that are homogeneous in terms of spectral reflectance. Since the optical images are employed by photo-interpreters in order to derive the forest LC, such segmentation is supposed to produce results similar to the forest LC.
- The segmentation is also applied to the rasterized normalized digital surface model (nDSM) (canopy height without ground relief). Such segmentation would produce "stands" that are homogeneous in term of height.

The results of the segmentation of the VHR optical image using the two segmentation algorithms is presented in Figure 4.32. In both cases, most borders found are not consistent with the forest LC DB, and no metrics can be computed. Visually, the hierarchical segmentation seems to be more relevant than the PFF segmentation. However, the hierarchical segmentation produces small segments due to high variation of illumination in the image, while the PFF segments are all relatively large (thanks to the m parameter). The high spectral variability due to the very high spatial resolution of optical is an hindering factor, that is not correctly handled with a direct segmentation technique

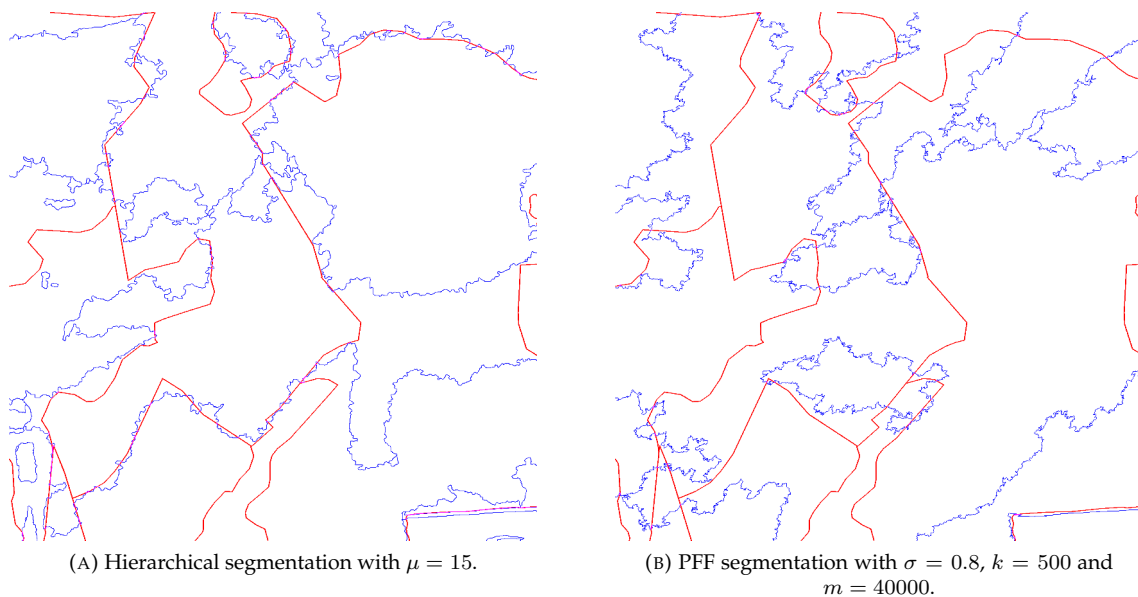


FIGURE 4.32: Result of the segmentation of the VHR optical image for the two segmentation algorithms. Blue lines correspond to the borders of the segments, red lines correspond to the borders of the forest LC.

The results of the segmentation of rasterized nDSM using the two segmentation algorithms is presented in Figure 4.33. Just like the segmentation of the VHR optical image, most of the found borders are not consistent with the forest LC database. Here, the PFF segmentation seems visually to perform better than the hierarchical segmentation since the retrieved borders are close to the borders of the Forest LC BD.

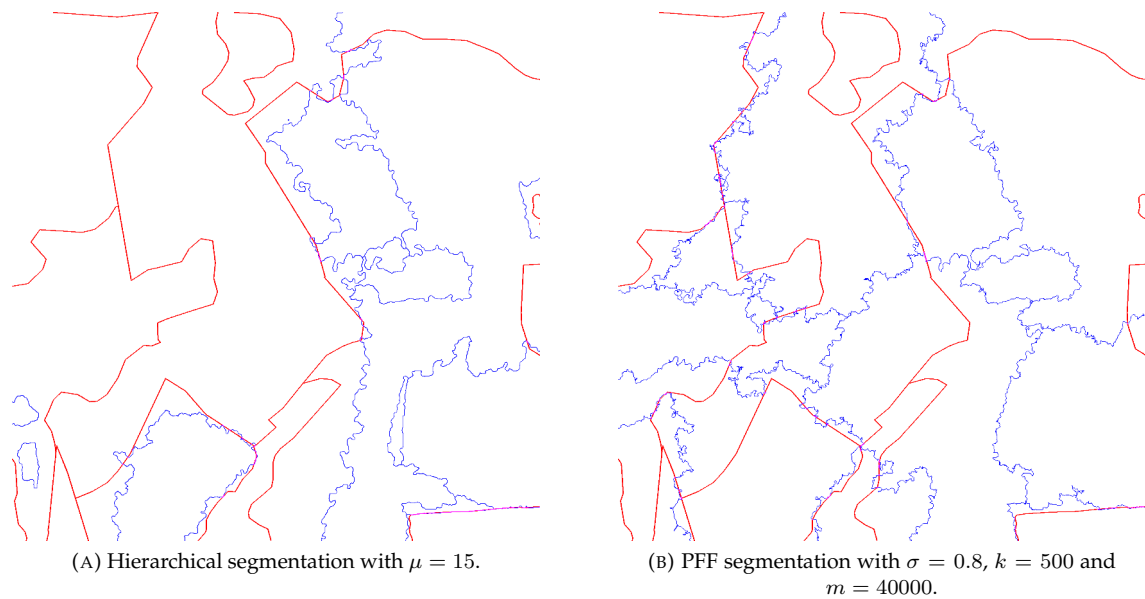


FIGURE 4.33: Result of the segmentation of the nDSM for the two segmentation algorithms. Blue lines correspond to the borders of the segments, red lines correspond to the borders of the forest LC.

Since the segmentation on the VHR optical image and the nDSM does not allow to retrieve directly the borders from the forest LC database, different values of the parameter μ have been tested for the hierarchical segmentation on the VHR optical images (see Figure 4.34) in order to determine if the choice of the parameters could be an issue for direct stand segmentation. It appears that decreasing μ does not allow to obtain the borders of the forest LC. It only leads to an over-segmentation of the image. It can be reminded that such over-segmentation is employed in the proposed framework but not as a relevant segmentation for stand delineation but as an input for object-based classification. It can be concluded that retrieving stands borders is not straightforward. The generalization of the Forest LC DB and the level of detail it offers are not the cause of the wrong delineation obtained by the direct segmentation of input data. The stands are not defined by their borders but by the tree species they are composed of, and segmentation algorithms do not grant access to such information.

The two proposed segmentation (Guigues et al., 2006; Felzenszwalb et al., 2004) algorithms are very efficient for image segmentation tasks but are not adapted to retrieve forest stands. However, they have shown their relevance at producing an interesting over-segmentation (see before).

4.4.2 Classification of the segments

The classification proposed in the framework gives information about the species at the object level. Since the segments extracted above are larger than the small classified objects/superpixels, a majority vote of pixel labels can be applied within each segment. The obtained label map could then be compared with the forest LC. The result of the classification for the area of interest is presented in Figure 4.35, the confusion matrix and other accuracy metrics for this classification are provided in Table C.34.

The results of the majority vote for the segmentation of the VHR optical images are presented in Figure 4.35, the confusion matrices and other metrics are presented in Tables C.35 and C.36.

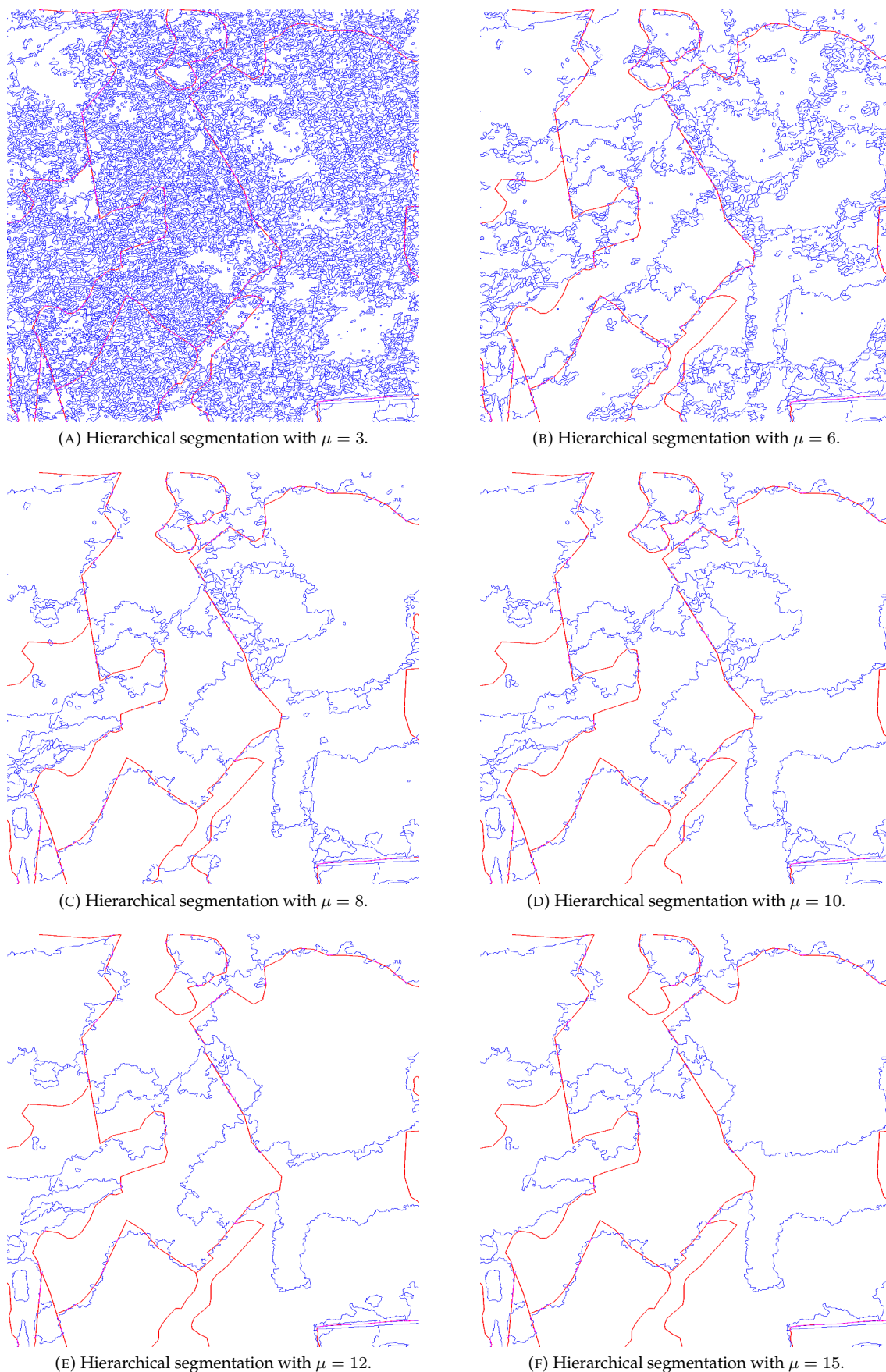


FIGURE 4.34: Results of the segmentation of the VHR optical image using different values of μ for the hierarchical segmentation. Blue lines correspond to the borders of the segments, red lines correspond to the borders of the forest LC.



(A) Forest LC.

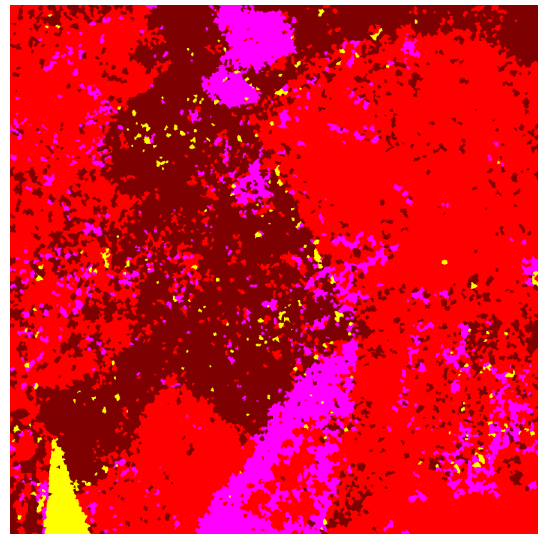
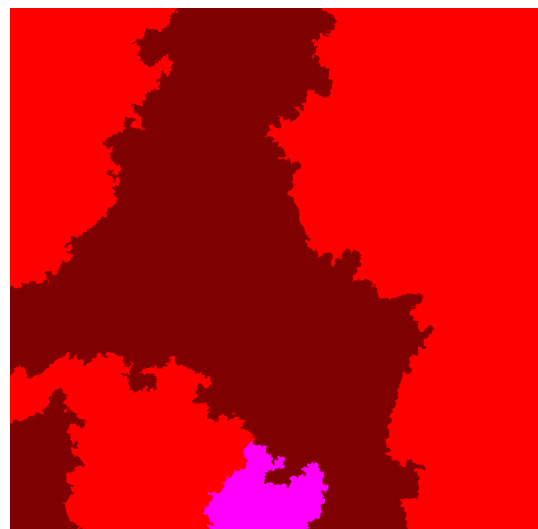
(B) Classification results (overall accuracy: 81.75%, κ : 0.67).(C) Semantic information for the hierarchical segmentation of the VHR optical image with $\mu = 15$.(D) Semantic information for the PFF segmentation of the VHR optical image with $\sigma = 0.8$, $k = 500$ and $m = 40000$.(E) Semantic information for the hierarchical segmentation of the nDSM with $\mu = 15$.(F) Semantic information for the PFF segmentation of the nDSM with $\sigma = 0.8$, $k = 500$ and $m = 40000$.

FIGURE 4.35: Forest LC and classification results.

From a visual inspection, it appears that adding semantic information to the obtained segments does not allow to retrieve a relevant stands since the obtained segments are too generalized. Indeed, some classes are not represented. When they are under-represented in a segment, they are not taken into account. Thus, they are removed from the final results. For the hierarchical segmentation, the overall accuracy increases when adding semantic information through majority vote (81.94%, versus 81.75% for the classification). The κ (0.64) is also revealing a good agreement. This shows the limitation of the two global metrics: the result appears to be good when inspecting them, while it is not. The IoU is more relevant in order to underline the irrelevance of the majority vote on a segmentation (41.1%, versus 56.4% for the classification). The mean F-score (50.68%, versus 70.6% for the classification) also concurs to this conclusion.

The results of the majority vote for the segmentation of the nDSM is presented in Figure 4.35, the confusion matrices and other metrics are provided in Tables C.37 and C.38.

The same results are observed, a majority vote applied to a segmentation equivalent to stands (in term of size) does not allow to retrieve a relevant mapping of forested areas.

From this section, a conclusions can be drawn: the direct segmentation of the data does not allow to directly retrieve relevant forest stands in terms of species. Even with an addition of semantic information from a classification, the results are not sufficient for a good mapping of the forest. **Thus, the proposed regularization framework is completely justified.**

Besides, some metrics are not relevant in order to evaluate the results. Indeed, the overall accuracy or the κ are not sufficient, other metrics, such as intersection over union and F-score, are needed for the correct evaluation of the results.

4.5 Derivation of other outputs

The outputs of the framework can be employed for further investigation in order to extract relevant information about forest. Firstly, the forest stands obtained can be employed for the semi-automatic update or geometric enrichment of the forest LC. Furthermore, the parameter γ allows to refine the Forest LB DB by providing small homogeneous forest stands (see Section 4.2.4). Secondly, the information extracted from the original data can be employed to enrich forest inventory, in case of multi-source inventory.

4.5.1 Semi-automatic update process

The semi-automatic update of the forest LC DB can be performed by the joint analysis of the final stand segmentation and the existing forest LC DB.

A change map can be derived from the stand segmentation results and changes can be prioritized according to their size and shape. Here 3 criteria are employed to characterize a change area:

- The number of pixels,
- The size of the rectangular bounding box (see Figure 4.36b),
- The size of the circular bounding box (see Figure 4.36c).

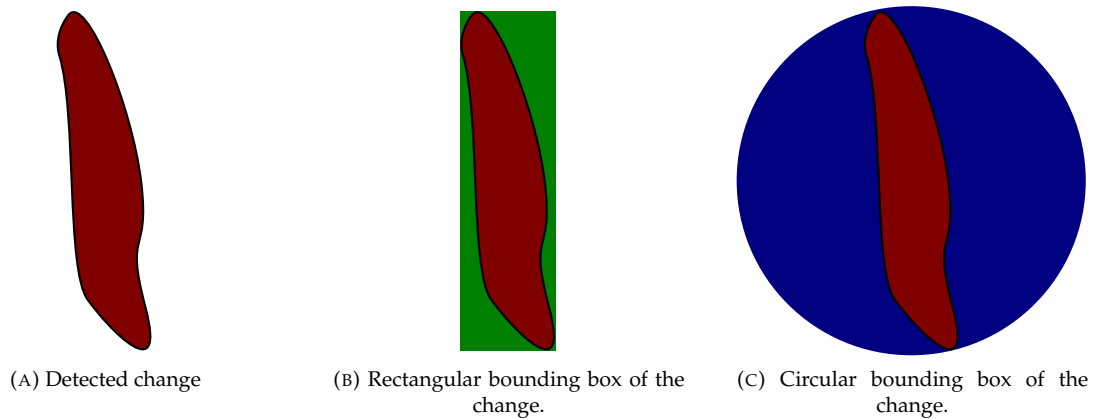


FIGURE 4.36: Bounding boxes of the change detection.

The changes are classified using only thresholds based on the size (in pixels) of the change s , the ratio between the size of the change and the size of the rectangular bounding box r_1 and the ratio between the size of the change and the size of the circular bounding box r_2 . A change is defined as major when one of the arbitrary conditions (4.2) or (4.3) is verified.

$$s \geq 100 \text{ and } r_1 \geq 0.3, \quad (4.2)$$

$$s \geq 100 \text{ and } r_2 \geq 0.2. \quad (4.3)$$

These thresholds have been empirically defined since they produce visually consistent results. Further investigations are needed for a better description of change areas.

An "updated" forest LC can then be created, the major change areas being labeled as *no data* (see Figure 4.37). A difference map is also produced with two kinds of changes:

- The minor changes; they correspond to areas where the borders of the Forest LC DB do not exactly fit the borders obtained by the framework. This type of error is common because of the generalization level of the DB, since the border from the Forest LC DB are mostly straight lines while the proposed framework tends to follow the natural borders of the forest because of the regularization.
- The major changes; they correspond to large patches that differ from the Forest LC DB. Two cases can be differentiated:
 - The obtained results can be wrong.
 - The forest has changed or has been exploited (cut or plantation).

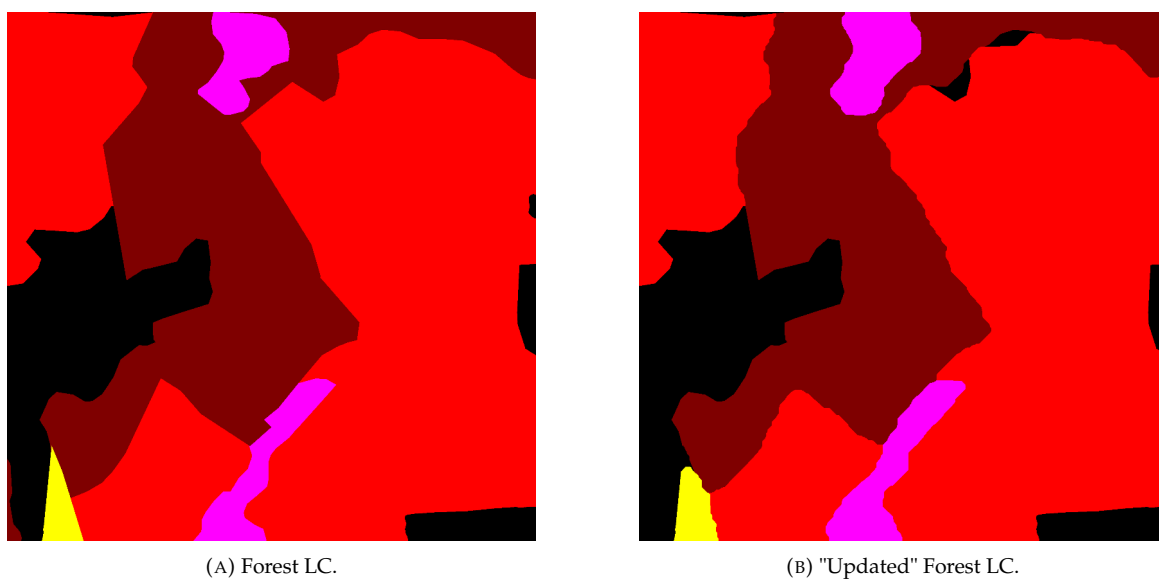


FIGURE 4.37: Automatic update of the Forest LC.

In case of major change, the decision can be taken to keep the original forest LC or to employ the obtained results. A human operator can also focus on the change and correct it manually (see Figure 4.38)

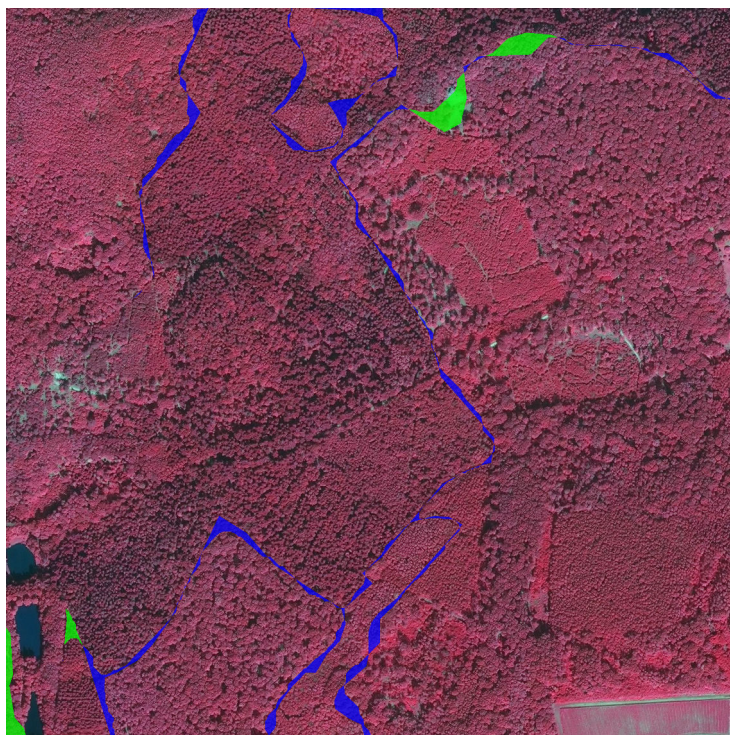


FIGURE 4.38: Detected changes of the Forest LC, green corresponds to major changes, blue corresponds to minor changes.

4.5.2 Data enrichment for inventory

Features derived from Lidar and the final stand segmentation can be jointly used in order to extract additional information. The features can be also used to define a "new" probability map.

Additional information extraction

The extraction of information from the lidar features is performed for each stand obtained from the framework. On each stand, the number of trees can be counted and statistics on the height of the stand can be derived. Such information are easy to extract and might be useful for forest understanding (see Figure 4.39).

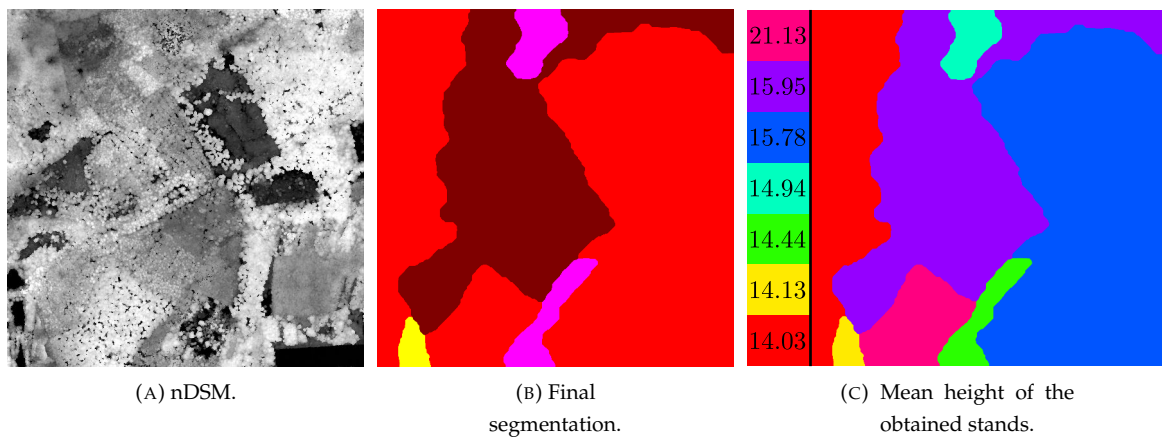


FIGURE 4.39: Extraction of mean stand height combining nDSM obtained from lidar final stand segmentation.

The extracted trees can also be used in order to extract the tree density for each segment (see Figure 4.40).

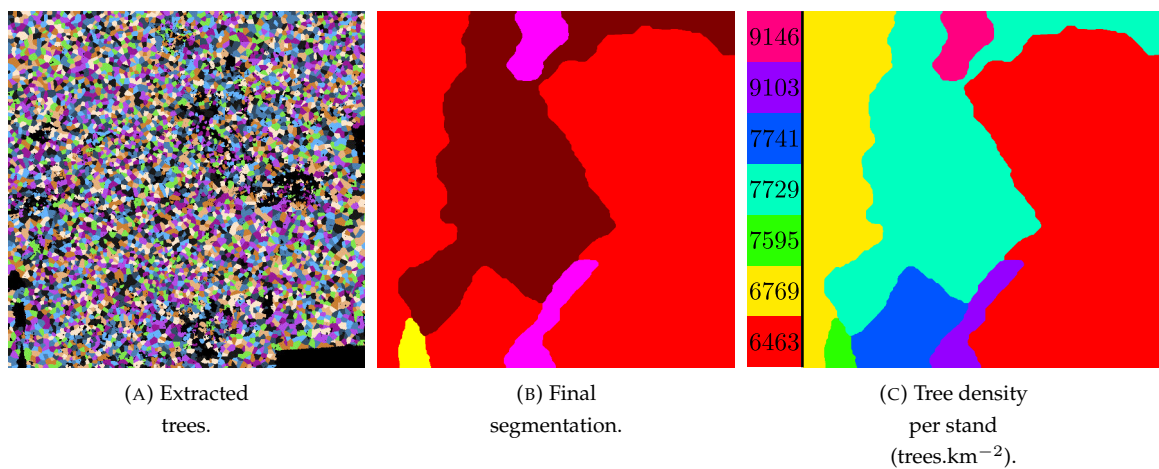


FIGURE 4.40: Extraction of tree stand density trees extracted from lidar data and final stand segmentation.

Segmentation using complementary criteria

It consists in an unsupervised segmentation of the nDSM according to the height distribution. The histogram of height distribution is computed and the different modes are extracted. Each original specie class is then splitted into m subclasses. At the end, $n \times m$ classes are generated from the original n classes. The probabilities of the classification are then redistributed according to the n classes and the m modes at the pixel level. The previous probability map has n components. The new probability map has $n \times m$ components. In the new probability map, the component k has the probability p_k such as:

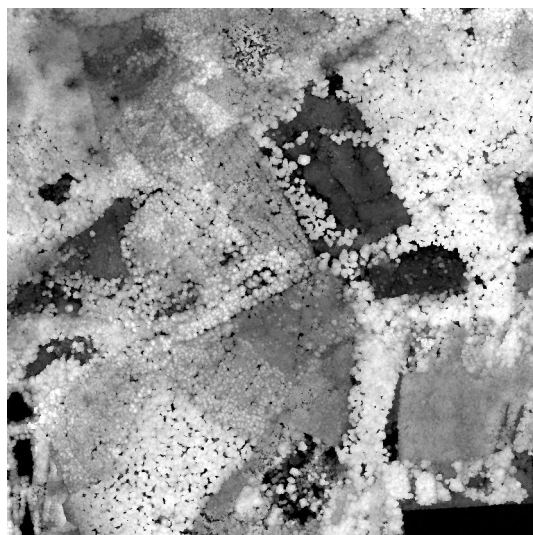
$$k = m_k \times n + n_k, \quad (4.4)$$

$$p_k = p_{n_k}. \quad (4.5)$$

Where n_k is the target class of the pixel in the original probability map, m_k is the mode of the pixel, and p_{n_k} is the probability of the of the class n_k in the original probability map.

Such probability map can then be regularized in order to obtain a map of homogeneous stands both in terms of height and species (see Figure 4.41).

The redistribution of the probabilities according to the modes of the height distribution has two advantages. Firstly, it produces results similar to the regularization without the redistribution (i.e., when considering original classes). Thus, such process does not drastically affect the final segmentation results. Secondly, the obtained results are coherent with the nDSM. It therefore is possible to obtain stands of different maturity, and such information is very interesting for statistical inventory. The main drawback is that "new" classes are created, thus the computation times are increased.



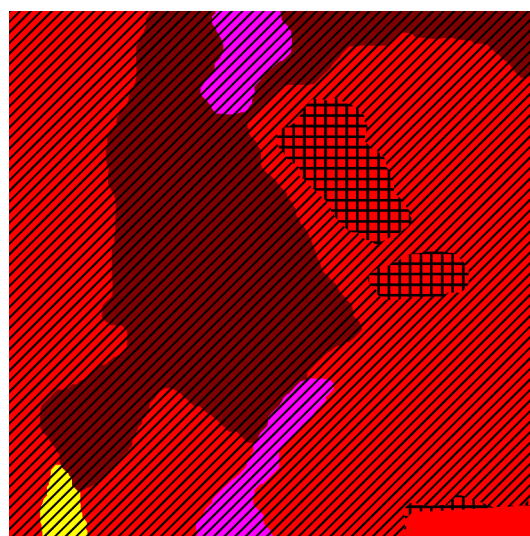
(A) nDSM.



(B) VHR CIR Orthoimage.



(C) Forest LC.



(D) Regularization with redistributed probabilities.

FIGURE 4.41: Redistribution of the probabilities.

4.6 Conclusion and perspectives

Multiple experiments have been conducted in order to validate the choices operated in the framework:

- The extraction of small objects is necessary for efficient classification. The method employed for such extraction has a significant impact on the classification but is less pronounced on the final results (after regularization).
- The feature selection allows to validate the complementarity of both data sources. It also improves the classification results. Furthermore an optimal feature set can be obtained with the feature selection.
- The design of the training set from the forest LC DB is interesting since it improves the classification results.
- The global regularization allows to obtain relevant segments (with respect to the Forest LC DB). Furthermore, tuning the parameter γ gives different segmentation that exhibit levels of details that are interesting for a finest forest analysis. The regularization is further discussed in Chapter 5.

In order to validate the operated choices, the framework has been applied on areas from different regions of France. It comes out that our proposal is relevant, being not specific to a single area.

The framework has been compared to a naive segmentation of the input data, enlightening that the extraction of forest stand is not a straightforward problem.

Finally, other contributions of the proposed framework are proposed. It consists in extracting quantitative informations (such as stands tree density, of stands mean height) from the obtained stands. It fully benefits from the integration of lidar data in the framework.

The proposed framework has been employed using VHR optical images and low density lidar point cloud. It could be interesting to investigate the use of hyperspectral or multispectral imagery since it might provide more spectral information that has shown to be relevant for tree species classification (Dalponte et al., 2013; Dalponte et al., 2012; Clark et al., 2005). Finally, the use of high density lidar could also be interesting. Firstly, with more points, trees could be delineated more precisely and structural shape features at the tree level could be derived. For instance, a convex envelope can be extracted for each tree and the volume of the envelope can be obtained as a new feature (Lin et al., 2016; Li et al., 2013a; Ko et al., 2013).

Regularization: How to obtain smooth relevant stands?

5.1	Local methods	108
5.1.1	Filtering	108
5.1.2	Probabilistic relaxation	109
5.2	Global smoothing	109
5.2.1	Unary/data term	112
5.2.2	Pairwise/prior term	112
5.3	Energy minimization	114
5.4	Constraints integration	115
5.4.1	Size constraint	115
5.4.2	Border constraint	115
5.5	How to efficiently smooth a classification?	116
5.5.1	Local methods	116
5.5.2	Global methods	117
5.5.3	Constraints integration	120
5.6	Conclusion	122

In this chapter, several methods for classification smoothing are proposed and compared. The classification process has been previously presented, and it provides a label map for the areas of interest, accompanied with a class membership probability map, which provides, for each pixel of the image, the posterior class membership for all classes of interest. These are the necessary inputs for the methods that are described below.

However, whereas the classification has been performed at the object level, this label map remains quite noisy. Forest stands often do not clearly appear on this map as it can be seen in Figures 4.19c and 4.19d. Thus, an additional smoothing step is necessary to obtain relevant forest stands. As shown in Section 4.4.2 the simple method consisting in a majority vote within regions from a segmentation do not lead to desired results. Thus, soft labeling smoothing methods (Schindler, 2012) have been considered. This chapter aims at exploring several possible solutions, and at validating the retained ones.

Here, both local and global methods are investigated. For local techniques, majority voting and probabilistic relaxation are selected. For global methods, various energy formulations based on a feature-sensitive Potts model are proposed.

5.1 Local methods

5.1.1 Filtering

An easy way to smooth a classification label map or probability map is to filter it. All the pixels in a $r \times r$ pixels moving window \mathcal{W} are combined in order to generate an output label for the central pixel. The most popular of which is the majority vote filter. First of all, the class probabilities are converted into labels, assuming that the label of pixel \mathbf{x} is the label of the most probable class.

$$\begin{aligned} \mathcal{L}: & \text{ a set of labels,} \\ C(\mathbf{x}) &= \operatorname{argmax}_{c \in \mathcal{L}} P(\mathbf{x}, c). \end{aligned} \tag{5.1}$$

From this label image, the final smoothed result is obtained by selection the majority vote in local neighborhood \mathcal{W} :

$$C_{smooth}(\mathbf{x}) = \operatorname{arg max}_{c_i \in \mathcal{L}} \left[\sum_{\mathbf{u} \in \mathcal{W}} [C(\mathbf{u}) = c_i] \right]. \tag{5.2}$$

The majority filter does not take into account the original class posterior likelihoods (it works directly on labels). This explains the poor observed results, since uncertainty and heterogeneity are not taken into account. However, it is a straightforward method that can give a baseline for the analysis of further results. The main issue is for example, if in a 5×5 neighborhood, 13 pixels have a probability of 51% for class *Douglas fir*, and the 12 other pixels have a 99% probability for class *beech*, the voting will, however, prefer *Douglas fir*. There are variants which give pixels closer to the center more voting power, but typically yield similar results. Other filters have been developed such as the Gaussian Filter, the Bilateral Filter (Paris et al., 2009b; Paris et al., 2009a) and the Edge-Aware Filter (Chen et al., 2007) but they are taking into account object contours, which are not fully relevant for unstructured environment like forests. In addition, filters applied to class probability maps can be an alternative.

5.1.2 Probabilistic relaxation

The probabilistic relaxation aims at homogenizing probabilities of a pixel according to its neighboring pixels. The relaxation is an iterative algorithm in which the probability at each pixel is updated at each iteration in order to make it closer to the probabilities of its neighbors (Gong et al., 1989). It was adopted for simplicity reasons. First, good accuracies have been reported with decent computing time, which is beneficial over large scale (Smeeckaert et al., 2013). Secondly, it offers an alternative to edge-aware/gradient-based techniques that may not be adapted in semantically unstructured environments like forests. The probability $P_k^t(\mathbf{x})$ of class k at a pixel \mathbf{x} at the iteration t is updated by $\delta P_k^t(\mathbf{x})$ which depends on:

- The distance $d_{\mathbf{x},\mathbf{u}}$ between the pixel \mathbf{x} and its neighbor \mathbf{u} . Neighbors are defined as the pixels that are distant of less than r pixels from \mathbf{x} .
- A co-occurrence matrix $T_{k,l}$ defining *a priori* correlation between the probabilities of neighboring pixels. The local co-occurrence matrix has been tuned arbitrarily. It can also be estimated using training pixels (Volpi et al., 2015) if dense training is available. The matrix is expressed as follows:

$$T_{k,l} = \begin{bmatrix} 0.8 & p & \cdots & p \\ p & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & p \\ p & \cdots & p & 0.8 \end{bmatrix}, \text{ with } p = \frac{0.2}{n_c - 1}.$$

The update factor is then defined as:

$$\delta P_k^t(\mathbf{u}) = \sum_{\mathbf{u} \in \mathcal{W}_{\mathbf{x}}} d_{\mathbf{x},\mathbf{u}} \sum_{l=1}^{n_c} T_{k,l}(\mathbf{x}, \mathbf{u}) \times P_l^t(\mathbf{u}). \quad (5.3)$$

In order to keep the probabilities normalized, the update is performed in two steps using the unnormalized probability $Q_k^{t+1}(\mathbf{x})$ of class k at a pixel \mathbf{x} at the iteration $t + 1$:

$$Q_k^{t+1}(\mathbf{x}) = P_k^t(\mathbf{x}) \times (1 + \delta P_k^t(\mathbf{x})), \quad (5.4)$$

$$P_k^{t+1}(\mathbf{x}) = \frac{Q_k^{t+1}(\mathbf{x})}{\sum_{l=1}^{n_c} Q_l^{t+1}(\mathbf{x})}. \quad (5.5)$$

5.2 Global smoothing

Before moving on to the formulation of the global smoothing method, some notations and terminology need to be introduced first. As usual, the pixel values of an image with k features are viewed as samples of a non-parametric function $I : \mathbb{R}^2 \rightarrow \mathbb{R}^k$. The number of pixels is denoted by n , and individual pixel locations are referred to by two-dimensional vectors, denoted with \mathbf{x} . The aim of classification is to assign each image pixel one of l possible class labels c_i , to obtain a the final thematic label map $C : \mathbb{R}^2 \rightarrow \{c_1 \dots c_l\}$.

Finding the thematic map \hat{C} with the highest probability consist in searching the labeling which maximizes the likelihood $P(C|I) \sim P(I|C)P(C)$. This is to say,

$$\begin{aligned}\hat{C} &= \arg \max_C P(C|I) \\ \text{with,} \\ P(C|I) &= \prod_{\mathbf{x}} P(C(\mathbf{x})|I(\mathbf{x})) \\ &\sim \prod_{\mathbf{x}} P(I(\mathbf{x})|C(\mathbf{x})) \prod_{\mathbf{x}} P(C(\mathbf{x})).\end{aligned}\tag{5.6}$$

It respectively corresponds to the minimization of its negative log-likelihood,

$$\begin{aligned}\hat{C} &= \arg \min_C -\log P(C|I) \\ \text{with,} \\ -\log P(C|I) &= -\log P(I|C) - \log P(C) + \text{const.} \\ &= \sum_{\mathbf{x}} -\log P(I(\mathbf{x})|C(\mathbf{x})) + \sum_{\mathbf{x}} -\log P(C(\mathbf{x})) + \text{const.}\end{aligned}\tag{5.7}$$

Thus, minimizing the log-likelihood amounts at finding the minimum of an energy $E(I, C)$ defined as:

$$\begin{aligned}\hat{C} &= \arg \min_C E(I, C) \\ \text{with,} \\ E(I, C) &= E_{\text{data}}(I, C) + E_{\text{smooth}}(I, C).\end{aligned}\tag{5.8}$$

The energy consists of two parts:

- a "data term" which describes how likely a certain label is at each pixel given the observed data (this is actually the output of the classification), and decreases as the labeling fits the observed data better.
- a "prior term", which introduces a prior concerning the label configuration, and decreases as the labeling gets smoother.

Without a smoothness prior the second term vanishes and classification decomposes into per-pixel decisions which can be taken individually ($P(C|I) \sim P(I|C) P(C)$). That is to say, maximizing the probability $P(C|I)$ amounts to simply assign each pixel the most probable class.

If smoothness is included, the labels at different locations \mathbf{x} are no longer independent, but form a random field. The energy of a given pixel depends not only on its data $I(\mathbf{x})$, but also on the labels of other pixels in its neighborhood (4-connectivity or 8-connectivity) or composed of cliques. Since different cliques interact through common pixels, they can no longer be treated independently. The smoothness can be integrated with values chosen arbitrary or that depends on the features corresponding to the pixel. **Thus, it is possible to add contextual information from the features.** For Markovian

Random Fields, it is assumed that the label of each pixel only depends of its neighborhood, i.e.,

$$\begin{aligned} P(C(\mathbf{x})) &= \prod_{\mathbf{u} \in \mathcal{N}(\mathbf{x})} P(C(\mathbf{x})|C(\mathbf{u})) \\ &= \prod_{\mathbf{u} \in I} P(C(\mathbf{x})|C(\mathbf{u})) \end{aligned} \quad (5.9)$$

where $\mathcal{N}(\mathbf{x})$ is a neighborhood of \mathbf{x} (not especially the 4-connexity or 8-connexity).

In general, finding the labeling that globally minimizes $E(I, C)$ is intractable since there is no factorization into smaller problems one would, at least conceptually, have to check all l^n possible labelings.

For random fields with only pairwise cliques (called first-order random fields), efficient approximation methods can find good minima. Such random fields are often represented as graphs: every pixel corresponds to a node with an associated unary potential (corresponding to the data term), and each neighbor pair corresponds to an edge linking the corresponding node pair, with an associated pairwise potential (corresponding to the prior term).

Over the entire resulting first order random field, the maximization of the posterior probability leads to a smoother result. This can be done by finding the minimum of the energy ($\arg \min_C (E(I, C, A))$) defined for an image I as follows:

$$\begin{aligned} E(I, C, A) &= \sum_{\mathbf{u} \in I} E_{\text{data}}(\mathbf{u}, P(C(\mathbf{u}))) + \\ &\gamma \sum_{\mathbf{u} \in I, \mathbf{v} \in \mathcal{N}_{\mathbf{u}}} E_{\text{pairwise}}(\mathbf{u}, \mathbf{v}, C(\mathbf{u}), C(\mathbf{v}), A(\mathbf{u}), A(\mathbf{v})), \end{aligned} \quad (5.10)$$

where $A(\mathbf{u})$ is a vector of values of the features at pixel \mathbf{u} to be selected so as to constrain the problem according to a given criterion (height for instance). $\mathcal{N}_{\mathbf{u}}$ is the 8-connexity neighborhood of the pixel \mathbf{u} (only the 8-connected neighborhood has been investigated). When $\gamma = 0$, the pairwise prior term has no effect in the energy formulation; the most probable class is attributed to the pixel, leading to the same result as the classification output. When $\gamma > 0$, the resulting label map becomes more homogeneous, and, according to $A(\mathbf{u})$ and $A(\mathbf{v})$, the borders of the segments/stands are smoother. However, if γ is too high, the small areas are bound to be merged into larger ones, removing part of the useful information provided by the classification step. The automatic tuning of the parameter γ has been addressed for instance in (Moser et al., 2013) but is not adopted here. There is a range of γ values that produce relevant results. This parameter permits to control the level of detail and its value should be chosen regarding the expected results. Discussions about the choice of γ are detailed in Section 5.5.2.

Despite connections limited to local neighbors, the optimization propagates information over large distances (Schindler, 2012). The problem is NP-hard, but efficient approximate optimization algorithms exist (Boykov et al., 2001; Kolmogorov, 2006; Felzenszwalb et al., 2006).

Here, two formulations of E_{data} (unary term) and four formulations of E_{pairwise} (pairwise prior term) are investigated.

5.2.1 Unary/data term

The data term E_{data} expresses how the integration of the classification results is performed in the energy formulation. It is expressed according to the posterior class probabilities estimated by the classifier. This term decreases according to $P(\mathbf{u})$ (i.e., the higher the posterior probability, the smaller the energy).

- A widely used formulation for the unary term is the log-inverse formulation using the natural logarithm. It corresponds to the information content in Information Theory. It is derived directly from the log-likelihood (i.e., $-\log P(C|I)$). It is formulated as follows:

$$E_{\text{data}} = -\log(P(\mathbf{u})). \quad (5.11)$$

It highly penalizes the low-probability classes but can increase the complexity with potential infinite values.

- An other simple formulation for the unary term is the linear formulation,

$$E_{\text{data}} = 1 - P(\mathbf{u}). \quad (5.12)$$

It penalizes low probabilities less than the log-inverse formulation, but has the advantage of having values lying in $[0, 1]$.

5.2.2 Pairwise/prior term

The pairwise prior term integrates some *a priori* constraints on the neighborhood, weighting the relationship between two neighboring pixels. This weight is expressed according to contextual information from the concerned pixels such as the class assignment of two neighboring pixels. Furthermore, other information, such as features values, can also be taken into account.

In this work, the prior energy term has a value depending on the class of neighboring pixels. In the four proposed formulations, two neighboring pixels pay no penalty if they are assigned to the same class. First, two basic and popular priors, the Potts model and the contrast-sensitive Potts model (called here z-Potts model), are investigated.

- In the Potts model, two neighboring pixels pay the same penalty if they are assigned to different labels. Thus, the prior for the Potts model is:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 1. \end{aligned} \quad (5.13)$$

- In the z-Potts model, the penalty for a label change depends on the height gradient between two neighboring pixels. The z-Potts model is a standard contrast-sensitive Potts model applied to the height estimated from the lidar point cloud. Here, since we are dealing with forest stands that are likely to exhibit distinct heights, the gradient of the height map is computed for each of the four directions separately. The maximum M_g over the whole image in the four directions is used to compute the

final pairwise energy. A linear function has been used: the penalty is highest when the gradient is 0, and decreases until the gradient reaches its maximum value. The prior term of the *z-Potts model* is defined as:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 1 - \frac{|z(\mathbf{u}) - z(\mathbf{v})|}{M_g}, \end{aligned} \quad (5.14)$$

where $z(\mathbf{u})$ is the height of pixel \mathbf{u} and $z(\mathbf{v})$ the height of pixel \mathbf{v} .

- An other investigated pairwise energy is formulated according to the values of the features of the neighboring pixels. It is called here *Exponential-feature model*. The pairwise energy is computed with respect to a pool of n features. When the features have close values, the penalty is high and decreases when the features tends to be very different. The pairwise energy in this case is expressed as follows:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= \frac{1}{n} \sum_{i=1}^n \exp(-\lambda_i |A_i(\mathbf{u}) - A_i(\mathbf{v})|), \end{aligned} \quad (5.15)$$

where $A_i(\mathbf{u})$ is the value of the i^{th} feature at the pixel \mathbf{u} and $\lambda_i \in \mathbb{R}^{+*}$ the importance given to feature i in the regularization. To compute such energy, the features need to be first normalized (i.e., zero mean, unit standard deviation) in order to ensure that they all have the same range.

- The last formulation investigated is also formulated according to the values of the features of the neighboring pixels. It is called here *Distance-feature model*. The pairwise energy is still computed with respect to a pool of n features. In this case, the energy is computed according to the Euclidean distance between the two neighboring pixels in the feature space, the penalty is high when the pixels are close in the feature space and decrease when they get distant. The pairwise energy in this case is expressed as follows:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 1 - \|A(\mathbf{u}); A(\mathbf{v})\|_{n,2}, \end{aligned} \quad (5.16)$$

with:

$$\|A(\mathbf{u}); A(\mathbf{v})\|_{n,2} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n \lambda_i (A_i(\mathbf{u}) - A_i(\mathbf{v}))^2}. \quad (5.17)$$

$\lambda_i \in]0; 1]$ the importance given to feature i in the regularization.

To compute such energy, the features need to be first normalized (i.e., zero mean, unit standard deviation) in order to ensure that they all have the same range. They are then rescaled between 0 and 1 to ensure that $\|A(\mathbf{u}); A(\mathbf{v})\|_{n,2}$ lies in $[0; 1] \forall (\mathbf{u}, \mathbf{v})$.

A high number of features was extracted from available lidar and optical images, the number n of features employed in the last two formulations can then be very important. Since a feature selection has been previously carried out, it is interesting to only use these selected features. Furthermore, they can also be weighted according to their importance (through λ_i). The Random Forest classification process is here very interesting since it natively gives the feature importance. Since the most

important features (20) are almost all equally weighted, it does not bring additional discriminative information for these two models. The λ_i in the *Exponential-feature model* and *Distance-feature model* were therefore set to $1 \forall i$.

5.3 Energy minimization

The random field can be expressed as a graphical model. Thus, the energy minimization can be performed using graph-cut methods. These methods are based on the observation that the binary labeling problem with only two labels (i.e., $C(\mathbf{x}) \in \{0, 1\}$) can be solved to global optimality. To that end the graph of the random field is augmented with a source and a sink node, which represent the two labels and are connected to all pixels by edges representing the unary potentials. A large additive constant on those terms guarantees that the minimum cut of the augmented graph into two unconnected parts leaves each node connected to only the source or the sink. Computing the minimum cut is equivalent to finding the maximum flow from source to sink, for which fast algorithms exist.

The graph-cut algorithm employed here is the Quadratic Pseudo-Boolean Optimization (QPBO). The QPBO is a popular and efficient graph-cut method that efficiently solves energy minimization problems (such as the proposed ones). Thus, the problem is expressed as a graph and the optimal cut is computed over it (Kolmogorov et al., 2007). Furthermore, standard graph-cut methods can only handle simple cases of pairwise/prior term expressed as follows:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= 0, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= f(A(\mathbf{u}), A(\mathbf{v})), \end{aligned} \quad (5.18)$$

where f is a function expressed according to the features values at pixels \mathbf{u} and \mathbf{v} . The QPBO can integrate more constraints since it can solve problems with pairwise/prior term expressed as follows:

$$\begin{aligned} E_{\text{pairwise}}(0 = C(\mathbf{u}) = C(\mathbf{v}) = 0) &= f_1(A(\mathbf{u}), A(\mathbf{v})), \\ E_{\text{pairwise}}(1 = C(\mathbf{u}) = C(\mathbf{v}) = 1) &= f_2(A(\mathbf{u}), A(\mathbf{v})), \\ E_{\text{pairwise}}(0 = C(\mathbf{u}) \neq C(\mathbf{v}) = 1) &= f_3(A(\mathbf{u}), A(\mathbf{v})), \\ E_{\text{pairwise}}(1 = C(\mathbf{u}) \neq C(\mathbf{v}) = 0) &= f_4(A(\mathbf{u}), A(\mathbf{v})), \end{aligned} \quad (5.19)$$

where f_1, f_2, f_3 and f_4 are functions expressed according to the features values at pixels \mathbf{u} and \mathbf{v} .

With the binary graph cut algorithm as a building block, multi-label problems can be solved approximately using α -expansion moves (Kolmogorov et al., 2004) and can be adopted for QPBO. In this method, each label α is visited in turn and a binary labeling is solved between that label and all others, thus flipping the labels of some pixels to α ; these expansion steps are iterated until convergence. The algorithm directly returns a labeling C of the entire image which corresponds to a minimum of the energy $E(I, C, A)$.

Such method gives a hard assignment (a single label for each pixel) while other methods assign probabilities to pixels (soft assignment) which is also interesting (Landrieu et al., 2016).

5.4 Constraints integration

The proposed energy minimization algorithm (namely QPBO) can solve the problem when the energy is modified in order to add more constraint to the problem. Such investigation have been envisaged and two additional constraints have been tested.

5.4.1 Size constraint

Firstly, the size of a segment can be taken into account in the energy formulation. This is done by setting the pairwise term to an a very big value v_s (ideally $v_s = \infty$) when a pixel belongs to a segment defined as too small by the user (in forested areas, a minimum stand size is usually 0.5 ha, further discussions are proposed in Section 5.5.3). Thus the Equation 5.13 becomes:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= f(\mathbf{u}, \mathbf{v}), \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 1 + f(\mathbf{u}, \mathbf{v}), \end{aligned} \quad (5.20)$$

with

$$\begin{aligned} f(\mathbf{u}, \mathbf{v}) &= 0 \text{ if } \mathbf{u} \text{ and } \mathbf{v} \text{ are not in a small segment,} \\ f(\mathbf{u}, \mathbf{v}) &= v_s \text{ if } \mathbf{u} \text{ or } \mathbf{v} \text{ are in a small segment.} \end{aligned} \quad (5.21)$$

Such constraint can be applied in addition to the other proposed priors.

5.4.2 Border constraint

The second constraint is related to the borders. Indeed, it is possible the set the energy to a specific value in order to ensure that a border will be created. An a priori border needs to be defined. Let $b(\mathbf{u}, \mathbf{v})$ be a binary function that defines if a border between pixel \mathbf{u} and \mathbf{v} is to be set. If $b(\mathbf{u}, \mathbf{v}) = 0$, no borders want to be set and $b(\mathbf{u}, \mathbf{v}) = 1$ means that a border wants to be set. Equation 5.13 becomes:

$$\begin{aligned} E_{\text{pairwise}}(C(\mathbf{u}) = C(\mathbf{v})) &= b(\mathbf{u}, \mathbf{v}) \times v_b, \\ E_{\text{pairwise}}(C(\mathbf{u}) \neq C(\mathbf{v})) &= 0, \end{aligned} \quad (5.22)$$

with v_b an important value (ideally $v_b = \infty$).

Adding constraints increases the computational load and time but might be interesting in order to refine even more the results. However, adding such constraints also leads to some issues. Firstly, even if the minimum size of a forest stand is clearly defined in the specifications of the forest LC database, forcing segment to have a minimum size could suppress some information (e.g. small pure segments) that are relevant in many thematical and ecological application. In practice, such generalization could also be obtained by increasing the γ parameter. Secondly, adding borders means that predetermined relevant borders could be retrieved. However, in practice, such borders can not be straightforward extracted for this stand segmentation problem (as it has been presented in Section 4.4.1 of Chapter 4). Thus, adding these constraint have been considered in a limited way (validations have been performed on synthetic images).

5.5 How to efficiently smooth a classification?

After having presented several soft labeling smoothing methods, let us test them and define the optimal smoothing procedure. The classification result has already been presented in Chapter 4. The classification results are illustrated as a reminder in Figure 5.1.

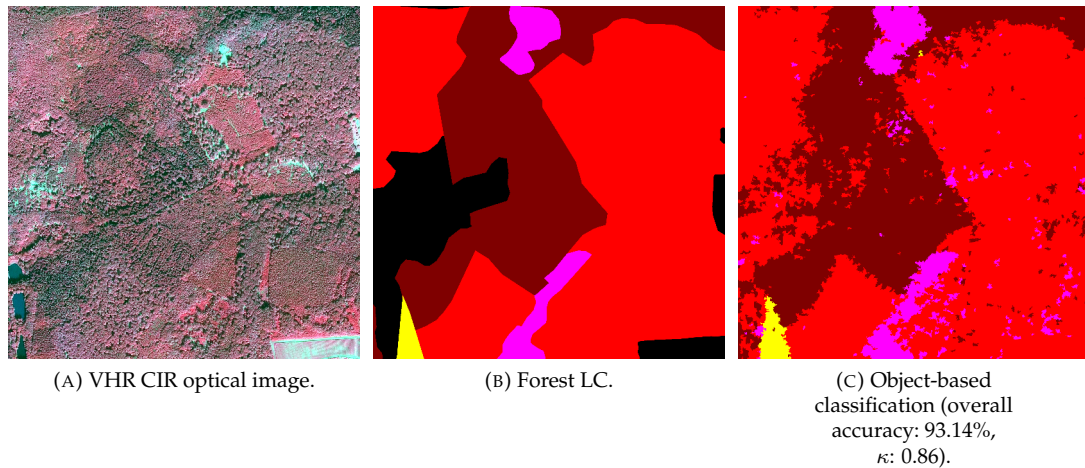


FIGURE 5.1: Result of the classification.

5.5.1 Local methods

Two methods have been investigated for the local smoothing of the classification. They are very easy to implement and have low computation times. Firstly, the use of a majority filter on labels has been employed. Since it does not take into account the probabilities, a probabilistic relaxation has also been tested. For both method, the main problem is to define a relevant neighborhood size. If it is too small, the results will remain noisy, and if it is too important, the results will be over generalized and the computation times will explode.

Majority filter.

The results for the majority filter are presented in Figure 5.2. The filtering method performs the worse with a gain of less than 1% compared to the classification, even with large filters. Furthermore, the larger the filter, the longer the computation times.

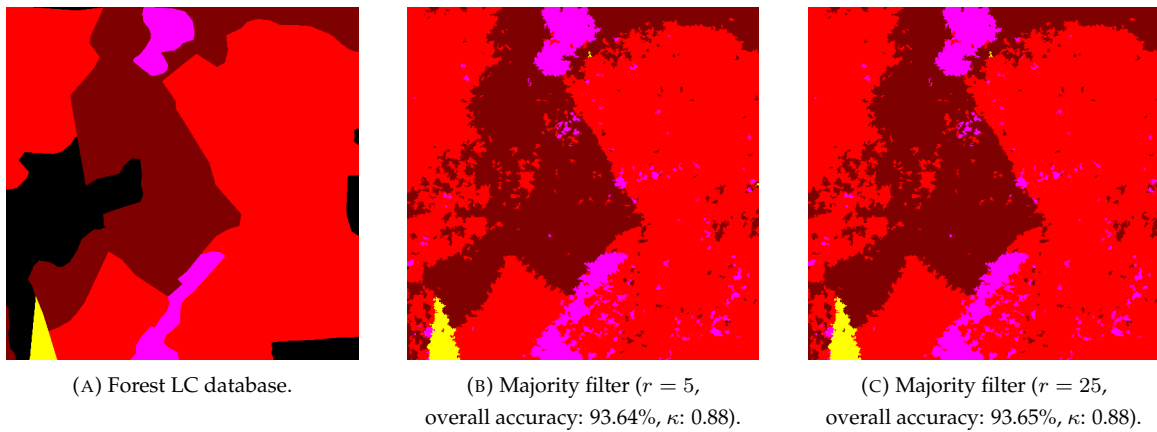


FIGURE 5.2: Smoothing with the majority filters.

Probabilistic relaxation.

The results for the probabilistic relaxation are presented in Figure 5.3. The probabilistic relaxation has also poor results (+5% than the classification) and has also important computation times, since the iterative process has to converge.

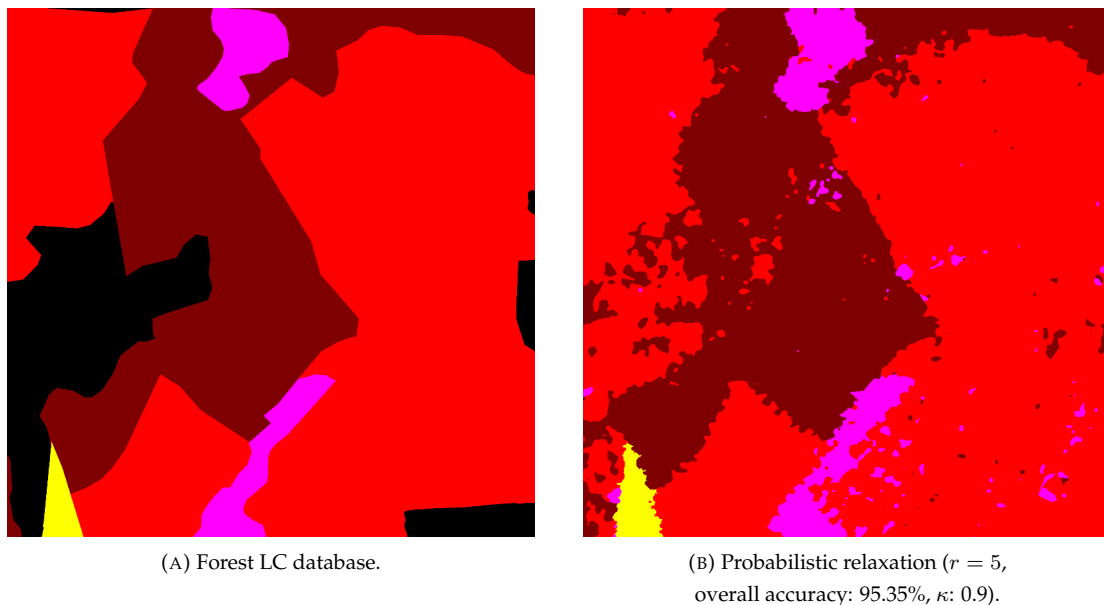


FIGURE 5.3: Smoothing with probabilistic relaxation.

5.5.2 Global methods

In the formulation of the energy, three aspects are taken into account. The two most important are the integration of the information from the classification (unary term) and how the features are integrated into the smoothing process (prior). The last aspect is the integration of other constraints, allowed by the use of the QPBO algorithm.

The global methods produce significantly better results than the local methods with an average gain of 2% in terms of overall accuracy.

Which unary/data term?

The choice of the unary/data term (fit-to-data term) has a major impact on the regularization results (see Figure 5.4). Indeed, the *log-inverse* formulation highly penalizes pixels with low probability, thus, small areas with high probabilities will be kept, even for an important γ . In contrast, the *linear* formulation has a stronger smoothing effect.

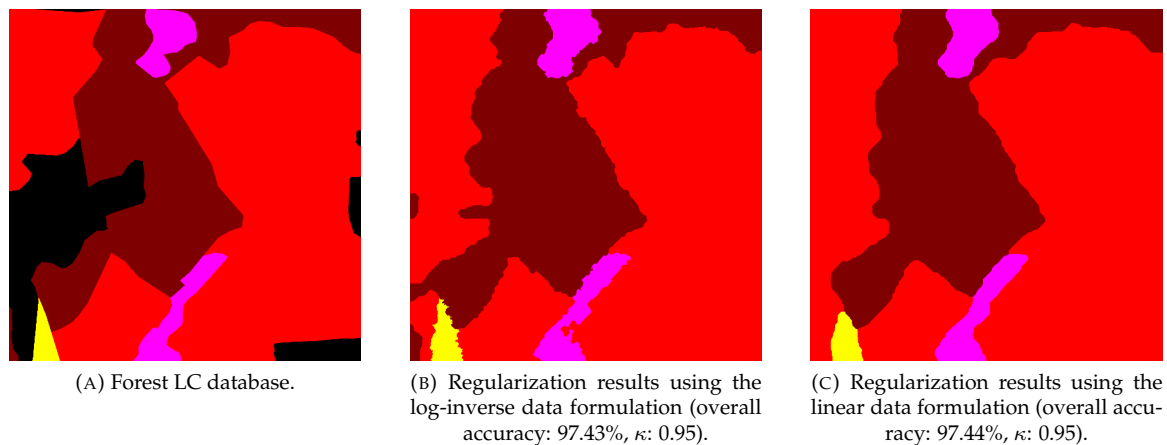


FIGURE 5.4: Impact of the formulation of the unary/data term.

Both are interesting for forest analysis. The *log-inverse* allows to obtain small areas that keep an important class confidence. Such areas are useful for forest inventory or DB enrichment since they give information about large forested areas with small segments of pure species. Conversely, the *linear* formulation produces smooth segments that are more compliant to the present specifications of the forest LC DB.

In terms of accuracy, the two formulation lead to similar results. The *log-inverse* formulation and the *linear* formulation have similar overall accuracy (97.43% and 97.44% respectively) and κ (0.95). The other scores (mean F-score and IoU) show that the *log-inverse* formulation is slightly better than the *linear* formulation (94.18% vs 94.04% for the mean F-score and 89.21% vs 88.97% for the IoU). Since the results are very similar, the *linear* formulation is preferred since it produces smoother segments.

Which prior energy term?

The choice of a prior has a weaker impact on the regularization results (see Figures 5.5 and 5.6).

All proposed models lead to similar results with an overall accuracy greater than 97.3%, a very strong agreement ($\kappa > 0.94$). The mean F-score ($\sim 93\%$) and the IoU ($\sim 88\%$) confirm that few confusions are reported.

After a more thorough inspection of the results, it appears that the *Exponential-feature model* leads to slightly better results (gain of 0.1% in term of overall accuracy, 0.002 for the κ , 0.3% for the mean F-score and 0.5% for the IoU). This improvement is not impressive, showing that a simple *Potts model* is sufficient. However, it also shows that adding feature information can improve the final results.

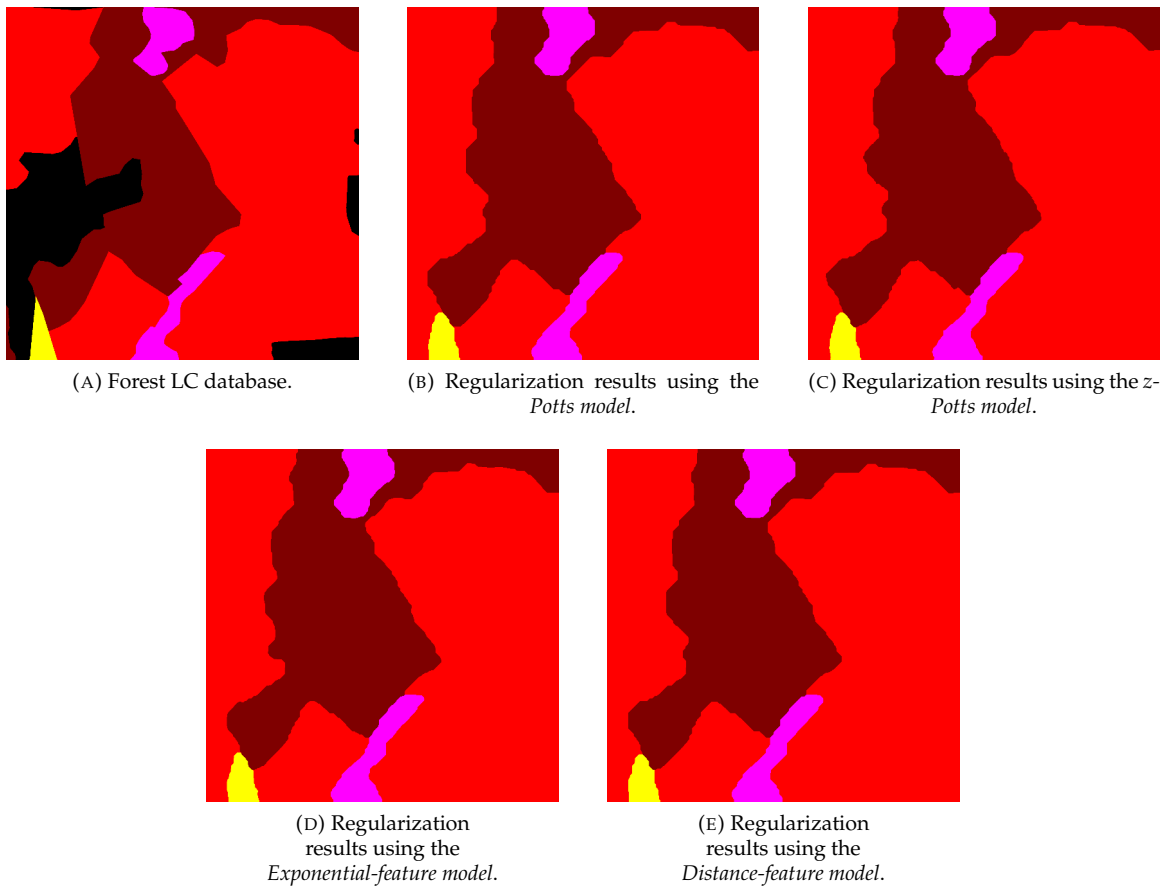


FIGURE 5.5: Impact of the formulation of the pairwise/prior term.

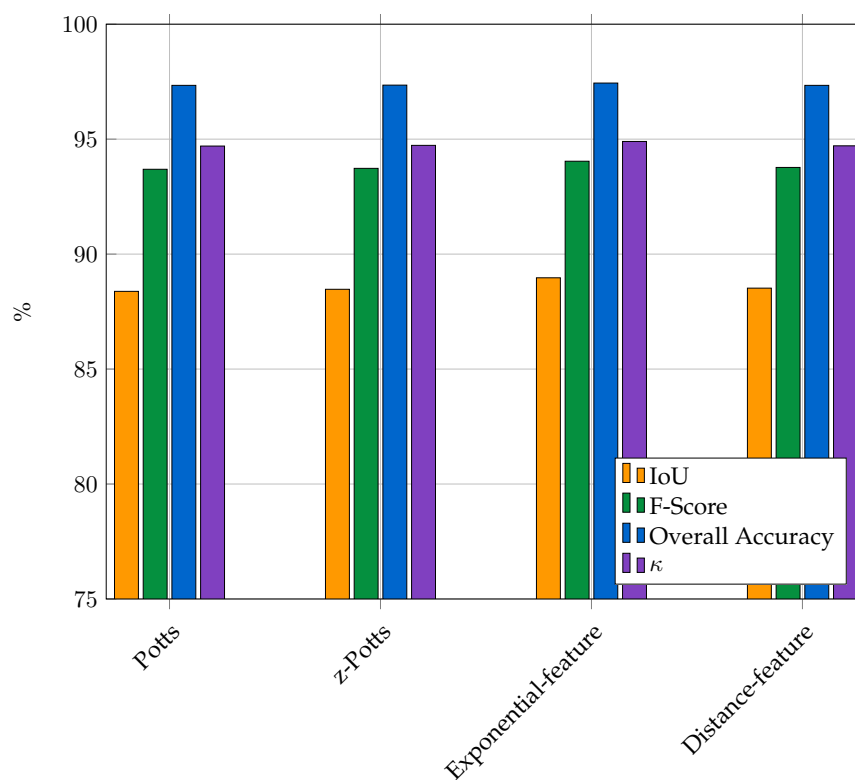


FIGURE 5.6: Regularization accuracy metrics using different formulations for the pairwise/prior term.

5.5.3 Constraints integration

The addition of constraints could be easily carried out with the QPBO algorithm. Constraints about strong borders can be added in order to retrieve them in the final segmentation. Unfortunately, such borders can not be easily extracted (see Section 4.4.1). Thus, in practice, it is not very relevant to add such constraints for this stand segmentation problem since they will not lead to accurate results.

A second constraint can be employed in order to ensure a minimal size of the final segments. However, the minimal size of a segment should be defined. Even if such minimal size is defined by specification of forest stands (0.5 ha), a segment with a size of 0.5 ha+1 pixels will not be considered as a small segment.

In order to validate the formulation of the constraints, experiments have been tried on synthetic data. An object-based probability map has been generated as it is the needed input for the regularization. Here the unary term employed is the linear one and the prior is the *Potts model* (thus, no features are needed in order to compute the energy). Only a qualitative evaluation of the formulation is proposed here. The results using the standard energy formulation (with $\gamma = 1$) is proposed in Figure 5.7.

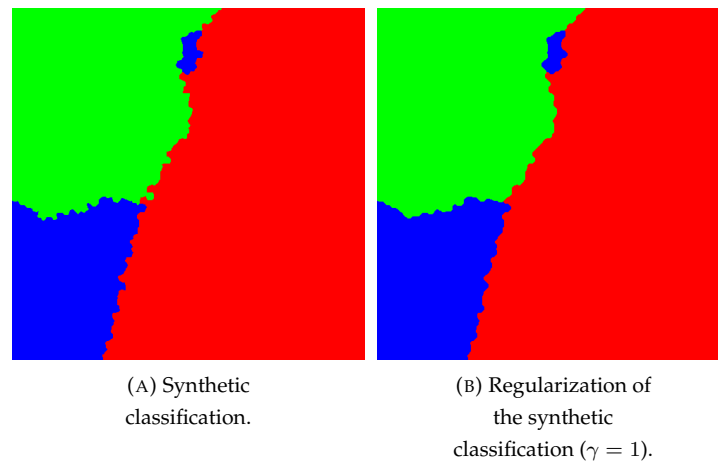


FIGURE 5.7: Results of the method on synthetic data: each color corresponds to a class.

We assume that strong borders (i.e., borders that we want to retrieve in the final result) are provided. Small segments are retrieved from the classification (in the proposed example, there is only one small segment). The constraints that will be applied are presented in Figure 5.8

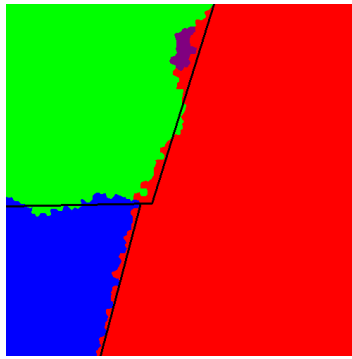
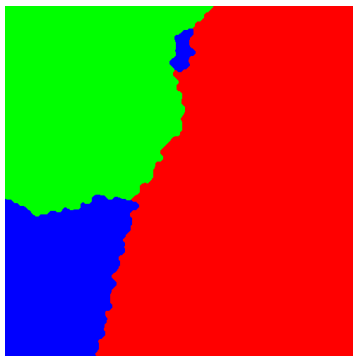
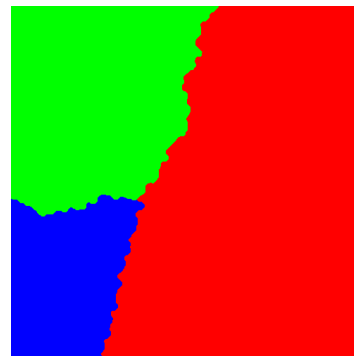


FIGURE 5.8: Constraints desired for the synthetic data set superposed with the classification: the black lines corresponds to borders that we want to retrieve in the final segmentation. The purple area corresponds to the small segment we want to remove.

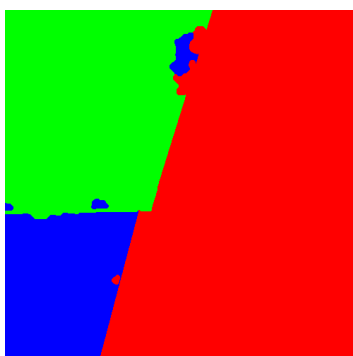
The results obtained when adding constraints are illustrated in Figure 5.9. Three cases have been tested depending on whether we want to retrieve borders, remove small segments, or both. The value of the constraints (v_s and v_b) are set to 100.



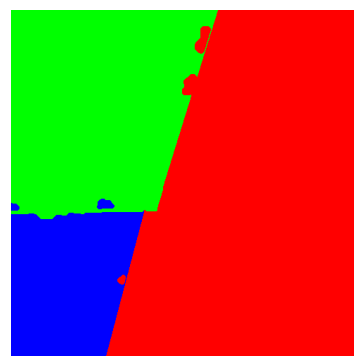
(A) Regularization of the synthetic classification ($\gamma = 1$) without constraints.



(B) Regularization of the synthetic classification ($\gamma = 1$) with size constraint.



(C) Regularization of the synthetic classification ($\gamma = 1$) with border constraint.



(D) Regularization of the synthetic classification ($\gamma = 1$) with both constraint.

FIGURE 5.9: Results of the method on synthetic data when integrating constraint, each color correspond to a class.

The results show that the proposed model for the integration of constraint works well. The size constraint allows to remove the small blue segment, as expected. When adding the border constraint,

they are also more or less retrieved. Indeed, sometimes, it does not work. However, when adding such constraint, new small segments are created, thus, the constraint map must be updated and the regularization should be processed again.

5.6 Conclusion

In this chapter, different methods for the spatial regularization of a classification have been proposed and evaluated.

- Local methods, such as filtering or probabilistic relaxation, are easy to implement but lead to poor results. The resulting segmentation is often not smooth enough.
- Global methods are more adapted to the problematic of forest stand segmentation. The control of the level of smoothing can be performed using the parameter γ . The integration of external data (such as features derived from both data sources) allows to constraint the problem. Other constraints can be added to the model.

Furthermore, in global methods, constraints can be added thanks to the QPBO algorithm. However, such constraints are not straightforward to integrate since they require *a priori* knowledges on the studied area.

- Defining a minimal size for stand segments is possible, but if a segment is slightly more important than the defined threshold, it will no be taken into account.
- Obtaining persistent borders from the remote sensing data is not relevant, mainly because of the shadowing effect and canopy holes. They can also be obtained from other DB such as the roads but they are not relevant borders for forest stand. Indeed, a road can pass through a forest stand. The forest cadastre could also be used, but it appears to be highly fragmented and most of the forest owners do not exploit the forest, leading to very heterogeneous stands.

Data fusion

6.1	Levels of fusion	126
6.1.1	<i>Object-level fusion</i>	126
6.1.2	<i>Classification-level fusion</i>	127
6.1.3	<i>Regularization-level fusion</i>	128
6.2	Designing the best fusion scheme	128
6.3	Optimal fusion scheme	130
6.4	Conclusion	134

In the previous chapters, a framework has been proposed for the extraction of forest stands by joint use of airborne lidar and VHR imagery. The standard tuning brings relevant results. Its modularity allows to investigate where the fusion is mandatory. This chapter focuses on the cooperative use of lidar and VHR optical image. Indeed, both data can be employed at different levels within the framework. So several fusion scenarii are possible. Further experiments are proposed for the evaluation of the fusion process in the proposed framework. The possible levels of fusion are firstly analyzed. It permits the design of a limited number of experiment in order to assess the contribution of the two remote sensing data at each step of the framework. Once such experiments have been processed, it is possible to define several fusion schemes producing results with different quality according to the desired level of detail/computation times.

6.1 Levels of fusion

In this framework, the fusion between lidar and optical images informations can be performed at multiple levels (see Figure 6.1;

- Data employed for the over-segmentation (called here *object-level fusion*),
- Features employed for the classification (called here *classification-level fusion*),
- Features employed for the regularization (called here *regularization-level fusion*).

Since 95 features are available and 6 over-segmentation methods have been proposed, the combinatory is important and thousands of scenarii can be envisaged only at the *object-level fusion*. The *classification-level fusion* and the *regularization-level fusion* can be both performed involving:

- only spectral features,
- only Lidar features,
- both spectral and Lidar features.

6.1.1 Object-level fusion

The over-segmentation can be performed on the lidar data, on the optical data or on a combination of both of them. It is difficult to define an accurate data that combines both lidar and optical image and that is relevant for an over-segmentation. Thus, the over-segmentation is only performed on the VHR RGB optical image or on the nDSM. Besides, as it is shown in Section 4.2.1, the solution for this segmentation step has a little influence on the final results (after regularization).

The Quickshift and the SLIC segmentation algorithms have been developed specifically for RGB images, thus these two methods have been employed here only for the over-segmentation of VHR RGB optical images. The PFF algorithm have also shown good results on the segmentation of RGB images (Felzenszwalb et al., 2004). It is therefore the last segmentation algorithm employed for the over-segmentation of VHR RGB optical image in these experiments.

The tree extraction can be performed either on the VHR optical images or using the 3D information of the point cloud. The tree extraction out of the VHR optical images is quite complex since it needs to take into account enlightened and shadowed tree parts. The tree extraction from 3D point cloud can be easily performed. Thus, only tree extraction out of lidar points cloud was employed.

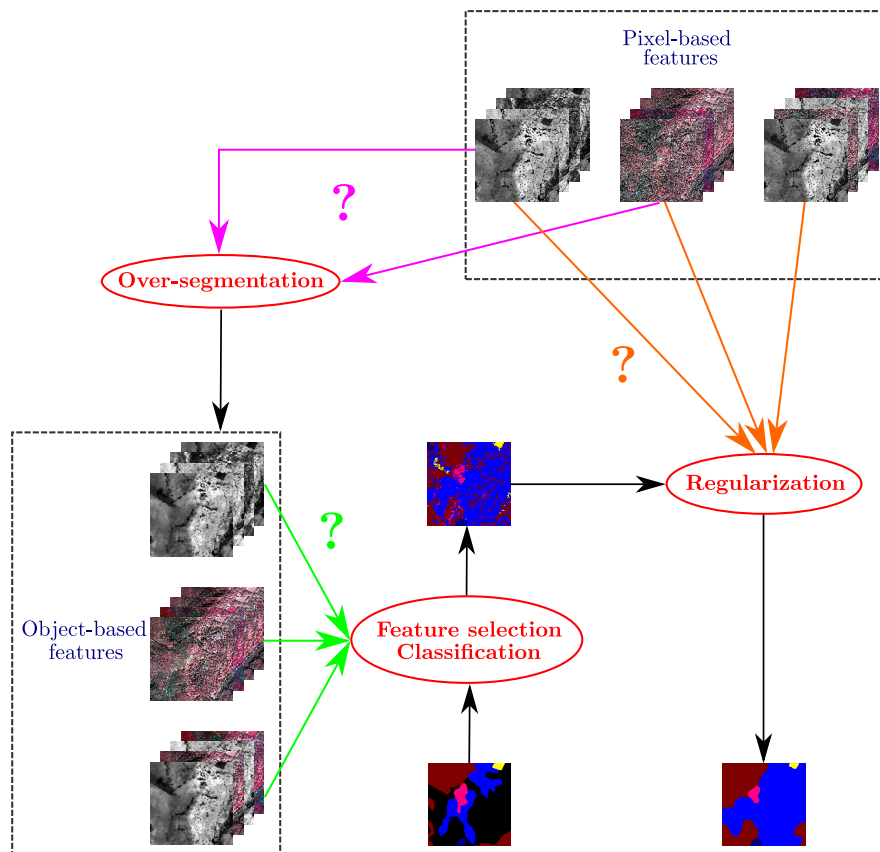


FIGURE 6.1: General flowchart of the proposed method with all the possible fusion schemes.

The watershed and the hierarchical segmentation are employed for the over-segmentation of the nDSM. Since, both are relevant for the segmentation of grayscale images (such as the nDSM).

To sum it up, among the many possible scenarii for the over-segmentation, only 6 are retained:

- Over-segmentation of VHR RGB optical image:
 - PFF
 - Quickshift
 - SLIC
- Over-segmentation of lidar point cloud:
 - Tree extraction
 - Over-segmentation of the nDSM:
 - * Watershed
 - * Hierarchical segmentation

6.1.2 Classification-level fusion

The classification can be performed in three different ways using different features:

- only lidar features,

- only spectral features,
- both lidar and spectral features.

In the classification process, a feature selection is supposed to have been carried out. However, in the case of the classification using only lidar features, no feature selection has been performed since only 25 lidar features are available (furthermore, Lidar features are less significant than spectral features for tree species classification, as it will be presented later in Section 6.2). In the other cases, 20 features are selected.

6.1.3 Regularization-level fusion

In the proposed method, the energy models that take into account the values of the features offer an other level of fusion. Two prior models integrate the features, namely the *Exponential-features model* and the *Distance-features model*. Here, only the *Exponential-features model* has been envisaged, since it has been shown to produce slightly better results.

In this model, one can choose to integrate:

- only Lidar features,
- only spectral features,
- both Lidar and spectral features.

6.2 Designing the best fusion scheme

Regarding the different fusion schemes that can be envisaged, only a limited number of scenarii are investigated. The idea is to define the fusion scheme giving the best results with respect to the Forest LC DB.

The impact of the choice of the data for the over-segmentation has already been investigated in Section 4.2.1. Employing an over-segmentation from VHR optical images (e.g. using PFF, Quickshift or SLIC) or from lidar data (e.g. using hierarchical segmentation on nDSM or tree extraction) produces similar results after final regularization.

In order to evaluate the impact of the data modality choice on the classification and regularization, 5 scenarii have been investigated:

- A *lidar* scenario: here, all lidar features (25) are employed, the feature selection is not carried out:
 - Hierarchical segmentation on the nDSM,
 - Object-based classification (with selection of training samples to cope with Forest LC DB errors) using the 25 lidar features,
 - Regularization using the 25 lidar features.
- A *spectral* scenario:
 - PFF segmentation of the RGB VHR optical image,

- Object-based classification (with selection of training samples to cope with Forest LC DB errors) using a selection of 20 spectral features,
 - Regularization using a selection of 20 spectral features.
- 3 interleaved scenarii at regularization level the same over-segmentation and classification scenario is kept:
 - PFF segmentation of the VHR optical image,
 - Object-based classification (with selection of training samples to cope with Forest LC DB errors) using a selection of 20 features (lidar and spectral),
 - Regularization using
 1. the 25 lidar features,
 2. a selection of 20 spectral features,
 3. a selection of 20 features (lidar and spectral).

Many other scenarii can be proposed. However, the 5 investigated allow to define a critical fusion path that produces the best results in term of retrieval of the forest stands.

The overall accuracies reached for these 5 scenarii are presented in Table 6.1 for the area *Vosges1* but similar results have been observed other areas.

Scenario	Overall accuracy (%)		
	Classification	Regularization	Gain
Full lidar	74.8	92.2	17.4
Full spectral	79.1	95.2	16.1
Regularization lidar	87.6	96.0	8.4
Regularization spectral		96.1	8.5
Regularization lidar + spectral		96.2	8.6

TABLE 6.1: Classification and stand segmentation accuracies for the 5 scenarii investigated (*Vosges1*).

The impact of the choice of the data on the classification step is obvious; the classification using a single data source performs worse than when using both. Furthermore, the spectral information tends to be more relevant than the lidar information. Such results are coherent with the results of the feature selection since the spectral feature are selected for 61% and for 39% for the lidar. Indeed, these two data sources are complementary since the spectral information can efficiently discriminate the tree species while lidar gives additional information about the vertical structure of the forest, that is also helpful for a better discrimination of tree species.

The impact of the choice of the data on the regularization is less significant. We can first note that even in the case of the use of a single modality, the regularization can greatly improve the results starting from a poor classification (+17.4% for lidar and +16.1% for spectral). However, when the classification is already of high quality, the improvement is more or less the same whatever the scenario. The spectral information is a bit more beneficial (gain of 8.5% in terms of overall accuracy) than the lidar information (gain of 8.4% in terms of overall accuracy) at the regularization step. Fusing the two data sources or using only the spectral information in the regularization step does not significantly change the final results.

However, it is important to notice that the energy models that take into account the features values are only slightly more efficient than the *Potts model*. Thus it is not absolutely necessary to integrate features in this step to obtain consistent results.

6.3 Optimal fusion scheme

Regarding the different results, an optimal fusion scheme can be defined. Such scheme is obviously not unique:

- The over-segmentation can be performed on the Lidar data or on the VHR optical images. At this step, the choice of the modality does not impact the final result.
- The classification step is the most crucial. The classification employing only Lidar features leads to poor results, while using VHR optical images produces better results. Here, both Lidar and VHR optical features are needed to obtain the best classification scores.
- When employing a feature sensitive regularization method (i.e., that can take into account features), it appears that using both lidar and VRH optical features leads to the best results. However, using a single data source still produces good results.

Thus, three schemes are proposed for different levels of details and computation times:

- A low cost scheme in terms of data required and computational load, no fusion is operated, the idea is to only use VHR optical images. This scheme allows to extract forest stands with a relatively good accuracy. The forest stands can be retrieved when no Lidar information is available:
 - 75 spectral features can derived,
 - small objects are extracted using the PFF algorithm on the VHR RGB optical images,
 - a selection of 20 spectral features is operated,
 - the classification is performed with the selected features and a optimized training set (to cope with the Forest LC DB errors),
 - the regularization is performed for a linear unary term and the *Potts model* for prior.
- A time cost effective fusion scheme, the idea is to compute the minimum amount of features in order to reduce the feature computation times and to suppress the feature selection:
 - 20 pre-defined features (spectral-based and lidar-based) are computed ¹,
 - small objects are extracted using the PFF algorithm on the VHR RGB optical images,
 - the classification is performed using the 20 features and a optimized training set (to cope with the Forest LC DB errors),
 - the regularization is performed for a linear unary term and the *Potts model* for prior.
- An efficient fusion scheme, the best results are reported with high computation times and load:
 - all the 95 features (spectral-based and Lidar-based) are computed.
 - small object are extracted using the PFF algorithm on the VHR RGB optical images,

¹The selected features are: minimum of the green band, minimum of the blue band, maximum of the green band, maximum of the NIR band, median of the green band, standard deviation the the red band, standard deviation of the blue band, *meanADmed* of the red band, *medADmean* of the blue band, standard deviation of the NDVI, minimum of the DVI, mean of the RVI, density \mathcal{D}_2 , planarity, standard deviation of the height, *medADmed* of the height, 30thpercentile, 50thpercentile, 90thpercentile and mean of the Lidar intensity.

- a selection of 20 features (among the 95) is operated (the features are adapted to the current area),
- the classification is performed using the selected feature and a optimized training set (to cope with the Forest LC DB errors),
- the regularization is performed for a linear unary and the *Exponential-feature model* for prior.

The results of the different schemes are presented in Figures 6.2, 6.3 and 6.4 and in Tables 6.2, 6.3 and 6.4. They have been conducted on a single area (*Vosges1*).

The low cost scheme produces the worse results, especially for the classification (overall accuracy: 79.1%, κ : 0.63, mean F-score: 65.77%, IoU: 51.97%). However, the regularization allows to greatly smooth the classification, leading to final results that are satisfactory (overall accuracy: 95.66%, κ : 0.91, mean F-score: 86.55%, IoU: 77.65%). This scheme has also the advantage of being faster than the standard scheme presented before since fewer feature are derived and processed. This scheme runs in about 2 h 30 for a 1 km² compared to 4 h (1.6 times faster). Here, since the features are only obtained from VHR optical images, several confusions are observed for *Chestnut* and *Robinia*, leading to relatively poor results. However, this scheme is very interesting since it shows that it is possible to obtain a relevant segmentation, according to the forest LC DB, when only using VHR optical images.

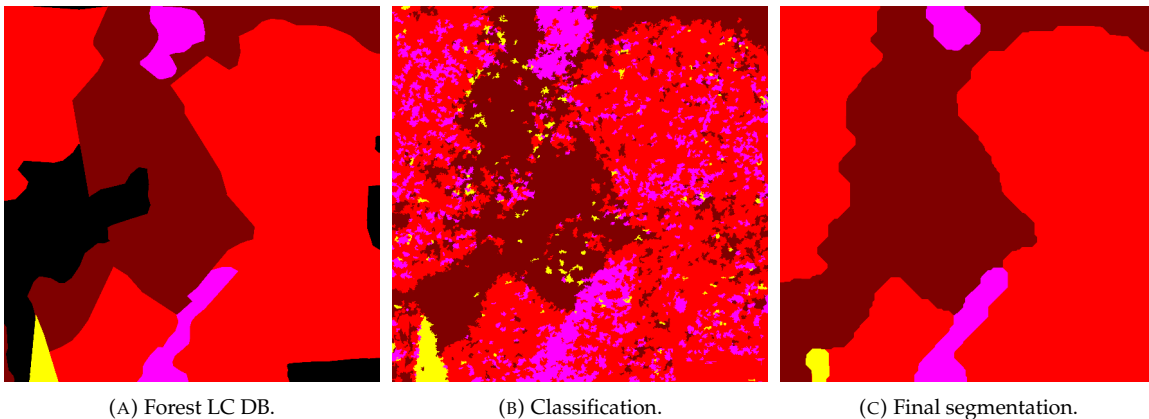


FIGURE 6.2: Result of the low cost scheme (scheme using only VHR optical images) (1 km²).

Accuracy metrics of the classification					
Label	● 1	● 4	● 5	● 13	Overall
IoU	71.71	38.21	24.81	73.17	51.97
F-score	83.52	55.29	39.76	84.51	65.77
Accuracy	81	98.64	88.17	90.39	79.1
P0	0.81	0.99	0.88	0.9	0.79
Pe	0.51	0.97	0.82	0.57	0.43
κ	0.6162	0.5473	0.3519	0.7754	0.6319

Accuracy metrics of the regularization					
Label	● 1	● 4	● 5	● 13	Overall
IoU	95.16	56.93	68.97	89.53	77.65
F-score	97.52	72.55	81.63	94.48	86.55
Accuracy	96.9	99.62	98.36	96.43	95.66
P0	0.97	1	0.98	0.96	0.96
Pe	0.53	0.99	0.91	0.56	0.5
κ	0.9338	0.7238	0.8078	0.9184	0.9136

TABLE 6.2: Accuracy metrics of the low cost scheme (scheme using only VHR optical images).

The time cost effective fusion scheme has better results than the low cost scheme. The classification shows relevant results (overall accuracy: 92.04%, κ : 0.84, mean F-score: 86.17%, IoU: 76.92%) and the regularization result are very close to the Forest LC DB (overall accuracy: 96.24%, κ : 0.92, mean F-score: 90.93%, IoU: 83.93%). It has an other major advantage; only a limited number of features are computed (it takes about 30 minutes to extract them for a 1 km²), and no feature selection is carried out. Thus, only "generic features" (i.e. features that are relevant for different geographical regions) are employed leading to small confusions for *Robinia*. The results are under-optimal but still very good regarding the computation times. Indeed, the whole algorithm runs in about 1 h 30 for a 1 km² area (2.6 times faster than the standard proposed procedure).

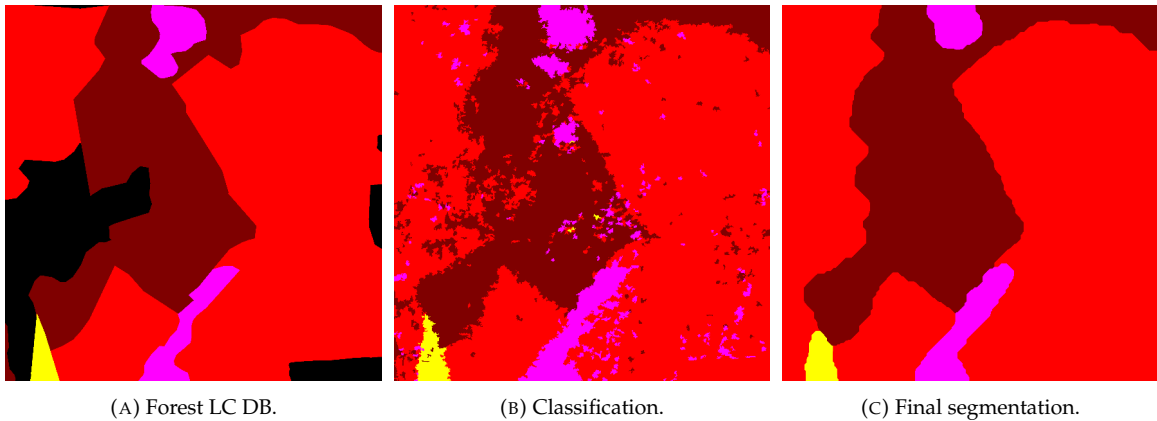


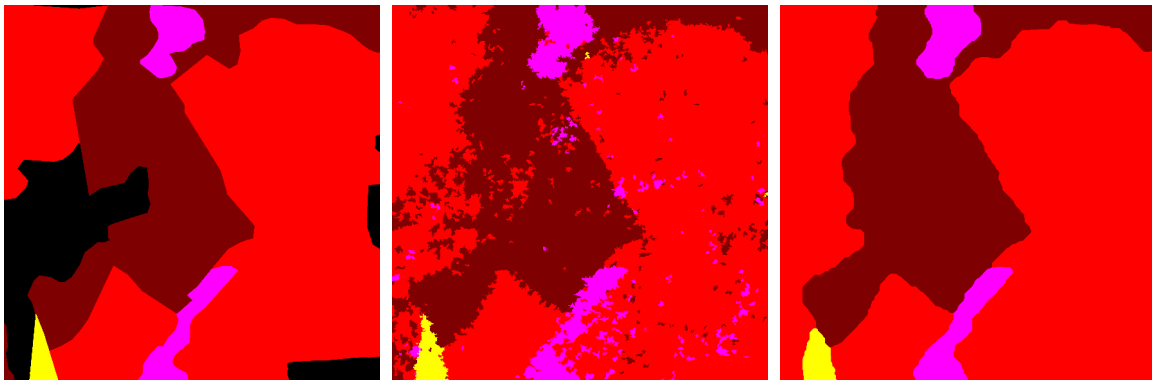
FIGURE 6.3: Result of the time cost effective fusion scheme (1 km²).

Accuracy metrics of the classification					
Label	● 1	● 4	● 5	● 13	Overall
IoU	90.09	81.11	53.24	83.24	76.92
F-score	94.78	89.57	69.49	90.85	86.17
Accuracy	93.46	99.81	96.39	94.43	92.04
P0	0.93	1	0.96	0.94	0.92
Pe	0.53	0.98	0.89	0.58	0.49
κ	0.8602	0.8947	0.6767	0.8685	0.8442

Accuracy metrics of the regularization					
Label	● 1	● 4	● 5	● 13	Overall
IoU	95.64	78.97	69.84	91.28	83.93
F-score	97.77	88.25	82.25	95.44	90.93
Accuracy	97.17	99.79	98.34	97.17	96.24
P0	0.97	1	0.98	0.97	0.96
Pe	0.54	0.98	0.91	0.57	0.5
κ	0.9391	0.8814	0.8138	0.9339	0.9247

TABLE 6.3: Accuracy metrics of the time cost effective fusion scheme.

The efficient fusion scheme produces the best results, close to a perfect fit with the Forest LC DB (overall accuracy: 97.44%, κ : 0.95, mean F-score: 94.04%, IoU: 88.97%). This scheme should be preferred when Lidar and VHR optical images are available and if no time constraint for the production of the results are required. Indeed, relevant features regarding the area of interest are employed in the classification and regularization process leading to optimal results. Running such scheme takes 4 h instead of 1 h 30 for the time cost effective fusion scheme (2.6 times slower).







(A) Forest LC DB.

(B) Classification.

(C) Final segmentation.

FIGURE 6.4: Result of the efficient fusion scheme (1 km²).

Accuracy metrics of the classification					
Label	 1	 4	 5	 13	Overall
IoU	90.63	83.03	58.98	86.52	79.79
F-score	95.08	90.73	74.2	92.77	88.19
Accuracy	93.89	99.83	97.09	95.48	93.14
P0	0.94	1	0.97	0.95	0.93
Pe	0.53	0.98	0.89	0.57	0.49
κ	0.8701	0.9064	0.7271	0.8948	0.8662





Accuracy metrics of the regularization					
Label	 1	 4	 5	 13	Overall
IoU	96.67	81.36	83.89	93.94	88.97
F-score	98.31	89.72	91.24	96.88	94.04
Accuracy	97.87	99.82	99.13	98.06	97.44
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9542	0.8963	0.9079	0.9547	0.949

TABLE 6.4: Accuracy metrics of the efficient fusion scheme.

6.4 Conclusion

In this chapter, the different levels of fusion have been more precisely presented and analyzed. From the different conducted experiments, three schemes appear to be relevant for forest stand delineation.

- When no lidar is available, the single use of VHR optical image is possible in order to obtain quite relevant forest stands.
- For large scale results (and when lidar is available), only a limited number of features need to be computed, leading to generalized results. However, such scheme produces good results with very decent computation times.
- When precise mapping is needed, the employment of all the steps of the proposed framework allows to obtain a very relevant stand segmentation (according to the Forest LC DB).

Conclusions and perspectives

7.1	Conclusion	138
7.1.1	Extraction of forest stands	138
7.1.2	Fusion schemes	140
7.1.3	Quantitative features	141
7.2	Perspectives	141
7.2.1	Relevance of other remote sensing data sources	142
7.2.2	Improvement	142
7.2.3	Application to other land-cover problems	143
7.3	Final outlook	143

7.1 Conclusion

The framework proposed in this thesis allows to draw strong conclusions on forest stand segmentation and fusion of optical spectral images and Lidar point clouds for such a task. Furthermore, Lidar might not be limited to a labeling data source since it can also provide quantitative information and biophysical features about the vertical structure of forest.

An automatic framework, composed of several steps that can be optimized independently has been proposed. It integrates operational constraints, such as errors in the Forest LC DB and draws the best of all the data (namely VHR optical images, lidar and Forest LC DB). The proposed framework has been validated on different areas with various landscapes. Several variants were also tested so as to identify the best "frameworks". The contribution of the thesis are the following:

- the development of a modular and versatile framework with few parameters,
- each step have been justified through multiple experiments,
- special attention is paid to the regularization of the classification, standard formulation for global method have been proposed and additional constraints have been investigated,
- a study of the best cooperative use of lidar and VHR optical images and of their relevance for forest stand retrieval, leading to three possible variant of the pipeline.

7.1.1 Extraction of forest stands

The proposed framework is composed of four main steps that can be divided in sub-steps:

- The feature computation step that aims at the extraction of:
 - Spectral image based features: it mainly corresponds to the computation of statistical features (minimum, maximum, mean ...) using different neighborhood sizes, but also vegetation indices that have shown their relevance in vegetation discrimination.
 - Lidar features at the point level, these features are mainly related to the height and the spatial distribution of the points. They have also shown relevance for classification tasks in forested areas. The features derived at the point level are, like the spectral image based features, extracted according to statistical functions. The point features are then rasterized at the spatial resolution of the optical images.
 - Small objects that have a size and shape similar to trees (trees are the main components of forest). They are needed in order to obtain features at the object level. Such objects are coarsely delineated, since they are only employed for an object based classification that will be refined after. The extraction of objects can be performed directly on the Lidar point cloud (e.g., extraction of trees) or using over-segmentation methods (e.g., segmentation algorithms with adapted parameters or superpixels algorithms) on rasterized Lidar features (mainly the nDSM) or the optical images.
 - Features at the object level. It is a simple averaging of the pixel-based features for all extracted objects but it has proven to increase classification performances.
- A supervised forward feature selection step using the κ of the Random Forest as feature set relevance score. The feature selection is driven by three main objectives. Firstly, it greatly

decreases the computation times (instead of processing the entire feature set, only an optimal subset is needed). Secondly, it allows to avoid the curse of high dimensionality, that decreases the classification performance when the number of features increases. Lastly, since our features are derived from different remote sensing modalities, the feature selection is a way to assess the relevance of each modalities in the classification process.

- The supervised classification step:
 - The training is based on the forest LC DB, which is natively generalized. Such generalization results in borders that do not follow the natural borders of the forest. Furthermore, forest stands that are not 100% pure. A k-means is employed in order to retrieve only the main component of the forest stands that is bound to correspond to the genuine tree specie while keeping a certain level of variability to avoid over-fitting.
 - The classification itself is performed using the Random Forest. This state of the art algorithm has shown good discrimination performances. It can handle important amount of features from different data sources. Furthermore, it generates the posterior probabilities for each target class and a feature importance score is also natively obtained. The posterior probabilities can be employed for subsequent smoothing.
- The regularization step that can be envisaged at 2 different levels:
 - The local level; in this case only, a local neighborhood is considered to smooth the final results. It can run from a simple majority vote to the iterative probability relaxation technique. Such local methods are straightforward to implement but do not lead to relevant results for forest stands. The results remain noisy.
 - The global level; in this case, all the pixels of the image are taken into account. The aim is to minimize an energy composed of two terms, one related to the classification (unary or data term) and one related to the context (prior or pairwise term). Furthermore, information and strong thematical constrains can be added in such model. The choice of the unary term exhibits the strongest influence while the prior term has a weak impact on the final results. The regularization level (i.e., how smooth the results are) is controlled through the parameter γ (the only impacting parameter of the framework). This parameter could be tuned automatically. However, it appears that $\gamma = 10$ produces the most accurate results compared to the forest LC DB regardless of the concerned area. Besides, this parameter can be tuned differently depending on the expected outputs. When $\gamma < 10$, the resulting segmentation will be less smooth and small pure segments will be conserved (the accuracy will decrease but small pure segments will be detected). If $\gamma > 10$, the results will be over-generalized. Such generalization can be interesting for a national mapping, since only the dominant species will be highlighted.

The proposed framework allows to extract homogeneous forest stands that are relevant according to the French Forest LC DB. It has only **few parameters**; the most impacting being the γ of the regularization. **The modularity of the framework is also a great advantage** since it can be fed with different inputs for comparison or improvements. Thus when trying different configurations (i.e., testing different inputs for each step), one can seek for the best fusion scheme.

7.1.2 Fusion schemes

The fusion is operated at different levels in each step of the framework.

- In the feature computation step, the data and algorithm employed for the object extraction correspond to a medium level of fusion and more precisely to a cooperative understanding of the data.
- In the classification step, the fusion is operated at both low and medium level. The feature selection has selected a relevant amount of features among the 95 available (70 spectral-based + 25 Lidar-based). The selected features are employed for the supervised classification and show the complementarity of both data sources.
- In the regularization step, the fusion can be performed at the high and medium level (since the output of the classification and features are jointly employed) when a global model that takes into account features is employed. In such model, the classification output (posterior probabilities) and the features values are both employed in order to obtain a final smooth decision.

Here, the modularity of the framework allows to efficiently assess where the fusion is crucial. We can conclude that there is not a single best fusion scheme;

- In the feature computation step, the object extraction has not an important impact on the final results. It only slightly impacts the classification results. Thus any over-segmentation method can be employed. For better classification results, an over-segmentation based on VRH optical image is recommended.
- In the classification step, both spectral and Lidar features are required. Indeed, even with an unbalanced proportion of spectral features (70 over 95) and Lidar features (25 over 95), 60% of the selected features are spectral-based features and 40% are Lidar-based features. When employing only a single data source in the classification, the accuracy of the results greatly decreases. The fusion is crucial in the classification step. The two remote sensing modalities are here complementary.
- In the regularization step, when not taking into account the features through global methods, the obtained results already report very good accuracies. Slightly better results can be obtained when employing feature values in the energy formulation. Here again, employing both data sources leads to the best results (but is not mandatory).

The proposed framework allows to determine how the fusion should be carried out in order to obtain the best results. It is also possible to define what kind of results can be expected at different levels or when employing only one data source.

- Employing only Lidar-based features is not consistent for forest stand segmentation in this context.
- Employing only image-based features allows to obtain relevant forest stands, however, confusion might be reported.

At the end, three schemes are proposed:

- S1:** A low cost fusion scheme that permits to obtain exploitable results with limited computing times but limited accuracy. This scheme only use the spectral features of the VHR optical images. Thus no fusion is operated but the results are exploitable. It is straightforward to implement. It can not be employed for a precise delineation but would give satisfactory results if no lidar is available.
- S2:** A time cost effective fusion scheme, very satisfactory results are reported for decent computation times:
- Only a subset of relevant features (spectral and Lidar) are computed (they have been previously defined through a global feature selection).
 - The objects are extracted employing an efficient segmentation algorithm (PFF) on the VHR optical images.
 - The classification is performed using the object-based features.
 - The regularization is performed using a global model with linear unary term and *Potts model* prior, which is a standard formulation in many remote sensing application cases.
- S3:** An efficient fusion scheme. The best results are reported coupled with high computation times and load:
- All the 95 features are computed.
 - The object are extracted employing an efficient segmentation algorithm (PFF) on the VHR optical images.
 - The classification is performed using a selection of the object-based features.
 - The regularization is performed using a global model with linear unary term and *Exponential-feature model* prior.

S3 is recommended for higher accuracies, **S2** may be more suitable for scalability purposes and **S1** in case of absence of lidar data.

7.1.3 Quantitative features

In the previous workflow, it appears that Lidar is only useful in the classification process. This is probably due to the low point cloud density. If the spectral information allows to discriminate efficiently the tree species, lidar provides information about the vertical structure of the forest. Thus, it is possible to extract other relevant forest indicators employing Lidar. Once the stands are delineated, one can count the tree density per stand. Such information can then be employed for forest exploitation or to derive statistics and to help/improve inventory tasks. Furthermore, once the height and the tree species are determined, allometric equations can be employed in order to derive other relevant information from the stands (Muukkonen, 2007). Here, Lidar appears as an meaningful measurement tools for tree height that can be employed for other purposes than mapping (Hyypä et al., 2001).

7.2 Perspectives

The proposed framework allows to efficiently extract forest stands. However, some improvements can be envisaged. Firstly, other remote sensing data sources can be employed. The second

idea would be to improve the different steps of the framework. Finally, experiments could be conducted employing the proposed framework to different target classes.

7.2.1 Relevance of other remote sensing data sources

In this work, only standard VHR spectral optical images and low density (1-5 point/m²) Lidar point cloud were used.

Employing hyperspectral images could be interesting, since the high spectral resolution would allow to discriminate more precisely tree species (Dalponte et al., 2014; Liu et al., 2011; Torabzadeh et al., 2015; Clark et al., 2005). Such data mostly has **lower spatial resolution** or have sufficient spatial resolution but on **very limited area**. Thus statistical features that have here proven to be efficient for tree specie classification could not be derived. Feature selection would allow to estimate the impact of higher spectral resolution despite inferior spatial one.

A higher density lidar acquisition would also be interesting. Firstly, with more points, trees could be delineated more precisely and structural shape features at the tree level could be derived. For instance, a convex envelope can be extracted for each tree and the volume of the envelope or penetration indices can be derived (Lin et al., 2016; Li et al., 2013a; Ko et al., 2013). Again, feature selection would allow to conduct a precise and objective assessment of the relevance of such high density lidar. Multiple wavelength lidar also appears to be very promising for forest analysis (Budei et al., 2017), since it provides "spectral" and spatial information.

7.2.2 Improvement

As explained before, new features could be derived when employing other data sources. However, even with VHR optical images and low density lidar, other meaningful features that have not been tested in this work could be envisaged. The most interesting and obvious possibility would be to derive deep-based features. Indeed, deep-learning algorithms allow to directly learn features. It may appear interesting in such an unstructured environment. Deep-based features are optimal for specific classification tasks and can be employed as an input of traditional classification algorithms (such as RF or SVM) (Kontschieder et al., 2015).

The classification could also be performed directly using deep neural networks (Paisitkriangkrai et al., 2016; Paisitkriangkrai et al., 2015; Wegner et al., 2016; Workman et al., 2017). They have shown to deliver better results than traditional classification algorithms. The main advantage of the deep methods is that once a model is learned, it can be applied to other areas without being retrained. Postadjian et al. (2017) have shown that it is possible to automatically derive training samples from noisy LC DB for large scale classification. No preprocessing is required and the training procedure is validated at very large scales (> 10000 km²). Indeed, the model can be refined for small costs, leading to better results. The main drawback of these methods is their training times and system requirement; even if the transfer is important, the training of such model needs specific architecture and graphic cards.

Finally, efforts could be envisaged on the smoothing methods. One can consider fusion of different complementary classification outputs (Ouerghemmi et al., 2017). Thus, the integration of a classification from different data sources at different spatial resolutions could be assessed.

7.2.3 Application to other land-cover problems

The proposed framework has been developed specifically for forest stand segmentation and especially to retrieve smooth quite important stands segments, but could be applied to other semantic segmentation problems. Indeed, at the end, no *a priori* knowledge on forested areas has been inserted and all labels are considered equally and in an agnostic way. Thus, our pipeline exhibits rather general applicability. Experiments have been conducted on an small urban area (see Figure C.1) with exactly the same pipeline, data sources and data specifications.

The results are presented in Figure C.2 and Table C.39. Here the framework has been applied using the same standard parameters as the ones employed in forested areas. Only the parameter γ has been tuned in order to have consistent segments regarding the urban LC DB. Indeed, the use of a too important γ leads to over-smoothed results. Consequently, smaller elements such as roads are removed.

It appears that the framework also seems to be adapted to urban semantization. Indeed, all the global quality metrics indicate that the framework performs well in the discrimination of the urban classes, even with an important γ . However, some classes are poorly retrieved (e.g. *road*). Since such classes are under-represented, a loss of accuracy then does not decrease the global results. A visual inspection shows that the results do not give relevant description of the scene (e.g. the roads are not continuous).

The results could be improved with two procedures:

- Extracting features that are more related to urban environments,
- Propose a new formulation of the energy that can take into account the variation in the gradient of height (i.e. integrate borders constraints) and/or continuity in classes (e.g. roads are connected to each other) (Wegner et al., 2015; Wegner et al., 2013), and/or meaningful transition between classes (Volpi et al., 2015).

7.3 Final outlook

A fully automatic modular workflow has been proposed for the extraction of tree species forest stands using airborne VHR optical images and airborne lidar point cloud. It involves different image processing algorithms, such as segmentation, classification and smoothing. An attention was also paid to feature extraction; meaningful features that shown their efficiency for tree specie classification were extracted. The proposed framework produces excellent results for forest stand segmentation while highlighting the complementarity of both remote sensing data sources. The framework fulfills most of the operational constraints for a national mapping agency: no critical parameters, decent computing times, automation, selection of the level of detail and quantitative metrics for more in depth forest analysis.



Color code





















Label	Color	Espèce	Vegetation type	Scientific name
1		<i>Chênes décidus</i>	Deciduous oaks	<i>Deciduous quercus</i>
2		<i>Chênes sempervirents</i>	Evergreen oaks	<i>Quercus ilex</i>
3		Hêtre	Beech	<i>Fagus</i>
4		Châtaignier	Chestnut	<i>Castanea sativa</i>
5		Robinier	Robinia	<i>Robinia pseudacacia</i>
6		<i>Autre feuillu pur</i>	Other hardwood	-
7		<i>Pin maritime</i>	Maritime pine	<i>Pinus pinaster</i>
8		<i>Pin sylvestre</i>	Scots pine	<i>Pinus sylvestris</i>
9		<i>Pin laricio ou pin noir</i>	Black pine	<i>Pinus nigra</i>
10		<i>Pin d'Alep</i>	Aleppo pine	<i>Pinus halepensis</i>
11		<i>Pin à crochet ou pin cembro</i>	Mountain pine or Swiss pine	<i>Pinus uncinata</i> or <i>Pinus cembra</i>
12		<i>Autre pin pur</i>	Other pine	-
13		<i>Sapin ou épicéa</i>	Fir or Spruce	<i>Abies</i> or <i>pinea</i>
14		Mélèze	Larch	<i>Pinaceae</i>
15		Douglas	Douglas fir	<i>Pseudotsuga menziesii</i>
16		<i>Autre conifère pur autre que pin</i>	Other conifer other than pine	-
17		<i>Lande ligneuse</i>	Woody heathland	-
18		<i>Formation herbacée</i>	Herbaceous formation	-
19		<i>Peupleraie</i>	Poplar	<i>Populus</i>

TABLE A.1: Vegetation type color code.


**B**

Publications


B.1 Journal articles	150
B.2 Peer-reviewed conference papers	150





B.1 Journal articles


C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery*. ISPRS Journal of Photogrammetry and Remote Sensing, 126 (2017), pp.129–145, 2017. 

B.2 Peer-reviewed conference papers

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *How to combine LIDAR and very high resolution multispectral images for forest stand segmentation?* Proc. of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, USA, July 2017. 

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *Semantic segmentation of forest stands of pure specie as a global optimisation problem*. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017. 

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet. *Segmentation sémantique de données de télédétection multimodale : application aux peuplements forestiers*. ORASIS, Colleville-sur-Mer, France, June 2017. 

C. Dechesne, C. Mallet, A. Le Bris, V. Gouet-Brunet, A. Hervieu. *Forest stand segmentation using airborne Lidar data and very high resolution multispectral imagery*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 41 (B3), pp 207-214 , ISPRS Congress, Prague, July 2016. 



Results - Confusion matrices and accuracy metrics

C.1 Accuracy metrics	154
C.1.1 Metrics at the class level	154
C.1.2 Metrics at the global level	156
C.2 Flowchart assessment	157
C.2.1 Over-segmentation	157
C.2.2 Classification	163
C.2.3 Feature selection	166
C.2.4 Regularization	167
Local methods	167
Global methods	168
C.3 Can forest stands be simply retrieved?	173
C.3.1 VHR optical images	174
C.3.2 nDSM	175
C.4 Test on urban area	175

C.1 Accuracy metrics

In this section, the different accuracy metrics presented in the following tables are detailed. Most are standard metrics for classification evaluation.

All are based on the confusion matrix. The metrics can be computed at the global level, or for each class. In this section, the metrics are defined using the confusion matrix presented in Table C.1 (problem with r classes). In order to compute the metrics for a single class, an other confusion matrix can be derived. In this section such confusion matrix is presented in Table C.2.

Confusion matrix					
Classes	1	2	...	r	Total
1	n_{11}	n_{12}	...	n_{1r}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2r}	$n_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
r	n_{r1}	n_{r2}	...	n_{rr}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.r}$	n

TABLE C.1: Confusion Matrix.

Confusion matrix			
	i	\hat{i}	Total
i	TP	FP	Pp
\hat{i}	FN	TN	Np
Total	P	N	n

TABLE C.2: Confusion Matrix for the class i detection.

The relation between the elements of Table C.1 and Table C.2 are the following:

$$\begin{aligned}
 \text{TP} &= n_{ii} & \text{FP} &= n_{i.} - n_{ii} & \text{Pp} &= n_{i.} \\
 \text{FN} &= n_{.i} - n_{ii} & \text{TN} &= n - n_{.i} - n_{i.} + n_{ii} & \text{Np} &= n - n_{i.} \\
 \text{P} &= n_{.i} & \text{N} &= n - n_{.i} & &
 \end{aligned}
 \tag{C.1}$$

C.1.1 Metrics at the class level

Precision or producer's accuracy

For the class $i \in [1, r]$, the precision (or producer's accuracy) p_i is defined as follows:

$$p_i = \frac{n_{ii}}{n_{i.}} = \frac{\text{TP}}{\text{Pp}}.
 \tag{C.2}$$

It is the accuracy from the point of view of the map maker (the producer). This is how often real samples on the ground are correctly shown on the classified map or the probability that a certain land cover of an area on the ground is classified as such.

The Producer's Accuracy is complement of the Omission Error, Producer's Accuracy = 100%-Omission Error. It is also the number of reference sites classified accurately divided by the total number of reference sites for that class

Recall or user's accuracy

For the class $i \in [1, r]$, the recall (or user's accuracy) r_i is defined as follow:

$$r_i = \frac{n_{ii}}{n_{.i}} = \frac{TP}{P}. \quad (C.3)$$

It is the accuracy from the point of view of a map user, not the map maker. The User's accuracy essentially tells use how often the class on the map will actually be present on the ground, that is to say how exhaustive the map is. This is referred to as reliability. The User's Accuracy is complement of the Commission Error, User's Accuracy = 100%-Commission Error. When a class is not represented in the classification map, the recall can not be computed.

Intersection over Union

The Intersection over Union (or Jaccard index) (Jaccard, 1912) measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. It has been designed for the evaluation of object detection. For a class i , the Intersection over Union (IoU_i) is defined as follow:

$$IoU_i = \frac{n_{ii}}{n_{.i} + n_{.i} - n_{ii}} = \frac{TP}{Pp + P - TP}. \quad (C.4)$$

F-score

It is the harmonic mean of precision and recall. It considers both the precision p and the recall r to compute the score. The F-score can be interpreted as a weighted average of the precision and recall, where an F-score reaches its best value at 1 and worst at 0. The F-score (F_1) of the class i is defined as follow

$$F_{1,i} = 2 \frac{p_i r_i}{p_i + r_i} \quad (C.5)$$

Accuracy

The accuracy (A_i) (or relative observed agreement among raters) of the class i is computed as follow:

$$A_i = \frac{TP + TN}{n}. \quad (C.6)$$

Kappa coefficient

The Kappa coefficient (Cohen, 1960) (κ_i) is generated from a statistical test to evaluate the accuracy of a classification. Kappa essentially evaluates how well the classification performs as compared to just randomly assigning values (i.e. did the classification do better than randomness.) The Kappa Coefficient can range from -1 to 1. A value of 0 indicates that the classification is no better than a random classification. A negative number indicates the classification is significantly worse than random. A value close to 1 indicates that the classification is significantly better than random. The

Kappa coefficient is computed as follow:

$$\kappa_i = \frac{P_{0,i} - P_{e,i}}{1 - P_{e,i}}, \quad (\text{C.7})$$

where P_0 is the relative observed agreement among raters, and P_e is the hypothetical probability of chance agreement. They are defined as follow:

$$P_{0,i} = \frac{TP + TN}{n}, \quad (\text{C.8})$$

$$P_{e,i} = \frac{P \times Pp + N \times Np}{n^2}. \quad (\text{C.9})$$

C.1.2 Metrics at the global level

Intersection over Union

The overall Intersection over Union score (IoU) is defined as the mean of the local Intersection over Union scores:

$$IoU = \frac{1}{r} \sum_{i=1}^r IoU_i. \quad (\text{C.10})$$

F-score

The overall F-score (F_1) is defined as the mean of the local F-scores. If a $F_{1,i}$ can not be computed, it is considered as zero.

$$F_1 = \frac{1}{r} \sum_{i=1}^r F_{1,i}. \quad (\text{C.11})$$

Overall Accuracy

The overall accuracy (OA) is defined as follows:

$$OA = \frac{1}{n} \sum_{i=1}^r n_{ii}. \quad (\text{C.12})$$

Kappa coefficient

The Kappa coefficient is computed as follows:

$$\kappa = \frac{P_0 - P_e}{1 - P_e}, \quad (\text{C.13})$$

where P_0 is the relative observed agreement among raters, and P_e is the hypothetical probability of chance agreement. They are defined as follow:

$$P_0 = \frac{1}{n} \sum_{i=1}^r n_{ii}, \quad (\text{C.14})$$

$$P_e = \frac{1}{n^2} \sum_{i=1}^r n_i \cdot n_{.i}. \quad (\text{C.15})$$









C.2 Flowchart assessment

In this section, the confusion matrices of the different conducted experiments are presented. They allow to estimate precisely where confusions are reported. Even if they are relevant for such analysis, the metrics at the class level and global level are sufficient to evaluate the reliability of the method. The confusion matrices are related to the area *Vosges1*.

C.2.1 Over-segmentation

The confusion matrices resulting from the different proposed over-segmentation methods are presented here.

Trees.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2020114	2556	113280	86058	90.91
 4	335	28093	354	1173	93.78
 5	2569	0	143267	15875	88.59
 13	36915	2886	11342	1041207	95.32
Recall	98.07	83.77	53.41	90.99	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	89.31	79.37	49.97	87.1	76.44
F-score	94.36	88.5	66.64	93.1	85.65
Accuracy	93.11	99.79	95.91	95.6	92.2
P0	0.93	1	0.96	0.96	0.92
Pe	0.52	0.98	0.88	0.57	0.48
κ	0.8553	0.8839	0.6461	0.8988	0.8507

TABLE C.3: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2168270	2468	22793	28477	97.58
 4	684	29235	0	36	97.6
 5	86	0	151836	9789	93.89
 13	13407	2457	5586	1070900	98.04
Recall	99.35	85.58	84.25	96.55	





Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	96.96	83.82	79.88	94.72	88.84
F-score	98.46	91.2	88.81	97.29	93.94
Accuracy	98.06	99.84	98.91	98.3	97.55
P0	0.98	1	0.99	0.98	0.98
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9585	0.9111	0.8824	0.9604	0.9515

TABLE C.4: Confusion Matrix and accuracy metrics of the regularization.

Watershed.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	1803817	11010	197844	209337	81.18
● 4	41	27438	175	2301	91.6
● 5	8018	1115	140922	11656	87.14
● 13	72314	15007	39069	965960	88.43
Recall	95.73	50.28	37.28	81.22	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	78.35	48.06	35.34	73.42	58.79
F-score	87.86	64.92	52.22	84.67	72.42
Accuracy	85.78	99.15	92.64	90.03	83.8
P0	0.86	0.99	0.93	0.9	0.84
Pe	0.51	0.98	0.86	0.56	0.45
κ	0.7098	0.6453	0.4892	0.773	0.7048









TABLE C.5: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	2078028	977	110274	32729	93.52
● 4	317	26414	0	3224	88.18
● 5	443	0	153631	7637	95
● 13	42882	1009	3336	1045123	95.68
Recall	97.94	93.01	57.49	96	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	91.72	82.7	55.8	92.01	80.56
F-score	95.68	90.53	71.63	95.84	88.42
Accuracy	94.65	99.84	96.53	97.41	94.21
P0	0.95	1	0.97	0.97	0.94
Pe	0.53	0.98	0.88	0.57	0.48
κ	0.8866	0.9045	0.699	0.9396	0.8879

TABLE C.6: Confusion Matrix and accuracy metrics of the regularization.

Hierarchical segmentation.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	1906809	2069	136358	176772	85.81
 4	317	26775	141	2722	89.38
 5	2540	0	154347	4824	95.45
 13	57939	4404	31601	998406	91.4
Recall	96.91	80.53	47.87	84.42	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	83.53	73.5	46.8	78.2	70.51
F-score	91.03	84.73	63.76	87.77	81.82
Accuracy	89.28	99.72	95	92.06	88.03
P0	0.89	1	0.95	0.92	0.88
Pe	0.52	0.98	0.87	0.56	0.47
κ	0.7783	0.8459	0.6139	0.8191	0.7762

TABLE C.7: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2164661	1350	27069	28928	97.42
 4	1019	27568	0	1368	92.03
 5	373	0	159813	1525	98.83
 13	35484	612	5944	1050310	96.15
Recall	98.32	93.36	82.88	97.06	





Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	95.83	86.37	82.07	93.43	89.43
F-score	97.87	92.69	90.15	96.6	94.33
Accuracy	97.31	99.88	99	97.89	97.04
P0	0.97	1	0.99	0.98	0.97
Pe	0.53	0.98	0.9	0.57	0.5
κ	0.9423	0.9263	0.8963	0.9508	0.9412

TABLE C.8: Confusion Matrix and accuracy metrics of the regularization.

PFF.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	2072223	2635	66560	80590	93.26
● 4	200	29345	119	291	97.96
● 5	12398	0	146668	2645	90.7
● 13	51976	2751	20296	1017327	93.13
Recall	96.98	84.49	62.77	92.41	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	90.63	83.03	58.98	86.52	79.79
F-score	95.08	90.73	74.2	92.77	88.19
Accuracy	93.89	99.83	97.09	95.48	93.14
P0	0.94	1	0.97	0.95	0.93
Pe	0.53	0.98	0.89	0.57	0.49
κ	0.8701	0.9064	0.7271	0.8948	0.8662









TABLE C.9: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	2174574	3039	14711	29684	97.87
● 4	1643	28105	0	207	93.82
● 5	2407	0	158293	1011	97.89
● 13	23327	1550	12265	1055208	96.6
Recall	98.76	85.96	85.44	97.15	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	96.67	81.36	83.89	93.94	88.97
F-score	98.31	89.72	91.24	96.88	94.04
Accuracy	97.87	99.82	99.13	98.06	97.44
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9542	0.8963	0.9079	0.9547	0.949

TABLE C.10: Confusion Matrix and accuracy metrics of the regularization.

Quickshift.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2010751	1970	85980	123307	90.49
 4	441	28932	420	162	96.58
 5	8280	0	137651	15780	85.12
 13	57765	4237	16873	1013475	92.78
Recall	96.8	82.34	57.13	87.92	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	87.86	80.01	51.95	82.29	75.53
F-score	93.54	88.89	68.38	90.28	85.27
Accuracy	92.08	99.79	96.37	93.78	91.01
P0	0.92	1	0.96	0.94	0.91
Pe	0.52	0.98	0.89	0.56	0.48
κ	0.8333	0.8879	0.6653	0.8571	0.8267

TABLE C.11: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2172018	1463	19015	29512	97.75
 4	536	28364	0	1055	94.69
 5	427	0	147858	13426	91.43
 13	23755	1499	4103	1062993	97.31
Recall	98.87	90.54	86.48	96.03	





Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	96.67	86.17	80	93.55	89.1
F-score	98.31	92.57	88.89	96.66	94.11
Accuracy	97.87	99.87	98.95	97.91	97.3
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9543	0.925	0.8833	0.9514	0.9462

TABLE C.12: Confusion Matrix and accuracy metrics of the regularization.

SLIC.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	1999280	2337	61150	159241	89.98
● 4	398	26354	653	2550	87.98
● 5	4645	0	152316	4750	94.19
● 13	44996	2776	18514	1026064	93.93
Recall	97.56	83.75	65.47	86.04	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	87.99	75.15	62.93	81.51	76.9
F-score	93.61	85.81	77.25	89.81	86.62
Accuracy	92.22	99.75	97.44	93.36	91.39
P0	0.92	1	0.97	0.93	0.91
Pe	0.52	0.98	0.89	0.56	0.48
κ	0.837	0.8569	0.7594	0.849	0.8345

TABLE C.13: Confusion Matrix and accuracy metrics of the classification.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	2179198	2066	15072	25672	98.07
● 4	33	28139	0	1783	93.94
● 5	911	0	158212	2588	97.84
● 13	16802	1696	6974	1066878	97.67
Recall	99.19	88.21	87.77	97.26	









Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	97.3	83.46	86.1	95.05	90.48
F-score	98.63	90.98	92.53	97.46	94.9
Accuracy	98.27	99.84	99.27	98.42	97.9
P0	0.98	1	0.99	0.98	0.98
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9629	0.909	0.9215	0.9631	0.9583

TABLE C.14: Confusion Matrix and accuracy metrics of the regularization.

C.2.2 Classification

The confusion matrices resulting from the different experiments of the classification are presented here.

Pixel-based classification versus object-based classification.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	1573764	74079	323206	250959	70.83
 4	568	23026	1642	4719	76.87
 5	18420	7765	119886	15640	74.14
 13	124530	134448	79061	754311	69.05
Recall	91.64	9.622	22.89	73.55	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	66.53	9.351	21.2	55.31	38.1
F-score	79.9	17.1	34.98	71.23	50.8
Accuracy	77.42	93.63	87.29	82.62	70.48
P0	0.77	0.94	0.87	0.83	0.7
Pe	0.5	0.92	0.82	0.58	0.41
κ	0.5508	0.1582	0.3005	0.588	0.5004

TABLE C.15: Confusion Matrix and accuracy metrics of the pixel-based classification.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2072223	2635	66560	80590	93.26
 4	200	29345	119	291	97.96
 5	12398	0	146668	2645	90.7
 13	51976	2751	20296	1017327	93.13
Recall	96.98	84.49	62.77	92.41	





Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	90.63	83.03	58.98	86.52	79.79
F-score	95.08	90.73	74.2	92.77	88.19
Accuracy	93.89	99.83	97.09	95.48	93.14
P0	0.94	1	0.97	0.95	0.93
Pe	0.53	0.98	0.89	0.57	0.49
κ	0.8701	0.9064	0.7271	0.8948	0.8662

TABLE C.16: Confusion Matrix and accuracy metrics of the object-based classification (PFF).

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	2032550	112	144846	44500	91.47
● 4	469	18581	0	10905	62.03
● 5	839	0	130259	30613	80.55
● 13	47199	689	2430	1042032	95.39
Recall	97.67	95.87	46.93	92.37	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	89.52	60.41	42.16	88.43	70.13
F-score	94.47	75.32	59.31	93.86	80.74
Accuracy	93.21	99.65	94.9	96.11	91.94
P0	0.93	1	0.95	0.96	0.92
Pe	0.53	0.99	0.88	0.57	0.48
κ	0.8571	0.7516	0.5679	0.9102	0.845









TABLE C.17: Confusion Matrix and accuracy metrics of the regularization after a pixel-based classification.

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	2174574	3039	14711	29684	97.87
● 4	1643	28105	0	207	93.82
● 4	2407	0	158293	1011	97.89
● 13	23327	1550	12265	1055208	96.6
Recall	98.76	85.96	85.44	97.15	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	96.67	81.36	83.89	93.94	88.97
F-score	98.31	89.72	91.24	96.88	94.04
Accuracy	97.87	99.82	99.13	98.06	97.44
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9542	0.8963	0.9079	0.9547	0.949

TABLE C.18: Confusion Matrix and accuracy metrics of the regularization after an object-based classification (PFF).

Training set design.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	1776715	14410	268660	162223	79.96
 4	167	26119	1203	2466	87.19
 5	10505	2383	135761	13062	83.95
 13	119353	84627	57826	830544	76.03
Recall	93.18	20.48	29.29	82.37	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	75.54	19.88	27.74	65.39	47.14
F-score	86.07	33.17	43.43	79.08	60.44
Accuracy	83.59	97	89.91	87.46	78.98
P0	0.84	0.97	0.9	0.87	0.79
Pe	0.51	0.96	0.83	0.58	0.44
κ	0.6639	0.3223	0.3928	0.7015	0.6242

TABLE C.19: Confusion Matrix and accuracy metrics of the classification without training set design (training pixels are randomly selected).

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2118007	1016	57835	45150	95.32
 4	1913	25818	0	2224	86.19
 5	517	0	152774	8420	94.47
 13	63771	1016	4869	1022694	93.62
Recall	96.97	92.7	70.9	94.83	





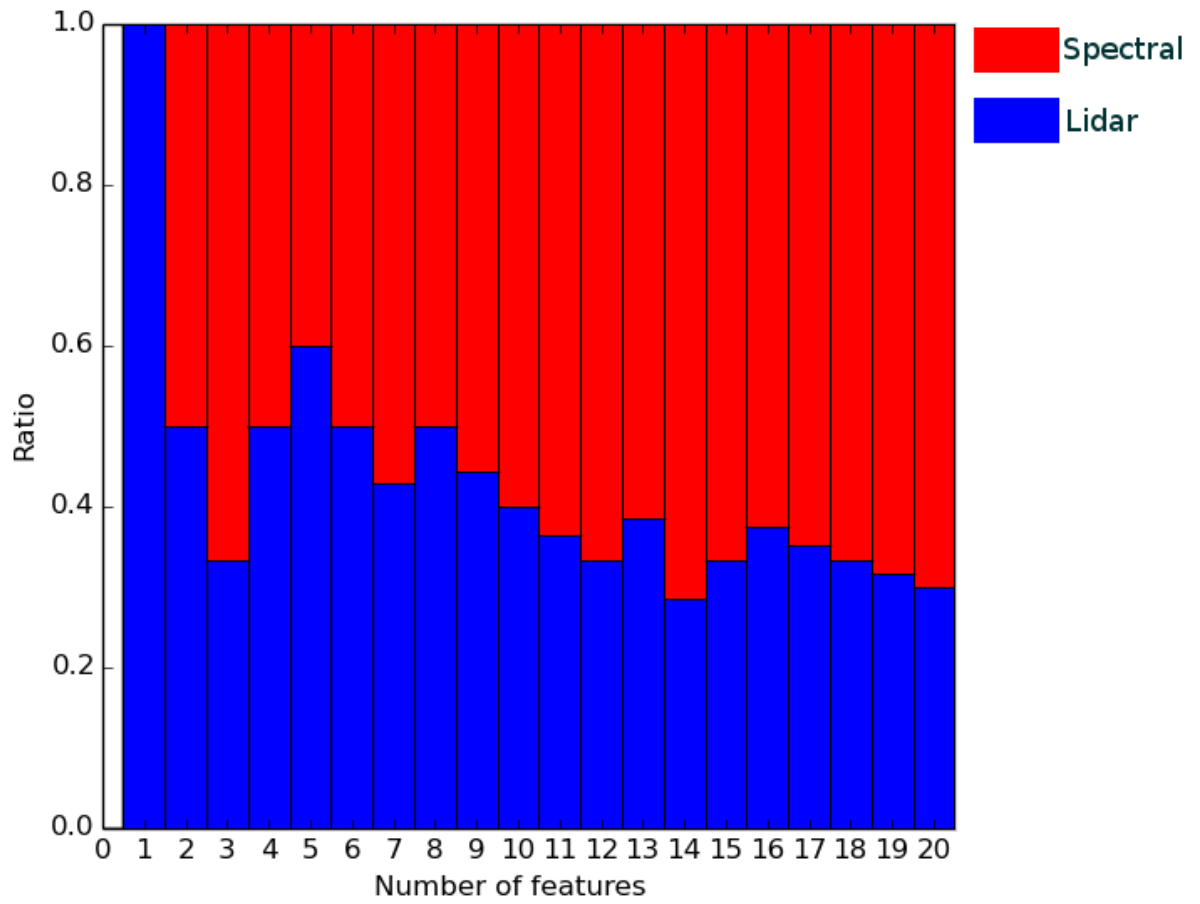
Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	92.56	80.71	68.08	89.07	82.61
F-score	96.14	89.33	81.01	94.22	90.17
Accuracy	95.15	99.82	97.96	96.42	94.67
P0	0.95	1	0.98	0.96	0.95
Pe	0.53	0.98	0.9	0.57	0.49
κ	0.8961	0.8924	0.7995	0.9163	0.8948

TABLE C.20: Confusion Matrix and accuracy metrics of the regularization without training set design (training pixels are randomly selected).

C.2.3 Feature selection











C.2.4 Regularization

The confusion matrices resulting from the different experiments of the regularization are presented here. Both local and global methods are presented here.

Local methods

Majority filter.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2083695	2660	61515	74138	93.78
 4	129	29478	86	262	98.41
 5	12020	0	147361	2330	91.13
 13	47586	2529	19812	1022423	93.6
Recall	97.21	85.03	64.41	93.02	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	91.32	83.88	60.61	87.46	80.82
F-score	95.46	91.23	75.48	93.31	88.87
Accuracy	94.35	99.84	97.27	95.82	93.64
P0	0.94	1	0.97	0.96	0.94
Pe	0.53	0.98	0.89	0.57	0.49
κ	0.8799	0.9115	0.7407	0.9027	0.8757

TABLE C.21: Confusion Matrix and accuracy metrics of the regularization ($r = 5$).

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2083643	2660	61569	74136	93.77
 4	129	29478	86	262	98.41
 5	12003	0	147401	2307	91.15
 13	47040	2526	19802	1022982	93.65
Recall	97.24	85.04	64.41	93.02	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	91.34	83.88	60.62	87.51	80.84
F-score	95.47	91.24	75.48	93.34	88.88
Accuracy	94.37	99.84	97.27	95.83	93.65
P0	0.94	1	0.97	0.96	0.94
Pe	0.53	0.98	0.89	0.57	0.49
κ	0.8802	0.9116	0.7408	0.9031	0.876

TABLE C.22: Confusion Matrix and accuracy metrics of the regularization ($r = 25$).

Probabilistic relaxation.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2122380	2802	43233	53593	95.52
 4	11	29880	0	64	99.75
 5	9862	0	150745	1104	93.22
 13	32943	2299	17229	1039879	95.2
Recall	98.02	85.42	71.37	95	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	93.71	85.24	67.85	90.65	84.36
F-score	96.75	92.03	80.85	95.1	91.18
Accuracy	95.94	99.85	97.96	96.94	95.35
P0	0.96	1	0.98	0.97	0.95
Pe	0.53	0.98	0.9	0.57	0.49
κ	0.9133	0.9196	0.7979	0.9287	0.9085

TABLE C.23: Confusion Matrix and accuracy metrics of the regularization ($r = 5$).**Global methods****Impact of the parameter γ .**

Confusion matrix					
Labels	 1	 4	 5	 13	Precision
 1	2170873	3034	16065	32036	97.7
 4	153	29715	0	87	99.2
 5	6914	0	154199	598	95.35
 13	17452	2019	12837	1060042	97.04
Recall	98.88	85.47	84.22	97.01	













Accuracy metrics					
Labels	 1	 4	 5	 13	Overall
IoU	96.63	84.88	80.9	94.22	89.16
F-score	98.29	91.82	89.44	97.02	94.14
Accuracy	97.84	99.85	98.96	98.15	97.4
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9537	0.9175	0.889	0.9568	0.9483

TABLE C.24: Confusion Matrix and accuracy metrics of the regularization ($\gamma = 5$).

Confusion matrix					
Labels	 1	 4	 5	 13	Precision
 1	2174574	3039	14711	29684	97.87
 4	1643	28105	0	207	93.82
 5	2407	0	158293	1011	97.89
 13	23327	1550	12265	1055208	96.6
Recall	98.76	85.96	85.44	97.15	













Accuracy metrics					
Labels	 1	 4	 5	 13	Overall
IoU	96.67	81.36	83.89	93.94	88.97
F-score	98.31	89.72	91.24	96.88	94.04
Accuracy	97.87	99.82	99.13	98.06	97.44
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9542	0.8963	0.9079	0.9547	0.949

TABLE C.25: Confusion Matrix and accuracy metrics of the regularization ($\gamma = 10$).

Confusion matrix					
Labels	 1	 4	 5	 13	Precision
 1	2184507	2975	5458	29068	98.31
 4	2514	27213	0	228	90.85
 5	22584	0	137459	1668	85
 13	25778	1754	10980	1053838	96.47
Recall	97.72	85.2	89.32	97.15	





Accuracy metrics					
Labels	 1	 4	 5	 13	Overall
IoU	96.11	78.46	77.16	93.82	86.39
F-score	98.02	87.93	87.11	96.81	92.47
Accuracy	97.48	99.79	98.84	98.02	97.06
P0	0.97	1	0.99	0.98	0.97
Pe	0.54	0.98	0.91	0.57	0.5
κ	0.9456	0.8782	0.865	0.9537	0.9409

TABLE C.26: Confusion Matrix and accuracy metrics of the regularization ($\gamma = 15$).

Confusion matrix					
Labels	● 1	● 4	● 5	● 13	Precision
● 1	2181685	2552	3839	33932	98.19
● 4	4080	25647	0	228	85.62
● 5	51623	0	106115	3973	65.62
● 13	27573	899	9543	1054335	96.52
Recall	96.32	88.14	88.8	96.51	

Accuracy metrics					
Labels	● 1	● 4	● 5	● 13	Overall
IoU	94.64	76.77	60.6	93.26	81.32
F-score	97.25	86.86	75.47	96.51	89.02
Accuracy	96.47	99.78	98.03	97.83	96.06
P0	0.96	1	0.98	0.98	0.96
Pe	0.54	0.98	0.92	0.57	0.51
κ	0.9235	0.8675	0.7447	0.9494	0.9198

TABLE C.27: Confusion Matrix and accuracy metrics of the regularization ($\gamma = 20$).**Unary term.**

Confusion matrix					
Labels	● 1	● 4	● 5	● 13	Precision
● 1	2168246	2951	15896	34915	97.58
● 4	26	29861	0	68	99.69
● 5	6376	0	154601	734	95.6
● 13	14023	2380	12605	1063342	97.34
Recall	99.07	84.85	84.43	96.75	

Accuracy metrics					
Labels	● 1	● 4	● 5	● 13	Overall
IoU	96.69	84.63	81.28	94.26	89.21
F-score	98.32	91.67	89.67	97.05	94.18
Accuracy	97.88	99.85	98.98	98.15	97.43
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9547	0.916	0.8914	0.957	0.9491

TABLE C.28: Confusion Matrix and accuracy metrics of the regularization using the log-inverse unary/data formulation.

Confusion matrix					
Labels	● 1	● 4	● 5	● 13	Precision
● 1	2174574	3039	14711	29684	97.87
● 4	1643	28105	0	207	93.82
● 5	2407	0	158293	1011	97.89
● 13	23327	1550	12265	1055208	96.6
Recall	98.76	85.96	85.44	97.15	

Accuracy metrics					
Labels	● 1	● 4	● 5	● 13	Overall
IoU	96.67	81.36	83.89	93.94	88.97
F-score	98.31	89.72	91.24	96.88	94.04
Accuracy	97.87	99.82	99.13	98.06	97.44
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9542	0.8963	0.9079	0.9547	0.949

TABLE C.29: Confusion Matrix and accuracy metrics of the regularization using the linear unary/data formulation.

Prior.

Confusion matrix					
Labels	● 1	● 4	● 5	● 13	Precision
● 1	2172746	3032	14419	31811	97.78
● 4	1700	27558	0	697	92
● 5	2673	0	157955	1083	97.68
● 13	24348	1737	11912	1054353	96.52
Recall	98.7	85.25	85.71	96.91	

Accuracy metrics					
Labels	● 1	● 4	● 5	● 13	Overall
IoU	96.54	79.36	84	93.64	88.38
F-score	98.24	88.49	91.3	96.72	93.69
Accuracy	97.78	99.8	99.14	97.96	97.34
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9522	0.8839	0.9085	0.9523	0.947

TABLE C.30: Confusion Matrix and accuracy metrics of the regularization the *Potts model*.

Confusion matrix					
Labels	● 1	● 4	● 5	● 13	Precision
● 1	2174110	3124	13860	30914	97.84
● 4	2033	27472	0	450	91.71
● 5	2416	0	158306	989	97.89
● 13	25814	1653	11583	1053300	96.43
Recall	98.63	85.19	86.15	97.02	









Accuracy metrics					
Labels	● 1	● 4	● 5	● 13	Overall
IoU	96.53	79.1	84.59	93.65	88.47
F-score	98.23	88.33	91.65	96.72	93.73
Accuracy	97.77	99.79	99.18	97.96	97.35
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9521	0.8822	0.9122	0.9524	0.9473

TABLE C.31: Confusion Matrix and accuracy metrics of the regularization the *z-Potts model*.

Confusion matrix					
Labels	● 1	● 4	● 5	● 13	Precision
● 1	2174574	3039	14711	29684	97.87
● 4	1643	28105	0	207	93.82
● 5	2407	0	158293	1011	97.89
● 13	23327	1550	12265	1055208	96.6
Recall	98.76	85.96	85.44	97.15	

Accuracy metrics					
Labels	● 1	● 4	● 5	● 13	Overall
IoU	96.67	81.36	83.89	93.94	88.97
F-score	98.31	89.72	91.24	96.88	94.04
Accuracy	97.87	99.82	99.13	98.06	97.44
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.9542	0.8963	0.9079	0.9547	0.949

TABLE C.32: Confusion Matrix and accuracy metrics of the regularization the *Exponential-feature model*.

Confusion matrix					
Labels	 1	 4	 5	 13	Precision
 1	2171863	2951	14706	32488	97.74
 4	2144	27578	0	233	92.06
 5	2633	0	157920	1158	97.66
 13	23487	1544	11976	1055343	96.61
Recall	98.72	85.99	85.55	96.89	













Accuracy metrics					
Labels	 1	 4	 5	 13	Overall
IoU	96.52	80.05	83.82	93.71	88.52
F-score	98.23	88.92	91.2	96.75	93.77
Accuracy	97.76	99.8	99.13	97.98	97.34
P0	0.98	1	0.99	0.98	0.97
Pe	0.53	0.98	0.91	0.57	0.5
κ	0.952	0.8882	0.9075	0.9528	0.9471

TABLE C.33: Confusion Matrix and accuracy metrics of the regularization the *Distance-features model*.

C.3 Can forest stands be simply retrieved?

A segmentation of the input data (i.e., VHR optical image or nDSM) is operated, and a majority vote is performed. The resulting confusion matrices are presented here.

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	1783941	13837	204918	219306	80.29
 4	63	29545	62	285	98.63
 5	6482	726	145449	9054	89.94
 13	113114	19865	52105	907266	83.06
Recall	93.71	46.18	36.13	79.87	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	76.18	45.89	34.73	68.68	56.37
F-score	86.48	62.91	51.56	81.43	70.59
Accuracy	84.09	99.01	92.2	88.2	81.75
P0	0.84	0.99	0.92	0.88	0.82
Pe	0.51	0.97	0.85	0.57	0.45
κ	0.6744	0.6247	0.4814	0.7279	0.6679

TABLE C.34: Confusion Matrix and accuracy metrics of the classification (object-based hierarchical).

C.3.1 VHR optical images

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	1869715	0	6022	346265	84.15
 4	8142	0	0	21813	0
 5	82695	0	44484	34532	27.51
 13	108307	0	25382	958661	87.76
Recall	90.37	-	58.62	70.42	













Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	77.22	0	23.03	64.13	41.1
F-score	87.15	-	37.44	78.14	50.68
Accuracy	84.27	99.15	95.76	84.7	81.94
P0	0.84	0.99	0.96	0.85	0.82
Pe	0.52	0.99	0.93	0.54	0.5
κ	0.6695	0	0.3555	0.6659	0.6417

TABLE C.35: Confusion Matrix and accuracy metrics when adding semantic information to a direct hierarchical segmentation ($\mu = 15$).

Confusion matrix					
Label	 1	 4	 5	 13	Precision
 1	2021033	0	0	200969	90.96
 4	0	0	0	29955	0
 5	83682	0	0	78029	0
 13	589638	0	0	502712	46.02
Recall	75.01	-	-	61.94	





Accuracy metrics					
Label	 1	 4	 5	 13	Overall
IoU	69.8	0	0	35.87	26.42
F-score	82.22	-	-	52.81	33.76
Accuracy	75.06	99.15	95.39	74.37	71.98
P0	0.75	0.99	0.95	0.74	0.72
Pe	0.57	0.99	0.95	0.6	0.56
κ	0.4176	0	0	0.3573	0.3644

TABLE C.36: Confusion Matrix and accuracy metrics when adding semantic information to a direct PFF segmentation ($\sigma = 0.8$, $k = 500$ and $m = 40000$).

C.3.2 nDSM

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	1526001	0	0	696001	68.68
● 4	0	0	0	29955	0
● 5	109	0	0	161602	0
● 13	105832	0	0	986518	90.31
Recall	93.51	-	-	52.64	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	65.55	0	0	49.83	28.84
F-score	79.19	-	-	66.51	36.43
Accuracy	77.13	99.15	95.39	71.67	71.66
P0	0.77	0.99	0.95	0.72	0.72
Pe	0.49	0.99	0.95	0.49	0.46
κ	0.5508	0	0	0.4477	0.4737

TABLE C.37: Confusion Matrix and accuracy metrics when adding semantic information to a direct hierarchical segmentation ($\mu = 15$).

Confusion matrix					
Label	● 1	● 4	● 5	● 13	Precision
● 1	1788281	0	59707	374014	80.48
● 4	5687	0	0	24268	0
● 5	2510	0	44950	114251	27.8
● 13	244694	0	2	847654	77.6
Recall	87.61	-	42.95	62.32	

Accuracy metrics					
Label	● 1	● 4	● 5	● 13	Overall
IoU	72.26	0	20.3	52.82	36.34
F-score	83.89	-	33.75	69.12	46.69
Accuracy	80.42	99.15	94.97	78.4	76.47
P0	0.8	0.99	0.95	0.78	0.76
Pe	0.52	0.99	0.93	0.54	0.49
κ	0.5903	0	0.3126	0.5282	0.5374

TABLE C.38: Confusion Matrix and accuracy metrics when adding semantic information to a direct PFF segmentation ($\sigma = 0.8$, $k = 500$ and $m = 40000$).

C.4 Test on urban area

In this section, results are presented for a small urban area. The main problem of the ground truth of urban area is that *road* and *railway* are represented as line instead of polygons. Thus only few pixels are available for training/validation (a dilatation can be envisaged but has not been carried out in our tests).

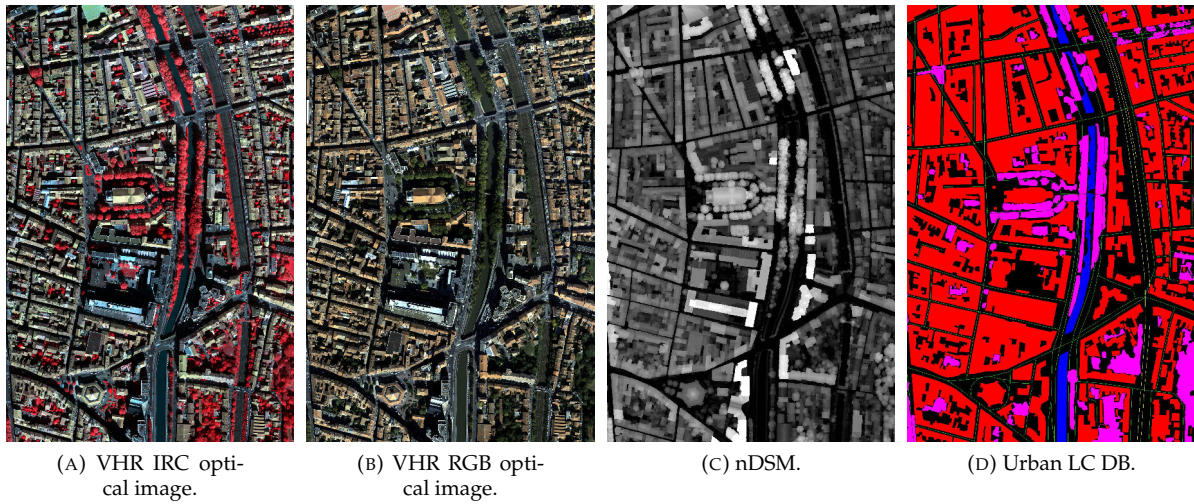


FIGURE C.1: VHR optical images, rasterized nDSM and urban LC of the urban area ($600\text{ m} \times 900\text{ m}$). Color code: ● building, ● road, ● water, ● railways, ● vegetation.

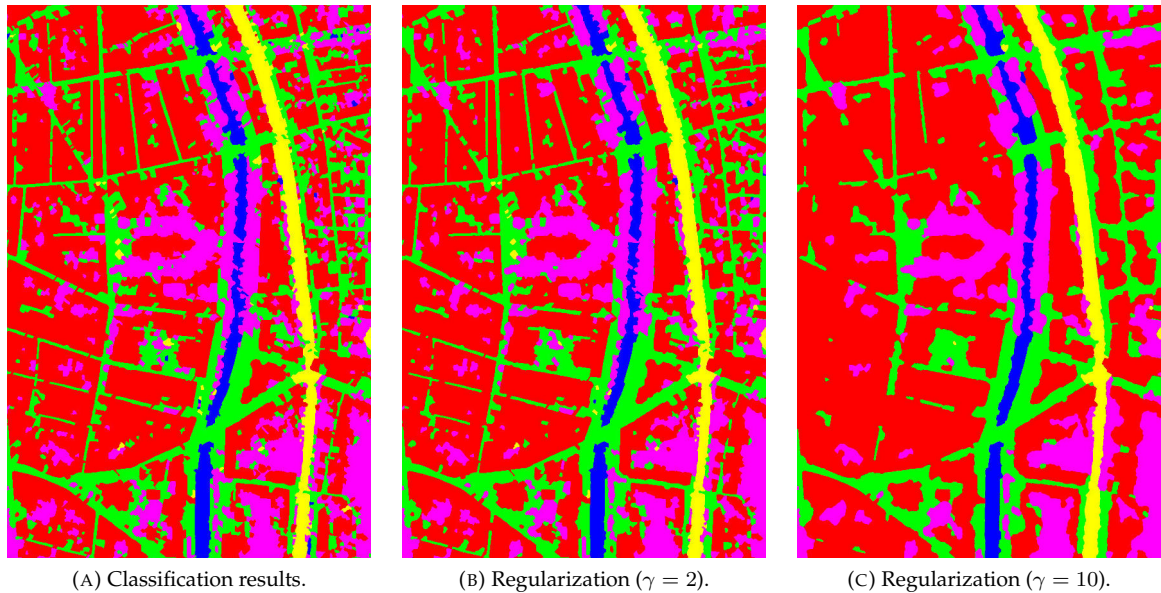


FIGURE C.2: Results of the proposed method on the urban area ($600\text{ m} \times 900\text{ m}$). Color code: ● building, ● road, ● water, ● railways, ● vegetation.

Metric	Classification	Regularization ($\gamma = 2$)	Regularization ($\gamma = 10$)
OA	86.48	88	90.16
κ	0.6887	0.7172	0.7551
F-score	74.57	77	78.66
IoU	64.33	67.6	70.25

TABLE C.39: Accuracy metrics of the different results of the method on the urban area.

Bibliography

- Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk (2012). "SLIC superpixels compared to state-of-the-art superpixel methods". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 34.11, pp. 2274–2282
(Cited on pages [24](#), [25](#), [45](#), and [46](#)).
- Achard, F. (2009). *Vital forest graphics*. UNEP/Earthprint
(Cited on page [2](#)).
- Albert, L., F. Rottensteiner, and C. Heipke (2016). "Contextual land use classification: how detailed can the class structure be?" *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B4*, pp. 11–18
(Cited on page [31](#)).
- Anderson, G. L. and J.D. Hanson (1992). "Evaluating hand-held radiometer derived vegetation indices for estimating above ground biomass". *Geocarto International* 7.1, pp. 71–78
(Cited on page [40](#)).
- Atkinson, P.M., M.E.J. Cutler, and H. Lewis (1997). "Mapping sub-pixel proportional land cover with AVHRR imagery". *International Journal of Remote Sensing* 18.4, pp. 917–935
(Cited on page [26](#)).
- Audebert, N., B. Le Saux, and S. Lefèvre (2016). "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks". In: *Asian Conference on Computer Vision*. Springer, pp. 180–196
(Cited on page [52](#)).
- Baatz, M., G. Binnig, P. Eschenbacher, A. Melchinger, and M. Sogtrop (2004). *Method of iterative segmentation of a digital picture*. US Patent 6,832,002
(Cited on page [23](#)).
- Bacour, C., F.-M. Bréon, and F. Maignan (2006). "Normalization of the directional effects in NOAA–AVHRR reflectance measurements for an improved monitoring of vegetation cycles". *Remote Sensing of Environment* 102.3, pp. 402–413
(Cited on page [40](#)).
- Bannari, A., D. Morin, F. Bonn, and A.R. Huete (1995). "A review of vegetation indices". *Remote Sensing Reviews* 13.1-2, pp. 95–120
(Cited on page [40](#)).
- Battiti, R. (1994). "Using mutual information for selecting features in supervised neural net learning". *IEEE Transactions on Neural Networks* 5.4, pp. 537–550
(Cited on page [29](#)).

- Beaudoin, A., T. Le Toan, S. Goze, E. Nezry, A. Lopes, E. Mougin, C.C. Hsu, H.C. Han, J.A. Kong, and R.T. Shin (1994). "Retrieval of forest biomass from SAR data". *International Journal of Remote Sensing* 15.14, pp. 2777–2796
(Cited on page 6).
- Belgiu, M. and L. Drăguț (2016). "Random Forest in remote sensing: A review of applications and future directions". *ISPRS Journal of Photogrammetry and Remote Sensing* 114, pp. 24–31
(Cited on pages xiii and 47).
- Bellman, R. E. (2015). *Adaptive control processes: a guided tour*. Princeton university press
(Cited on pages 28 and 69).
- Belward, A. S. and J. O. Skøien (2015). "Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites". *ISPRS Journal of Photogrammetry and Remote Sensing* 103, pp. 115–128
(Cited on page 5).
- Beucher, S. and C. Lantuéjoul (1979). "Use of watersheds in contour detection". In:
(Cited on page 44).
- Bosch, A., A. Zisserman, and X. Munoz (2007). "Image classification using random forests and ferns". In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8
(Cited on page 27).
- Boser, B. E., I. M. Guyon, and V. N. Vapnik (1992). "A training algorithm for optimal margin classifiers". In: *ACM Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152
(Cited on page 26).
- Boykov, Y., O. Veksler, and R. Zabih (2001). "Fast approximate energy minimization via graph cuts". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 23.11, pp. 1222–1239
(Cited on page 111).
- Bradski, G. and A. Kaehler (2008). *Learning OpenCV: Computer vision with the OpenCV library*. "O'Reilly Media, Inc."
(Cited on pages 47 and 74).
- Brandtberg, T. (2007). "Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar". *ISPRS Journal of Photogrammetry and Remote Sensing* 61.5, pp. 325–340
(Cited on page 22).
- Breiman, L. (2001). "Random forests". *Machine learning* 45.1, pp. 5–32
(Cited on pages 26, 27, and 29).
- Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and regression trees*. CRC press
(Cited on pages 26 and 27).
- Brooks, T. M., R. A. Mittermeier, G. A.B. Da Fonseca, J. Gerlach, M. Hoffmann, J. F. Lamoreux, C. G. Mittermeier, J. D. Pilgrim, and A. S.L. Rodrigues (2006). "Global biodiversity conservation priorities". *Science* 313.5783, pp. 58–61
(Cited on page 3).

- Bruzzone, L. and S. B. Serpico (2000). "A technique for feature selection in multiclass problems". *International Journal of Remote Sensing* 21.3, pp. 549–563
(Cited on page 29).
- Budei, B. C., B. St-Onge, C. Hopkinson, and F. A. Audet (2017). "Identifying the genus or species of individual trees using a three-wavelength airborne lidar system". *Remote Sensing of Environment*. ISSN: 0034-4257
(Cited on page 142).
- Camps-Valls, G. and L. Bruzzone (2009). *Kernel methods for remote sensing data analysis*. John Wiley & Sons
(Cited on page 26).
- Cang, S. and H. Yu (2012). "Mutual information based input feature selection for classification problems". *Decision Support Systems* 54.1, pp. 691–698
(Cited on page 29).
- Carlotto, M. J. (2009). "Effect of errors in ground truth on classification accuracy". *International Journal of Remote Sensing* 30.18, pp. 4831–4849
(Cited on page 48).
- Chan, T. F. and L. A. Vese (2001). "Active contours without edges". *IEEE Transactions on image processing* 10.2, pp. 266–277
(Cited on page 50).
- Chang, Y.-L., K.-S. Chen, B. Huang, W.-Y. Chang, J. A. Benediktsson, and L. Chang (2011). "A parallel simulated annealing approach to band selection for high-dimensional remote sensing images". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.3, pp. 579–590
(Cited on page 30).
- Chen, J., S. Paris, and F. Durand (2007). "Real-time edge-aware image processing with the bilateral grid". *ACM Transactions on Graphics (TOG)* 26.3, p. 103
(Cited on page 108).
- Cheriyadat, A. and L. M. Bruce (2003). "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 6, pp. 3420–3422
(Cited on page 28).
- Clark, M. L., D. A. Roberts, and D. B. Clark (2005). "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales". *Remote sensing of environment* 96.3, pp. 375–398
(Cited on pages 5, 105, and 142).
- Cohen, J. (1960). "A coefficient of agreement for nominal scales". *Educational and psychological measurement* 20.1, pp. 37–46
(Cited on page 155).
- Comaniciu, D. and P. Meer (2002). "Mean shift: A robust approach toward feature space analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5, pp. 603–619
(Cited on pages 24, 45, and 46).

- Dalponte, M., L. Bruzzone, and D. Gianelle (2012). "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data". *Remote Sensing of Environment* 123, pp. 258–270
(Cited on page 105).
- Dalponte, M., H O. Orka, T. Gobakken, D. Gianelle, and E. Næsset (2013). "Tree species classification in boreal forests with hyperspectral data". *IEEE Transactions on Geoscience and Remote Sensing* 51.5, pp. 2632–2645
(Cited on page 105).
- Dalponte, M., H. O. Ørka, L. T. Ene, T. Gobakken, and E. Næsset (2014). "Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data". *Remote Sensing of Environment* 140, pp. 306–317
(Cited on pages xii, 16, 26, 39, 43, and 142).
- Dalponte, M., L. T. Ene, C. Witte, M. Marconcini, T. Gobakken, and E. Næsset (2015). "Semi-supervised SVM for individual tree crown species classification". *ISPRS Journal of Photogrammetry and Remote Sensing* 110, pp. 77–87
(Cited on pages x and 16).
- De Backer, S., P. Kempeneers, W. Debruyn, and P. Scheunders (2005). "A band selection technique for spectral classification". *IEEE Geoscience and Remote Sensing Letters* 2.3, pp. 319–323
(Cited on pages 29 and 30).
- Dean, J., G. S. Corrado, R. Monga, K. Chen, M. Devin, Q. V. Le, M. Z. Mao, M.A. Ranzato, A. Senior, P. Tucker, K. Yang, and A. Y. Ng (2012). "Large scale distributed deep networks". In: *Advances in Neural Information Processing Systems*, pp. 1223–1231
(Cited on page 26).
- Dechesne, C., C. Mallet, A. Le Bris, V. Gouet-Brunet, and A. Hervieu (2016). "Forest stand segmentation using airborne lidar data and very high resolution multispectral imagery". *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41-B3, pp. 207–214
(Cited on pages xiii, 47, and 74).
- Dechesne, C., C. Mallet, A. Le Bris, and V. Gouet-Brunet (2017). "Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery". *ISPRS Journal of Photogrammetry and Remote Sensing* 126, pp. 129–145
(Cited on pages 16 and 22).
- Demantké, J., C. Mallet, N. David, and B. Vallet (2011). "Dimensionality based scale selection in 3D lidar point clouds". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38.Part 5, W12
(Cited on page 38).
- Demuth, H. B., M. H. Beale, O. De Jess, and M. T. Hagan (2014). *Neural network design*. Martin Hagan
(Cited on page 36).
- Derin, H. and H. Elliott (1987). "Modeling and segmentation of noisy and textured images using Gibbs random fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 1, pp. 39–55
(Cited on page 23).

- Díaz-Uriarte, R. and S. A. De Andres (2006). “Gene selection and classification of microarray data using random forest”. *BMC Bioinformatics* 7.1, p. 3
(Cited on page 29).
- Diedershagen, O., B. Koch, and H. Weinacker (2004). “Automatic segmentation and characterisation of forest stand parameters using airborne lidar data, multispectral and fogis data”. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36(8/W2), pp. 208–212
(Cited on pages x, 20, and 21).
- Estévez, P. A., M. Tesmer, Cl. A. Perez, and J. M. Zurada (2009). “Normalized mutual information feature selection”. *IEEE Transactions on Neural Networks* 20.2, pp. 189–201
(Cited on page 29).
- Eysn, L., M. Hollaus, K. Schadauer, and N. Pfeifer (2012). “Forest delineation based on airborne LIDAR data”. *Remote Sensing* 4.3, pp. 762–783
(Cited on pages 16, 19, 21, and 43).
- Fahey, T. J., P. B. Woodbury, J. J. Battles, C. L. Goodale, S. P. Hamburg, S. V. Ollinger, and C. W. Woodall (2010). “Forest carbon storage: ecology, management, and policy”. *Frontiers in Ecology and the Environment* 8.5, pp. 245–252
(Cited on page 3).
- Fauvel, M., C. Dechesne, A. Zullo, and F. Ferraty (2015). “Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.6, pp. 2824–2831
(Cited on pages 26 and 29).
- Felzenszwalb, P. F. and D. P. Huttenlocher (2004). “Efficient graph-based image segmentation”. *International Journal of Computer Vision* 59.2, pp. 167–181
(Cited on pages 24, 45, 46, 94, 96, and 126).
- (2006). “Efficient belief propagation for early vision”. *International Journal of Computer Vision* 70.1, pp. 41–54
(Cited on page 111).
- Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim (2014). “Do we need hundreds of classifiers to solve real world classification problems”. *J. Mach. Learn. Res* 15.1, pp. 3133–3181
(Cited on page 47).
- Ferraz, A., M. Mallet, and N. Chehata (2016a). “Large-scale road detection in forested mountainous areas using airborne topographic lidar data”. *ISPRS Journal of Photogrammetry and Remote Sensing* 112, pp. 23–36
(Cited on page 40).
- Ferraz, A., S. Saatchi, C. Mallet, and V. Meyer (2016b). “Lidar detection of individual tree size in tropical forests”. *Remote Sensing of Environment* 183, pp. 318–333
(Cited on page 6).
- Filin, S. and N. Pfeifer (2005). “Neighborhood systems for airborne laser data”. *Photogrammetric Engineering & Remote Sensing* 71.6, pp. 743–755
(Cited on page 38).

- Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems". *Annals of Human Genetics* 7.2, pp. 179–188
(Cited on pages 28 and 29).
- Foody, G. M. (2002). "Status of land cover classification accuracy assessment". *Remote Sensing of Environment* 80.1, pp. 185–201
(Cited on page 26).
- Franklin, S.E., R.J. Hall, L.M. Moskal, A.J. Maudie, and M.B. Lavigne (2000). "Incorporating texture into classification of forest species composition from airborne multispectral images". *International Journal of Remote Sensing* 21.1, pp. 61–79
(Cited on pages 5, 6, and 20).
- Fukunaga, K. and L. Hostetler (1975). "The estimation of the gradient of a density function, with applications in pattern recognition". *IEEE Transactions on Information Theory* 21.1, pp. 32–40
(Cited on page 46).
- Geurts, P., D. Ernst, and L. Wehenkel (2006). "Extremely randomized trees". *Machine Learning* 63.1, pp. 3–42
(Cited on page 27).
- Ghosh, A., N. R. Pal, and S. K. Pal (1991). "Image segmentation using a neural network". *Biological Cybernetics* 66.2, pp. 151–158
(Cited on page 23).
- Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson (2006). "Random forests for land cover classification". *Pattern Recognition Letters* 27.4, pp. 294–300
(Cited on page 27).
- Goldberg, D. E. (1989). "Genetic algorithms in search, optimization, and machine learning, 1989". *Reading: Addison-Wesley*
(Cited on page 30).
- Gong, P. and P.J. Howarth (1989). "Performance analyses of probabilistic relaxation methods for land-cover classification". *Remote Sensing of Environment* 30.1, pp. 33–42
(Cited on pages 31 and 109).
- Gressin, A., C. Mallet, N. Vincent, and N. Paparoditis (2013a). "Updating land cover databases using a single very high resolution satellite image". In: *The ISPRS Workshop on Image Sequence Analysis*. Vol. 2. 3, pp. 13–18
(Cited on page 10).
- (2013b). "Updating land cover databases using a single very high resolution satellite image". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W2, pp. 13–18
(Cited on page 48).
- (2014). "Updating the new French national land cover database". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3534–3537
(Cited on page 10).

- Guigues, L., J.-P. Cocquerez, and H. Le Men (2006). "Scale-sets image analysis". *IJCV* 68.3, pp. 289–317
(Cited on pages 23, 44, 45, 94, and 96).
- Guo, B., R. I. Damper, S. R. Gunn, and J. D.B. Nelson (2008). "A fast separability-based feature-selection method for high-dimensional remotely sensed image classification". *Pattern Recognition* 41.5, pp. 1653–1662
(Cited on page 29).
- Hansen, F.R. and H. Elliott (1982). "Image segmentation using simple Markov field models". *Computer Graphics and Image Processing* 20.2, pp. 101–132
(Cited on page 23).
- Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S.A. Turubanova, A. Tyukavina, D. Thau, S.V. Stehman, S.J. Goetz, T.R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C.O. Justice, and J. R. G. Townshen (2013). "High-resolution global maps of 21st-century forest cover change". *science* 342.6160, pp. 850–853
(Cited on page 4).
- Haralick, R. M., K. Shanmugam, and I. Dinstein (1973). "Textural features for image classification". *IEEE Transactions on Systems, Man and Cybernetics* 6, pp. 610–621
(Cited on page 41).
- Heinzel, J. and B. Koch (2012). "Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation". *International Journal of Applied Earth Observation and Geoinformation* 18, pp. 101–110
(Cited on page x).
- Hepner, G., T.S. Logan, N. Ritter, and N. Bryant (1990). "Artificial neural network classification using a minimal training set- Comparison to conventional supervised classification". *Photogrammetric Engineering and Remote Sensing* 56.4, pp. 469–473
(Cited on page 26).
- Hernando, A., D. Tiede, F. Albrecht, and S. Lang (2012). "Spatial and thematic assessment of object-based forest stand delineation using an OFA-matrix". *International Journal of Applied Earth Observation and Geoinformation* 19, pp. 214–225
(Cited on pages x, 18, and 21).
- Herold, M., M. E. Gardner, and D. A. Roberts (2003). "Spectral resolution requirements for mapping urban areas". *IEEE Transactions on Geoscience and Remote Sensing* 41.9, pp. 1907–1919
(Cited on page 29).
- Holmgren, J. (2004). "Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning". *Scandinavian Journal of Forest Research* 19.6, pp. 543–553
(Cited on page 5).
- Hovi, A., L. Korhonen, J. Vauhkonen, and I. Korpela (2016). "LiDAR waveform features for tree species classification and their sensitivity to tree-and acquisition related parameters". *Remote Sensing of Environment* 173, pp. 224–237
(Cited on page 22).

- Hughes, G. (1968). "On the mean accuracy of statistical pattern recognizers". *IEEE Transactions on Information Theory* 14.1, pp. 55–63
(Cited on pages 28 and 69).
- Hyyppä, J., O. Kelle, M. Lehtikoinen, and M. Inkinen (2001). "A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners". *IEEE Transactions on Geoscience and Remote Sensing* 39.5, pp. 969–975
(Cited on page 141).
- Indyk, P. and R. Motwani (1998). "Approximate nearest neighbors: towards removing the curse of dimensionality". In: *ACM Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 604–613
(Cited on page 26).
- Jaccard, P. (1912). "The distribution of the flora in the alpine zone." *New Phytologist* 11.2, pp. 37–50
(Cited on page 155).
- Jakubowski, M. K., Q. Guo, and M. Kelly (2013). "Tradeoffs between lidar pulse density and forest measurement accuracy". *Remote Sensing of Environment* 130, pp. 245–253
(Cited on page 6).
- John, G. H. (1997). "Enhancements to the data mining process". PhD thesis. Stanford University
(Cited on page 28).
- Jolliffe, I. (2011). "Principal component analysis". In: *International Encyclopedia of Statistical Science*. Springer, pp. 1094–1096
(Cited on page 28).
- Jordan, C. F. (1969). "Derivation of leaf-area index from quality of light on the forest floor". *Ecology* 50.4, pp. 663–666
(Cited on page 40).
- Jutten, C. and J. Herault (1991). "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture". *Signal Processing* 24.1, pp. 1–10
(Cited on page 28).
- Kaartinen, H., J. Hyyppä, X. Yu, M. Vastaranta, H. Hyyppä, A. Kukko, M. Holopainen, C. Heipke, M. Hirschmugl, F. Morsdorf, E. Næsset, Pitkänen J., S. Popescu, S. Solberg, B. M. Wolf, and J.-C. Wu (2012). "An international comparison of individual tree detection and extraction using airborne laser scanning". *Remote Sensing* 4.4, pp. 950–974
(Cited on pages 25 and 64).
- Kaasalainen, S., T. Lindroos, and J. Hyyppä (2007). "Toward hyperspectral lidar: Measurement of spectral backscatter intensity with a supercontinuum laser source". *IEEE Geoscience and Remote Sensing Letters* 4.2, pp. 211–215
(Cited on page 6).
- Kandare, K., M. Dalponte, D. Gianelle, and J. C.-W. Chan (2014). "A new procedure for identifying single trees in understory layer using discrete LiDAR data". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1357–1360
(Cited on pages 16, 26, and 43).

- Kangas, A. and M. Maltamo (2006). *Forest inventory: methodology and applications*. Vol. 10. Springer Science & Business Media
(Cited on pages ix, 4, 5, and 57).
- Keenan, R. J., G. A. Reams, F. Achard, J. V. de Freitas, A. Grainger, and E. Lindquist (2015). "Dynamics of global forest area: results from the FAO Global Forest Resources Assessment 2015". *Forest Ecology and Management* 352, pp. 9–20
(Cited on page 2).
- Khosravipour, A., A. K. Skidmore, M. Isenburg, T. Wang, and Y. A. Hussin (2014). "Generating pit-free canopy height models from airborne lidar". *Photogrammetric Engineering & Remote Sensing* 80.9, pp. 863–872
(Cited on pages xii and 39).
- Kim, M., M. Madden, and T.A. Warner (2009). "Forest Type Mapping using Object-specific Texture Measures from Multispectral Ikonos Imagery: Segmentation Quality and Image Classification Issues". *Photogrammetric Engineering & Remote Sensing* 75.7, pp. 819–829
(Cited on pages ix and 5).
- Kim, S., T. Hinckley, and D. Briggs (2011). "Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data". *Remote sensing of environment* 115.12, pp. 3329–3342
(Cited on page 39).
- Kirillov, A, D. Schlesinger, W. Forkel, A. Zelenin, S. Zheng, P. Torr, and C. Rother (2015). "A Generic CNN-CRF Model for Semantic Segmentation". *arxiv:1511.05067*
(Cited on page 31).
- Kittler, J. (1978). "Feature set search algorithms". *Pattern recognition and signal processing*
(Cited on page 30).
- Ko, C., G.o Sohn, and T. K. Rimmel (2013). "Tree genera classification with geometric features from high-density airborne LiDAR". *Canadian Journal of Remote Sensing* 39.s1, S73–S85
(Cited on pages 41, 105, and 142).
- Koch, B., C. Straub, M. Dees, Y. Wang, and H. Weinacker (2009). "Airborne laser data for stand delineation and information extraction". *International Journal of Remote Sensing* 30.4, pp. 935–963
(Cited on pages ix, 18, and 21).
- Kohavi, R. and G. H. John (1997). "Wrappers for feature subset selection". *Artificial intelligence* 97.1-2, pp. 273–324
(Cited on page 29).
- Kohli, P., L. Ladický, and P. Torr (2009). "Robust Higher Order Potentials for Enforcing Label Consistency". English. *International Journal of Computer Vision* 82.3, pp. 302–324. ISSN: 0920-5691
(Cited on page 31).
- Kolmogorov, V. (2006). "Convergent tree-reweighted message passing for energy minimization". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 28.10, pp. 1568–1583
(Cited on page 111).

- Kolmogorov, V. and R. Zabih (2004). "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.2, pp. 147–159
(Cited on pages [xv](#), [31](#), and [114](#)).
- Kolmogorov, V. and C. Rother (2007). "Minimizing non-submodular functions with graph cuts—a review". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.7, pp. 1274–1279
(Cited on pages [xv](#) and [114](#)).
- Kontschieder, P., M. Fiterau, A. Criminisi, and S. Rota Buló (2015). "Deep neural decision forests". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1467–1475
(Cited on pages [52](#) and [142](#)).
- Korpela, I., H. O. Ørka, M. Maltamo, T. Tokola, and J. Hyypä (2010). "Tree species classification using airborne LiDAR—effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type". *Silva Fennica* 44.2, pp. 319–339
(Cited on page [38](#)).
- Kumar, S. and M. Hebert (2006). "Discriminative random fields." *International Journal of Computer Vision* 68.2, pp. 179–201
(Cited on page [31](#)).
- Ladický, L., C. Russell, P. Kohli, and P. Torr (2012). "Inference Methods for CRFs with Co-occurrence Statistics". *International Journal of Computer Vision* 103.2, pp. 213–225
(Cited on page [31](#)).
- Landgrebe, D. A. (2005). *Signal theory methods in multispectral remote sensing*. Vol. 29. John Wiley & Sons
(Cited on page [26](#)).
- Landrieu, L. and G. Obozinski (2016). "Cut Pursuit: fast algorithms to learn piecewise constant functions". In: *Artificial Intelligence and Statistics*, pp. 1384–1393
(Cited on pages [24](#) and [114](#)).
- Larsen, R. (2002). "Decomposition using maximum autocorrelation factors". *Journal of Chemometrics* 16.8-10, pp. 427–435
(Cited on page [28](#)).
- Le Moan, S., A. Mansouri, Y. Voisin, and J. Y. Hardeberg (2011). "A constrained band selection method based on information measures for spectral image color visualization". *IEEE Transactions on Geoscience and Remote Sensing* 49.12, pp. 5104–5115
(Cited on page [29](#)).
- Le Toan, T., A. Beaudoin, J. Riom, and D. Guyon (1992). "Relating forest biomass to SAR data". *IEEE Transactions on Geoscience and Remote Sensing* 30.2, pp. 403–411
(Cited on page [6](#)).
- Lebart, L., A. Morineau, and M. Piron (1997). "Multidimensional exploratory statistics"
(Cited on page [28](#)).

- Leckie, D. G., F. A. Gougeon, N. Walsworth, and D. Paradine (2003). "Stand delineation and composition estimation using semi-automated individual tree crown analysis". *Remote Sensing of Environment* 85.3, pp. 355–369
(Cited on pages ix, x, xvii, 18, and 21).
- Lee, N. J. and K. Nakane (1997). "Forest vegetation classification and biomass estimation based on Landsat TM data in a mountainous region of west Japan". In: *The use of remote sensing in the modeling of forest productivity*. Springer, pp. 159–171
(Cited on page 40).
- Lemprière, T.C., P.Y. Bernier, A.L. Carroll, M.D. Flannigan, R.P. Gilsenan, D.W. McKenney, E.H. Hogg, J.H. Pedlar, and D. Blain (2008). "The importance of forest sector adaptation to climate change". *Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre*
(Cited on page 3).
- Leppänen, V. J., T. Tokola, M. Maltamo, L. Mehtätalo, T. Pusa, and J. Mustonen (2008). "Automatic delineation of forest stands from LIDAR data". *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(4/C1), pp. 5–8
(Cited on pages x, xvii, 20, and 21).
- Levinshtein, A., A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi (2009). "Turbopixels: fast superpixels using geometric flows". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 31.12, pp. 2290–2297
(Cited on pages 25 and 45).
- Li, J., B. Hu, and T. L. Noland (2013b). "Classification of tree species based on structural features derived from high density LiDAR data". *Agricultural and Forest Meteorology* 171, pp. 104–114
(Cited on page 22).
- (2013a). "Classification of tree species based on structural features derived from high density LiDAR data". *Agricultural and Forest Meteorology* 171–172, pp. 104–114
(Cited on pages xii, 39, 105, and 142).
- Li, S., H. Wu, D. Wan, and J. Zhu (2011). "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine". *Knowledge-Based Systems* 24.1, pp. 40–48
(Cited on page 29).
- Lim, K., P. Treitz, M.I. Wulder, B. St-Onge, and M. Flood (2003). "LiDAR remote sensing of forest structure". *Progress in Physical Geography* 27.1, pp. 88–106
(Cited on page 6).
- Lin, Y. and J. Hyypä (2016). "A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification". *International Journal of Applied Earth Observation and Geoinformation* 46, pp. 45–55
(Cited on pages 105 and 142).
- Liu, H. and R. Setiono (1997). "Feature selection and classification—a probabilistic wrapper approach". In: *Proceedings of 9th International Conference on Industrial and Engineering Applications of AI and ES*, pp. 419–424
(Cited on page 30).

- Liu, L., Y. Pang, W. Fan, Z. Li, and M. Li (2011). "Fusion of airborne hyperspectral and LiDAR data for tree species classification in the temperate forest of northeast China". In: *IEEE Conference on Geoinformatics*, pp. 1–5
(Cited on page 142).
- Lu, D. and Q. Weng (2007). "A survey of image classification methods and techniques for improving classification performance". *International Journal of Remote Sensing* 28.5, pp. 823–870
(Cited on page 26).
- Lucchi, A., Y. Li, X. Boix, K. Smith, and P. Fua (2011). "Are Spatial and Global Constraints Really Necessary for Segmentation?" In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 9–16
(Cited on page 31).
- Maas, A., F. Rottensteiner, and C. Heipke (2016). "Using label noise robust logistic regression for automated updating of topographic geospatial databases". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III-7, pp. 133–140
(Cited on page 48).
- Mallet, C., F. Bretar, M. Roux, U. Soergel, and C. Heipke (2011). "Relevance assessment of full-waveform lidar data for urban area classification". *ISPRS Journal of Photogrammetry and Remote Sensing* 66.6, S71–S84
(Cited on pages xii and 40).
- Marill, T. and D. Green (1963). "On the effectiveness of receptors in recognition systems". *IEEE transactions on Information Theory* 9.1, pp. 11–17
(Cited on page 30).
- Martínez-UsÓ, A., F. Pla, J. Martínez Sotoca, and P. García-Sevilla (2007). "Clustering-based hyperspectral band selection using information measures". *IEEE Transactions on Geoscience and Remote Sensing* 45.12, pp. 4158–4171
(Cited on page 29).
- Mather, P. and B. Tso (2016). *Classification methods for remotely sensed data*. CRC Press
(Cited on page 26).
- Means, J. E., S. A. Acker, B. J. Fitt, M. Renslow, L. Emerson, and C. J. Hendrix (2000). "Predicting forest stand characteristics with airborne scanning lidar". *Photogrammetric Engineering & Remote Sensing* 66.11, pp. 1367–1372
(Cited on pages ix, 4, and 5).
- Mellor, A., A. Haywood, C. Stone, and S. Jones (2013). "The performance of random forests in an operational setting for large area sclerophyll forest classification". *Remote Sensing* 5.6, pp. 2838–2856
(Cited on page 48).
- Mellor, A., S. Boukir, A. Haywood, and S. Jones (2014). "Using ensemble margin to explore issues of training data imbalance and mislabeling on large area land cover classification". In: *IEEE International Conference on Image Processing (ICIP)*, pp. 5067–5071
(Cited on page 48).

- (2015). “Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin”. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 155–168
(Cited on page 48).
- Moore, A. P., S. J.D. Prince, J. Warrell, U. Mohammed, and G. Jones (2008). “Superpixel lattices”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8
(Cited on pages 25 and 45).
- Mora, B., M. A. Wulder, and J. C. White (2010). “Segment-constrained regression tree estimation of forest stand height from very high spatial resolution panchromatic imagery over a boreal environment”. *Remote Sensing of Environment* 114.11, pp. 2474–2484
(Cited on pages 18 and 21).
- Moritz, P., R.t Nishihara, I. Stoica, and M. I. Jordan (2015). “Sparknet: Training deep networks in spark”. *arXiv preprint arXiv:1511.06051*
(Cited on page 26).
- Morsdorf, F., E. Meier, B. Allgöwer, and D. Nüesch (2003). “Clustering in airborne laser scanning raw data for segmentation of single trees”. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34.part 3, W13
(Cited on page 43).
- Moser, G., S.B. Serpico, and J.A. Benediktsson (2013). “Land-cover mapping by Markov modeling of spatial contextual information in Very-High-Resolution Remote Sensing Images”. *Proceedings of the IEEE* 101.3, pp. 631–651
(Cited on pages 31 and 111).
- Muukkonen, P. (2007). “Generalized allometric volume and biomass equations for some tree species in Europe”. *European Journal of Forest Research* 126.2, pp. 157–166
(Cited on page 141).
- Næsset, E. (2002). “Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data”. *Remote Sensing of Environment* 80.1, pp. 88–99
(Cited on page 57).
- Nguyen, A. and B. Le (2013). “3D point cloud segmentation: a survey”. In: *IEEE 6th Conference on Robotics, Automation and Mechatronics (RAM)*, pp. 225–230
(Cited on pages 22 and 25).
- Nitzberg, M., D. Mumford, and T. Shiota (1993). *Filtering, segmentation and depth*
(Cited on page 22).
- Ouerghemmi, W., A. Le Bris, N. Chehata, and C. Mallet (2017). “A two-step decision fusion strategy: application to hyperspectral and multispectral images for urban classification.” *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 42
(Cited on page 142).
- Paisitkriangkrai, S., J. Sherrah, P. Janney, and V.D. Hengel (2015). “Effective semantic pixel labelling with convolutional networks and conditional random fields”. In: *Proceedings of the IEEE Conference*

- on *Computer Vision and Pattern Recognition (CVPR)*, pp. 36–43
(Cited on page 142).
- Paisitkriangkrai, S., J. Sherrah, P. Janney, and A. Van Den Hengel (2016). “Semantic labeling of aerial and satellite imagery”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.7, pp. 2868–2881
(Cited on pages 52 and 142).
- Pal, M. (2005). “Random forest classifier for remote sensing classification”. *International Journal of Remote Sensing* 26.1, pp. 217–222
(Cited on page 27).
- Pal, N. R. and S. K. Pal (1993). “A review on image segmentation techniques”. *Pattern Recognition* 26.9, pp. 1277–1294
(Cited on page 22).
- Pan, Y., R. A. Birdsey, O. L. Phillips, and R. B. Jackson (2013). “The structure, distribution, and biomass of the world’s forests”. *Annual Review of Ecology, Evolution, and Systematics* 44, pp. 593–622
(Cited on pages 2 and 3).
- Paola, J.D. and R.A. Schowengerdt (1995). “A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery”. *International Journal of remote sensing* 16.16, pp. 3033–3058
(Cited on page 26).
- Paris, S., P. Kornprobst, J. Tumblin, and F. Durand (2009a). “Bilateral filtering: theory and applications”. *Foundations and Trends in Computer Graphics and Vision* 4.1, pp. 1–73
(Cited on page 108).
- Paris, Sylvain and Frédo Durand (2009b). “A fast approximation of the bilateral filter using a signal processing approach”. *International Journal of Computer Vision* 81.1, pp. 24–52
(Cited on page 108).
- Patenaude G. and Milne, R. and T. P. Dawson (2005). “Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol”. *Environmental Science & Policy* 8.2, pp. 161–178
(Cited on page 5).
- Peli, T. and D. Malah (1982). “A study of edge detection algorithms”. *Computer Graphics and Image Processing* 20.1, pp. 1–21
(Cited on page 23).
- Pelletier, C., S. Valero, J. Inglada, N. Champion, and G. Dedieu (2016). “Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas”. *Remote Sensing of Environment* 187, pp. 156–168
(Cited on page 48).
- Perona, P. and J. Malik (1990). “Scale-space and edge detection using anisotropic diffusion”. *IEEE Transactions on pattern analysis and machine intelligence* 12.7, pp. 629–639
(Cited on page 30).

- Pommerening, A. (2002). "Approaches to quantifying forest structures". *Forestry: An International Journal of Forest Research* 75.3, pp. 305–324
(Cited on page 4).
- Postadjian, T., A. Le Bris, H. Sahbi, and C. Mallet (2017). "Investigating the potential of deep neural networks for large-scale classification of very high resolution satellite images." *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 4
(Cited on page 142).
- Pudil, P., J. Novovičová, and J. Kittler (1994). "Floating search methods in feature selection". *Pattern Recognition Letters* 15.11, pp. 1119–1125
(Cited on pages xiii, 30, and 47).
- Radoux, J. and P. Defourny (2007). "A quantitative assessment of boundaries in automated forest stand delineation using very high resolution imagery". *Remote Sensing of Environment* 110.4, pp. 468–475
(Cited on page 16).
- Radoux, J., C. Lamarche, E. Van Bogaert, S. Bontemps, C. Brockmann, and P. Defourny (2014). "Automated Training Sample Extraction for Global Land Cover Mapping". *Remote Sensing* 6.5, p. 3965. ISSN: 2072-4292. DOI: 10.3390/rs6053965
(Cited on page 48).
- Ramage, M. H., H. Burrige, M. Busse-Wicher, G. Fereday, T. Reynolds, D. U. Shah, G. Wu, L. Yu, P. Fleming, D. Densley-Tingley, J. Allwood, P. Dupree, P.F. Linden, and Scherman O. (2017). "The wood from the trees: The use of timber in construction". *Renewable and Sustainable Energy Reviews* 68, pp. 333–359
(Cited on page 3).
- Reitberger, J., C. Schnörr, P. Krzystek, and U. Stilla (2009). "3D segmentation of single trees exploiting full waveform LIDAR data". *ISPRS Journal of Photogrammetry and Remote Sensing* 64.6, pp. 561–574
(Cited on page 26).
- Rodriguez, J. J., L. I. Kuncheva, and C. J. Alonso (2006). "Rotation forest: A new classifier ensemble method". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 28.10, pp. 1619–1630
(Cited on page 27).
- Ronfard, R. (1994). "Region-based strategies for active contour models". *International Journal of Computer Vision* 13.2, pp. 229–251
(Cited on page 50).
- Salembier, P. and L. Garrido (2000). "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval". *IEEE Transactions on Image Processing* 9.4, pp. 561–576
(Cited on page 24).
- Schindler, K. (2012). "An overview and comparison of smooth labeling methods for land-cover classification". *IEEE Transactions on Geoscience and Remote Sensing* 50.11, pp. 4534–4545
(Cited on pages xiv, 10, 30, 31, 51, 108, and 111).

- Scholkopf, B. and A. J. Smola (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press
(Cited on page 26).
- Schuck, A., R. Päivinen, T. Hytönen, and B. Pajari (2002). *Compilation of forestry terms and definitions*. European Forest Institute
(Cited on page 2).
- Serpico, S. B. and G. Moser (2007). "Extraction of spectral channels from hyperspectral images for classification purposes". *IEEE Transactions on Geoscience and Remote Sensing* 45.2, pp. 484–495
(Cited on page 29).
- Shan, J. and C. K. Toth (2008). *Topographic laser ranging and scanning: principles and processing*. CRC Press
(Cited on page 6).
- Sheikh, Y. A., E. A. Khan, and T. Kanade (2007). "Mode-seeking by medoidshifts". In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8
(Cited on page 46).
- Shi, J. and J. Malik (2000). "Normalized cuts and image segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 22.8, pp. 888–905
(Cited on pages 24 and 45).
- Smeeckaert, J., C. Mallet, N. David, N. Chehata, and A. Ferraz (2013). "Large-scale classification of water areas using airborne topographic lidar data". *Remote Sensing of Environment* 138, pp. 134–148
(Cited on pages 31 and 109).
- Smith, J. M. and D. Smith (1977). "Database abstractions: aggregation and generalization". *ACM Transactions on Database Systems (TODS)* 2.2, pp. 105–133
(Cited on page 10).
- Smith, W. H. (2012). *Air pollution and forests: interactions between air contaminants and forest ecosystems*. Springer Science & Business Media
(Cited on page 3).
- Sotoca, J. M. and F. Pla (2010). "Supervised feature selection by clustering using conditional mutual information-based distances". *Pattern Recognition* 43.6, pp. 2068–2081
(Cited on page 29).
- Souchon, J.-P., C. Thom, C. Meynard, and O. Martin (2012). "A large format camera system for national mapping purposes". *Revue Française de Photogrammétrie et de Télédétection* 200, pp. 48–53
(Cited on pages xv, 11, and 56).
- Sterrett, F. S. (1994). *Alternative fuels and the environment*. CRC Press
(Cited on page 3).
- Strahler, A. H. (1980). "The use of prior probabilities in maximum likelihood classification of remotely sensed data". *Remote sensing of Environment* 10.2, pp. 135–163
(Cited on page 26).

- Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn (2007). "Bias in random forest variable importance measures: Illustrations, sources and a solution". *BMC Bioinformatics* 8.1, p. 25
(Cited on page 27).
- Strunk, J., H. Temesgen, H.-E. Andersen, J. P. Flewelling, and L. Madsen (2012). "Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables". *Canadian Journal of Remote Sensing* 38.05, pp. 644–654
(Cited on page 6).
- Stutz, D., A. Hermans, and B. Leibe (2017). "Superpixels: An evaluation of the state-of-the-art". *Computer Vision and Image Understanding*
(Cited on page 22).
- Sullivan, A. A, R. J. McGaughey, H.-E. Andersen, and P. Schiess (2009). "Object-oriented classification of forest structure from light detection and ranging data for stand mapping". *Western Journal of Applied Forestry* 24.4, pp. 198–204
(Cited on pages ix, 19, and 21).
- Tang, J., S. Alelyani, and H. Liu (2014). "Feature selection for classification: A review". *Data Classification: Algorithms and Applications*, p. 37
(Cited on page 29).
- Taxt, T., P. J. Flynn, and A. K. Jain (1989). "Segmentation of document images". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.12, pp. 1322–1329
(Cited on page 23).
- Thelandersson, S. and H. J. Larsen (2003). *Timber engineering*. John Wiley & Sons
(Cited on page 3).
- Tiede, D., T. Blaschke, and M. Heurich (2004). "Object-based semi automatic mapping of forest stands with Laser scanner and Multi-spectral data". *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36(8/W2), pp. 328–333
(Cited on pages x, 19, and 21).
- Tochon, G. (2015). "Hierarchical analysis of multimodal images". PhD thesis. Université Grenoble Alpes
(Cited on page 41).
- Tochon, G., J.B. Féret, S. Valero, R.E. Martin, D.E. Knapp, P. Salembier, J. Chanussot, and G.P. Asner (2015). "On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images". *Remote Sensing of Environment* 159, pp. 318–331
(Cited on page 24).
- Tokarczyk, P., J. D. Wegner, S. Walk, and K. Schindler (2015). "Features, color spaces, and boosting: New insights on semantic classification of remote sensing images". *IEEE Transactions on Geoscience and Remote Sensing* 53.1, pp. 280–295
(Cited on page 28).
- Tokola, T. (2015). "Remote Sensing Concepts and Their Applicability in REDD+ Monitoring". *Current Forestry Reports* 1.4, pp. 252–260
(Cited on page 5).

- Tomppo, E., H. Olsson, G. Ståhl, M. Nilsson, O. Hagner, and M. Katila (2008). "Combining national forest inventory field plots and remote sensing data for forest databases". *Remote Sensing of Environment* 112.5, pp. 1982–1999
(Cited on page 5).
- Torabzadeh, H., F. Morsdorf, and M. E. Schaepman (2014). "Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems—A review". *ISPRS Journal of Photogrammetry and Remote Sensing* 97, pp. 25–35
(Cited on pages ix, 5, and 57).
- Torabzadeh, H., R. Leiterer, M.E. Schaepman, and F. Morsdorf (2015). "Optimal structural and spectral features for tree species classification using combined airborne laser scanning and hyperspectral data". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5399–5402
(Cited on pages xii, 39, and 142).
- Trias-Sanz, R. (2006). "Semi-automatic rural land cover classification from high-resolution remote sensing images". PhD thesis. Université Paris 5
(Cited on page 23).
- Tucker, C. J. (1979). "Red and photographic infrared linear combinations for monitoring vegetation". *Remote sensing of Environment* 8.2, pp. 127–150
(Cited on page 40).
- Tuia, D., F. Pacifici, M. Kanevski, and W. J. Emery (2009). "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines". *IEEE Transactions on Geoscience and Remote Sensing* 47.11, pp. 3866–3879
(Cited on page 29).
- Tuia, D., M. Volpi, and G. Moser (2016). "Getting pixels and regions to agree with conditional random fields". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5399–5402
(Cited on page xiv).
- Ullrich, A. and M. Pfennigbauer (2016). "Linear LIDAR versus Geiger-mode LIDAR: impact on data properties and data quality". In: *SPIE Defense+ Security*. International Society for Optics and Photonics, pp. 983204–983204
(Cited on page 6).
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media
(Cited on pages 48 and 74).
- Vauhkonen, J., I. Korpela, M. Maltamo, and T. Tokola (2010). "Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics". *Remote Sensing of Environment* 114.6, pp. 1263–1276
(Cited on page 41).
- Vauhkonen, J., T. Hakala, J. Suomalainen, S. Kaasalainen, O. Nevalainen, M. Vastaranta, M.s Holopainen, and J. Hyyppä (2013). "Classification of spruce and pine trees using active hyper-spectral LiDAR". *IEEE Geoscience and Remote Sensing Letters* 10.5, pp. 1138–1141
(Cited on page 38).

- Vedaldi, A. and S. Soatto (2008). "Quick shift and kernel methods for mode seeking". *Computer vision–ECCV 2008*, pp. 705–718
(Cited on pages 24, 45, and 46).
- Véga, C., A. Hamrouni, S. El Mokhtari, J. Morel, J. Bock, J.-P. Renaud, M. Bouvier, and S. Durrieu (2014). "PTrees: a point-based approach to forest tree extraction from lidar data". *International Journal of Applied Earth Observation and Geoinformation* 33, pp. 98–108
(Cited on pages 16, 26, and 43).
- Veksler, O., Y. Boykov, and P. Mehrani (2010). "Superpixels and supervoxels in an energy optimization framework". *Computer Vision–ECCV 2010*, pp. 211–224
(Cited on pages 25 and 45).
- Vincent, L. and P. Soille (1991). "Watersheds in digital spaces: an efficient algorithm based on immersion simulations". *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6, pp. 583–598
(Cited on pages 23, 24, 44, and 45).
- Viterbini, M., A. Adriani, and G. Di Donfrancesco (1987). "Single photon detection and timing system for a lidar experiment". *Review of scientific instruments* 58.10, pp. 1833–1839
(Cited on page 6).
- Vogt, K. A., J.M. Honea, D.J. Vogt, T. Patel-Weynand, R.L. Edmonds, R. Sigurdardottir, D.G. Briggs, and M.G. Andreu (2006). "Global societies and forest legacies creating today's forest landscapes." *Forests and society: sustainability and life cycles of forests in human landscapes*, pp. 30–59
(Cited on page 3).
- Volpi, M. and V. Ferrari (2015). "Semantic segmentation of urban scenes by learning local class interactions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–9
(Cited on pages xiv, 109, and 143).
- Vosselman, G. and H.-G. Maas (2010). *Airborne and terrestrial laser scanning*. CRC Press
(Cited on page 6).
- Wang, Y., J. Hyypä, X. Liang, H. Kaartinen, X. Yu, E. Lindberg, J. Holmgren, Y. Qin, C. Mallet, A. Ferraz, H. Torabzadeh, F. Morsdorf, L. Zhu, J. Liu, and P. Alho (2016). "International Benchmarking of the Individual Tree Detection Methods for Modeling 3D Canopy Structure for Silviculture and Forest Ecology Using Airborne Laser Scanning". *IEEE Transactions on Geoscience and Remote Sensing*, 154.9, pp. 501–502
(Cited on pages 25, 43, and 64).
- Wang, Z., R. Boesch, and C. Ginzler (2012). "Forest delineation of aerial images with Gabor wavelets". *International Journal of Remote Sensing* 33.7, pp. 2196–2213
(Cited on page 16).
- Wasiq, M. and M. Ahmad (2004). *Sustaining forests: A development strategy*
(Cited on page 3).

- Wegner, J. D., J. A. Montoya-Zegarra, and K. Schindler (2013). "A higher-order CRF model for road network extraction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1698–1705
(Cited on page 143).
- (2015). "Road networks as collections of minimum cost paths". *ISPRS Journal of Photogrammetry and Remote Sensing* 108, pp. 128–137
(Cited on page 143).
- Wegner, J. D., S. Branson, D. Hall, K. Schindler, and P. Perona (2016). "Cataloging public objects using aerial and street-level images-urban trees". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6014–6023
(Cited on page 142).
- Weinmann, M. (2016). "Reconstruction and Analysis of 3D Scenes". *Springer International Publishing, Cham*. doi 10, pp. 978–3
(Cited on pages xii and 40).
- Weinmann, M., B. Jutzi, S. Hinz, and C. Mallet (2015). "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers". *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 286–304
(Cited on pages xi and 39).
- White, J. C., N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski (2016). "Remote sensing technologies for enhancing forest inventories: A review". *Canadian Journal of Remote Sensing* 42.5, pp. 619–641
(Cited on pages xv and 5).
- Whitney, A. W. (1971). "A direct method of nonparametric measurement selection". *IEEE Transactions on Computers* 100.9, pp. 1100–1103
(Cited on page 30).
- Wilson, R. and M. Spann (1988). *Image segmentation and uncertainty*. John Wiley & Sons, Inc.
(Cited on page 22).
- Workman, Scott, Menghua Zhai, David J Crandall, and Nathan Jacobs (2017). "A Unified Model for Near and Remote Sensing". In: *IEEE International Conference on Computer Vision (ICCV)*
(Cited on pages 52 and 142).
- Wu, Z., V. Heikkinen, M. Hauta-Kasari, J. Parkkinen, and T. Tokola (2014). "ALS data based forest stand delineation with a coarse-to-fine segmentation approach". In: *IEEE Congress on Image and Signal Processing (CISP)*, pp. 547–552
(Cited on pages x, 19, and 21).
- Wulder, M. A., C. W. Bater, N. C. Coops, T. Hilker, and J. C. White (2008a). "The role of LiDAR in sustainable forest management". *The Forestry Chronicle* 84.6, pp. 807–826
(Cited on page 5).
- Wulder, M. A., J. C. White, G. J. Hay, and G. Castilla (2008b). "Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery". *The Forestry Chronicle* 84.2,

- pp. 221–230
(Cited on pages 5 and 18).
- Yang, H.-C., S.-B. Zhang, K.-Z. Deng, and P.-J. Du (2007). “Research into a feature selection method for hyperspectral imagery using PSO and SVM”. *Journal of China University of Mining and Technology* 17.4, pp. 473–478
(Cited on pages 29 and 30).
- Yanowitz, S. D. and A. M. Bruckstein (1989). “A new method for image segmentation”. *Computer Vision, Graphics and Image Processing* 46.1, pp. 82–95
(Cited on page 23).
- Zargar, A., R. Sadiq, B. Naser, and F. I. Khan (2011). “A review of drought indices”. *Environmental Reviews* 19, pp. 333–349
(Cited on pages xii and 40).
- Zhang, L., Y. Zhong, B. Huang, J. Gong, and P. Li (2007). “Dimensionality reduction based on clonal selection for hyperspectral imagery”. *IEEE Transactions on Geoscience and Remote Sensing* 45.12, pp. 4172–4186
(Cited on page 29).
- Zhang, Y.-J. (2006). *Advances in image and video segmentation*. IGI Global
(Cited on page 22).
- Zhuo, L., J. Zheng, X. Li, F. Wang, B. Ai, and J. Qian (2008). “A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine”. In: *Geoinformatics and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images*, 71471J–71471J
(Cited on page 29).