



**Estimation et suivi de la ressource en bois en France
métropolitaine par valorisation des séries
multi-temporelles à haute résolution spatiale d'images
optiques (Sentinel-2) et radar (Sentinel-1,
ALOS-PALSAR)**

David Morin

► **To cite this version:**

David Morin. Estimation et suivi de la ressource en bois en France métropolitaine par valorisation des séries multi-temporelles à haute résolution spatiale d'images optiques (Sentinel-2) et radar (Sentinel-1, ALOS-PALSAR). Milieux et Changements globaux. Université Paul Sabatier - Toulouse III, 2020. Français. NNT : 2020TOU30079 . tel-03098487

HAL Id: tel-03098487

<https://theses.hal.science/tel-03098487>

Submitted on 5 Jan 2021

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.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par

David MORIN

Le 6 février 2020

Titre :

Estimation et suivi de la ressource en bois en France métropolitaine par valorisation des séries multi-temporelles à haute résolution spatiale d'images optiques (Sentinel-2) et radar (Sentinel-1, ALOS-PALSAR)

École doctorale et discipline ou spécialité :

ED SDU2E : Surfaces et interfaces continentales, Hydrologie

Unité de recherche :

CESBIO UMR 5126

Directeur/trice(s) de Thèse :

Gérard DEDIEU

Milena PLANELLS

Jury

Pierre-Louis FRISON
Guerric LE MAIRE
Dominique GUYON
Anne JOLLY
Mehrez ZRIBI
Nicolas BAGHDADI
Éric CESCHIA
Gérard DEDIEU
Milena PLANELLS
Miriam BUITRAGO

Maître de conférence, Université Paris-Est-Marne
Chercheur, CIRAD UMR Eco&Sols, Montpellier
Ingénieure de Recherche, INRA UMR ISPA, Bordeaux
Chargée de Recherche, ONF, Nancy
Directeur de Recherche, CNRS UMR CESBIO, Toulouse
Directeur de Recherche, IRSTEA UMR TETIS, Montpellier
Directeur de Recherche, INRA UMR CESBIO, Toulouse
Directeur de Recherche, CNES UMR CESBIO, Toulouse
Ingénieure, CNES UMR CESBIO, Toulouse
Ingénieure, ADEME, Angers

Rapporteur
Rapporteur
Examinatrice
Examinatrice
Examinateur
Examinateur
Examinateur
Directeur de thèse
Invitée, co-directrice de thèse
Invitée



Ce travail de thèse a été réalisé au Centre d'Études Spatiales de la BIOSphère (CESBIO) avec le soutien financier de l'Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME) et du Centre National d'Études Spatiales (CNES).

L'Office Nationale des Forêts (ONF) et l'Institut National de Recherche Agronomique (INRA) ont permis l'obtention des données terrain utilisées dans la thèse.



Remerciements

Ce travail de thèse a été réalisé grâce aux financements de l'ADEME et du CNES. Je remercie Gueric Lemaire et Pierre-Louis Frison d'avoir été rapporteurs, Éric Céschia qui a présidé le jury, et les examinateurs Mehrez Zribi, Anne Jolly, Nicolas Baghdadi et Dominique Guyon. Merci à Miriam Buitrago qui a suivi mon travail pour l'ADEME. Enfin, un grand merci à Gérard Dedieu et Milena Planells qui ont dirigé ma thèse.

Mon travail d'apprentissage de la recherche a commencé quelques années plus tôt en stage au pôle satellitaire du CETE (actuellement CEREMA) puis par un peu plus de trois ans de contrat au CESBIO. J'adresse un grand merci à tous ceux qui m'ont fait confiance pour travailler avec eux, découvrir le monde de la recherche, et l'apprécier jusqu'à vouloir concrétiser cela par une thèse de doctorat : Dominique Hebrard, Jacques Bouffier, Valérie Demarez, Olivier Hagolle, Silvia Valero, Jordi Inglada, et Gérard Dedieu.

La construction de ce projet de thèse n'aurait pu se faire sans Jean-François Dejoux, Gérard Dedieu et Éric Paillassa qui m'ont grandement aidé à monter le programme scientifique et établir les contacts nécessaires. Les membres du Comité de Pilotage de la thèse ont contribué au bon déroulement de mon travail, particulièrement pour éviter trop dispersions : David Sheeren, Éric Paillassa, Cédric Véga, Jean-Pierre Vigneron et Jean-François Dejoux. Je remercie les membres de l'équipe BIOMASS pour leur appui et leur suivi, notamment Stéphane Mermoz, Thuy Le Toan, Alexandre Bouvet, Milena Planells et Ludovic Villard. Enfin, un grand merci à ceux qui m'ont permis d'avoir des données de terrain pour travailler : l'ONF avec Alexandre Piboule, Jérôme Bock, Anne Jolly et al. ; l'INRA avec Dominique Guyon, Sylvia Dayau, Bernard Issenhut et al.

Je remercie également tous ceux avec qui j'ai pu échanger durant ces années pour le travail ou la vie personnelle, dans le désordre et de manière non exhaustive : Mohamed, Isabel, Amanda, Stéphane, Marcela, Renaud, Juliette, Arthur, Charlotte, Lucas, Hoa, Arnaud, les Émilies, les Hervés, Marvin, Beatriz, Christophe, Marie, Dawa... J'en passe et des meilleurs.

Enfin, merci à Anne Marie qui m'a soutenu pendant cette trentaine d'années dans mes réussites comme dans mes échecs, merci aussi à Thomas, merci à Seb et Alexis qui m'ont permis de m'aérer l'esprit et décompresser autour de bonnes bières belges à parler politique, jeux vidéo, psychologie et lignes de code informatique, et merci à Christina qui m'a soutenu au jour le jour et a réussi à me supporter (dans les deux sens du terme) en fin de thèse.

Table des matières

Remerciements.....	v
Liste des figures.....	ix
Liste des tableaux.....	xii
Liste des sigles et des acronymes.....	xiii
Résumé.....	xv
Abstract.....	xvii
Chapitre I : Introduction.....	1
1.1. Contexte.....	1
1.2. Évaluation des ressources forestières.....	2
Mesures terrain.....	3
Télédétection de la structure 3D des forêts.....	3
Télédétection spatiale à couverture mondiale.....	4
1.4. Objectifs de recherche et structure du manuscrit.....	13
Chapitre II : Construction de la méthode.....	15
2.1. Site d'étude.....	15
2.2. Mesures de référence terrain.....	17
Protocole.....	17
Statistiques descriptives.....	18
2.3. Données explicatives.....	19
Rétrodiffusion radar en bande L (ALOS2-PALSAR2).....	20
Rétrodiffusion radar en bande C (Sentinel-1).....	21
Réflectance optique (Sentinel-2).....	21
Indices de textures.....	22
Imagerie THRS (SPOT-6/7).....	23
2.4. Pré-analyse des données.....	24
2.4.1. Analyse des paramètres forestiers.....	24
2.4.2. Analyse des séries temporelles.....	26
2.4.2.1. Analyse du profil temporel Sentinel-2.....	26
2.4.2.2. Analyse du profil temporel Sentinel-1.....	27
2.4.3. Analyse des textures.....	29
2.4.4. Relations entre les paramètres forestiers et les données satellitaires.....	32
2.4.5. Corrélations entre primitives (variables explicatives).....	33
2.5. Approche méthodologique.....	35
2.6. Extraction des primitives.....	37
2.7. Réduction de dimensions.....	39
2.7.1. Analyses en Composantes Principales.....	39
2.7.2. Méthodes de sélection de primitives « stepwise ».....	41
2.7.3. Comparaison entre l'ACP et les méthodes de sélection de primitives.....	42
2.8. Algorithmes de régression.....	43
2.8.1 Comparaison des algorithmes.....	44
2.8.2. Paramétrisation du SVR.....	45
2.8.3. Évaluation des modèles.....	47
2.9. Résultats.....	49
2.9.1. Hauteur moyenne (Height).....	51
2.9.2. Diamètre moyen à 130 cm / mean DBH.....	52
2.9.3. Âge des peuplements.....	53
2.9.4. Densité d'arbres.....	53
2.9.5. Surface terrière / Basal Area (BA).....	54
2.9.6. Aboveground Biomass (AGB).....	55
2.10. Analyse de l'apport des types de données satellitaires.....	56
Ajout de textures à très haute résolution spatiale.....	58
2.11. Conclusion.....	60
Chapitre III : Généralisation de la méthode.....	61

3.1. Application à une même essence.....	61
3.1.1. Présentation des données.....	61
3.1.2. Estimation des paramètres forestiers.....	65
1) Application de la méthode au site Landes-ONF.....	65
2) Application des modèles du site Landes-INRA au site Landes-ONF.....	67
3) Meilleurs résultats possibles du site Landes-INRA vers le site Landes-ONF.....	71
4) Application de la méthode aux deux sites réunis.....	72
Conclusions sur cette expérience.....	73
3.2. Application à plusieurs essences forestières.....	74
3.2.1. Présentation des nouveaux sites d'étude.....	74
Forêt de Saint Gobain.....	74
Forêt de Les Bordes (Orléans).....	75
Surface représentative pour l'extraction des primitives.....	75
3.2.2. Analyse des jeux de données.....	76
Distribution des paramètres forestiers.....	76
Relations entre paramètres forestiers et données satellitaires.....	78
3.2.3. Application de la méthode.....	79
3.2.4. Impact de la nomenclature des types de forêts.....	83
Exemple chêne.....	83
Toutes les essences.....	84
Séparation entre feuillus et conifères.....	84
3.3. Conclusion.....	86
Chapitre IV : Cartographie des forêts.....	87
4.1. Cartes des types de forêts.....	87
4.1.1. Données existantes.....	87
Masques des forêts.....	87
Cartes des types/essences des forêts.....	88
4.1.2. Production de cartes des forêts.....	89
Méthode de classification.....	89
Cartes de classification.....	89
4.2. Spatialisation des modèles.....	91
4.2.1. Méthodes de spatialisation.....	91
Approche pixel.....	91
Approche objet (segmentation).....	91
4.2.2. Évaluation des différents approches.....	92
Distribution des valeurs.....	92
Validation des estimations sur de placettes de référence.....	93
Aspect visuel des cartes.....	94
4.3. Analyse des cartes de paramètres forestiers.....	95
Site Orléans.....	95
Site Saint-Gobain.....	98
Sites Landes.....	98
4.4. Conclusion.....	100
Chapitre V : Conclusion.....	101
5.1. Conclusions.....	101
5.2. Perspectives.....	103
Bibliographie.....	105
Annexes.....	119
Annexe A : Mesures et seuil de pré-comptage.....	120
Annexe B : Variation du signal radar avec l'humidité.....	122
Annexe C : Communication scientifique.....	125
C.1. Journal international à comité de lecture.....	125
C.2. Conférence internationale à comité de lecture.....	126
C.3. Conférence nationale.....	126

Liste des figures

Figure I.1 : Évolution de la surface forestière entre 1990 et 2015 (FAO, 2015).....	2
Figure I.2 : Illustration de la sensibilité des signaux optiques et radar aux différents éléments des arbres.....	6
Figure I.3 : Mesure de télédétection dans le domaine optique (Bacour, 2001).....	6
Figure I.4 : Signature spectrale optique de différents types de végétation et d'occupation du sol (Nicolas and Dusseux, 2016).....	7
Figure I.5 : Représentation simplifiée d'un système de radar à synthèse d'ouverture (SAR).....	8
Figure I.6 : Représentation de l'angle d'incidence θ et des polarisations incidentes horizontale (champ E_h) et verticale (champ E_v). (Proisy, 1999).....	8
Figure I.7 : Sensibilité du signal radar aux feuilles, branches et troncs selon la bande de fréquence SAR utilisée. (Le Toan, 2007).....	9
Figure I.8 : Visualisation d'une image de texture spatiale calculée selon l'indice d'homogénéité (textures de Haralick) à partir de le l'indice de brillance de l'image. (a) image SPOT-XS sur des parcelles de pin maritime ; (b) image de texture (homogénéité) sur la même zone. Les pixels les plus brillants reflètent une plus grande homogénéité dans la répartition des niveaux de gris.....	10
Figure I.9 : Exemple d'un calcul d'une matrice de co-occurrence avec une fenêtre de 5×5 pixels centrée sur le pixel cible, une orientation de 0° , un décalage (offset) de 1 pixel, et 4 niveaux de gris (de 0 à 3). Les paires de pixels sont comptabilisées dans la matrice de co-occurrence, les indices statistiques de Haralick sont ensuite calculés à partir de cette matrice.....	11
Figure II.1 : Localisation du site d'étude dans les Landes de Gascogne. Fond de carte : occupation du sol en 2016 (Inglada et al., 2017).....	16
Figure II.2 : Visualisation du site d'étude sur une image optique (Sentinel-2). Les parcelles échantillonnées apparaissent en rouge.....	17
Figure II.3 : Relations au niveau placette entre les paramètres forestiers. La transversale contient les histogrammes (10 classes) de distribution de chaque paramètre.....	25
Figure II.4 : Exemple de deux parcelles avec la même biomasse aérienne, mais différentes relations allométriques. (a) AGB=118 tonnes/ha, âge=19 ans, DBH=18 cm, densité=1463 arbres/ha ; (b) AGB=116 tonnes/ha, âge=68 ans, DBH=57 cm, densité=92 arbres/ha.....	25
Figure II.5 : Profils temporels NDVI, BI et NDWI sur des peuplements de pins maritime (« Conifer ») et de feuillus (« Broadleaved »), site des Landes.....	26
Figure II.6 : Profils temporels Sentinel-1 VH, VV et le ratio VH/VV. (a) Site d'étude des Landes, avec indices spectraux pour la comparaison ; (b) autre site d'étude dans la forêt d'Orléans (chênes) avec une plus grande profondeur temporelle.....	27
Figure II.7 : Profils temporels Sentinel-1 VH, VV et le ratio VH/VV sur des parcelles de pin maritime (« Conifer ») et de feuillus (« Broadleaved »), par date et avec moyenne mensuelle.....	28
Figure II.8 : Visualisation d'une image Sentinel-1 VH en juillet 2016 (a) avant filtrage, (b) après filtrage multi-image, et (c) avec moyenne des images de juillet et août 2016.....	28
Figure II.9 : Problème d'hétérogénéité dans les images de textures. (a) indice de texture corrélation sur une image Sentinel-1, (b) spatialisation d'un modèle de prédiction du DBH où l'image de gauche a été utilisée.....	29
Figure II.10 : Distribution des valeurs de l'indice de texture « homogénéité » (niveau pixel et moyenné par polygone) selon les paramètres de calcul (de haut en bas : rayon de fenêtre « rad » de 1 à 5 ; de gauche à droite : nombre de niveau de gris « nbins » de ~0.3 et ~0.6 fois le nombre de pixels) et à droite, la corrélation linéaire avec les paramètres forestiers (au niveau polygone). Image d'entrée : C-VH-win (tableau II.6).....	31
Figure II.11 : Coefficients de détermination r^2 entre les 31 primitives satellitaires et les 6 paramètres forestiers. Les valeurs des primitives ont été extraites sur des polygones autour des placettes (voir section 2.6).....	33
Figure II.12 : Inter-corrélations des primitives satellitaires au niveau polygone.....	34
Figure II.13 : Schéma résumé de l'approche méthodologique.....	36
Figure II.14 : Parcelles de pin maritime sur la zone d'étude. (a) Photographie aérienne de l'IGN, 2015 ; (b) Sentinel-2 fausses couleurs, été 2016 ; (c) Sentinel-1 VH, été 2016 ; (d) ALOS2-PALSAR2 HV, 2016.....	37
Figure II.15 : Différentes surfaces d'intérêt au sein des peuplements de pin maritime.....	37
Figure II.16 : Différences entre l'extraction des primitives au niveau polygone et au niveau placette pour les relations linéaires avec l'AGB et la hauteur. Les barres vers le haut indiquent un r^2 plus fort au niveau polygone, les barres vers le bas indiquent un r^2 plus fort au niveau placette.....	38
Figure II.17 : RMSEs relatifs des estimations des 6 paramètres forestiers avec un SVR et une sélection Forward, selon l'extraction des valeurs de primitives sur la placette mesurée (« Real plots ») ou sur une zone élargie (« Polygons »).....	39

Figure II.18 : Analyse en composantes principales sur l'ensemble des 31 primitives.....	40
Figure II.19 : RMSEs relatifs des estimations des 6 paramètres forestiers avec un SVR et une sélection Forward, avec une ACP à partir de l'ensemble des 31 primitives (PCA all-features) ou par capteur (PCA sensors).....	40
Figure II.20 : Profil de sélection Forward des primitives pour l'estimation du DBH moyen avec un SVR. Les primitives indiquées en abscisse sont celles qui sont ajoutées à chaque itération.....	42
Figure II.21 : Comparaison des performances pour l'estimation des 6 paramètres forestiers selon la méthode de réduction de dimension et l'algorithme qui sont utilisés.....	43
Figure II.22 : Comparaison des performances des algorithmes de régression MLR, RF et SVR avec une sélection Forward sur les 6 paramètres forestiers. Les algorithmes RF et SVR ont fait l'objet d'une optimisation des paramètres.....	45
Figure II.23 : Évolution du RMSE relatif obtenu pour l'estimation du DBH et de la hauteur en faisant varier les paramètres C, G et E du SVR. Résultats obtenus avec 2/3 des échantillons pour l'apprentissage (Train) et 1/3 pour la validation (Test), sur 100 itérations. Chaque paramètre est analysé en fixant les deux autres paramètres (valeurs de l'optimisation).....	46
Figure II.24 : Effet de l'évolution conjointe des paramètres C et G du SVR sur le RMSE relatif pour l'estimation du DBH. Le paramètre E est fixé à 0.001.....	47
Figure II.25 : Évolution des performances (RMSE relatif) pour les 6 paramètres forestiers en fonction du ratio apprentissage(train)/validation(test). La moyenne (ligne) et l'écart type (zones pleines autour) sont calculés sur 500 divisions aléatoires. La validation LOO est calculée sur la même échelle de 10 à 100% des échantillons utilisés en entrée, avec 500 divisions aléatoires.....	48
Figure II.26 : Profils de sélection Forward pour l'estimation des paramètres forestiers avec un SVR. L'appartenance des primitives aux 5 différents types est indiquée par les couleurs des carrés.....	49
Figure II.27 : Graphiques de validation des résultats d'estimation de la hauteur moyenne (Height).....	52
Figure II.28 : Graphiques de validation des résultats d'estimation du diamètre moyen à 130 cm (DBH).....	52
Figure II.29 : Graphiques de validation des résultats d'estimation de l'âge.....	53
Figure II.30 : Graphiques de validation des résultats d'estimation de la densité d'arbres (Density).....	54
Figure II.31 : Graphiques de validation des résultats d'estimation de la surface terrière (BA).....	54
Figure II.32 : Graphiques de validation des résultats d'estimation de la biomasse aérienne (AGB).....	55
Figure II.33 : Photographies aériennes des zones de mesures dans deux peuplements de pin maritime. Les cercles jaunes représentent la zone de mesure, les zones illustrées font 200 × 170 m. (a) AGB = 51 tonnes/ha (estimée à 81 tonnes/ha), âge = 51 ans, densité = 91 arbres/ha ; (b) AGB = 56 tonnes/ha (estimée à 82 tonnes/ha), âge = 14 ans, densité = 1370 arbres/ha.....	56
Figure II.34 : Comparaison des RMSE relatifs obtenus pour l'estimation des 6 paramètres forestiers avec chaque différent type de primitives, les données Sentinel-1 et Sentinel-2, et la combinaison des 31 primitives. Utilisation d'un algorithme SVR avec une sélection Forward.....	57
Figure II.35 : RMSE relatifs en tableaux croisés des combinaisons entre types de données satellitaires. Les meilleures combinaisons sont encadrées en blanc.....	58
Figure II.36 : Exemple des problèmes radiométriques pour les mosaïques SPOT. (a) mosaïque couleur sur Landes de Gascogne, (b) limite de tuiles PAN (étirements identiques) sur la forêt d'Orléans.....	59
Figure II.37 : Comparaison des RMSE relatifs obtenus avec un SVR et une sélection Forward, sur les 31 primitives (ALOS2-PALSAR2, Sentinel-1 et Sentinel-2) ou avec l'ajout des textures spatiales à meilleure résolution (SPOT-6 multi-spectral et panchromatique).....	59
Figure III.1 : Localisation des sites d'étude Landes-INRA et Landes-ONF.....	62
Figure III.2 : Distribution des valeurs d'angle d'incidence local (par pixel) d'une image Sentinel-1 en orbite descendante (008) sur les placettes du site Landes-ONF.....	63
Figure III.3 : Illustrations des angles d'incidence Sentinel-1 sur une même emprise dans la zone dunaire.....	63
Figure III.4 : Distributions (fréquence en %) des paramètres forestiers sur les deux sites d'étude des Landes.....	64
Figure III.5 : Coefficients de détermination r^2 entre les primitives et les paramètres forestiers sur les deux sites d'étude des Landes, calculés sur les polygones.....	65
Figure III.6 : Graphiques de validation pour l'estimation des paramètres forestiers du site Landes-ONF, coloration selon les classes d'angle d'incidence local (IA).....	66
Figure III.7 : Inter-corrélations entre paramètres forestiers pour les placettes de pin maritime du site Landes-INRA et du site Landes-ONF.....	68
Figure III.8 : Corrélations entre les primitives et la surface terrière pour les polygones de pin maritime du site Landes-INRA et du site Landes-ONF.....	70
Figure III.9 : Corrélations entre les primitives et la hauteur dominante pour les polygones de pin maritime du site Landes-INRA et du site Landes-ONF.....	70

Figure III.10 : Distribution des valeurs des primitives satellitaires au niveau polygone (moyennes) sur les sites Landes-INRA et Landes-ONF.....	71
Figure III.11 : Localisation des quatre sites d'étude et visualisation des nouveaux sites dans les forêts d'Orléans et de Saint Gobain sur des images optique en fausse couleur (PIR-R-V). Les points colorés représentent les placettes de mesure.....	74
Figure III.12 : Histogrammes des distributions du DBH moyen, de la hauteur dominante et de la surface terrière des placettes de référence selon les sites et les classes de forêt.....	77
Figure III.13 : Variabilité intra-placettes des mesures de DBH des arbres. Les boîtes à moustaches montrent la médiane et les quartiles de chaque classe de forêt.....	78
Figure III.14 : Coefficients de détermination r^2 entre les paramètres forestiers et les données satellitaires pour le site d'Orléans (classes chêne et pin sylvestre).....	79
Figure III.15 : Graphique de validation des estimations de surface terrière (BA) pour les 7 classes de forêts identifiées.....	80
Figure III.16 : Graphique de validation des estimations de DBH moyen pour les 7 classes de forêts identifiées.....	81
Figure III.17 : Graphique de validation des estimations de hauteur dominante (Hdom) pour les 7 classes de forêts identifiées.....	81
Figure III.18 : Graphique de validation des estimations de densité d'arbres pour les 7 classes de forêts identifiées. Le graphique est repositionné sur la plage de valeurs 0-800 pour une meilleure visibilité des mesures sur les sites Orléans et StGobain.....	82
Figure III.19 : Graphiques de validation des estimations de surface terrière (BA) des feuillus selon une approche avec des modèles différenciés par classe (a), ou avec un modèle unique pour les 4 classes (b).....	85
Figure IV.1 : Classification des types de forêt sur le site d'Orléans.....	90
Figure IV.2 : Classification des types de forêt sur le site de StGobain.....	90
Figure IV.3 : Distribution des valeurs de paramètres forestiers prédits sur l'ensemble des pixels de forêt du site d'Orléans (2.1 millions de pixels). Les valeurs des références sont superposées sur les mêmes intervalles. BA (surface terrière, m ² /ha) ; DBH (diamètre à 130cm, cm) ; Hdom (hauteur dominante, m) ; Density (densité d'arbres, arbres/ha).....	92
Figure IV.4 : Distribution des valeurs de paramètres forestiers prédits sur l'ensemble des pixels de forêt du site Landes-ONF (2.16 millions de pixels). Les valeurs des références sont superposées sur les mêmes intervalles. BA (surface terrière, m ² /ha) ; DBH (diamètre à 130cm, cm) ; Hdom (hauteur dominante, m) ; Density (densité d'arbres, arbres/ha).....	93
Figure IV.5 : Statistiques de validation (RMSE relatif) des cartes de prédictions sur les pixels des placettes de référence. BA (surface terrière, m ² /ha) ; DBH (diamètre à 130cm, cm) ; Hdom (hauteur dominante, m) ; Density (densité d'arbres, arbres/ha).....	94
Figure IV.6 : Illustrations des cartes de surface terrière (BA) produites sur le site Orléans avec les méthodes pixel et objet. La normalisation pour le SVR est faite sur les valeurs d'apprentissage.....	94
Figure IV.7 : Carte de prédiction de la surface terrière (BA, m ² /ha) sur le site Orléans.....	96
Figure IV.8 : Carte de prédiction du DBH (cm) sur le site Orléans.....	96
Figure IV.9 : Carte de prédiction de la densité d'arbres (arbres/ha) sur le site Orléans.....	97
Figure IV.10 : Carte de prédiction de la hauteur dominante (m) sur le site Orléans.....	97
Figure IV.11 : Carte de prédiction de la biomasse aérienne de bois (AGB) pour les pins maritimes du site Landes-INRA.....	99
Figure A.1. Graphiques de comparaison des valeurs mesurées avec des seuils de pré-comptage à 7.5 et 17 cm de DBH pour les paramètres de surface terrière (a), DBH (b), et densité d'arbres (c).....	121
Figure B.1 : Illustration des parcelles utilisées pour étudier les variations d'humidité. La parcelle (a) est déforestée et instrumentée pour mesurer l'humidité du sol.....	122
Figure B.2 : Corrélations entre les séries temporelles de SWI et de Sentinel-1 VV pour les quatre parcelles. (a) sol nu, (b) faible biomasse, (c) moyenne biomasse, (d) forte biomasse.....	123
Figure B.3 : Corrélations entre les séries temporelles de SWI et de Sentinel-1 VH pour les quatre parcelles. (a) sol nu, (b) faible biomasse, (c) moyenne biomasse, (d) forte biomasse.....	124
Figure B.4 : Profils temporels Sentinel-1 VH, VV et le ratio VH/VV sur la parcelle 1142 de pin maritime (553 arbres/ha et 131 tonnes/ha). Les profils temporels des produits SWI, GSMap et de la sonde MLX (sur la parcelle déboisée 1144) sont superposés.....	124

Liste des tableaux

Tableau II.1 : Statistiques descriptives des paramètres forestiers mesurés sur les placettes.....	19
Tableau II.2 : Caractéristiques des satellites utilisés dans l'étude.....	20
Tableau II.3 : Caractéristiques des bandes spectrales du capteur Sentinel-2.....	22
Tableau II.4 : Caractéristiques des bandes spectrales des capteurs SPOT-6/7.....	23
Tableau II.5 : Description des indices de texture de Haralick.....	29
Tableau II.6 : Primitives sélectionnées pour l'estimation des paramètres de structure forestière.....	32
Tableau II.7 : Paramètres des algorithmes de régression.....	44
Tableau II.8 : Résultats statistiques et liste des primitives utilisées pour l'estimation des paramètres forestiers. Les résultats sont obtenus avec un SVR et une sélection Forward sur les 31 primitives issues de ALOS2-PALSAR2, Sentinel-1 et Sentinel-2. Les numéros entre parenthèse indiquent l'ordre de sélection des primitives.....	50
Tableau II.9 : Tableau de contingence pour la fréquence de sélection des différents types de primitives satellitaires pour les 6 paramètres forestiers.....	51
Tableau II.10 : Tests du Khi2 pour chaque paramètre forestier entre la distribution des primitives sélectionnées et une distribution théorique déterminée (équation 9).....	51
Tableau II.11 : Statistiques de validation (RMSE relatif) pour les différentes configurations.....	60
Tableau III.1 : Statistiques descriptives des quatre paramètres forestiers sur les deux sites d'étude des Landes. Moy. : moyenne ; E. T. : écart type.....	64
Tableau III.2 : Statistiques de validation sur les sites Landes-INRA et Landes-ONF au niveau polygone (SVR, sélection Forward). La méthode est appliquée indépendamment sur chaque site.....	66
Tableau III.3 : Répartition des types de primitives sélectionnées dans les méthodes appliquées aux sites Landes-INRA et Landes-ONF. Le pourcentage est le nombre de primitive pour chaque type divisé par le nombre total de primitives sélectionnées.....	67
Tableau III.4 : Statistiques de validation de la méthode avec apprentissage sur les références du site Landes-INRA et validation sur le site Landes-ONF pour l'optimisation des paramètres du SVR et la sélection des primitives satellitaires.....	72
Tableau III.5 : Statistiques de validation des estimations à partir d'un jeu de données rassemblant les deux sites. Les lignes Landes-INRA et Landes-ONF correspondent aux statistiques différenciées obtenus par site avec le modèle combinant les deux sites.....	73
Tableau III.6 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers pour l'ensemble des classes de forêts sur les 4 sites d'étude.....	76
Tableau III.7 : Statistiques de validation des estimations de BA, DBH, Hdom et Densité pour l'ensemble des classes de forêts des 4 sites. Le pourcentage représente le RMSE relatif.....	79
Tableau III.8 : Répartition des types de primitives sélectionnées pour chaque classe de forêt et chaque paramètre forestier. Les pourcentages sont calculés par rapport au nombre de primitives sélectionnées par la méthode Forward.....	83
Tableau III.9 : Statistiques de validation des estimations de BA, DBH, Hdom et Densité pour l'ensemble des classes de forêts des 4 sites. Le pourcentage représente le RMSE relatif à chaque classe.....	84
Tableau III.10 : Statistiques de validation pour les placettes des 7 classes de forêts (hors classes « Autres ») selon une approche avec un modèle par classe ou un modèle unique construit pour l'ensemble des classes.....	84
Tableau III.11 : Statistiques de validation pour les placettes des 3 classes de conifères selon une approche avec un modèle par classe ou un modèle construit pour l'ensemble placettes de conifères.....	85
Tableau III.12 : Statistiques de validation pour les placettes des 4 classes de feuillus (hors classe « Autre ») selon une approche avec un modèle par classe ou un modèle construit pour l'ensemble des 4 classes feuillus.....	85
Tableau IV.1 : Produits disponibles pour la cartographie des forêts à haute résolution spatiale.....	87
Tableau IV.2 : Informations sur la méthode de classification pour la cartographie des essences forestières.....	89
Tableau IV.3 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers estimés sur l'ensemble des pixels de forêt du site Orléans.....	95
Tableau IV.4 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers estimés sur l'ensemble des pixels de forêt du site StGobain.....	98
Tableau IV.5 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers estimés sur l'ensemble des pixels de conifères des sites Landes INRA et ONF (pin maritime).....	99
Tableau IV.6 : Répartition de la biomasse aérienne (AGB) par tranche de DBH moyen.....	100

Liste des sigles et des acronymes

ACP : Analyse en composantes principales
AGB : Aboveground biomass
BA : Basal area
BI : Brightness index
CNPF : Centre national de la propriété forestière
CPU : processeur
DBH : Diameter at breast height
ENL : Equivalent number of looks
ESA : European space agency
FAO : Food and agriculture organisation of the United Nations
FBD : Fine beam dual polarization
FOTO : Fourier-based textural ordination
GDAL : Geospatial data abstraction library
GLCM : Grey level co-occurrence matrix
GPS : Global positioning system
GRD : Ground range detected
IA : Incidence angle
IGN : Institut national d'information géographique et forestière
INRA : Institut national de la recherche agronomique
JAXA : Japan aerospace exploration agency
LOO : Leave-one-out cross validation
MAE : Mean absolute error
MAYA : MACCS ATCOR joint algorithm
MIR : Moyen infra-rouge
MLR : Multi-linear regression
MMU : Minimum map unit
MNT : Modèle numérique de terrain
NDVI : Normalized difference vegetation index
NDWI : Normalized difference wetness index
ONF : Office nationale des forêts
OTB : Orfeo Toolbox
PIR : Proche infra-rouge
PSG : Plan simple de gestion
RBF : Radial basis function
RF : Random forest
RMSE : Root mean square error
SAR : Synthetic aperture radar
SLC : Single look complex
SVR : Support vector regression
SWI : Soil water index
THRS : Très haute résolution spatiale
TLS : Terrestrial laser scanning
TOA : Top of atmosphere
TOC : Top of canopy

Résumé

L'estimation et le suivi du carbone et de la ressource forestière sont des enjeux majeurs pour la gestion des territoires. Les forêts ont un rôle important dans les plans nationaux et internationaux pour l'atténuation du changement climatique (stockage du carbone, régulation du climat, biodiversité, énergies renouvelables). Dans les forêts tempérées, de nombreuses mesures des paramètres de structure forestière sont acquises sur des petites zones, des statistiques au niveau national ou sur de larges zones administratives sont délivrées annuellement par les organismes gouvernementaux. Les forêts tempérées sont fortement anthropisées (forte variabilité spatiale et fractionnement des peuplements), il y a actuellement un besoin fort d'une spatialisation plus fine et continue des ressources forestières dans ces régions.

Les images satellitaires optiques et radar apportent des informations sur l'état de la végétation, la structure des arbres et l'organisation spatiale des forêts. Dans un contexte exceptionnel de disponibilité mondiale et gratuite, de diversité, de qualité des images à haute résolution spatiale et temporelle, le travail de thèse a pour objectif de mettre en place les bases méthodologiques et scientifiques pour une production nationale semi-automatique d'une cartographie des paramètres forestiers (biomasse, diamètre, hauteur, etc.).

Nous avons évalué le potentiel des séries temporelles Sentinel-1 (radar en bande C), Sentinel-2 (optique), et des mosaïques annuelles ALOS2-PALSAR2 (radar, bande L) pour estimer les paramètres de structure forestière. Ces données satellitaires ont été combinées à l'aide d'algorithmes d'apprentissage supervisé et de mesures terrain pour construire des modèles d'estimation de la biomasse, du diamètre moyen des arbres (DBH), de la hauteur et d'autres paramètres de structure. Ces modèles peuvent ensuite être spatialisés sur l'ensemble du territoire à l'aide des images satellitaires, et fournir une information continue à la résolution spatiale des images utilisées (10 à 20 mètres). Notre approche a été conçue et testée sur quatre sites d'étude avec des essences forestières et des propriétés structurales et environnementales différentes : la zone intérieure et la zone dunaire de la forêt des Landes (pins maritimes), la forêt d'Orléans (chênes et pins sylvestres), et la forêt de Saint-Gobain (chênes, charmes et hêtres).

Les principaux développements portent sur les données satellitaires à utiliser, la sélection des variables explicatives, le choix des algorithmes de régression et leur paramétrisation, la différenciation des types de forêt et la cartographie des estimations de paramètres forestiers. Les primitives issues des données satellitaires fournissent des informations sur les propriétés optiques du sol et de la végétation, l'organisation spatiale des arbres, la structure et le volume de bois vivant des houppiers et des troncs. L'utilisation d'algorithmes de régression multivariée non-linéaire permet d'obtenir des estimations des paramètres forestiers avec des performances en termes d'erreur relative de l'ordre de 15 à 35 % pour la surface terrière (~2.8 à 5.9 m²/ha) selon les types de forêt, 5 à 20 % pour la hauteur (~1.3 à 3 m), et de 5 à 25 % pour le DBH (~1.5 à 8 cm). Les résultats montrent l'apport de la combinaison de plusieurs types de données satellitaires (optique, radar multi-fréquence et indices de texture spatiale) ainsi que l'importance de différencier les types de forêt pour la construction des modèles. L'application des modèles sur les images satellitaires permet de produire des cartes à haute résolution spatiale de ces paramètres forestiers, utilisables de l'échelle locale à l'échelle régionale/nationale.

La possibilité de produire une cartographie régulière à haute résolution de la ressource forestière est très prometteuse pour aider à améliorer la mise en place et le suivi des stratégies territoriales et nationales pour la filière bois, la biodiversité et le stockage du carbone.

Abstract

The estimation and monitoring of forest resources and carbon stocks are major issues for wood industry and public bodies. Forests play an important role in national and international plans for climate change mitigation (carbon storage, climate regulation, biodiversity, renewable energy). In temperate forests, monitoring is done at two different levels: on one hand, at local level, in small areas by the acquisition of many measures of forest structure parameters, and, on the other hand, by statistics at national level or in large administrative areas that are provided annually by public bodies. Temperate forests are highly anthropogenic (high spatial variability and fragmentation of stands), so there is currently a strong need for a more refined and regular maps of forest resources in these regions.

Optical and radar satellite images provide information on the state of vegetation, tree structure and spatial organization of forests. In an exceptional context of free global availability, diversity, and quality of images with high spatial and temporal resolution, the aim of this PhD work is to set up the methodological bases for a generic and semi-automatic production of forest parameters mapping (biomass, diameter, height, etc.). We have assessed the potential of Sentinel-1 (C-band radar), Sentinel-2 (optical) time series, and ALOS2-PALSAR2 (radar, L-band) annual mosaics to estimate forest structure parameters. These satellite data are combined, using supervised learning algorithms and field measurements, to construct models for estimating aboveground biomass (AGB), mean tree diameter (DBH), height, basal area and tree density. These models can then be spatially applied over the entire territory by using satellite images, providing thus continuous information on the spatial resolution of the images used (10 to 20 meters). This approach has been conceived and tested on four study sites with different forest species and structural and environmental properties: the inner and the dune zone of the Landes forest (maritime pines), the Orléans forest (oak and Scots pines), and the forest of Saint-Gobain (oaks, hornbeams and beeches). The investigated issues are the satellite data to be used, the selection of explanatory variables, the choice of regression algorithms and their parameterization, the differentiation of forest types and the spatialization of forest parameter estimates. The primitives derived from satellite data provide information on the optical properties of soil and vegetation, the spatial organization of trees, the structure and volume of live wood of crowns and trunks. The use of nonlinear multivariate regression algorithms allows to obtain forest parameter estimates with relative error performance in the order of 15 to 35 % for the basal area (~ 2.8 to $5.9 \text{ m}^2/\text{ha}$) depending on forest types, 5 to 20 % for height (~ 1.3 to 3 m), and 5 to 25 % for DBH (~ 1.5 to 8 cm). The results highlight the improvement by combining several types of satellite data (optical, multi-frequency radar and spatial texture indexes), as well as the importance of differentiating forest types for the construction of models.

This high-resolution, regular mapping of the forest resource is very promising to help improving the monitoring and policy of territorial and national strategies for the timber sector, biodiversity and carbon storage.

Chapitre I : Introduction

1.1. Contexte

Les écosystèmes forestiers apportent des services essentiels aux populations humaines et à la planète, en fournissant des moyens d'existence, en contribuant à la purification de l'eau et de l'air, en préservant la biodiversité et en offrant des solutions pour faire face au changement climatique. Les rapports de l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO) mettent en évidence le rôle prépondérant des forêts en quelques chiffres (FAO, 2015, 2018) :

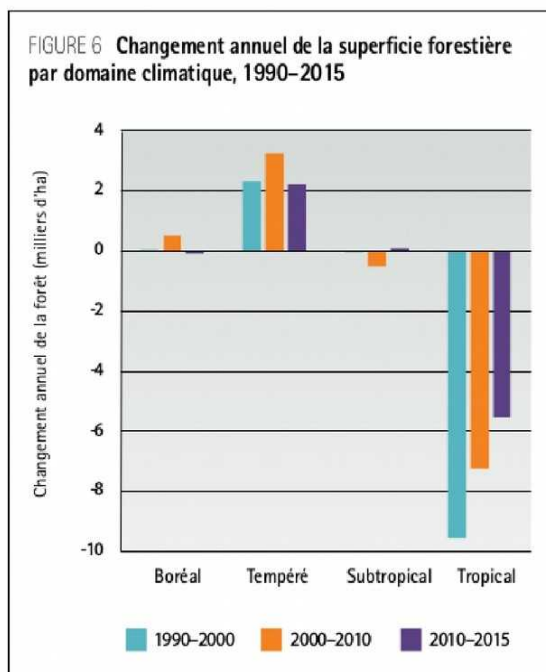
- Les forêts représentent 30.6 % des terres émergées.
- Un stock de carbone de 296 Gt dans la biomasse aérienne et souterraine. Lorsqu'elles sont gérées durablement, les forêts peuvent absorber environ 10 % des émissions mondiales de carbone. (actuellement ~ 0.5 Gt équivalent carbone sont absorbées). Les émissions d'origine anthropique sont estimées à ~10 Gt de carbone en 2018 (Le Quéré et al., 2018).
- Les combustibles ligneux répondent aujourd'hui à 40 % de la demande mondiale d'énergie renouvelable.
- Les forêts abritent plus des trois quarts de la biodiversité terrestre.
- On estime que 1/3 de la population mondiale dépend de produits et services forestiers tels que les aliments, le bois de feu, les médicaments, les opportunités d'emploi et les revenus.

Ces trente dernières années, l'étendue des forêts a régressé tandis que la population humaine continue à augmenter, avec une croissance de la demande en nourriture et en terres. La FAO comptabilise une perte de 129 millions d'hectares de forêts entre 1990 et 2015 (FAO, 2015, 2018). La principale cause de la déforestation est le changement d'utilisation des terres pour l'agriculture, mais on observe aussi de nombreux cas de dégradations des forêts à cause de maladies, de ravageurs, d'incendies ou de tempêtes. Ces éléments sont aggravés et plus fréquents ces dernières années à cause du réchauffement climatique.

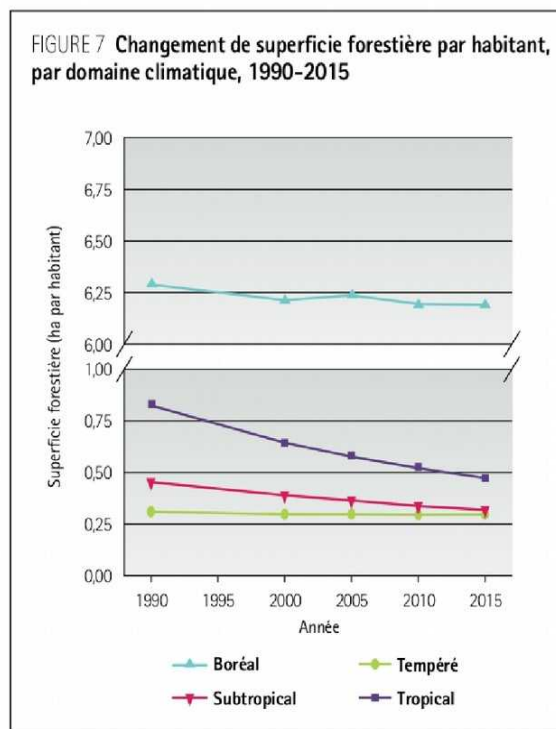
La figure I.1.a présente les changements annuels de la superficie forestière par domaine climatique entre 1990 et 2015. Ces statistiques révèlent une forte baisse de la surface forestière en zone tropicale, cette baisse diminue avec le temps mais reste très importante (5.5 millions ha/an de 2010 à 2015). À l'inverse, les statistiques de la FAO montrent une augmentation de la surface forestière dans les zones tempérées (~2 millions ha/an entre 2010 et 2015). La figure I.1.b montre cette fois l'évolution de la superficie forestière par habitant pour les mêmes domaines climatiques. Cette surface par habitant a quasiment été divisée par deux dans les zones tropicales entre 1990 et 2015. Pour les zones tempérées, malgré l'augmentation annuelle de la surface forestière estimée par la FAO et d'autres études, la surface forestière par habitant reste de 40 % inférieure à celle des zones tropicales en 2015. En raison de l'augmentation de la population dans les zones tempérées, cette surface forestière par habitant est restée stable entre 1990 et 2015.

La situation des forêts tempérées est donc un peu particulière puisque d'un côté, l'attention est surtout portée sur la déforestation des forêts tropicales, à juste titre puisqu'il s'agit d'immenses réservoirs de biodiversité et de carbone qui sont aujourd'hui menacés. Mais de l'autre côté, les forêts tempérées en Europe ont déjà subi cette déforestation massive (1 à 2 % des forêts tempérées mondiales n'ont jamais été coupées) (Currie and Bergen, 2008) et

subissent encore aujourd'hui une forte pression anthropique. L'intégration de la ressource en bois comme une énergie renouvelable et un puits de carbone dans les plans nationaux et européens, pour l'atténuation du changement climatique, nécessite une gestion durable des forêts (Colin A. and Thivolle-Cazat A., 2016; Dixon et al., 1994; Flamenco et al., 2019; Luyssaert et al., 2008).



(a)



(b)

Figure 1.1 : Évolution de la surface forestière entre 1990 et 2015 (FAO, 2015).

La France compte 16.9 millions d'hectares de forêts en métropole (31 % du territoire métropolitain). L'institut national d'information géographique et forestière (IGN) estime à 92 millions de m³ de bois la production annuelle de ces forêts, contre 45.2 millions de m³ prélevés annuellement, soit un accroissement du stock de bois sur pied. La loi de transition énergétique prévoit de renforcer la part des énergies renouvelables pour atteindre 32 % de la consommation d'énergie en 2030 (contre 13.7 % en 2012). La biomasse forestière (« bois énergie ») devrait représenter 40 % de la part des énergies renouvelable. Bien que le ratio production/prélèvements des forêts soit favorable à une augmentation des prélèvements, ce scénario accroît considérablement la pression sur les forêts françaises. Dans ce contexte, il y a un fort besoin de suivi des forêts afin de contrôler la durabilité de la gestion forestière et maintenir la biodiversité, la qualité des paysages, les stocks de carbone et la capacité de résilience des forêts.

1.2. Évaluation des ressources forestières

L'évaluation, le suivi et la cartographie des ressources forestières est donc un enjeu majeur pour permettre et contrôler la gestion durable des forêts. Actuellement, cette évaluation repose principalement sur un inventaire statistique de l'ensemble du territoire avec des placettes (~15 m de rayon) où un ensemble de mesures et de relevés sont effectués concernant la structure des arbres et leur environnement. En plus de cet inventaire forestier national réalisé par l'IGN, de nombreuses mesures et estimations des paramètres forestiers sont réalisées localement par

des organismes et propriétaires forestiers, des collectivités territoriales, ou encore des organismes de recherche.

Mesures terrain

La première méthode pour estimer la ressource forestière est d'aller mesurer les arbres directement dans les forêts sur des petites surfaces (placettes de 15 à 50 m de rayon). Plusieurs mesures sont utiles à la caractérisation des forêts :

- Le nombre d'arbres par hectare (densité).
- Le diamètre des arbres à hauteur d'épaule, ~130 cm de hauteur (diameter at breast height, DBH). On peut ensuite dériver un diamètre moyen ou un diamètre dominant (de manière générale, il s'agit du diamètre moyen des 100 plus gros arbres à l'hectare)
- La hauteur des arbres, qui peut être mesurée sur l'arbre en entier jusqu'au sommet de la canopée, ou bien sur la partie du tronc considérée comme exploitable pour du bois d'œuvre. De même que pour le DBH, la hauteur peut être dérivée en hauteur moyenne ou dominante.
- La surface terrière (G, ou BA pour « basal area » en anglais) est la surface des troncs (à 130 cm de hauteur) à l'hectare, en m²/ha.

Ces paramètres peuvent être ensuite utilisées pour estimer le volume de bois ou de grume (m³/ha) et la biomasse aérienne (AGB, « aboveground biomass », en tonnes/ha ou Mg/ha) à l'aide d'équations allométriques ou de tarifs de cubage. Les relations entre le volume ou la biomasse et les paramètres mesurés sur les arbres dépendent des essences et des structures des arbres (profil vertical des arbres, densité du bois, etc.), elles peuvent être très variables selon les forêts et les conditions environnementales. Des études analysent les différentes équations allométriques et proposent des relations préétablies pour certains types de forêts ou essences afin d'harmoniser les calculs (Chave et al., 2005; Deleuze et al., 2014; Shaiek et al., 2011), mais les équations restent souvent différentes entre les études et peuvent être sources d'erreurs.

Ce type de mesure directement dans les forêts est utilisé notamment pour les campagnes annuelles d'inventaire forestier national (IGN), ou par les coopératives et les gestionnaires forestiers. On peut trouver ces informations dans les plans simples de gestion (PSG) des forêts privées ou les plans d'aménagement forestiers des forêts publiques (par l'office nationale des forêts, ONF), mais ces données ne sont pas tout le temps numérisées ni centralisées. Pour obtenir une estimation des paramètres forestiers à l'échelle d'un peuplement, on procède souvent par échantillonnage avec un certain nombre de placettes de mesure par hectare en fonction de l'homogénéité du peuplement. Les mesures terrain sont très coûteuses en temps et en ressources humaines, mais constituent souvent le moyen le plus fiable d'estimer les paramètres forestiers sur les zones ciblées même s'il peut y avoir des erreurs ou des biais selon les protocoles de mesure et d'échantillonnage (Annexe A).

De nouvelles technologies sont développées pour améliorer la qualité ou l'efficacité des mesures terrain, notamment les Lidar terrestres et scanners laser terrestres (terrestrial laser scanning, TLS), permettant la reconstitution numérique de la structure 3D des forêts (Aubry-Kientz et al., 2019; Dassot et al., 2011; Donager et al., 2018; Trochta et al., 2017).

Télédétection de la structure 3D des forêts

Plusieurs données et méthodes permettent de caractériser la structure tridimensionnelle (3D) des forêts. Les données Lidar (Light Detection and Ranging), aussi appelées ALS (Airborne

Laser Scanning), sont les plus répandues de manière opérationnelle pour aider aux inventaires forestiers (Jochem et al., 2010; X. Yu et al., 2015). Ces données permettent d'obtenir des mesures précises de la hauteur de canopée et des estimations de la biomasse aérienne et d'autres paramètres à l'aide de relations allométriques ou de modèles (Andersen et al., 2009; Bouvier et al., 2015; d'Oliveira et al., 2012; Ho Tong Minh et al., 2016; Jochem et al., 2010).

Les couples stéréo d'images optiques à haute résolution spatiale permettent de faire de la photogrammétrie et d'obtenir des estimations de hauteur de canopée à l'aide d'un modèle numérique de hauteur (MNT), d'extraire les tailles des houppiers et d'estimer d'autres paramètres comme la biomasse aérienne à l'aide de relations allométriques. Cette méthode est de plus en plus utilisée, soit en utilisant les recouvrements d'images lors de l'acquisition de couvertures régionales et nationales d'images aériennes, ou en utilisant des couples d'images satellitaires (notamment GeoEye-1, Ikonos, Pléiades ou WorldView-2) sur de plus petites zones. Bien que plusieurs études montrent des résultats légèrement moins précis qu'avec le Lidar, la photogrammétrie permet d'apporter des informations utiles à la gestion des forêts et l'estimation des stocks de carbone (Jayathunga et al., 2019; Lisein et al., 2014; Noordermeer et al., 2019; Pearse et al., 2018; Persson, 2016; Piermattei et al., 2019).

Enfin, les données radar peuvent également être utilisées pour obtenir des données 3D sur les forêts, avec les méthodes d'interférométrie (Abdullahi et al., 2016; Karila et al., 2015) et de radargrammétrie (Persson and Fransson, 2014; Vastaranta et al., 2014b, 2014a).

Les précisions obtenues avec ces données 3D pour l'estimation des paramètres forestiers varient d'une étude à l'autre, selon le type de forêt, la qualité des données et l'échelle utilisée. D'après les études précédemment citées ainsi que les études comparatives et les revues (Holopainen et al., 2015; Pearse et al., 2018; Rahlf et al., 2014; Vastaranta et al., 2013; X. Yu et al., 2015), les erreurs relatives peuvent aller de 16 à 40% pour le volume aérien (VOL) ; de 14 à 35% pour la biomasse aérienne (AGB) ; de 12 à 36% pour la surface terrière (BA, G) ; de 11 à 25% pour le diamètre moyen (DBH) ; de 5 à 28% pour la hauteur moyenne ou dominante (H) ; et de 34 à 36% pour la densité d'arbres (moins souvent estimée). De manière générale, les données Lidar/ALS ont les meilleures précisions, suivies des données optiques stéréo (photogrammétrie), puis les méthodes radar d'interférométrie et de radargrammétrie. L'ordre de cette hiérarchisation n'est pas systématique, et les résultats peuvent changer avec l'évolution des méthodes et des technologies propres à chaque type de données.

Malgré les résultats très prometteurs pour intensifier et spatialiser les estimations des paramètres de structure forestière, ces données 3D et leurs traitements restent coûteux et complexes. Cela peut restreindre leur utilisation pour des acquisitions fréquentes ou sur de grands territoires, et rendre difficile l'accès à ces techniques pour les propriétaires ou organismes forestiers qui n'en auraient pas les moyens.

Télédétection spatiale à couverture mondiale

Actuellement, plusieurs satellites fournissent des images avec une couverture globale des terres émergées, et de manière continue (c'est-à-dire qu'une information est disponible pour chaque point de l'espace). Les images satellitaires à haute résolution spatiale (10 à 30 m) peuvent aider produire des cartes des paramètres forestiers. La télédétection spatiale a été utilisée soit 1) pour aider à spatialiser des inventaires forestiers et/ou des données de télédétection 3D des forêts, soit 2) en relation directe avec les paramètres forestiers.

Spatialisation des inventaires et des données 3D

Les données optiques (voir section suivante, Optique) ont depuis longtemps été utilisées pour fournir des indices de végétation et décrire les paramètres biophysiques de la canopée tels que la surface foliaire (LAI, Leaf area index) ou l'activité photosynthétique (FAPAR, fraction de rayonnement solaire absorbé par les plantes) (Brede et al., 2020; Fassnacht et al., 1997; Melnikova et al., 2018; Turner et al., 1999). Ces caractéristiques sont reliées à la productivité primaire de la végétation et peuvent être reliées ainsi à la biomasse ou à la hauteur des forêts (Kempes et al., 2011; Zhang et al., 2014; Zhang and Kondragunta, 2006). Plusieurs études ont mis en place des classifications non supervisées des forêts à partir de ces données satellitaires, puis ont relié les différentes classes de forêts aux mesures terrain afin d'assigner des plages de valeurs de biomasse ou de hauteur à chaque classe, et ainsi spatialiser ces mesures avec des cartes continues sur de grands territoires (Irulappa-Pillai-Vijayakumar et al., 2019; Koch, 2013; Mcroberts and Tomppo, 2007; Mitchard et al., 2012; Tomppo, 2004; Zhang et al., 2019).

Le satellite IceSAT, avec son instrument Lidar GLAS, a produit des données 3D sur les forêts du monde entier de 2003 à 2009 et permis d'estimer la hauteur et la biomasse des forêts le long de traces (Baghdadi et al., 2014; Harding and Carabajal, 2005; Lefsky et al., 2005; Sun et al., 2008; Y. Yu et al., 2015). La limite de GLAS est le manque de capacité d'imagerie et le fait qu'il fournit des informations sur des traces limitées spatialement (Sun et al., 2008). La couverture mondiale et la répétitivité des acquisitions permet cependant d'avoir un très gros échantillonnage des forêts. De nombreuses études ont mises à profit les corrélations observées entre la biomasse ou la hauteur estimée par les données 3D de GLAS ou de Lidar aéroporté, avec des indices de végétation provenant des images optiques MODIS (Chi et al., 2015; Nelson et al., 2009; Saatchi et al., 2011) ou Landsat (Bolton et al., 2018; Chi et al., 2017; Deo et al., 2017; Duncanson et al., 2010; Hansen et al., 2016; Li et al., 2011; Liu et al., 2017; Matasci et al., 2018; Wang et al., 2018; Zhang et al., 2014) pour produire des cartes à moyenne ou haute résolution de la hauteur et de la biomasse des forêts à l'échelle nationale ou continentale. Ces cartes ont des précisions moindres en dehors des traces Lidar, mais permettent néanmoins d'apporter une information importante sur la variabilité spatiale des forêts. L'arrivée prochaine des données de Lidar spatial GEDI puis IceSAT-2, ainsi que les nouveaux capteurs Sentinelles vont permettre de reprendre ces travaux de spatialisation des données 3D. L'utilisation des cartes d'élévation de Tandem-x (mosaïques globale) pour l'estimation de la hauteur des forêts (Rizzoli et al., 2017; Wessel et al., 2018), bien que payantes hors cadre de recherche, fait aussi déjà l'objet de travaux avec des simulations pour spatialiser les futures données 3D GEDI (Choi et al., 2019; Lee et al., 2019; Qi et al., 2019a, 2019b; Qi and Dubayah, 2016).

Téledétection optique et radar des paramètres forestiers

Concernant l'étude des relations directes entre les données satellitaires et les paramètres forestiers mesurés sur le terrain, de nombreuses études ont été faites dans les domaines optiques et radar. Les images optiques et radar sont sensibles à différents éléments et propriétés de la végétation des forêts tels qu'illustrés dans la figure I.2. Même si chaque capteur et bande de fréquence est plus sensible à une propriété ou élément de l'arbre donné, l'ensemble des éléments de l'arbre sont reliés (hauteur de canopée, diamètre de l'arbre, activité photosynthétique, etc.) et il existe des relations indirectes entre les différents signaux satellitaires. Ce domaine d'étude est complémentaire avec les approches de spatialisation des

données 3D, puisque ces dernières reposent justement sur la capacité des données optique (et parfois radar) à estimer la productivité primaire des forêts pour les relier aux paramètres forestiers issus du Lidar.

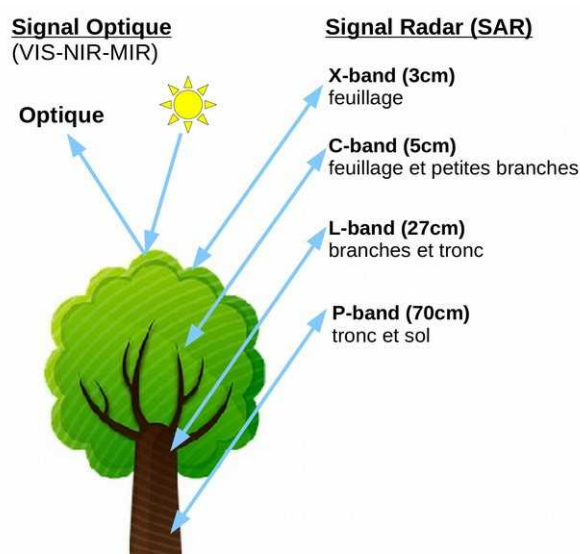


Figure I.2 : Illustration de la sensibilité des signaux optiques et radar aux différents éléments des arbres.

Données optiques

La télédétection optique pour l'observation de la terre est dite passive puisque le capteur optique enregistre le rayonnement dans certaines longueurs d'ondes sans diffuser lui-même un signal. La végétation reçoit le rayonnement solaire, en absorbe une partie et réfléchit l'autre partie qui sera mesurée par les capteurs optiques satellitaires après avoir retraversé l'atmosphère. Ce rayonnement réfléchi sera appelé réflectance par la suite (figure I.3).

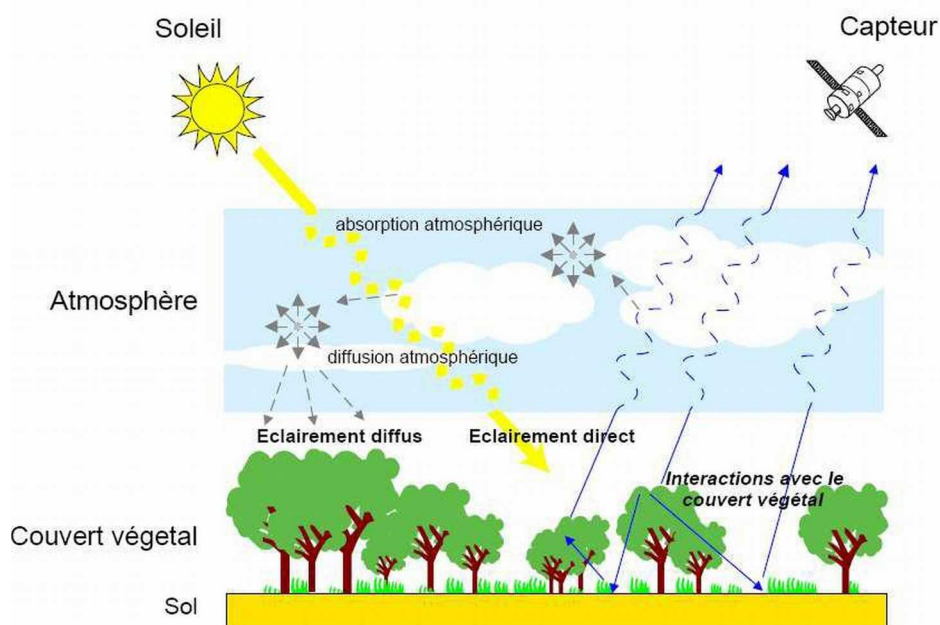


Figure I.3 : Mesure de télédétection dans le domaine optique (Bacour, 2001).

Selon les capteurs, on peut mesurer la réflectance dans différentes longueurs d'ondes : le visible (VIS : bleu, vert, rouge), mais aussi des longueurs d'onde invisible à l'œil humain telles que le proche infrarouge (PIR, ou NIR en anglais) et le moyen infrarouge (MIR, ou SWIR en anglais), qui sont des longueurs d'ondes importante pour mesurer l'activité et le

contenu en eau de la végétation. Il existe donc des capteurs panchromatiques, enregistrant la réflectance dans une seule bande spectrale relativement large et généralement dans le visible, des capteurs multispectraux enregistrant le signal dans plusieurs bandes spectrales, et des capteurs hyperspectraux pouvant enregistrer le signal de manière beaucoup plus précise dans les longueurs d'ondes (parfois des centaines de bandes spectrales).

Dans une forêt, la réflectance mesurée par satellite est fonction des propriétés spectrales des éléments constituant le couvert (dans le visible et dans l'infrarouge qui caractérisent les propriétés biochimiques de la végétation et du sol), de la structure du couvert (nombre et position des arbres, organisation des branches et des feuilles), mais aussi des conditions d'acquisition des images (heure de la journée, angle de rayonnement incident, angle de prise de vue, effets atmosphériques, etc). La figure I.4 illustre les spectres de réflectance que l'on pourrait observer en fonction de l'occupation du sol et du type de végétation.

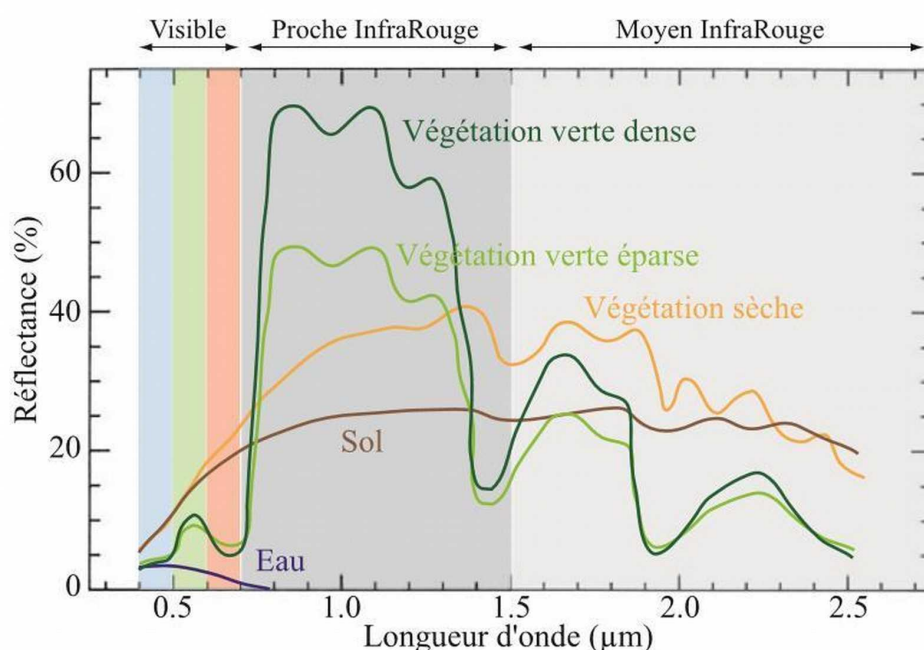


Figure I.4 : Signature spectrale optique de différents types de végétation et d'occupation du sol (Nicolas and Dusseux, 2016).

Les capteurs multispectraux ont ainsi été utilisés pour estimer certaines caractéristiques structurales et biochimiques des couverts végétaux et leur évolution dans le temps (phénologie, stress hydrique, croissance, coupe, etc.). Plusieurs études ont estimé la surface foliaire (LAI) et l'activité photosynthétique de la végétation (FAPAR) dans différentes forêts tempérées de plaine ou de montagne à partir d'images optiques Landsat ou Sentinel-2 (Brede et al., 2020; Fassnacht et al., 1997; Melnikova et al., 2018; Turner et al., 1999). Ces paramètres peuvent ensuite être reliés à la productivité primaire de la végétation (Kempes et al., 2011; Marsden et al., 2010; Zhang et al., 2014; Zhang and Kondragunta, 2006) et permettre d'estimer indirectement la biomasse et les paramètres de structure des forêts. Des études ont montré qu'il était possible de relier directement le signal optique (indices de végétations à une ou plusieurs dates) avec la biomasse aérienne (Castillo et al., 2017; Dong et al., 2003; Dube et al., 2017; Dube and Mutanga, 2015a; le Maire et al., 2011; Macedo et al., 2018; Marsden et al., 2010; Zhu and Liu, 2015) ou d'autres paramètres de structure des forêts (densité d'arbres, surface terrière, hauteur moyenne, etc.) (Castillo-Santiago et al., 2010; Freitas et al., 2005; Woodcock et al., 1994). Ces études montrent des résultats intéressants, les

meilleurs r^2 allant de 0.55 à 0.7 pour la biomasse, et de 0.7 à 0.8 pour les paramètres forestiers. Malgré ces bons résultats, l'utilisation des données optiques comme proxy de la biomasse utilise des relations souvent spécifiques à chaque jeu de données. De plus, la relation entre les propriétés optiques des forêts et leur biomasse tend à rapidement saturer quand la biomasse augmente (Dong et al., 2003; Dube et al., 2017; Duncanson et al., 2010; Lu, 2005; Mutanga and Skidmore, 2004).

Données radar

Les systèmes radar (RADio Detection and Ranging) consistent à émettre une onde électromagnétique et à mesurer l'onde réfléchiée par la surface observée. Les SARs (radar à synthèse d'ouverture, figure I.5) sont des systèmes imageurs. Ils permettent de localiser les sources de rétrodiffusion sur une surface bidimensionnelle, et donc de constituer une image de rétrodiffusion.

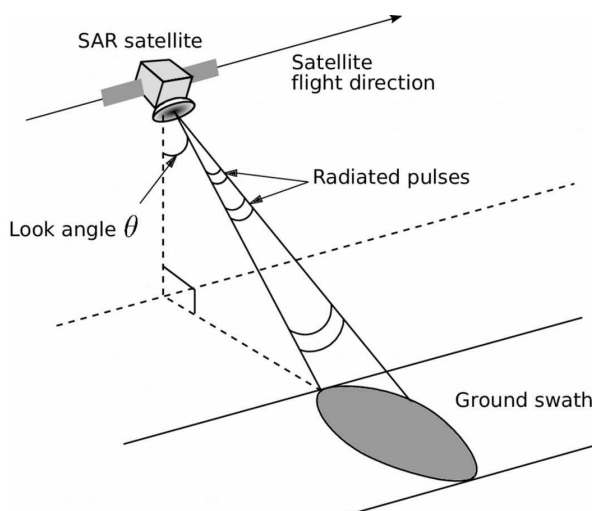


Figure I.5 : Représentation simplifiée d'un système de radar à synthèse d'ouverture (SAR).

La longueur d'onde émise peut varier de 1 cm à 1m. Parmi les satellites fréquemment utilisés on peut avoir de la bande X (2.5 à 3.85 cm), C (3.75 à 7.5 cm), L (15 à 30 cm) et prochainement P (~70 cm). L'onde émise est polarisée par rapport à la surface terrestre : horizontale (H) ou verticale (V) (figure I.6). La polarisation peut changer entre l'onde émise et l'onde reçu selon les propriétés de la surface, et l'intensité de l'onde reçu caractérise la réflectivité de la cible. En général, les capteurs visent la surface de la terre avec un angle d'incidence de 20 à 40° par rapport à la verticale (figure I.6).

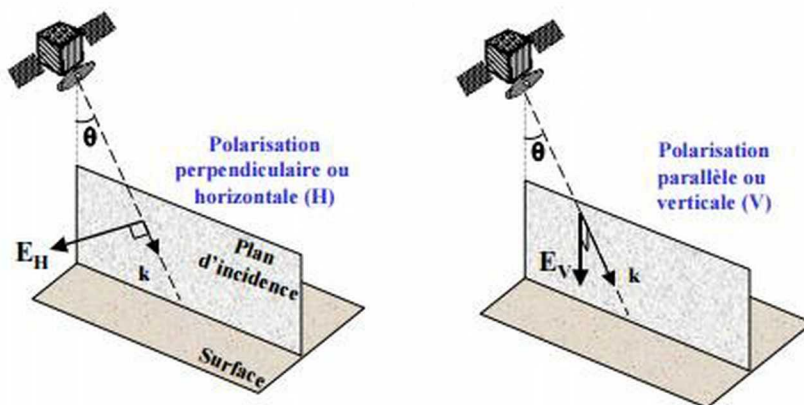


Figure I.6 : Représentation de l'angle d'incidence θ et des polarisations incidentes horizontale (champ E_H) et verticale (champ E_V). (Proisy, 1999).

Les images SAR fournissent des informations sur les propriétés géométriques et diélectriques de la surface ou du volume étudié, qui dépendent essentiellement de la rugosité de la surface, du type de matériau et de sa teneur en humidité. Concrètement, une surface rugueuse aura un signal retour plus élevé qu'une surface plane et lisse (qui ne renverra pratiquement pas de signal) ; un fort volume de végétation renverra un signal plus fort qu'un faible volume ; et la végétation humide renverra plus de signal que la même végétation sèche. Pour les forêts, le signal de rétrodiffusion n'est donc pas directement lié à la biomasse des arbres d'un point de vue physique, mais il va être influencé par la structure du couvert et la variabilité spatiale (Dobson et al., 1995; Le Toan et al., 1992; Mitchard et al., 2011; Rosenqvist et al., 2007). Le signal pénètre plus ou moins la canopée selon sa longueur d'onde (Figure I.2). Les longueurs d'onde plus courtes X et C sont sensibles aux petits éléments tels que les feuilles et les petites branches, alors que les plus grandes longueurs d'onde L et P sont sensibles aux grosses branches et aux troncs. La figure I.7 illustre les éléments de l'arbre observés théoriquement par les différentes fréquences radar. Par conséquent, les bandes L et P sont mieux reliées à la biomasse des forêts (Beaudoin et al., 1994; Le Toan et al., 2011, 1992; Santoro and Cartus, 2018).

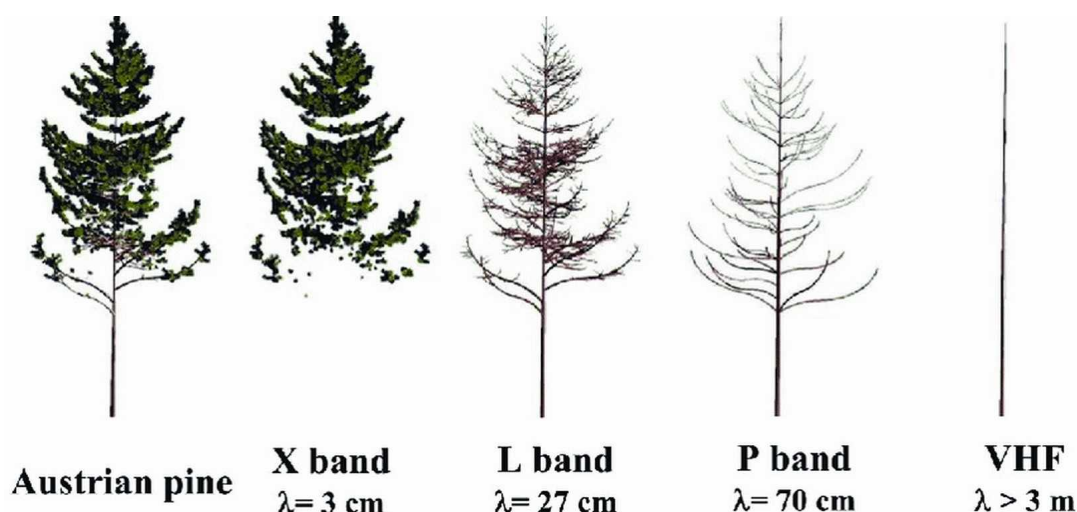


Figure I.7 : Sensibilité du signal radar aux feuilles, branches et troncs selon la bande de fréquence SAR utilisée. (Le Toan, 2007)

De nombreuses études ont utilisé les images SAR pour estimer la biomasse et les paramètres forestiers, à l'aide de modèles physiques ou de régressions. Ces travaux utilisent des images d'amplitudes uniquement (GRD, Ground Range Detection) ou des images complexes (SLC, Single Look Complex) avec les informations d'amplitude et de phase, qui sont plus lourdes à utiliser mais qui peuvent apporter des informations supplémentaires. Les images SAR ont ainsi permis d'estimer la biomasse aérienne et les paramètres forestiers sur des forêts tropicales denses et les mangroves (Austin et al., 2003; Berninger et al., 2019, 2018; Carreiras et al., 2012; Castillo et al., 2017; Luckman et al., 1998; Urbazaev et al., 2018; Wijaya et al., 2015), sur des savanes avec des biomasses plus faibles (Bouvet et al., 2018; Carreiras et al., 2012; Mermoz et al., 2014; Mitchard et al., 2011, 2009), sur des forêts boréales (Dobson et al., 1995; Joshi et al., 2015; Santoro et al., 2019, 2015), sur des tempérées plus diversifiées (Abdullahi et al., 2016; Austin et al., 2003; Dobson et al., 1995), et sur des plantations de conifères et de feuillus (Baghdadi et al., 2015; Castel, 2002; Harrell et al., 1997; Le Toan et al., 1992; Santoro et al., 2015). Sur certains jeux de données, les données SAR (principalement en bande L) expliquent jusqu'à 90-95% de la variance de la biomasse. Plusieurs de ces études montrent cependant une saturation du signal radar avec l'augmentation

de la biomasse (Mermoz et al., 2015, 2014; Mitchard et al., 2009; Yu and Saatchi, 2016), la valeur de seuil dépend des essences observées et augmente avec la longueur d'onde (Dobson et al., 1995; Imhoff, 1993). Avec l'apparition des séries temporelles à haute répétitivité comme Sentinel-1 ou la disponibilité (gratuité) des différents capteurs radar, les résultats tendent à montrer que ces approches multi-fréquences radar et/ou multi-temporelles permettent de mieux estimer les paramètres forestiers (Berninger et al., 2018; Dobson et al., 1995; Harrell et al., 1997; Santoro and Cartus, 2018).

Texture spatiales

La texture spatiale d'une image est une fonction de la variation spatiale des niveaux de gris, elle peut être calculée à partir d'images optiques ou radar. Elle renseigne sur la distribution spatiale des objets, et peut être reliée à des paramètres biophysiques. Par exemple, (Davi et al., 2006) ont trouvé une corrélation entre l'indice foliaire d'une parcelle de forêt plantée et l'écart type des NDVI (indice de végétation normalisé produit à partir des données optiques) des pixels de la parcelle. L'écart-type du NDVI diminue avec le LAI moyen de la parcelle. La corrélation avec les paramètres biophysiques peut être meilleure avec l'information de texture qu'avec la réflectance ou des indices spectraux directement, car les textures présentent moins de saturation (Castillo-Santiago et al., 2010; le Maire et al., 2006; Proisy et al., 2007).

L'information texturée sur une image de forêt est donc formée par l'alternance de pixels représentant la canopée éclairée, la canopée ombragée, les ombres projetées et la réponse des autres éléments de la surface (sous-bois, sol, etc.) (Ni et al., 1999). À chaque pixel sont associées des valeurs de texture calculée par rapport à la variation des pixels dans une fenêtre centrée sur lui. La figure I.8 montre la différence de texture spatiale dans des forêts de pin maritime avec un exemple d'indice de texture de Haralick. Les pixels les plus brillants reflètent une structure « lisse » avec une homogénéité de la canopée.

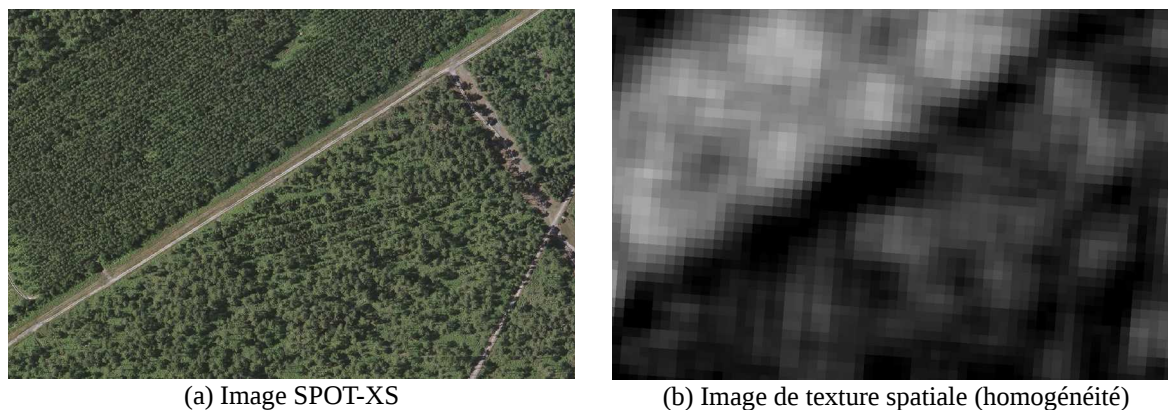


Figure I.8 : Visualisation d'une image de texture spatiale calculée selon l'indice d'homogénéité (textures de Haralick) à partir de l'indice de brillance de l'image. (a) image SPOT-XS sur des parcelles de pin maritime ; (b) image de texture (homogénéité) sur la même zone. Les pixels les plus brillants reflètent une plus grande homogénéité dans la répartition des niveaux de gris.

Il existe plusieurs types d'indices de texture spatiale, le plus simple étant les statistiques locales comme la variance évoquée ci-dessus. Des chercheurs ont développé une méthode basée sur l'utilisation d'un périodogramme de Fourier 2D pour modéliser la texture de la canopée, dite méthode FOTO (Fourier-based textural ordination) (Couteron et al., 2005; Proisy et al., 2007). Il existe également des indices statistiques de second ordre définis par Haralick (Haralick et al., 1973), basé sur les matrices de co-occurrence (GLCM, Grey Level Co-occurrence Matrix), construites en comptabilisant le nombre de paires de pixels (pour toutes les combinaisons de paires de valeurs disponibles) séparés d'une même distance

suivant une certaine orientation et ceci dans une fenêtre de taille à définir (figure I.9). Les indices de Haralick sont les plus utilisés dans la littérature de télédétection forestière (Boukir et al., 2013; Franklin et al., 2001) et leur implémentation est moins complexe que d'autres représentations de la texture comme la méthode FOTO, plusieurs logiciels et bibliothèques open-source intègrent les textures de Haralick.

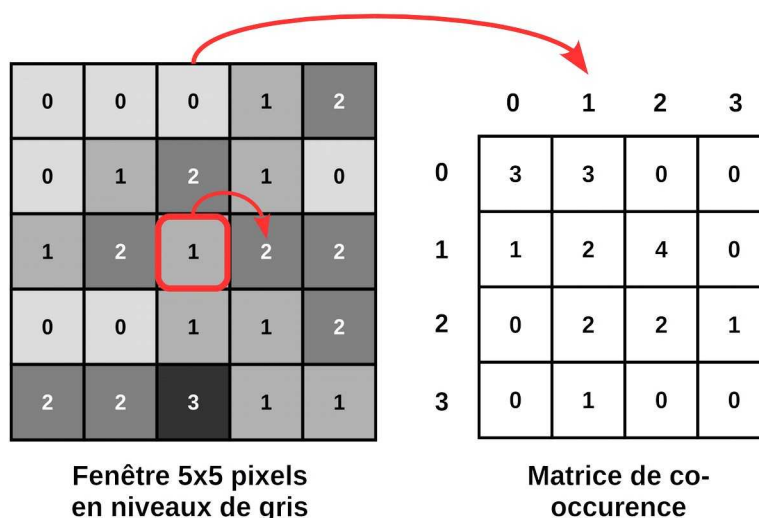


Figure I.9 : Exemple d'un calcul d'une matrice de co-occurrence avec une fenêtre de 5×5 pixels centrée sur le pixel cible, une orientation de 0° , un décalage (offset) de 1 pixel, et 4 niveaux de gris (de 0 à 3). Les paires de pixels sont comptabilisées dans la matrice de co-occurrence, les indices statistiques de Haralick sont ensuite calculés à partir de cette matrice.

Plusieurs études ont évalué la capacité des indices de texture spatiale pour estimer les principaux paramètres de structure et la biomasse aérienne sur différents types forêts. Une grande part de ces recherches ont utilisé des données optiques en raison de la plus grande précision spatiale de plusieurs capteurs optiques. La très haute résolution spatiale optique a permis d'estimer les paramètres forestiers dans des forêts tempérées et boréales (Beguet et al., 2014; Kayitakire et al., 2006; Ozdemir and Karnieli, 2011; Song et al., 2010; Tuominen and Pekkarinen, 2005) et la biomasse aérienne dans des forêts tempérées et tropicales (Bastin et al., 2014; Coutron et al., 2005; Meng et al., 2016; Proisy et al., 2007; Singh et al., 2015) avec des coefficients r^2 parfois supérieurs à 0.9, combinant les informations sur la distribution des arbres et sur la structure même des houppiers grâce aux résolutions spatiales inférieures à 2 m. La très haute résolution spatiale ($< 1-2$ m) est cependant difficile à obtenir de manière globale et annuelle. Les images avec une moins bonne résolution spatiale donnent des résultats d'estimation légèrement moins bons (Proisy et al., 2007; Singh et al., 2015) mais peuvent être intéressants pour une utilisation plus élargie de ce type de méthodes. Plusieurs études ont utilisé des images à haute résolution spatiale (10-30 m) optique et radar pour calculer les indices de texture et estimer les paramètres forestiers et la biomasse dans divers types de forêts (Castillo-Santiago et al., 2010; Dube and Mutanga, 2015b; Sarker and Nichol, 2011; Sarker et al., 2012; Wolter et al., 2009).

Combinaison de capteurs

La disponibilité croissante des images optiques et radar partout dans le monde, fréquemment et avec une haute résolution spatiale permet de mener de plus en plus d'études multi-capteurs en combinant les images optiques, les images SAR multi-fréquence et les textures spatiales. Ces études ont montré que la combinaison de plusieurs types de données satellitaires pouvait améliorer les estimations des paramètres forestiers sur différents types de forêts avec

différentes données satellitaires et différentes méthodes de régression (Attarchi and Gloaguen, 2014; Dobson et al., 1995; Vafaei et al., 2018). Les travaux combinent souvent les capteurs ALOS-PALSAR et Landsat en utilisant l'intensité, la texture ou les deux (Attarchi and Gloaguen, 2014; Gao et al., 2016; Vaglio Laurin et al., 2016; Zhao et al., 2016). Plus récemment, des études intègrent les données Sentinel-1 ou Sentinel-2 avec ou sans ALOS-PALSAR (Chen et al., 2019, 2018; Debastiani et al., 2019; Haywood et al., 2018; Laurin et al., 2018; Liu et al., 2019; Vafaei et al., 2018). Ces études, sur des forêts tropicales, tempérées et en zone montagneuse, permettent d'estimer la biomasse aérienne avec statistiques de validation indépendantes globalement bonnes mais variables d'un site à l'autre : r^2 de 0.54 à 0.76 ; RMSE de 24 à 47 tonnes/ha, soit 15 à 40 % d'erreur relative. Les erreurs se situent à peu près dans la même plage de valeurs que celles observées dans la littérature pour l'estimation de la biomasse aérienne à partir de données 3D.

Modèles

Relier les paramètres forestiers à des données issues d'images optiques, SAR et de textures spatiale, sur différents types de forêts, est difficile à réaliser avec une approche de modélisation, c'est pourquoi la plupart des études utilisent une approche empirique ou non-paramétrique. L'établissement d'équation empiriques entre les paramètres forestiers et chacune des données spatiale permet d'interpréter plus facilement les estimations et les erreurs associées, mais la tâche devient très chronophage si on combine de l'optique, du radar, des textures spatiales, et sur différentes forêts aux propriétés et aux relations allométriques totalement différentes. Une grande partie des études utilisent une régression multilinéaire, qui est souvent rangée dans la catégorie des méthodes non-paramétriques d'apprentissage automatique, mais qui en réalité suppose des relations linéaires entre chaque variable explicative (donnée satellitaire ou transformation log/exp) et le paramètre forestier à estimer. Certaines études utilisent des méthodes non-paramétriques telles que la régression en forêt aléatoire (RF, random forest regression) (Svetnik et al., 2003), qui est plus connu pour les tâches de classification mais qui permet des régressions non linéaires, et se construit à partir d'arbres de décision binaires. Les machines à vecteur de support (SVM, support vector machine) sont aussi principalement connues pour les tâches de classifications, mais peuvent être utilisées pour la régression (SVR) (Cortes and Vapnik, 1995; Drucker et al., 1996). La régression RF est plus adaptée à des grands jeux de données. Les SVR se basent sur des régressions linéaires, mais opèrent des changements de dimension qui permettent de faire des régressions non linéaires. Comparé au RF, le SVR nécessite moins d'échantillons d'apprentissage et pourrait être plus adapté aux études sur les paramètres forestiers et la biomasse pour lesquelles le nombre de mesures terrain est souvent limité. Enfin, on voit maintenant beaucoup de littérature en télédétection utilisant les réseaux de neurones. Ces méthodes sont connues pour être très adaptées à la reconnaissance de forme, mais leur utilisation s'élargit ces dernières années à d'autres domaines. Dans le cadre de l'estimation des paramètres forestiers ou de la biomasse aérienne, il semble cependant que la majeure partie des jeux de données disponibles actuellement soient trop réduits pour pouvoir utiliser des réseaux de neurones qui demandent de larges jeux de données.

1.4. Objectifs de recherche et structure du manuscrit

Il y a une demande pressante d'informations supplémentaires sur la ressource forestière, sa répartition sur les territoires et son évolution dans le temps avec de meilleures résolutions spatiales et temporelles. Le contexte actuel de la télédétection spatiale est exceptionnel, avec la disponibilité en libre accès de données globales, fréquentes, à haute résolution spatiale et issues de différents types de capteurs. Parmi ces données, les satellites Sentinel-1, Sentinel-2 et ALOS2-PALSAR2 font partie de programmes de long terme, et peuvent ainsi répondre au besoin de suivi des forêts avec une méthode générique, pour avoir des produits consistants utilisables à l'échelle du peuplement comme à l'échelle régionale ou nationale.

Le potentiel de ces données de type Sentinel-1, Sentinel-2 et ALOS2-PALSAR2 pour l'estimation de la biomasse et des paramètres de structure forestière a été validé dans plusieurs études pour chaque capteur (optique ou radar) séparément. Par la suite, de multiples méthodes ont été utilisées pour combiner ces données satellitaires sur différentes forêts et améliorer les estimations. L'utilisation de données complémentaires permet de dépasser les limites propres à chaque capteur (saturation, conditions climatiques, fréquence des données, etc.). Mais chaque étude a développé une méthodologie et une combinaison de capteurs spécifique au site d'étude. De plus, à l'heure actuelle, les relations entre la biomasse et le signal radar/optique doivent être étalonnées pour chaque type de forêt (essences, mode de gestion, nature du sous-bois, etc.). Un des enjeux de la thèse est donc de combiner des données optiques et radar (leur synergies) afin de réduire ce travail d'étalonnage et de permettre l'estimation de la biomasse sur de grands territoires sur lesquels divers types de forêts sont présents. Ensuite, la thèse a pour ambition de développer les bases scientifiques et les méthodes d'exploitation de données de télédétection satellitaires pour le suivi et l'estimation de la biomasse des forêts, avec une méthodologie générique et automatique. D'après les résultats observés dans la littérature pour les données 3D et les études multi-capteurs, l'objectif en terme de RMSE relatif pour les principaux paramètres forestiers serait de 25-30 % pour la biomasse aérienne, le volume et la surface terrière ; 20-25 % pour le diamètre moyen et l'âge (dans le cas de plantations) ; et 15 à 20 % pour l'estimation de la hauteur.

Le manuscrit de thèse est organisé de la manière suivante. Dans le Chapitre II, nous présentons la majeure partie du travail méthodologique et d'analyse pour la combinaison des données Sentinel et ALOS-PALSAR pour l'estimation des paramètres forestiers sur un jeu de données de plantations de pins maritimes dans les Landes. Le Chapitre III présente la validation de la méthodologie sur différentes forêts en France. Dans le Chapitre IV, nous présentons les bases méthodologiques de la spatialisation des modèles pour la production de cartes de paramètres forestiers. Enfin, le Chapitre V présente les conclusions et les perspectives de ce travail de thèse.

Chapitre II : Construction de la méthode

Chapitre II : Construction de la méthode.....	15
2.1. Site d'étude.....	15
2.2. Mesures de référence terrain.....	17
2.3. Données explicatives.....	19
2.4. Pré-analyse des données.....	24
2.4.1. Analyse des paramètres forestiers.....	24
2.4.2. Analyse des séries temporelles.....	26
2.4.3. Analyse des textures.....	29
2.4.4. Relations entre les paramètres forestiers et les données satellitaires.....	32
2.4.5. Corrélations entre primitives (variables explicatives).....	33
2.5. Approche méthodologique.....	35
2.6. Extraction des primitives.....	37
2.7. Réduction de dimensions.....	39
2.7.1. Analyses en Composantes Principales.....	39
2.7.2. Méthodes de sélection de primitives « stepwise ».....	41
2.7.3. Comparaison entre l'ACP et les méthodes de sélection de primitives.....	42
2.8. Algorithmes de régression.....	43
2.8.1 Comparaison des algorithmes.....	44
2.8.2. Paramétrisation du SVR.....	45
2.8.3. Évaluation des modèles.....	47
2.9. Résultats.....	49
2.9.1. Hauteur moyenne (Height).....	51
2.9.2. Diamètre moyen à 130 cm / mean DBH.....	52
2.9.3. Âge des peuplements.....	53
2.9.4. Densité d'arbres.....	53
2.9.5. Surface terrière / Basal Area (BA).....	54
2.9.6. Aboveground Biomass (AGB).....	55
2.10. Analyse de l'apport des types de données satellitaires.....	56
Ajout de textures à très haute résolution spatiale.....	58
2.11. Conclusion.....	60

2.1. Site d'étude

Pour la construction de la méthodologie, nous avons utilisé un site d'étude situé dans la forêt des Landes de Gascogne dans le Sud-Ouest de la France (Figure II.1). La forêt des Landes s'étend sur environ 1 400 000 hectares, composés à 85 % de pin maritime (*Pinus pinaster* (Ait.)). Il s'agit de la plus grande forêt plantée en Europe. La zone présente très peu de relief, à l'exception de la zone dunaire située à l'Ouest près de la côte. Le sol est de type podzol (lessivé et acide). Le climat est de type océanique. La moyenne annuelle des températures est de 12.6 °C, et il tombe en moyenne 1015 mm de pluie par an.

Le choix de cette forêt a été fait sur la base des données disponibles, mais aussi en raison des caractéristiques très particulières de la forêt des Landes, propices à la construction de notre méthode avec l'utilisation d'images satellites et d'algorithmes d'apprentissage automatique. En effet, les peuplements des Landes sont souvent gérés à la parcelle, ce qui fait que l'âge et la gestion sont identiques à l'intérieur de chaque parcelle. De plus, la gestion intensive et le cycle de coupes rases et de renouvellement/plantation sur 35 à 50 ans permettent de pouvoir observer dans une même zone restreinte l'ensemble des stades de développement de ces forêts de pin maritime. Ce facteur est très important lorsqu'on utilise des algorithmes d'apprentissage automatique, car les données de références doivent être représentatives de tout la dynamique de valeurs des paramètres forestiers.

La sylviculture du pin maritime est généralement assez similaire d'un peuplement à l'autre, mais avec un rythme et une intensité qui peuvent varier. Un ou deux dépressages ont lieu durant les 10 premières années (ou jusqu'à une hauteur de 4 mètres), il s'agit d'éliminer la végétation concurrente au pin maritime. Ensuite, la première éclaircie a lieu entre 15 et 20 ans (~12 mètres de hauteur), avec un prélèvement de 30 à 40 % des tiges. Puis deux à trois éclaircies sont à nouveau effectuées tous les 7 à 10 ans. La coupe rase a lieu entre 35 et 50 ans, généralement lorsque les tiges atteignent un diamètre à hauteur d'épaule (Diameter at Breast Height, DBH) de 45 centimètres et un volume de bois sur pied entre 300 et 400 m³/ha équivalent à ~120-140 tonnes/ha de biomasse sèche aérienne (AGB). Certains peuplements peuvent être gardés plus longtemps (~60 ans) si l'objectif est de produire un bois d'œuvre de haute qualité.

Le site d'étude est situé dans la partie Nord de la forêt des Landes à l'est et au sud du bassin d'Arcachon (Figures II.1 et II.2) en dehors de la zone dunaire, sur une étendue d'environ 20 × 25 km. Les parcelles de pin maritimes de la partie Sud-Ouest (Nezer) sont gérées de manière intensive, la plupart sont issues de plantations. Les parcelles de la partie Nord-Est (Tagon et Marcheprime) ont des gestions plus diversifiées, souvent issues de semis.

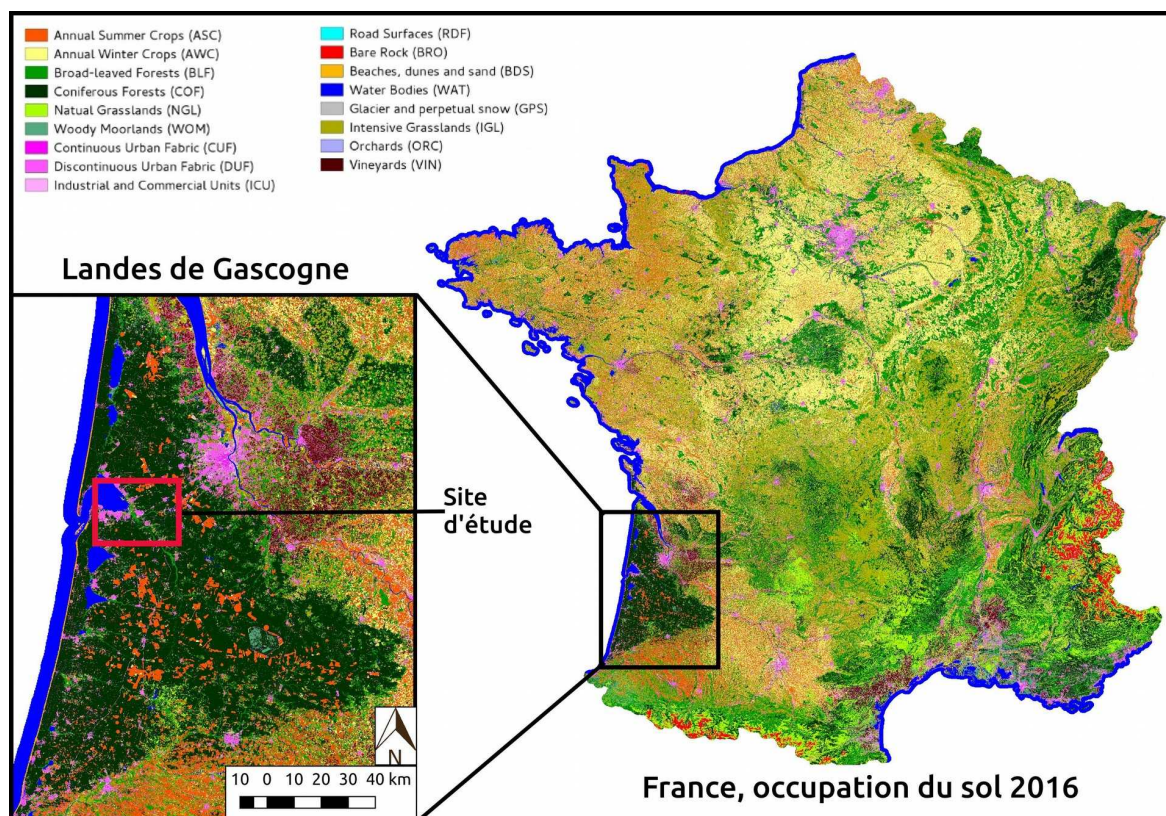


Figure II.1 : Localisation du site d'étude dans les Landes de Gascogne. Fond de carte : occupation du sol en 2016 (Inglada et al., 2017).

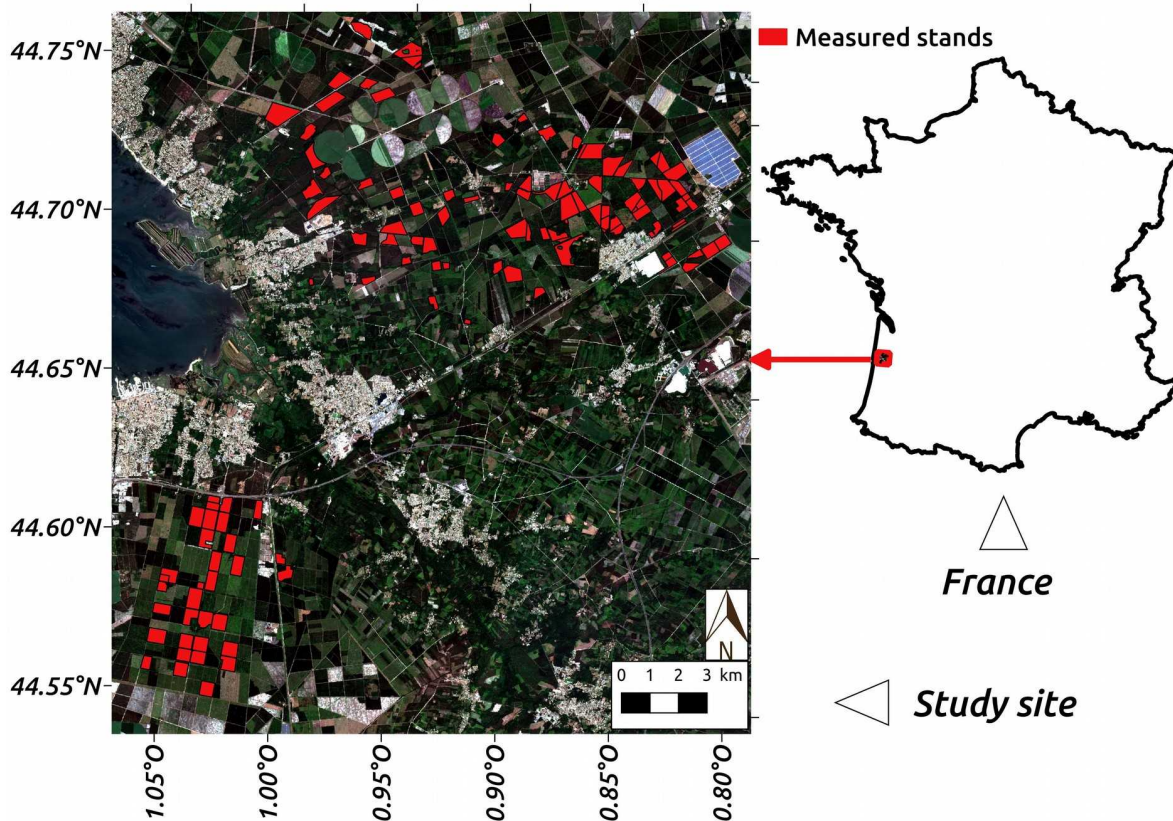


Figure II.2 : Visualisation du site d'étude sur une image optique (Sentinel-2). Les parcelles échantillonnées apparaissent en rouge.

2.2. Mesures de référence terrain

Les mesures terrain ont été faites par l'INRA (ISPA) dans le cadre du projet GLORI (Motte et al., 2016; Zribi et al., 2019). 83 peuplements de pin maritime ont été sélectionnés afin de couvrir toute la dynamique des paramètres de structure forestière pour ce type de peuplement dans la forêt des Landes. Les mesures ont été faites début 2016 avant la période de croissance des arbres.

Protocole

Au sein de chaque peuplement, les mesures ont été faites dans une placette d'une surface de 0.05 à 0.2 hectares (12 à 25 m de diamètre). Chaque placette est située dans une zone homogène éloignée de la bordure du peuplement, et le centre est géolocalisé avec un GPS. Six paramètres forestiers mesurés ont un fort intérêt pour la gestion forestière ainsi que le suivi du bilan carbone :

- **Âge (années)** : donné par les gestionnaires forestiers ou si besoin par comptage du nombre apparent de cycles annuels de croissance en hauteur, par repérage des verticilles de branche.
- **Densité (arbres/ha)** : dans chaque placette, tous les arbres sont comptés pour calculer la densité d'arbres par hectare.
- **DBH (cm)** : mesuré pour les 10 arbres vivants les plus proches du centre de la placette. On utilise ensuite un DBH moyen (moyenne des 10 arbres) pour la placette.
- **Hauteur (m)** : mesurée pour les 5 arbres vivants les plus proches du centre de la placette, elle peut ensuite être dérivée en hauteur moyenne ou en hauteur dominante pour la placette.

- **Surface terrière (Basal Area, BA) (m²/ha)** : surface de toutes les sections transversales des troncs à 1m30, calculée à partir du DBH pour les 10 arbres du centre de la placette. Elle est ensuite ramenée à une unité surfacique à la placette avec la densité d'arbres.
- **AGB (tonnes/ha)** : estimée avec une équation allométrique au niveau de l'arbre en utilisant le DBH seul (pour chacun des 10 arbres), ou le DBH et la hauteur (pour chacun des 5 arbres) comme variables d'entrée (équations 1 et 2), puis ramenée à une unité surfacique à la placette avec la densité.

Les estimations de biomasse sont issues d'une étude spécifique au pin maritime (Shaiek et al., 2011) pour déterminer les équations allométriques et leurs coefficients :

$$AGB_i = a * DBH_i^b \quad (1)$$

Avec $a = 5.013$ tonnes/m et $b = 2.48$. AGB_i est la biomasse aérienne individuelle de chaque arbre en tonnes de matière sèche, et DBH_i est le diamètre individuel de chaque arbre en mètre.

$$AGB_i = a * DBH_i^b * H_i^c \quad (2)$$

Avec $a = 3.177$ tonnes/m, $b = 2.38$ et $c = 0.12$. AGB_i est la biomasse aérienne individuelle de chaque arbre en tonnes, et DBH_i est le diamètre individuel de chaque arbre en mètres. Étant donné le protocole de mesure où seules les hauteurs des 5 arbres les plus proches du centre sont mesurées, contre 10 arbres pour le DBH, nous choisissons dans la suite de l'étude d'utiliser l'équation (1) avec $AGB=f(DBH)$ seul afin d'agrandir l'échantillonnage, d'autant plus que le DBH et la hauteur des arbres sont très corrélés sur des plantations de pin maritime. Une équation allométrique (Deleuze et al., 2014) a également été utilisée pour calculer le volume au niveau arbre :

$$VolTot = \frac{H_{tot} C_{130}^2}{4\pi \left(1 - \frac{1.3}{H_{tot}}\right)^2} \left(a + \frac{b * \sqrt{C_{130}}}{H_{tot}} + \frac{c * H_{tot}}{C_{130}} \right) \quad (3)$$

Où $VolTot$ est le volume de ligneux aérien, H_{tot} est la hauteur totale de l'arbre, et C_{130} est la circonférence de l'arbre à 130 cm de hauteur. Les constantes ont été déterminées dans Deleuze et al. (2014) pour le pin maritime (*Pinus pinaster*) avec $a = 0.396$, $b = 1.756$ et $c = -0.002$. Les valeurs d'AGB et de volume étant très corrélés ($r^2 > 0.99$), nous n'estimerons pas le volume dans la suite de l'étude.

Statistiques descriptives

Le tableau II.1 montre les minimums, maximums, moyennes et écart types des différents paramètres forestiers sur les placettes de mesure de notre site d'étude. Les minimums et maximums sont représentatifs des valeurs que l'on peut trouver en début et en fin de cycle pour les parcelles de pin maritimes dans les Landes, et les valeurs sont bien distribuées (Tableau II.1 et Figure II.3). La biomasse de ce type de forêt est inférieure à ce qu'on pourrait trouver dans les forêts matures de chêne ou d'autres plantations de résineux (épicéas commun, sapin pectiné), mais le volume moyen de notre échantillon (183 m³/ha) est comparable au volume moyen des forêts en France métropolitaine (170 m³/ha) d'après les données de l'inventaire forestier national.

Tableau II.1 : Statistiques descriptives des paramètres forestiers mesurés sur les placettes.

	Minimum	Maximum	Moyenne	E. T.
AGB (tonnes/ha)	1	136.2	70	40.2
Volume (m ³ /ha)	4.1	370.7	183.0	108.1
BA (m ² /ha)	0.8	44.1	21.2	11.0
DBH moyen (cm)	3	57	21.5	12.4
Age (années)	3	72	22.6	14.7
Densité (arbres/ha)	87	2622	835	537
Hauteur dominante (m)	2	24.7	14.4	6.6

2.3. Données explicatives

Les forêts françaises, et de nombreuses forêts tempérées en général, sont fragmentées soit par leur surface, soit par leur mode de gestion. En effet, on estime que seul 1 à 2 % des forêts tempérées n'ont jamais été coupées depuis le 15^e siècle (Currie and Bergen, 2008), ce sont donc des milieux fortement anthropisés. On peut observer côte à côte, sur quelques centaines de mètres, des essences et des biomasses très différentes. Pour cette raison, nous avons choisi de nous orienter vers de l'imagerie satellitaires à haute résolution spatiale (~10-20 m). De plus, plusieurs programmes spatiaux fournissent actuellement en libre accès des images sur l'ensemble des terres immergées, à haute résolution spatiale et temporelle avec une programmation assurée sur le long terme (Copernicus¹). Ces caractéristiques sont importantes pour un objectif de reproductibilité et de généricité. Dans ce contexte, nous avons choisi comme base de travail les capteurs Sentinel-1, Sentinel-2 et ALOS2-PALSAR2 (mosaïques annuelles). D'autres types d'images ont des conditions d'accès similaires à ces données (Landsat-8 par exemple, avec une résolution spatiale et temporelle inférieure à Sentinel-2), ou bien pourraient apporter de nouvelles informations sur la végétation au sol, mais ne remplissent pas les conditions de libre accès ou de couverture systématique du globe (la très haute résolution spatiale, l'hyper-spectral, le Lidar, etc.). D'autres types d'images ou de produits (SMOS, MODIS, etc.) peuvent être aussi utilisés pour aider au pré ou post-traitement des images et des cartographies à haute résolution. Ces images ne seront pas intégrées par défaut à la construction de notre méthodologie pour la cartographie des paramètres forestiers, mais pourraient être ajoutées dans le cadre d'études spécifiques. Dans cette étape de construction de la méthode, nous avons ajouté en dernière partie les images SPOT-6/7, disponibles annuellement sur la France pour les services publics, afin de tester l'utilité d'images optiques avec une meilleure résolution spatiale (6 m pour le mode multi-spectral, 1.5 m pour le mode panchromatique).

Le tableau II.2 montre les principales caractéristiques des 4 capteurs retenus. Les données explicatives (aussi appelés primitives) que nous allons utiliser dans l'algorithme d'apprentissage sont composés par les bandes spectrales de base (rétrodiffusion radar et réflectance optique) ainsi que des indices spectraux et des indices de textures calculés sur ces bandes spectrales. Les caractéristiques détaillées et les traitements appliqués sont présentées par la suite.

1 <https://www.copernicus.eu/fr>

Tableau II.2 : Caractéristiques des satellites utilisés dans l'étude.

Capteur	Produit	Résolution et couverture spatiale	Résolution temporelle	Donnée
ALOS2-PALSAR2 / ALOS-PALSAR (JAXA)	Mosaïque annuelle	25 × 25 m Mondiale	Annuel 2007-2010 2015- maintenant	SAR Polarisations HV et HH
Sentinel-1 (ESA)	Séries temporelles GRD	10 × 20 m Mondiale	12 à 6 jours 2015- maintenant	SAR Polarisations VH et VV
Sentinel-2 (ESA)	Séries temporelles	10 × 10 et 20 × 20 m Mondiale	5 à 10 jours 2016- maintenant	Optique VIS-PIR-MIR
SPOT-6/7 (CNES/Airbus)	Mosaïque annuelle	6 × 6 m France métropolitaine	Annuel 2013- maintenant	Optique VIS-PIR
	Mosaïque annuelle	1.5 × 1.5 m France métropolitaine	Annuel 2013- maintenant	Optique Panchromatique

Rétrodiffusion radar en bande L (ALOS2-PALSAR2)

Le satellite ALOS a été lancé en 2006 par la JAXA. Il est équipé notamment d'un radar à synthèse d'ouverture (SAR) en bande L (23cm, 1.2 GHz) et a fourni des images sur le monde entier entre 2006 et 2010. ALOS2 a été lancé en 2014, équipé d'un capteur similaire en bande L. Les images du mode FBD (pour *Fine Beam Dual polarization*) ont été utilisées par la JAXA pour faire des mosaïques annuelles couvrant le monde entier pour la période 2007-2010 et depuis 2015. Ces mosaïques sont disponibles gratuitement en libre accès². La JAXA livre les images par tuiles de 1 × 1° à 25 mètres de résolution spatiale avec un pré-traitement (algorithmes sigma-SAR IMAGE & MOSAIC) incluant l'orthorectification, la correction des pentes et la calibration radiométrique entre les tuiles adjacentes. Les bandes HH et HV sont disponibles. Nous avons téléchargé l'ensemble des années disponibles avec ALOS/ALOS2 sur notre site, soit au total 6 couples d'images annuelles HV et HH (2007-2010 et 2015-2016). Les valeurs sont converties d'abord en rétrodiffusion gamma, γ^0 (équation 4),

$$\gamma^0 = 10 * \log_{10}(DN^2) + CF \quad (4)$$

Où γ^0 est le coefficient de rétrodiffusion en décibel, DN est le *digital number* contenu dans les pixels, et CF est le facteur de calibration (-83.0 dB pour PALSAR/PALSAR2).

Les images radar sont caractérisées par un bruit avec un aspect « poivre et sel » appelé Chatoiement ou Speckle. Une des techniques pour réduire ce bruit est d'effectuer un multi-look temporel : moyenner des images ou vues (« looks »). Le nombre de « looks » est le nombre d'images indépendantes pour une zone cible. Étant donné que les images voisines dans le temps peuvent être corrélées entre elles, on définit un nombre de vues équivalent (« equivalent number of looks », ENL, équation 5) qui est généralement plus petit que le nombre d'images moyennées. L'ENL est le nombre de vues indépendantes qui, moyennées, donneraient le même bruit de Speckle. L'ENL décrit ainsi le degré de moyenne appliqué aux mesures SAR, il indique la qualité de l'image.

² https://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm

Nous avons donc appliqué un filtre multi-images (Bruniquel and Lopes, 1997; Quegan and Yu, 2001) afin de réduire le Speckle et augmenter l'ENL à 156 (équation 5).

$$ENL = \frac{M * N * L}{M + N - 1} \quad (5)$$

Où M est le nombre d'images utilisées en entrée, N est le nombre de pixels dans la fenêtre de filtrage, et L est le nombre de vues (*number of looks*) des images en entrée, soit L=16 pour les mosaïques annuelles ALOS-PALSAR et ALOS2-PALSAR2.

Le coefficient de rétrodiffusion en bande L (ALOS2-PALSAR2), avec une longueur d'onde d'environ 23 cm, est considéré comme étant sensible au volume de bois dans la canopée et donc aux paramètres AGB et BA. La polarisation croisée (HV) est normalement plus sensible au volume de bois, tandis que HH est plus sensible à la réflexion du sol et au double rebond sol-tronc.

Rétrodiffusion radar en bande C (Sentinel-1)

Sentinel-1A a été lancé en avril 2014, et Sentinel-1B en avril 2016. Ils fournissent des données en bande C (5.7 cm, 5.4 GHz). Les images sont disponibles en SLC (*Single Look Complex*) et en GRD (*Ground Range Detected*). Nous avons utilisé les images GRD (amplitude) avec deux polarisations VH et VV et une résolution spatiale de 20 m. Les informations de phase (produit SLC) n'ont pas été utilisées afin de réduire le volume de données et de rester sur des traitements simples et opérationnels.

Les données ont été téléchargées via la plateforme PEPS³ du CNES, et nous avons appliqué les pré-traitements avec la chaîne s1tiling⁴ développée par Thierry Kolléck au CESBIO : elle applique successivement les opérations de calibration, d'orthorectification, de découpage sur les tuiles Sentinel-2 traitées par la plateforme THEIA⁵ (100 × 100km), et enfin l'application d'un filtre multi-image pour réduire le Speckle (Bruniquel and Lopes, 1997; Quegan and Yu, 2001) en séparant les polarisations VV et VH. Chaque satellite acquiert une image tous les 12 jours, avec un décalage de 6 jours entre les deux satellites pour avoir un intervalle fixe de 6 jours entre chaque acquisition. Pour la majeure partie de 2016, seules les images d'un satellite sont disponibles. Avec environ 30 images dans l'année, l'ENL en sortie de la chaîne de pré-traitement est de ~87 (équation 5, L=4.9).

La bande C (Sentinel-1) a une longueur d'onde plus petite (~5.7 cm) que la bande L. Elle est sensible plutôt aux feuilles, aiguilles et petites branches. Ces données fournissent des informations sur la structure de la canopée, qui elle-même peut être reliée au volume de bois en particulier dans des plantations régulières.

Réflectance optique (Sentinel-2)

Les deux satellites Sentinel-2A et Sentinel-2B ont été lancés en juin 2015 et mars 2017 respectivement. Chaque satellite acquiert une image tous les 10 jours, avec un décalage de 5 jours entre les deux satellites pour avoir un intervalle fixe de 5 jours entre chaque acquisition. Sentinel-2 fournit des images optiques du visible au moyen infra-rouge (Tableau II.3.), sur une fauchée de 290 kilomètres avec une résolution de 10 à 20 m pour les bandes d'intérêt dans cette étude. Les images ont été téléchargées sur la plateforme THEIA avec orthorectification, correction des effets de pente, correction des effets atmosphériques et un masque de nuages avec la chaîne de traitement MAJA (Hagolle et al., 2015, 2010). Les images sont de niveau

3 <https://peps.cnes.fr>

4 <http://tully.ups-tlse.fr/kolleckt/s1tiling>

5 Pôle de données et de services surfaces continentales : <https://www.theia-land.fr>

2A « Top of Canopy » (TOC), livrées sur des tuiles de 100 × 100 km. Malgré l'acquisition d'une image tous les 10 jours avec Sentinel-2A, la présence régulière de nuages peut empêcher le traitement niveau 1A des images si plus de 90 % de la tuile 290 × 290 km est couverte de nuages (le géo-référencement et l'orthorectification ne sont pas possibles), ou le traitement niveau 2A si plus de 90% de la tuile 100 × 100 km est couverte de nuages. Ainsi, seules 18 dates sont disponibles sur notre site en 2016, parmi lesquelles on trouve 7 dates où l'ensemble des points de mesure sont visibles. Un seul satellite était en opération en 2016, les séries temporelles devraient être plus denses depuis le lancement de Sentinel-2B.

Tableau II.3 : Caractéristiques des bandes spectrales du capteur Sentinel-2.

Canal	Résolution spatiale	Bande spectrale
Aérosol	60 m	0.43 – 0.45 µm
Bleu	10 m	0.46 – 0.52 µm
Vert	10 m	0.54 – 0.58 µm
Rouge	10 m	0.65 – 0.68 µm
Red-edge 1	20 m	0.70 – 0.71 µm
Red-edge 2	20 m	0.73 – 0.74 µm
Red-edge 3	20 m	0.77 – 0.79 µm
Proche infra-rouge (PIR1)	10 m	0.78 – 0.90 µm
Proche infra-rouge (PIR2)	20 m	0.85 – 0.87 µm
Vapeur d'eau	60 m	0.93 – 0.95 µm
Moyen infra-rouge 1 (MIR1)	20 m	1.57 – 1.66 µm
Moyen infra-rouge 2 (MIR2)	20 m	2.10 – 2.28 µm
Cirrus	60 m	1.37 – 1.39 µm

La réflectance issue des données optiques (Sentinel-2) est sensible à la surface foliaire, la fraction de couvert (arbres et sous-bois) et les propriétés du sol. Ces données fournissent des informations sur le développement de la végétation et l'activité photosynthétique. Différents indices spectraux permettent de relier les réflectances à des propriétés biophysiques au sol : l'indice de végétation normalisé (NDVI, eq.6), l'indice de brillance (BI, eq.7), et l'indice d'humidité (NDWI, eq.8).

$$NDVI = \frac{R_{PIR1} - R_{Rouge}}{R_{PIR1} + R_{Rouge}} \quad (6)$$

$$BI = \frac{\sqrt{R_{Bleu}^2 + R_{Vert}^2 + R_{Rouge}^2}}{3} \quad (7)$$

$$NDWI = \frac{R_{PIR2} - R_{MIR1}}{R_{PIR2} + R_{MIR1}} \quad (8)$$

où R est la réflectance dans les différentes bandes spectrales, PIR étant le proche infra-rouge et MIR étant le moyen infra-rouge.

Indices de textures

Les indices de textures extraits des images Sentinel-1 (VH et VV) et Sentinel-2 (NDVI et BI) permettent de caractériser les pixels par l'organisation de leur environnement (patterns, homogénéité, etc.). Selon les types de forêts, l'organisation spatiale peut être reliée à la maturité des forêts (densité et homogénéité du couvert) ou bien à la taille des houppiers, fournissant ainsi une information complémentaire pour caractériser les paramètres de structure forestière.

Plusieurs méthodes de transformation texturale ont été développées, nous avons choisi d'utiliser les textures de Haralick (Haralick et al., 1973), calculées à partir des matrices de covariance, et qui ont souvent été utilisées dans la littérature et ont montré des résultats pertinents avec les images satellitaires. De plus, cette méthode est implémentée dans la bibliothèque OTB open-source (Grizonnet et al., 2017) et dans plusieurs autres librairies open source ou propriétaires. 8 indices statistiques de second ordre peuvent être extraits des matrices de covariance : énergie, entropie, corrélation, l'homogénéité (*inverse different moment*), inertie, *cluster shade*, *cluster prominence*, et corrélation de Haralick. Il y a plusieurs paramètres à fixer pour déterminer la base de calcul de ces indices : la taille de fenêtre (en pixels), l'orientation et le décalage (*offset*) des comparaisons entre pixels, et le nombre de classes à considérer pour la transformation en niveaux de gris (*number of bins*). L'étude de ces paramètres et des indices à utiliser est présentée dans la section 2.4.3.

La méthode FOTO (Proisy et al., 2007) pourrait également être utilisée, ayant donné de bons résultats pour l'estimation de la biomasse forestière (Meng et al., 2016; Proisy et al., 2007), mais cette méthode est plus lourde et développée au départ sur logiciel propriétaire (MatLab), ce qui rend actuellement son utilisation moins facile pour la reproductibilité de notre chaîne de traitement.

Nous n'avons pas extrait d'indices de structure spatiale des données en bande L en raison de la résolution spatiale des mosaïques (25 m). La taille de fenêtre nécessaire pour l'extraction de texture serait trop large comparée à la taille de certains peuplements. De plus, Sentinel-1 et Sentinel-2 offrent une résolution spatiale plus adaptée aux textures spatiales.

Imagerie THRS (SPOT-6/7)

SPOT-6 et SPOT-7 ont été lancés en 2012 et 2014 par Airbus. Les acquisitions d'images se font par programmation. Les deux satellites exploités simultanément ont une revisite de 1 jour. Des mosaïques annuelles sont programmées sur la France métropolitaine depuis 2013 et disponibles via la plateforme Géosud⁶ en libre accès pour les collectivités et missions de service public. Les images sont disponibles au niveau 1A : géo-référencées et ortho-rectifiées, réflectance « Top of Atmosphere » (TOA). Le mode multispectral permet d'avoir des images du visible au proche infrarouge à 6 m de résolution spatiale, et le mode panchromatique permet d'avoir une résolution de 1.5 m (tableau II.4.). Contrairement aux mosaïques annuelles ALOS-PALSAR/ALOS2-PALSAR2, aucun algorithme n'est appliqué pour harmoniser les réflectances entre images adjacentes.

Tableau II.4 : Caractéristiques des bandes spectrales des capteurs SPOT-6/7

	Canal	Résolution spatiale	Bande spectrale
Multispectral	Bleu	6 m	0.45 – 0.52 μm
	Vert	6 m	0.53 – 0.59 μm
	Rouge	6 m	0.62 – 0.69 μm
	Proche infra-rouge (PIR)	6 m	0.76 – 0.89 μm
Panchromatique	PAN	1.5 m	0.45 – 0.74 μm

⁶ ids.equipex-geosud.fr

2.4. Pré-analyse des données

2.4.1. Analyse des paramètres forestiers

Étant donné que ces forêts sont des peuplements équiennes et mono-spécifiques avec une sylviculture globalement similaire entre les différents peuplements de pin maritime, on s'attend à ce que les relations allométriques soient très fortes (fortes corrélations entre les paramètres forestiers). Cependant, l'analyse de ces relations à l'échelle de la placette de mesure (figure II.3) montre que les corrélations entre les paramètres forestiers ne sont pas tout le temps évidentes lorsqu'on étudie des valeurs surfaciques et non des valeurs à l'arbre :

- Les plus fortes relations linéaires sont entre le DBH, la hauteur et l'âge (r^2 de 0.74 à 0.91). Les valeurs de hauteurs saturent autour de 20-25 m pour les peuplements matures alors que l'âge et le DBH continuent de croître (à partir de DBH~45 cm et âge~35 ans).
- La biomasse (AGB) et la surface terrière (BA) sont aussi fortement corrélées ($r^2=0.89$), la relation entre l'AGB et le Volume n'est pas montré ici mais les bases de calculs étant similaires, les deux valeurs sont très fortement reliées ($r^2>0.99$).
- Les relations linéaires entre l'AGB et le DBH, la hauteur et l'âge sont assez fortes pour les jeunes peuplements. Jusqu'à ~10-15 ans, DBH~15cm, et hauteur~12-15m. Ensuite, il y a une rupture dans la relation, probablement due aux premières éclaircies qui interviennent lorsque les peuplements atteignent une hauteur dominante autour 12 mètres. On peut observer le même phénomène en comparant la surface terrière (BA) et ces trois mêmes paramètres (DBH, hauteur, âge).
- La densité d'arbres montre des relations linéaires relativement marquées avec le DBH, l'âge et la hauteur (r^2 de 0.54 à 0.64). Cela semble assez cohérent avec le fait que les éclaircies ont lieu en fonction de ces trois mêmes paramètres. Cela montre aussi la diversité des gestions (nombre d'arbres plantés/semés, intensité de l'éclaircie) entre les différentes parcelles. En revanche, les relations entre la densité et les paramètres plus globaux, tels que AGB et BA, sont plus difficiles à décrire.

Nous observons l'impact des dépressages et des éclaircies dans les relations entre les paramètres forestiers des peuplements : ruptures de continuité et dispersions des valeurs. En effet, lors d'une éclaircie, le DBH et la hauteur dominante des arbres restent stables ou augmentent, alors que la biomasse aérienne et la surface terrière, qui dépendent de la densité d'arbres, vont baisser. Cela peut perturber à la fois la construction de modèles statistiques, pour cartographier les paramètres forestiers de plusieurs peuplements à un moment donné, mais aussi le suivi de l'évolution de ces paramètres pour un même peuplement entre plusieurs années si des opérations sylvicoles sont effectuées dans cette période. Pour illustrer la complexité des valeurs surfaciques, la figure II.4 montre deux exemples de parcelles avec des biomasses similaires, mais des autres paramètres dendrométriques très différents.

En conclusion sur ces données terrain, nous pouvons dire qu'elles sont bien représentatives de la dynamique des forêts de pin maritime dans les Landes. Les relations allométriques sont cohérentes avec la sylviculture théorique pour ce type de forêt, mais montrent aussi la diversité dans la pratique de gestion entre les peuplements, due principalement à la densité d'arbres, la fréquence des éclaircies, et l'intensité des éclaircies.

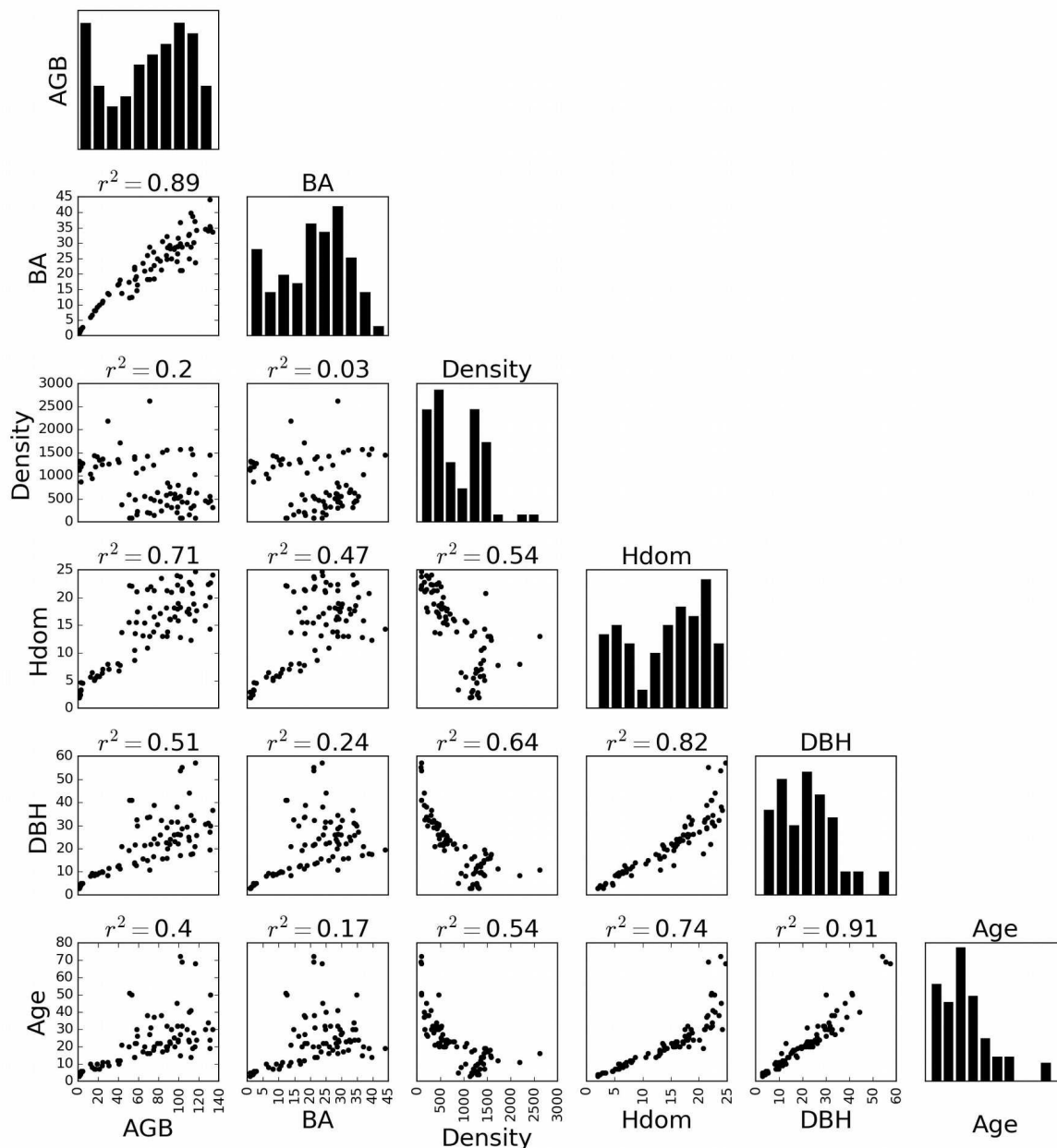


Figure II.3 : Relations au niveau placette entre les paramètres forestiers. La transversale contient les histogrammes (10 classes) de distribution de chaque paramètre.

DBH est le DBH moyen, Hdom la hauteur dominante



Figure II.4 : Exemple de deux parcelles avec la même biomasse aérienne, mais différentes relations allométriques. (a) AGB=118 tonnes/ha, âge=19 ans, DBH=18 cm, densité=1463 arbres/ha ; (b) AGB=116 tonnes/ha, âge=68 ans, DBH=57 cm, densité=92 arbres/ha.

2.4.2. Analyse des séries temporelles

2.4.2.1. Analyse du profil temporel Sentinel-2

Les réflectances évoluent avec la phénologie de la végétation au cours de l'année, l'information fournie par les images Sentinel-2 ne sera pas la même selon les dates d'acquisition. La figure II.5 montre les profils temporels (presque 3 années) des trois indices spectraux sur le site d'étude. Seul le capteur Sentinel-2A a été utilisé afin de pouvoir comparer l'ensemble du profil sur la même base, Sentinel-2B n'étant disponible qu'à partir de juillet 2017. Les profils sont moyennés sur des polygones de 1 ha en forêt fermée afin de réduire les effets liés à l'hétérogénéité ou à des défauts de recalage des images. 10 polygones ont été digitalisés sur des parcelles de pin maritime, et 10 polygones sur des peuplements de feuillus (majoritairement du chêne dans cette zone des Landes), soit environ 1000 pixels pour chaque classe (feuillus et conifères). Nous remarquons les points suivants :

- Abondance et dates d'acquisition des images différentes selon les années (7 en 2016 dont 4 en août/septembre ; 11 en 2017 dont 4 en avril/mai ; 10 en 2018 dont 6-7 de juillet à septembre).
- On observe bien la phénologie des feuillus, avec une augmentation du NDVI au printemps-été et une baisse à l'automne. Les conifères ont un signal NDVI sans variations saisonnières marquées.
- le NDWI semble être aussi un bon marqueur de différenciation entre conifères et feuillus, car le cycle est inversé entre les deux classes de forêt.
- Globalement, lorsqu'il y a un cycle visible, on observe deux plateaux en décembre-mars puis en juin-septembre, assez stables dans le temps et entre les années.

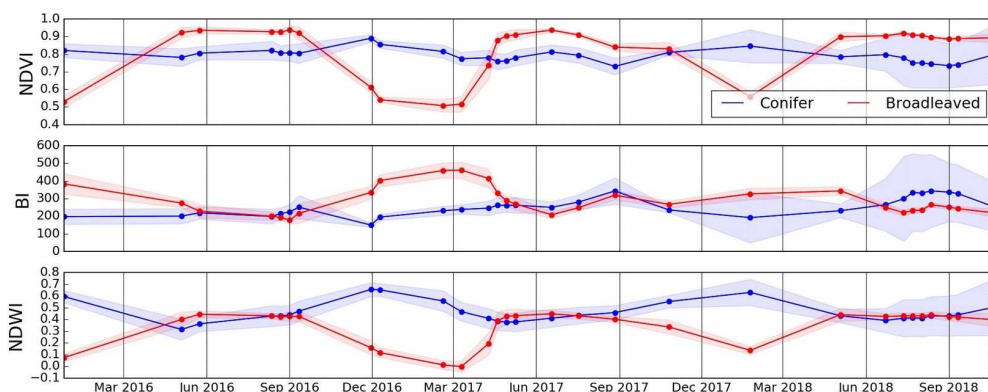


Figure II.5 : Profils temporels NDVI, BI et NDWI sur des peuplements de pins maritime (« Conifer ») et de feuillus (« Broadleaved »), site des Landes.

Les périodes d'acquisition les plus appropriées pour la méthodologie seraient plutôt en été et en hiver d'après la phénologie. Nous choisissons une date en hiver (décembre 2015) et une deuxième en été (août 2016), ce qui permet d'avoir des informations à deux différents stades de phénologie tout en gardant un nombre restreint de primitives. Sur ces peuplements fermés de pin maritime, les différences entre dates ne sont pas très marquées (réflectance et corrélations entre les réflectances et les paramètres forestiers), mais pour garder l'aspect générique de notre méthodologie (application aussi aux feuillus), nous gardons les deux dates.

2.4.2.2. Analyse du profil temporel Sentinel-1

Le signal Sentinel-1 est aussi affecté par la phénologie des arbres et du sous-bois. La figure II.6.a superpose les profils temporels de Sentinel-1 et de Sentinel-2 sur le site d'étude des Landes, toujours avec les peuplements de feuillus et de conifères. Le coefficient de rétrodiffusion VH augmente pendant l'hiver pour les feuillus (+0.05, équivalent à 2-3 dB), tandis que les conifères sont à peu près stables pendant tout l'année. La figure II.8.b montre le profil temporel Sentinel-1 sur des placettes de chêne d'un autre site d'étude dans la forêt d'Orléans, sur lequel nous avons presque 4 ans de profondeur temporelle (seule une année de données Sentinel-1 a été traitée sur le site des Landes). Malgré un fort bruit temporel, l'augmentation de VH l'hiver se vérifie tandis que VV est plus stable (hors bruit). En conséquence, le cycle annuel est très visible sur le ratio VH/VV. Des profils temporels similaires ont été observés dans (Frison et al., 2018), mais cette baisse du signal VH au printemps/été n'est pour l'instant pas expliquée. En effet, la perte des feuilles en hiver devrait plutôt provoquer la baisse de VH en hiver, sauf si le signal VH a une forte composante venant de la rétrodiffusion du sol (à faible angle d'incidence) (Ahern et al., 1993).

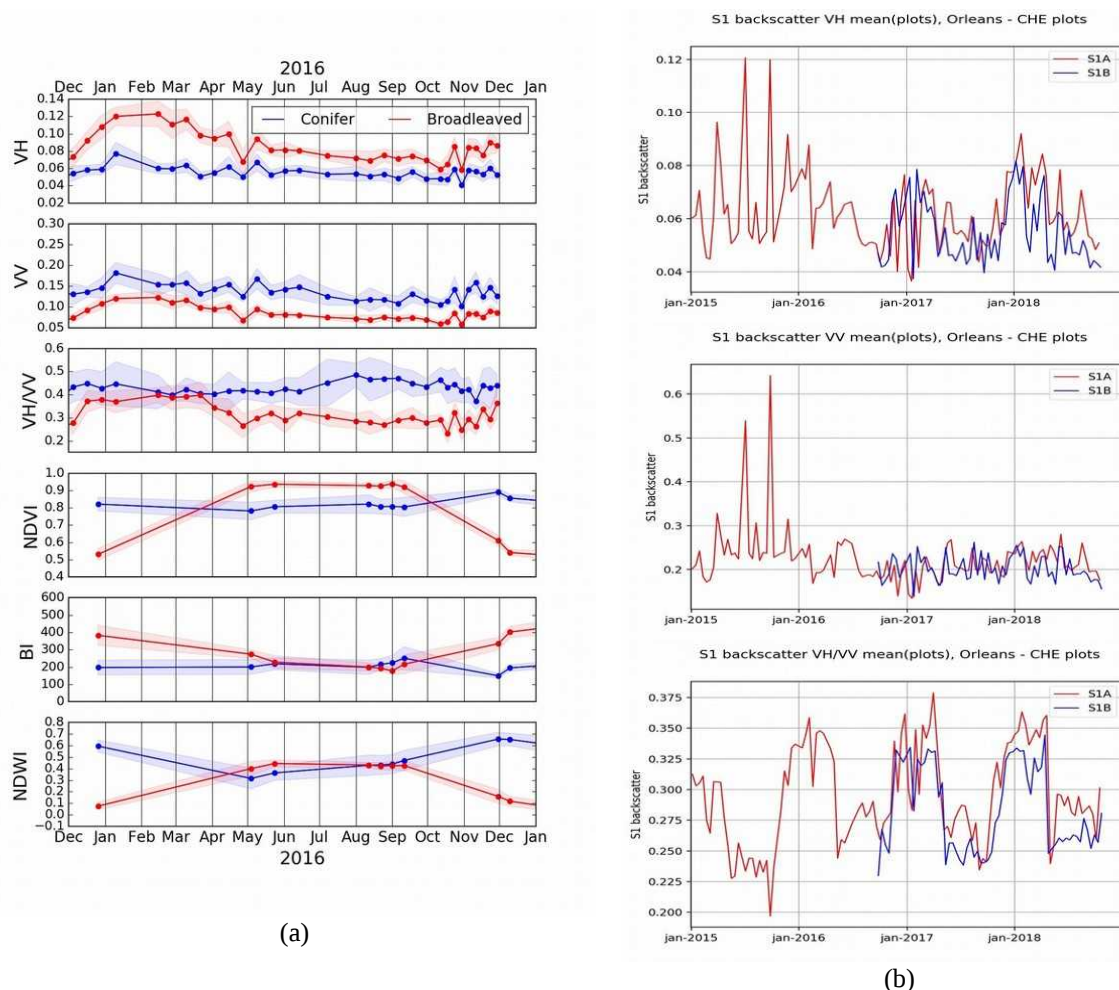


Figure II.6 : Profils temporels Sentinel-1 VH, VV et le ratio VH/VV. (a) Site d'étude des Landes, avec indices spectraux pour la comparaison ; (b) autre site d'étude dans la forêt d'Orléans (chênes) avec une plus grande profondeur temporelle.

Les variations temporelles observées dans la figure II.6 peuvent être gênantes pour la sélection des dates à intégrer dans la méthodologie. Nous nous sommes intéressés à l'origine de ces variations. L'hypothèse la plus évidente est que cette variation du signal radar est due aux variations d'humidité. Une analyse est faite en Annexe B.

Ces variations du signal radar avec l'humidité nous gênent pour retrouver une dynamique liée seulement aux variations de végétation. Cependant, les variations du signal dues à la pluie sont des variations rapides, tandis que les variations dues à un changement de biomasse sont plus lentes. Ainsi, nous avons opté pour moyenner les images de manière mensuelle ou bimensuelle pour réduire les variations temporelles. La figure II.7 montre l'évolution de VH, VV et VH/VV sur les précédentes parcelles de pin maritime et de feuillus (section 2.4.2.1), avec ou sans moyenne mensuelle. Le signal est lissé mais garde les grandes tendances de hausse et de baisse du signal.

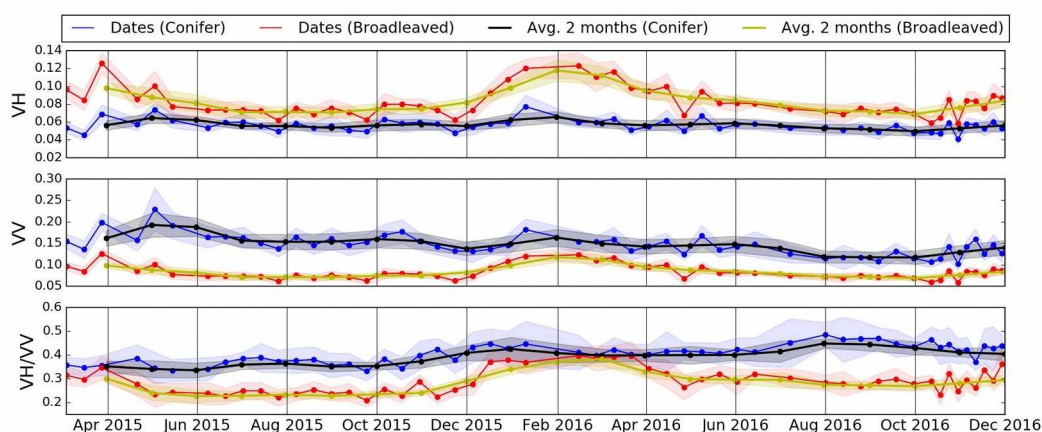


Figure II.7 : Profils temporels Sentinel-1 VH, VV et le ratio VH/VV sur des parcelles de pin maritime (« Conifer ») et de feuillus (« Broadleaved »), par date et avec moyenne mensuelle.

La figure II.8 montre une image Sentinel-1 selon les traitements sur extrait du site d'étude. On peut voir une claire réduction du bruit avec le filtrage multi-image (Fig. II.8.a) comparé à l'image sans filtrage (Fig. II.8.b). La moyenne bimensuelle (Fig. II.8.c) améliore encore la visualisation de la zone, notamment la réduction des variations dans chaque parcelle de pin maritime. Cet aspect visuel est important pour la dernière étape de spatialisation des modèles afin d'obtenir des cartes réalistes des paramètres forestiers.

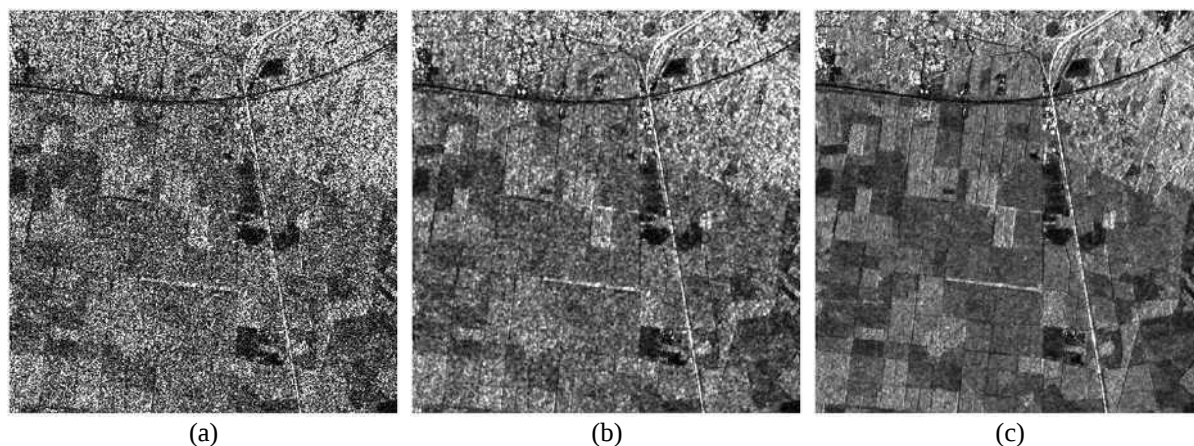


Figure II.8 : Visualisation d'une image Sentinel-1 VH en juillet 2016 (a) avant filtrage, (b) après filtrage multi-image, et (c) avec moyenne des images de juillet et août 2016.

2.4.3. Analyse des textures

Nous avons utilisé les textures de Haralick (Haralick et al., 1973), calculées à partir des matrices de covariance. Les 8 indices statistiques de second ordre extraits des matrices présentent des redondances dans les informations, confirmées par des corrélations entre images de texture. Afin de réduire cette redondance et le volume de données, nous avons sélectionné 4 indices (corrélacion, homogénéité, cluster shade et corrélacion de Haralick) présentés dans le tableau II.5.

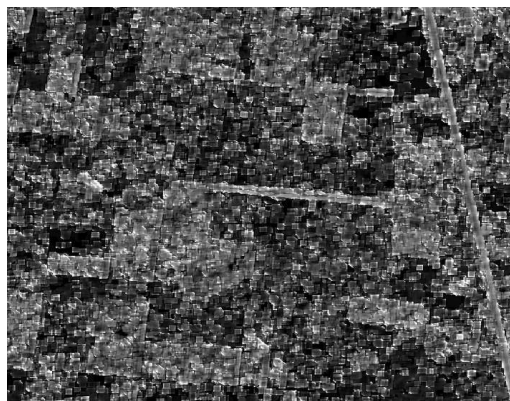
Tableau II.5 : Description des indices de texture de Haralick.

Nom	Équation	Signification
Corrélacion	$\sum_{i,j} \frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2}$	Dépendance linéaire entre pixels voisin
Homogénéité	$\sum_{i,j} \frac{1}{1+(i-j)^2} g(i,j)$	Mesure de l'homogénéité des valeurs de pixels
Cluster Shade	$\sum_{i,j} ((i-\mu)+(j-\mu))^3 g(i,j)$	Moment d'ordre supérieur (répétitivité des patterns)
Corrélacion de Haralick	$\sum_{i,j} \frac{(i,j)g(i,j)-\mu_t^2}{\sigma_t^2}$	Mesure de corrélacion

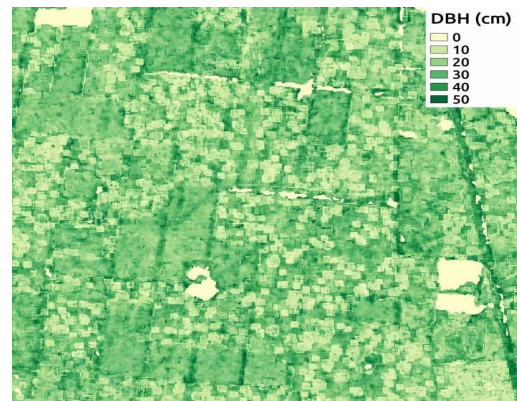
Avec :

- $g(i,j)$ élément de la matrice de cooccurrence normalisée,
- $\mu = \sum_{i,j} i \cdot g(i,j) = \sum_{i,j} j \cdot g(i,j)$ la moyenne pondérée des valeurs des pixels,
- $\sigma = \sum_{i,j} (i-\mu)^2 \cdot g(i,j) = \sum_{i,j} (j-\mu)^2 \cdot g(i,j)$ (matrice symétrique) la variance pondérée des valeurs des pixels

Ces 4 indices amènent des informations pertinentes pour la construction des modèles, tant à partir de Sentinel-1 que de Sentinel-2. Néanmoins, l'analyse des images puis des cartes de paramètres forestiers produites nous a amené à ne garder que l'homogénéité et la corrélacion de Haralick. Ces deux indices sont les plus pertinents pour les modèles, et surtout la visualisation des images d'indices cluster shade et corrélacion sur un logiciel SIG nous a révélé de très fortes hétérogénéités locales (Figure II.9.a). Cet élément n'était pas gênant pour travailler sur des moyennes au polygone sur nos peuplements de pin maritime, mais la spatialisation des résultats donne des résultats trop hétérogènes comparés à la réalité du terrain lorsque les modèles utilisent ce type de données (figure II.9..b).



(a) indice de texture « corrélacion »



(b) spatialisation des prédictions de DBH

Figure II.9 : Problème d'hétérogénéité dans les images de textures. (a) indice de texture corrélacion sur une image Sentinel-1, (b) spatialisation d'un modèle de prédiction du DBH où l'image de gauche a été utilisée.

L'extraction des textures de Haralick nécessite plusieurs paramètres à fixer : la taille de la fenêtre en pixels, l'orientation et l'offset (décalage) des comparaisons entre pixels pour calculer la matrice de covariance, et le nombre de classes à considérer pour la transformation en niveaux de gris (« nb.bins »). Nous avons laissé l'offset à 1, et les tests préliminaires (images Sentinel à 10 m et SPOT-6 à 6 m et 1.5 m de résolution spatiale) ont montré que la direction avait très peu d'influence sur le résultat même avec ce jeu de données de plantations/semis en lignes dont les orientations peuvent changer entre peuplements. Cette isotropie pourrait s'expliquer par l'espacement faible entre les rangées (2 à 4 mètres) et la taille des houppiers (6 mètres de large au maximum). Ces distances sont comparables ou inférieures à la résolution spatiale des images, ce qui fait que les patterns sont peu dépendants de l'orientation des rangées. Dans le cas d'une utilisation d'images à résolution plus fine (< 1 mètre), on pourrait être confronté à de l'anisotropie et il faudra faire attention à ce paramètre de direction pour le calcul des textures. Pour l'instant, avec nos images à 10 mètres de résolution, nous avons donc fixé ce paramètre à $1/1$ (45°) et concentré l'effort de paramétrage sur le nombre de niveaux de gris et la taille de la fenêtre. La figure II.14 montre la distribution des valeurs de pixels et des moyennes au niveau polygone pour les textures de Haralick (homogénéité et corrélation de Haralick) selon deux valeurs de paramètres, sur les 83 mesures terrain. On peut voir que dans le cas d'une fenêtre plus petite, les textures sont plus concentrées sur certaines valeurs. Si on agrandit la fenêtre, la distribution est plus large. Dans le but de construire ensuite des modèles statistiques, nous privilégions le fait d'avoir une distribution des valeurs assez large afin que l'indice de texture puisse être représentatif de la variabilité des paramètres de structure forestière. Cependant, l'élargissement de la fenêtre présente aussi le risque de ne plus être représentatif à cause des bordures des peuplements (ou parcelles forestières). Certains peuplements font 100 ou 200 m de large, on ne peut donc pas prendre une taille de 15 pixels (150 m) qui serait très vite impactée par les bordures du peuplement. Sur la figure II.10, on voit l'effet de la taille de fenêtre et du nombre de niveaux de gris sur la distribution des valeurs de l'indice d'homogénéité (au niveau pixel et polygone) et sur les relations linéaires entre cet indice au niveau polygone et les valeurs terrain des paramètres forestiers. Lorsqu'on augmente le rayon (rad) à 3, soit une fenêtre de 7×7 pixels, et avec 30 niveaux de gris (nbbins), soit $\sim 0.6 \times (7 \times 7 \text{ pixels})$, on a une distribution plus large et presque gaussienne des valeurs, et des relations significatives avec les paramètres forestiers (r^2 de 0.4 à 0.7). Si on augmente encore le rayon et le nombre de niveaux de gris, la distribution continue à tendre vers une gaussienne, mais on ne gagne pas d'informations supplémentaire en termes de relation linéaire avec les paramètres forestiers.

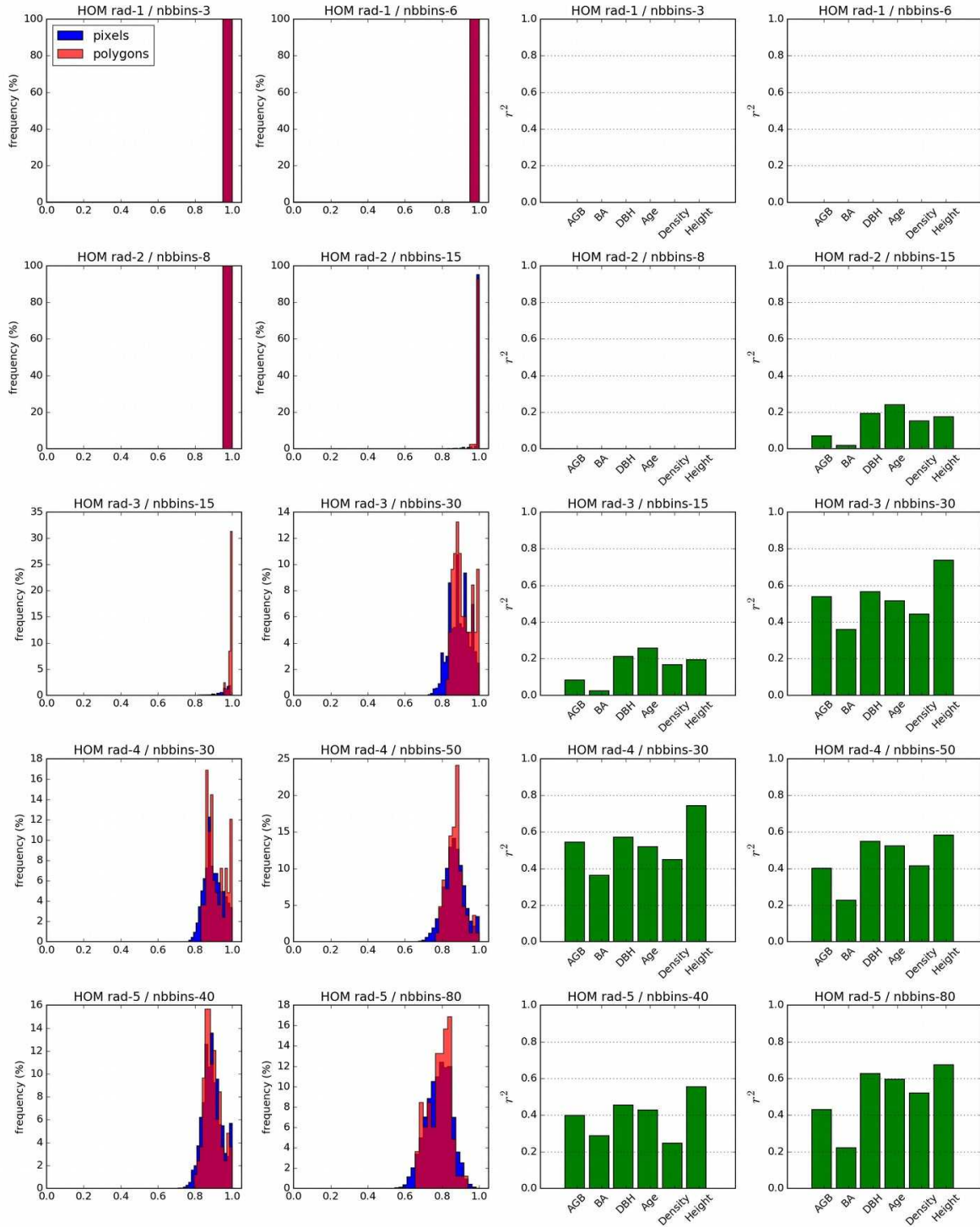


Figure II.10 : Distribution des valeurs de l'indice de texture « homogénéité » (niveau pixel et moyenné par polygone) selon les paramètres de calcul (de haut en bas : rayon de fenêtre « rad » de 1 à 5 ; de gauche à droite : nombre de niveau de gris « nbins » de ~0.3 et ~0.6 fois le nombre de pixels) et à droite, la corrélation linéaire avec les paramètres forestiers (au niveau polygone). Image d'entrée : C-VH-win (tableau II.6).

Suite à ces tests préliminaires sur différentes images Sentinel-1 et Sentinel-2 pour les combinaisons de nombre de niveaux de gris et de taille de fenêtre, nous avons fixé ces paramètres à 7×7 pixels pour la taille de fenêtre et 30 niveaux de gris pour les images Sentinel-1 et Sentinel-2 à 10 m de résolution. Pour l'ajout des indices de textures des images Spot-6 XS et PAN, nous avons considéré une taille de fenêtre d'environ 50 m (XS = 9×9 pixels ; PAN = 35×35 pixels) et toujours un ratio de nombre de niveaux de gris à ~0.6.

2.4.4. Relations entre les paramètres forestiers et les données satellitaires

Le tableau II.6. rassemble les 31 primitives retenues pour notre méthode d'estimation des paramètres forestiers.

Tableau II.6 : Primitives sélectionnées pour l'estimation des paramètres de structure forestière.

Sensors	Feature types		Features	Abbreviation
ALOS2-PALSAR2	L-band backscatter (L-band)		L-band HV	L-HV
			L-band HH	L-HH
			L-band HV/HH	L-Ratio
Sentinel-1	C-band backscatter (C-band) DES-33°IA orbit Summer/Winter averages	Summer	C-band VH	C-VH-sum
			C-band VV	C-VV-sum
			C-band VH/VV	C-Ratio-sum
		Winter	C-band VH	C-VH-win
			C-band VV	C-VV-win
			C-band VH/VV	C-Ratio-win
	C-band textural indexes (C-TI) DES-33°IA orbit Summer/Winter averages	Summer	VH-homogeneity	C-VH-sum-hom
			VH-Haralick correlation	C-VH-sum-Hcor
			VV-homogeneity	C-VV-sum-hom
			VV-Haralick correlation	C-VV-sum-Hcor
		Winter	VH-homogeneity	C-VH-win-hom
			VH-Haralick correlation	C-VH-win-Hcor
			VV-homogeneity	C-VV-win-hom
			VV-Haralick correlation	C-VV-win-Hcor
Sentinel-2	S2 spectral indexes (S2-SI)	Summer (Jully)	NDVI July	S2-NDVI-sum
			BI July	S2-BI-sum
			NDWI July	S2-NDWI-sum
		Winter (December)	NDVI December	S2-NDVI-win
			BI December	S2-BI-win
			NDWI December	S2-NDWI-win
	S2 textural indexes (S2-TI)	Summer (Jully)	BI-homogeneity	S2-BI-sum-hom
			BI-Haralick correlation	S2-BI-sum-Hcor
			NDVI-homogeneity	S2-NDVI-sum-hom
			NDVI-Haralick correlation	S2-NDVI-sum-Hcor
		Winter (December)	BI-homogeneity	S2-BI-win-hom
			BI-Haralick correlation	S2-BI-win-Hcor
			NDVI-homogeneity	S2-NDVI-win-hom
			NDVI-Haralick correlation	S2-NDVI-win-Hcor

Nous avons calculé les r^2 entre les 31 primitives satellitaires et les 6 paramètres forestiers. On trouve deux grandes tendances : les paramètres plus globaux d'un côté (AGB et BA, qui dépendent de plusieurs paramètres de structure), et d'un autre côté les paramètres spécifiques mesurés (DBH, hauteur, densité d'arbres et âge). La figure II.11 illustre ces statistiques entre paramètres forestiers et primitives. Les coefficients r^2 vont de ~ 0 au minimum à $\sim 0.7-0.75$ au maximum. Les relations sont assez similaires pour AGB et BA, cependant les primitives issues de Sentinel-1 (coefficient de rétrodiffusion et indices de texture) décrivent mieux AGB que BA, alors que les indices spectraux de Sentinel-2 décrivent mieux BA que AGB. Les plus fortes relations linéaires pour AGB et BA sont trouvées avec la bande L de ALOS2-PALSAR2 (HV et le ratio HV/HH, $r^2 \sim 0.6$). Pour le DBH (qui est relié à l'âge, la densité et la hauteur), les relations linéaires avec les primitives satellitaires sont globalement plus fortes avec Sentinel-1 et les indices de texture de Sentinel-2 (r^2 jusqu'à 0.57), mais plus faibles avec ALOS2-PALSAR2 ($r^2 < 0.25$) et les indices spectraux de Sentinel-2 ($r^2 < 0.1$). Les trois autres

paramètres (densité, âge et hauteur) ont des relations linéaires avec les primitives très proches de celles du DBH décrites ci-dessus. La hauteur a les plus fortes relations linéaires. Ces différences dans les relations entre paramètres forestiers et primitives satellitaires s'expliquent par les différents types d'informations apportés par les images satellitaires. La bande L et les indices spectraux donnent une information plus globale sur la quantité de végétation (volume, végétation active, etc.), alors que la bande C et les indices de textures de Sentinel-2 donnent une information plutôt sur la structure de la forêt (taille et forme des houppiers, espacement des arbres, etc.).

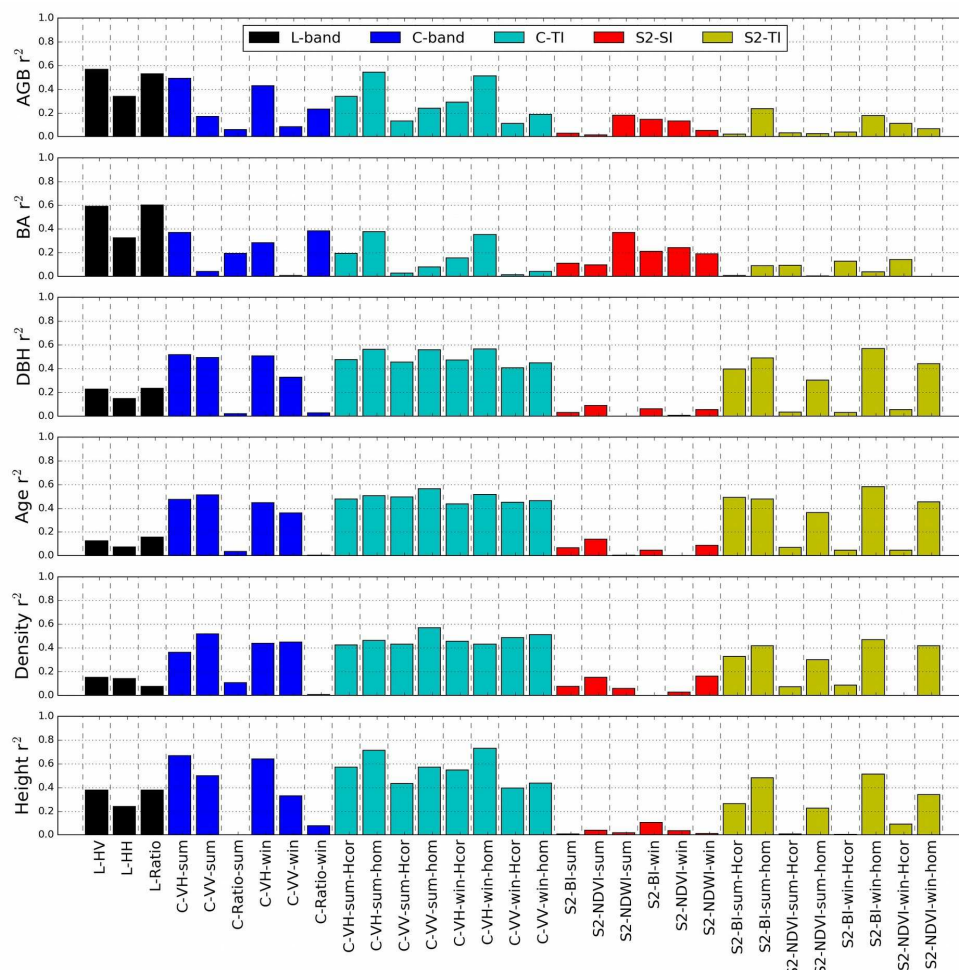


Figure II.11 : Coefficients de détermination r^2 entre les 31 primitives satellitaires et les 6 paramètres forestiers. Les valeurs des primitives ont été extraites sur des polygones autour des placettes (voir section 2.6).

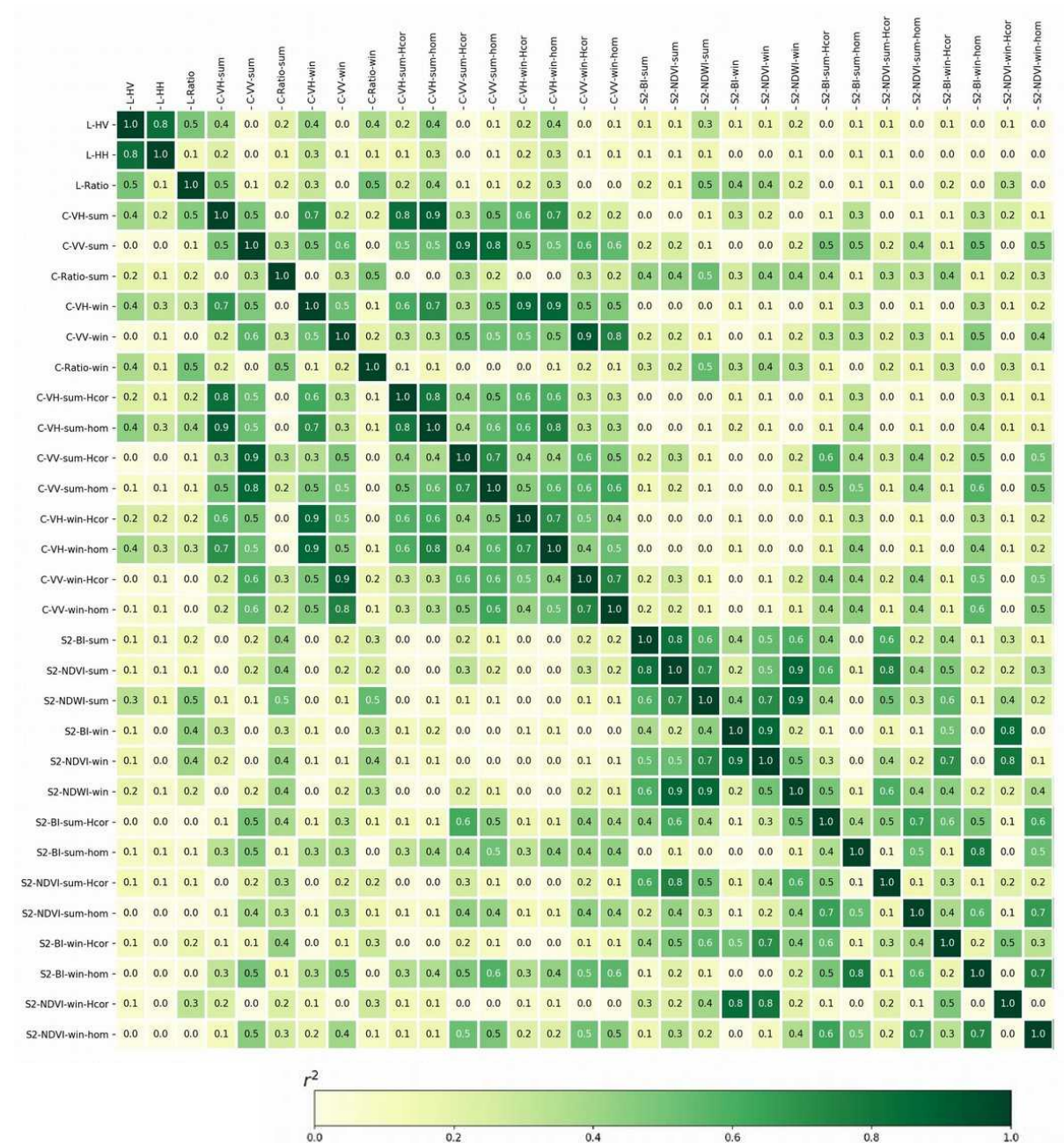
2.4.5. Corrélations entre primitives (variables explicatives)

La figure II.12 présente les coefficients de détermination r^2 entre les 31 primitives :

- **Rétrodiffusion bande L** : les bandes HV et HH sont fortement corrélées
- **Rétrodiffusion bande C** : il y a des fortes relations linéaires entre les images été et hiver dans la même polarisation (r^2 de 0.5 à 0.7), et de manière moins marquée entre polarisations VH et VV dans la même saison ($r^2=0.5$).
- **Indices de textures en bande C** : les relations linéaires entre indices de textures sont aussi marquées, de 0.3 à 0.8. Ces indices de textures sont fortement reliés aux images de rétrodiffusion dont ils sont extraits (r^2 de 0.8 à 0.9).
- **Indices spectraux** optique : forte corrélation entre index de la même saison. Forte corrélation entre le NDWI hiver et le NDVI été.

- Indices de **textures en optique** : homogénéité été et hiver très corrélées pour les deux indices BI et NDWI. Hcor corrélé avec les indices spectraux de la même saison.
- **Entre capteurs** : les relations linéaires entre primitives de capteurs différents sont moins fortes. ALOS2-PALSAR2 montre des r^2 au maximum de 0.5 avec les autres capteurs (entre le ratio HV/HH et celui de Sentinel-1 en hiver, et avec le NDWI en été). Les relations entre S1-TI et S2-indices spectraux sont faibles ($r^2 \leq 0.3$), un peu plus fortes entre S1-backscatter et S2-indices spectraux ($r^2 \leq 0.5$), et plus fortes entre les indices de texture des deux capteurs ($r^2 \leq 0.6$).

Ainsi, il y a apparemment des redondances dans l'information contenue dans l'ensemble de ces primitives, même si cette analyse ne vaut que pour le cas de ce site d'étude sur des peuplements de pin maritime dans les Landes.



2.5. Approche méthodologique

Dans la littérature, trois grandes approches ont été utilisées pour estimer les paramètres de structure forestière (Santoro et al., 2015) : 1) les modèles empiriques avec régression paramétrique, 2) les modèles physiques et semi-empiriques, et 3) les modèles non paramétriques.

Les deux premières approches reposent sur des concepts physiques et des relations pré-établies entre des variables explicatives (données de télédétection, données environnementales et/ou climatiques) et les paramètres forestiers.

Les modèles non paramétriques reposent sur des concepts moins concrets, ils utilisent les relations statistiques reliant les paramètres de structure forestière aux variables explicatives, dans notre cas les données satellitaires optiques et radar. L'apprentissage statistique supervisé, ou apprentissage automatique (« machine learning », intelligence artificielle, etc.), permet d'apprendre automatiquement des règles sur des échantillons connus puis de prédire les paramètres forestiers sur de nouveaux échantillons inconnus. De cette manière, à partir d'un échantillonnage sur une zone, on peut cartographier de manière continue l'ensemble de la zone. Les avantages de ce type de méthode sont la capacité à gérer différents types de données en entrée sans a priori sur leur relation avec les paramètres forestiers, ainsi que d'être adaptable et reproductible sur des différents types de forêts (au niveau de la méthode, mais pas forcément des modèles).

Notre objectif est de construire une chaîne de traitement automatique pour la cartographie des paramètres de structure forestière. Elle doit être reproductible, c'est-à-dire qu'elle doit être applicable à n'importe quelle forêt sans changement majeur dans la méthode. Les données prédictives (images satellitaires) sont disponibles partout, en libre accès, et les bibliothèques utilisées sont open source. La chaîne doit également être adaptable pour, d'une part, se spécialiser sur le jeu de données qui lui est fourni, et d'autre part, l'utilisateur doit pouvoir modifier des éléments à la marge (algorithme d'apprentissage, données prédictives) s'il pense que ces modifications peuvent améliorer les résultats sur ses données. La figure II.13 représente l'organisation globale de notre chaîne de traitement ainsi que les différents tests et analyses méthodologiques faites sur les étapes de la chaîne.

Approche méthodologique

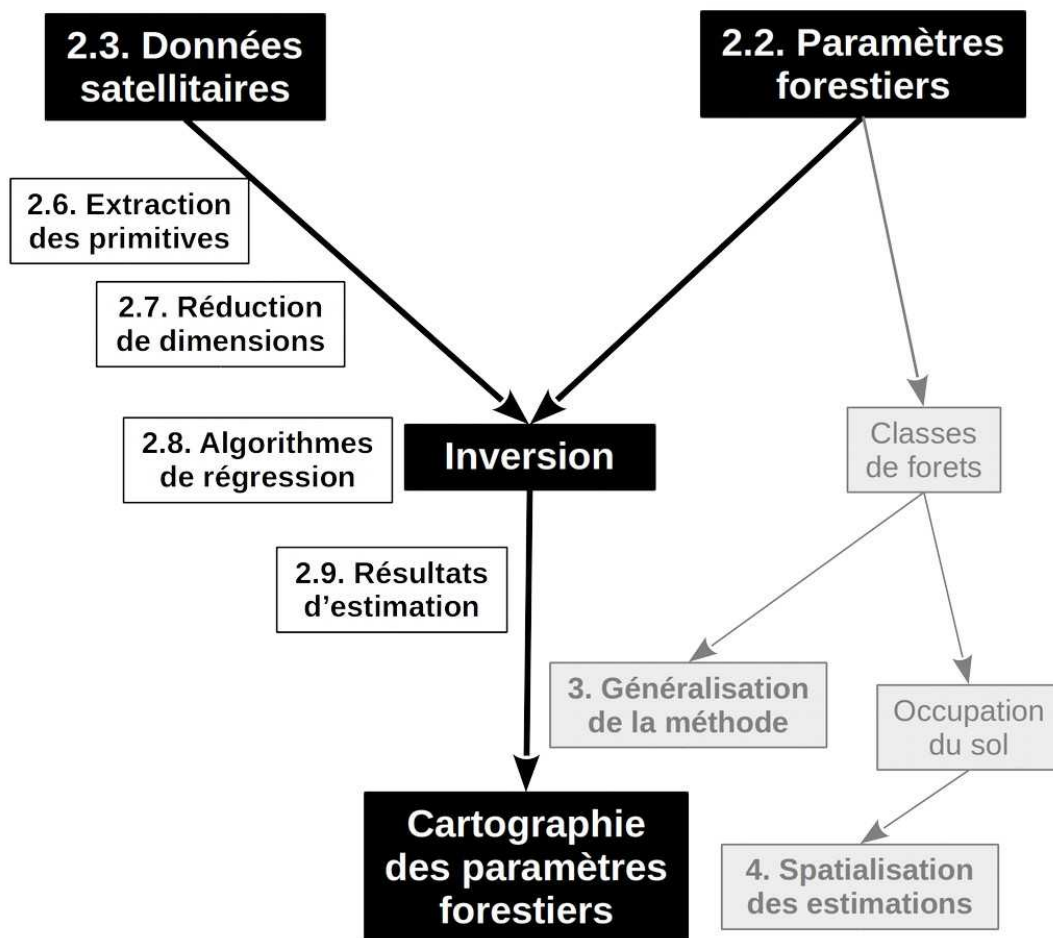


Figure II.13 : Schéma résumé de l'approche méthodologique.

Ainsi, plusieurs grandes étapes sont nécessaires pour mettre au point une méthodologie :

- Sélection des primitives : il est nécessaire de faire un pré-tri des données satellitaires et d'analyser les capacités de chacune à fournir des informations sur les paramètres forestiers afin de pouvoir interpréter les résultats avant de les reproduire.
- Algorithmes de régression et réduction des dimensions : différents algorithmes sont disponibles avec leurs propres avantages et inconvénients (capacité à gérer un grand nombre de primitives ou un petit nombre références, paramétrage, vitesse de calcul, etc.). Nous avons testé la régression linéaire multiple (MLR, multi-linear regression), le Random Forest (RF) et le Support Vector Regression (SVR).
- Validation et analyse des résultats d'estimation : compréhension des limites de la méthode, importance des différents types de données satellitaires, transférabilité des modèles, etc.

Les points spécifiques à la généralisation de la méthode et à la spatialisation des estimations seront abordés dans les Chapitres III et IV du manuscrit.

2.6. Extraction des primitives

Les mesures terrain ont été faites sur des placettes d'une surface de 0.05 à 0.2 ha, ce qui équivaut pour les plus petites à 4 pixels à la résolution spatiale de 10 m. Une surface trop petite peut poser des problèmes pour les indices de texture spatiale calculés sur des fenêtres de 7×7 pixels, et pour les images radar qui sont très sensibles aux effets de bords (Chemins, route au bord de la parcelle) et aux hétérogénéités locales. La figure II.14 illustre les hétérogénéités observées sur les images radar, y compris sur des couverts forestiers très homogènes gérées de la même manière à la parcelle (donc avec sous-bois probablement très homogène à l'intérieur de chaque parcelle). Dans un premier temps, nous préférons travailler sur des surfaces plus larges qui nous permettent de travailler avec des valeurs moyennes moins bruitées et plus stables.

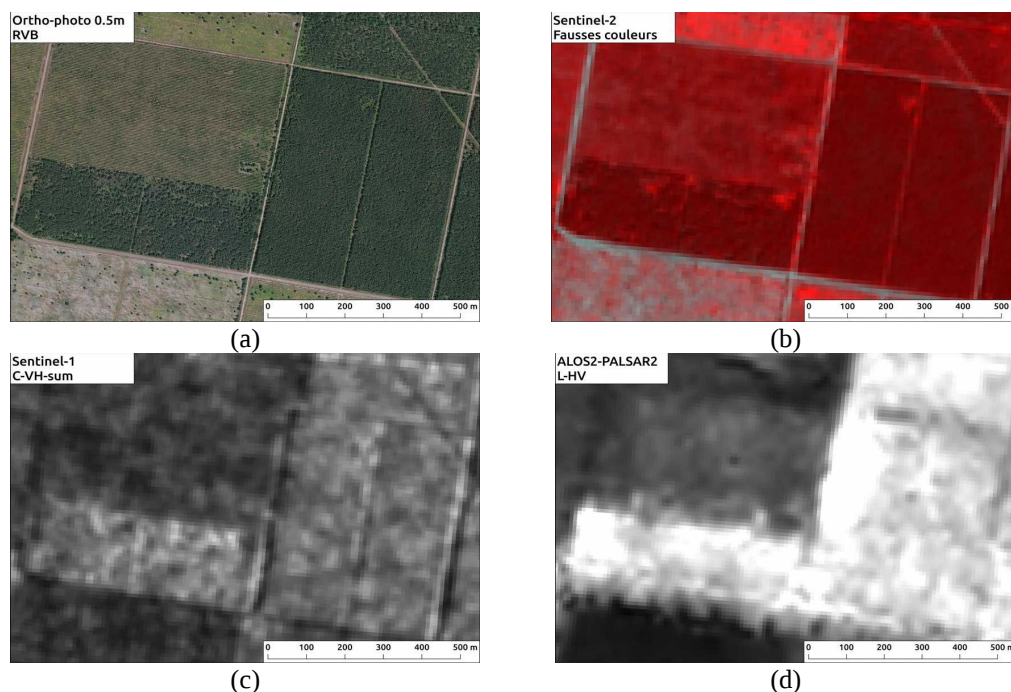


Figure II.14 : Parcelles de pin maritime sur la zone d'étude. (a) Photographie aérienne de l'IGN, 2015 ; (b) Sentinel-2 fausses couleurs, été 2016 ; (c) Sentinel-1 VH, été 2016 ; (d) ALOS2-PALSAR2 HV, 2016.

Nous travaillons sur des plantations de pin maritime homogènes en termes de gestion, avec seulement quelques hétérogénéités locales. Ainsi, nous avons digitalisé des zones plus larges autour des placettes de mesure, avec un risque réduit de couvrir des zones avec des paramètres forestiers différents de ceux mesurés. La taille moyenne des nouveaux polygones est de 2.7 hectares. La figure II.15 montre ces nouvelles unités d'extraction (Polygones), comparés aux unités de gestion (Peuplement) et aux surfaces réellement mesurées (Placettes).

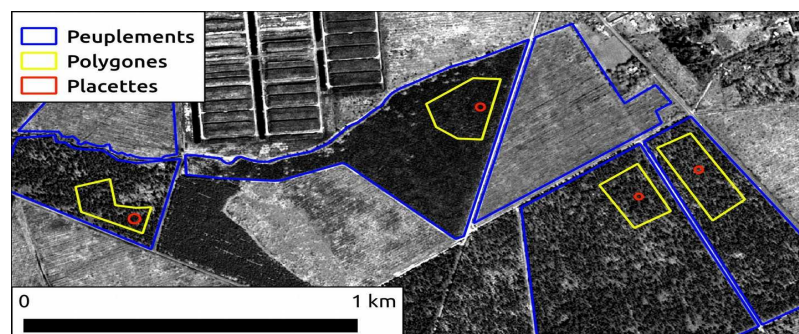


Figure II.15 : Différentes surfaces d'intérêt au sein des peuplements de pin maritime.

Les relations linéaires entre les primitives satellitaires et les paramètres forestiers sont plus fortes lorsqu'on extrait les valeurs au niveau polygones qu'au niveau des surfaces réelles des placettes mesurées. Cela confirme le bénéfice de travailler sur des zones plus larges lorsque le type de forêt le permet. La figure II.16 montre les différences de r^2 observées pour l'AGB et la hauteur. L'impact est particulièrement visible pour les données radar en bande C (10 m de résolution spatiale), pour lesquelles le moyennage au niveau polygone permet de mieux réduire le bruit de Speckle.

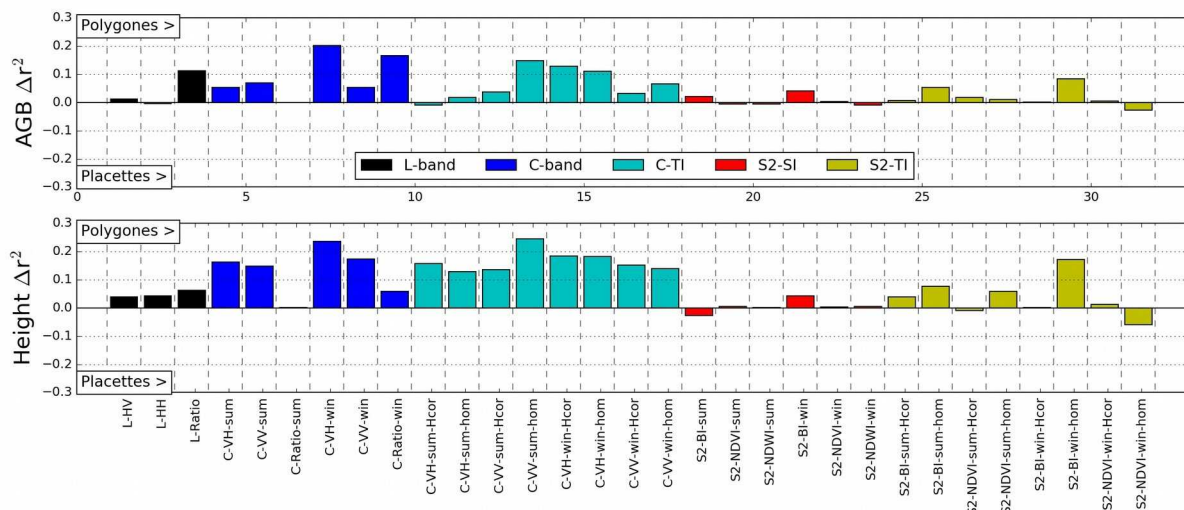


Figure II.16 : Différences entre l'extraction des primitives au niveau polygone et au niveau placette pour les relations linéaires avec l'AGB et la hauteur. Les barres vers le haut indiquent un r^2 plus fort au niveau polygone, les barres vers le bas indiquent un r^2 plus fort au niveau placette.

L'impact du niveau d'extraction est visible sur les résultats d'estimation. La figure II.17 montre les statistiques de validation (RMSE relatif) obtenues pour les six paramètres forestiers avec un algorithme de régression SVR et une méthode de sélection des primitives « stepwise » (en mode « forward »).

L'utilisation des polygones, plutôt que les placettes, améliore légèrement les résultats sur les paramètres DBH, age, densité et hauteur. En effet, ces paramètres sont plus fortement reliés linéairement avec les primitives issues de Sentinel-1, elles-mêmes plus affectées par le bruit. L'élargissement de la zone d'extraction des valeurs réduit le bruit et améliore au final les résultats d'estimation de ces paramètres. On note cependant que cette amélioration est inférieure à 5 % de RMSE relatif, les résultats obtenus en utilisant les surfaces réelles des placettes restent acceptables. Ainsi, cette méthode d'estimation peut quand même donner de bons résultats même lorsqu'il n'est pas possible de retravailler les données de référence en élargissant les zones d'extraction des valeurs. Ce cas de figure peut intervenir lorsque les forêts sont plus hétérogènes ou si le temps et les compétences manquent pour photo-interpréter les zones de mesure.

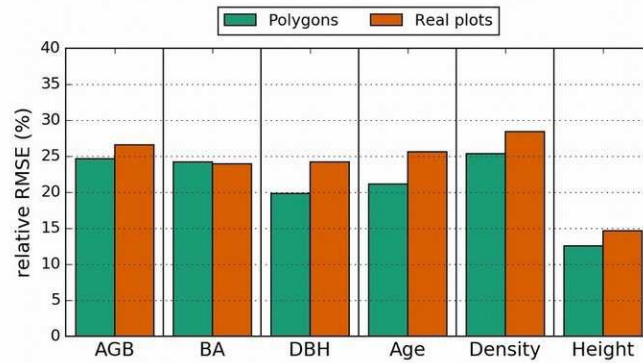


Figure II.17 : RMSEs relatifs des estimations des 6 paramètres forestiers avec un SVR et une sélection Forward, selon l'extraction des valeurs de primitives sur la placette mesurée (« Real plots ») ou sur une zone élargie (« Polygons »).

2.7. Réduction de dimensions

La littérature est parfois contradictoire en matière de gestion de la dimension pour les problèmes de classification et de régression en apprentissage automatique. En effet, l'apprentissage automatique est très utilisé dans le domaine du « Big Data », et les études laissent à penser que plus on a de variables explicatives, meilleur sera le résultat, étant donné que les algorithmes font automatiquement le « tri » des variables explicatives. Cependant, dans certains cas, des variables supplémentaires peuvent apporter un bruit qui va gêner la mise en place de règles statistiques reliant les variables explicatives et la variable dépendante. Ce phénomène apparaîtra d'autant plus dans le cas où l'échantillonnage des données de référence est réduit en nombre, comme pour cette étude.

Dans notre méthode, nous combinons plusieurs types de primitives satellitaires avec une approche par apprentissage automatique pour mieux estimer les paramètres forestiers. On a vu que plusieurs primitives ont de fortes relations linéaires avec les paramètres forestiers, mais aussi que les 31 primitives satellitaires peuvent avoir de la redondance entre elles. Cette redondance, en plus du nombre de primitives comparé au nombre d'échantillons de références, peut nuire à la performance des algorithmes de régression. Nous avons vérifié cet effet et testé deux approches pour réduire le nombre de primitives : appliquer une Analyse en Composantes Principales (ACP) ou sélectionner les meilleures primitives avec des méthodes dites « stepwise » (pas à pas).

2.7.1. Analyses en Composantes Principales

Nous avons analysé la variance du jeu de données de primitives satellitaires avec une ACP. La figure II.18 montre les représentations 2D des variances expliquées par les 31 primitives dans les composantes 1, 2 et 3, ainsi que la variance expliquée des 10 premières composantes (jusqu'à 95 % de la variance expliquée en cumulé).

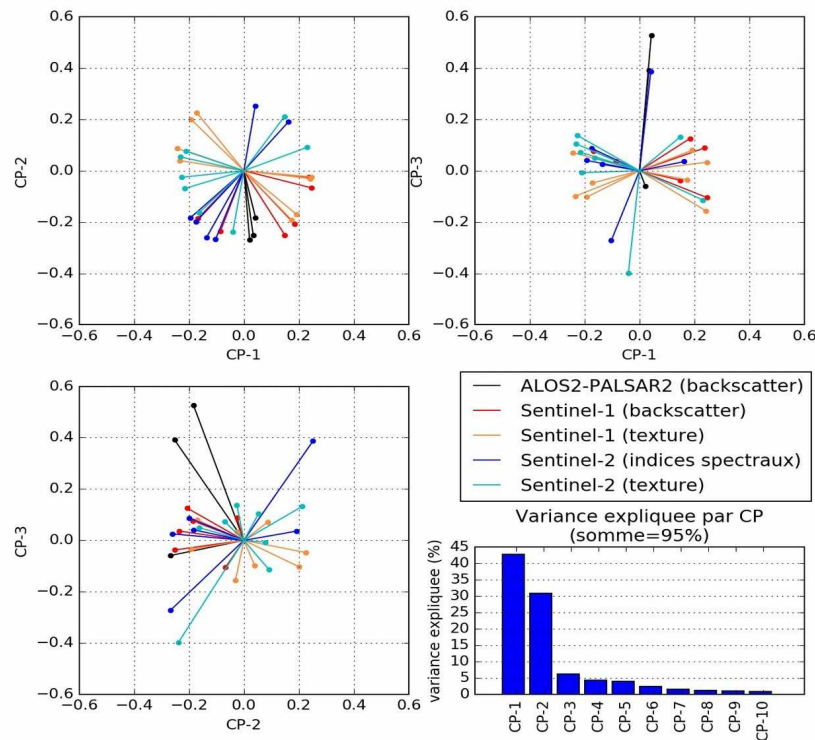


Figure II.18 : Analyse en composantes principales sur l'ensemble des 31 primitives.

L'analyse par ACP n'est pas très concluante. Toutes les primitives apparaissent comme génératrices de variance, avec peut-être une légère différence pour les données de rétrodiffusion en bande L et les index spectraux d'un côté, et les indices de texture de l'autre côté.

Pour tester l'efficacité de cette approche par ACP, nous avons également appliqué l'ACP capteur par capteur (donc 3 ACP) en gardant à chaque fois les composantes principales qui représentent 95 % de la variance. On a au final 13 composantes, contre 10 lorsqu'on applique l'ACP à l'ensemble des 31 primitives. Les résultats de régression avec un SVR montrent des statistiques de validation assez similaires (Figure II.19). Les estimations sont légèrement meilleures avec ACP-All pour BA et AGB (2 à 3 % de rRMSE), mais ACP-Sensor obtient des résultats un peu meilleurs pour le DBH, la densité et la hauteur (1-2 %), et particulièrement pour l'âge (7 %). Nous avons choisi pour les tests suivants de considérer l'ACP faite séparément sur les 3 capteurs (PCA sensors avec 13 composantes).

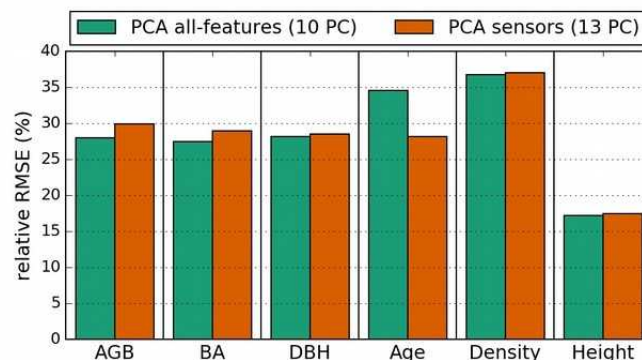


Figure II.19 : RMSEs relatifs des estimations des 6 paramètres forestiers avec un SVR et une sélection Forward, avec une ACP à partir de l'ensemble des 31 primitives (PCA all-features) ou par capteur (PCA sensors).

2.7.2. Méthodes de sélection de primitives « stepwise »

Les méthodes dites « stepwise » consistent à avancer par itérations pour atteindre un objectif. Dans le cas de la sélection des primitives pour le modèle, à chaque étape on construit notre modèle dans plusieurs configurations de primitives et on sélectionne le meilleur résultat selon un critère statistique (RMSE), puis on passe à l'itération suivante jusqu'à déterminer pas à pas la configuration qui donne le meilleur résultat. Deux approches ont été testées, on les appellera sélections « Forward » et « Backward ».

- Forward : on commence par tester l'apprentissage et la validation du modèle avec chaque primitive une à une afin de sélectionner celle qui donne le meilleur résultat. Ensuite, on garde cette primitive et on ajoute chacune des autres une à une pour sélectionner celle qui donne le meilleur résultat en combinaison avec cette première primitive, et on répète l'opération jusqu'à ce que toutes les primitives aient été ajoutées. Enfin, on sélectionne la combinaison qui donne le meilleur résultat au total.
- Backward : on commence par tester l'apprentissage et la validation du modèle avec l'ensemble des primitives toutes ensemble, puis on enlève une à une chacune des primitives pour sélectionner la primitive qui, lorsqu'elle est enlevée du modèle, donne le meilleur résultat. On enlève cette primitive et on recommence l'opération autant de fois qu'il y a de primitives jusqu'à ce qu'il ne reste qu'une primitive pour le modèle. Enfin, comme pour la sélection Forward on sélectionne la combinaison qui donne le meilleur résultat sur l'ensemble des tests.

La figure II.20 montre un exemple du profil de sélection Forward avec un algorithme SVR pour l'estimation du DBH. Le processus ajoute 9 primitives (indiquées en abscisse) jusqu'à obtenir le meilleur RMSE relatif sur la série. Ainsi, sur cet exemple on a pu réduire le nombre de dimensions de 31 à 9 de manière automatique et améliorer le RMSE relatif de 7 % par rapport à l'utilisation directe des 31 primitives. Les profils de sélection Backward sont assez similaires, on trouve le minimum de RMSE autour de 10-15 primitives. Dans l'approche Forward, on privilégie d'abord les primitives qui ont des fortes relations avec les paramètres forestiers, puis celles qui ont des synergies avec ces premières primitives. Dans le cas de l'approche Backward, on supprime les primitives qui ne sont pas utiles au modèle pour estimer les paramètres forestiers, ou qui apportent du bruit qui dégrade les résultats. En théorie, la sélection Backward devrait donner des résultats meilleurs et plus stables, car on est sûr de garder les synergies multiples entre les primitives. La sélection Forward ne teste à chaque fois que l'apport d'une seule primitive. Cependant, en pratique sur notre étude, les résultats sont très proches, et la sélection Forward est plus facile à utiliser dans notre chaîne de traitement : le rapport entre le nombre de primitives et le nombre d'échantillons de référence n'affecte pas la sélection Forward, qui commence le procédé en ajoutant des primitives. Par contre, le fonctionnement de la sélection Backward peut être perturbé par un trop grand nombre de primitives comparé au nombre d'échantillons, qui fausserait les premières itérations (problème de dimension et de sur-apprentissage), avec le risque, dans les premières itérations, de supprimer des primitives qui seraient en réalité utiles au modèle.

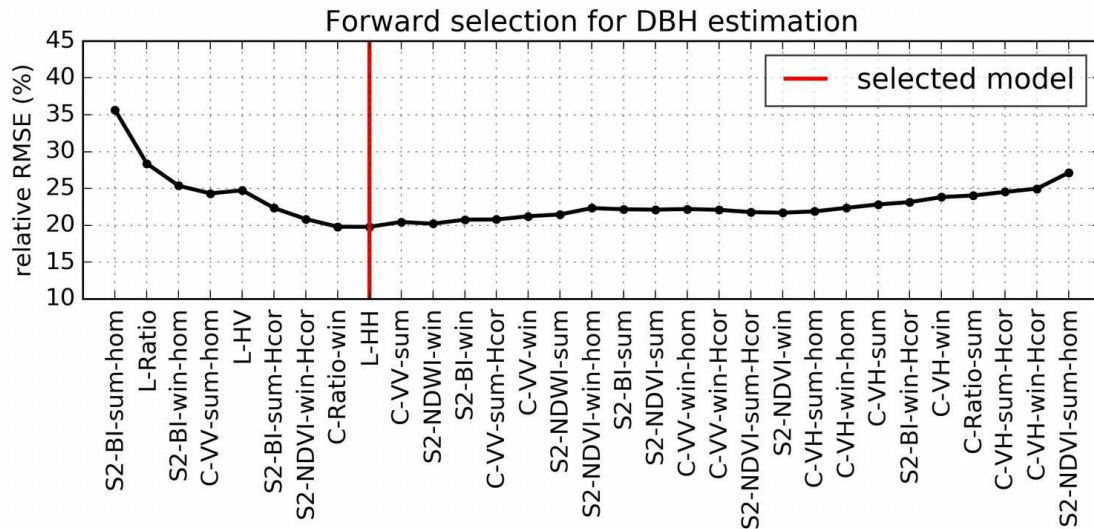
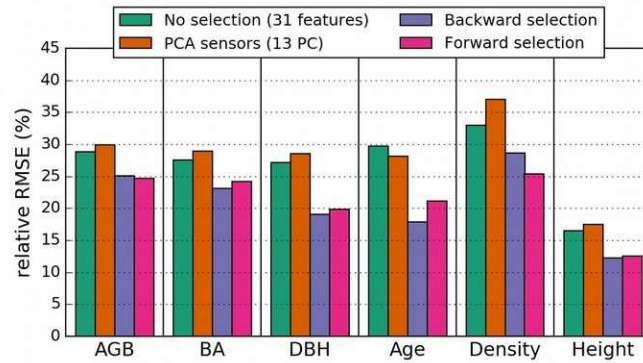


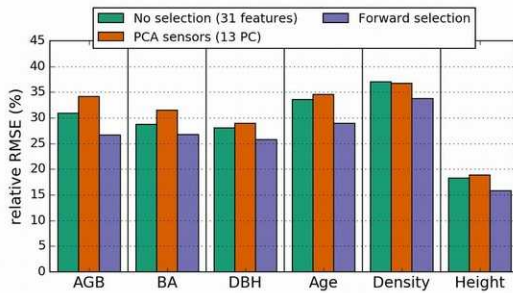
Figure II.20 : Profil de sélection Forward des primitives pour l'estimation du DBH moyen avec un SVR. Les primitives indiquées en abscisse sont celles qui sont ajoutées à chaque itération.

2.7.3. Comparaison entre l'ACP et les méthodes de sélection de primitives

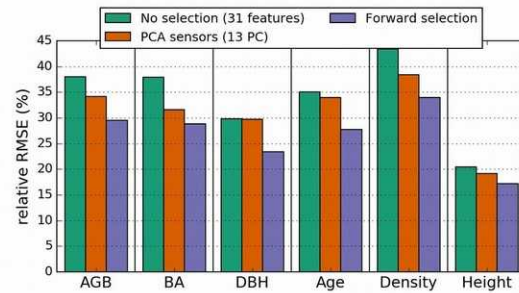
Nous avons fait des comparaisons de statistiques pour savoir si la réduction de dimension pouvait améliorer les résultats, et quelle méthode utiliser entre l'ACP et les méthodes stepwise pour la sélection des primitives les plus utiles aux modèles. Nous avons donc comparé : 31 primitives sans sélection, avec ACP (13 composantes), avec sélection Forward et Backward (9 à 16 primitives sélectionnées). La figure II.21.a présente la comparaison des RMSE relatifs, obtenus avec ces 4 méthodes, pour l'estimation des 6 paramètres forestier avec un SVR. Les figures II.21.b et c présentent cette même comparaison avec les algorithmes RF et MLR, sans la sélection Backward qui a un fonctionnement et des résultats similaires à la sélection Forward sur ce jeu de données. Avec le SVR, on observe clairement une meilleure performance des méthodes stepwise (Forward et Backward) alors que l'utilisation d'une ACP n'améliore pas les résultats (voir, les dégrade) comparée à l'utilisation directe des 31 primitives. L'amélioration, en termes de baisse du RMSE relatif, est de 4 à 10 % selon les paramètres forestiers. Ces résultats confirment d'une part le fait que le modèle est plus performant lorsqu'on lui fournit uniquement les primitives utiles dans ce cas-là, et d'autre part que la réduction de dimension avec une ACP ne permet pas d'obtenir un meilleur résultat, et dans certains cas, peut dégrader l'information contenue dans les primitives. Cette sous-performance de l'ACP pourrait être liée au fait que cette méthode utilise des plans linéaires, alors que les relations entre primitives et paramètres forestiers ne sont pas toutes linéaires, et les algorithmes Random Forest et SVR sont capables de décrire ces relations linéaires donc de l'information est « perdue » lorsqu'on utilise la transformation ACP.



(a) SVR



(b) RF



(c) MLR

Figure II.21 : Comparaison des performances pour l'estimation des 6 paramètres forestiers selon la méthode de réduction de dimension et l'algorithme qui sont utilisés.

Dans le cas de l'utilisation d'un RF ou d'un MLR pour la régression, les conclusions sont les mêmes quant aux meilleures performances des sélections stepwise. Cependant, dans le cas de la MLR, l'ACP améliore les résultats par rapport à l'utilisation directe des 31 primitives. Dans le cas du RF, l'ACP dégrade les résultats par rapport aux 31 primitives. Aussi, dans le cas où on utilise directement les 31 primitives, le SVR et le RF ont des résultats similaires tandis que le MLR est moins performant. En effet, le MLR gère moins bien l'information « inutile » que les deux autres algorithmes. Le SVR profite le plus de la sélection des primitives les plus utiles au modèle.

2.8. Algorithmes de régression

Les approches par apprentissage automatiques sont abondamment utilisées de nos jours, avec des familles d'algorithmes qui permettent de répondre à différentes problématiques. Plusieurs aspects sont à prendre en compte lorsqu'on utilise l'apprentissage automatique : le nombre d'échantillons et le nombre de variables explicatives, la compréhension du modèle et sa paramétrisation, et la capacité de généralisation des modèles.

Nous avons sélectionné trois algorithmes à tester dans notre étude, chacun étant documenté et utilisé efficacement dans la littérature (Belgiu and Drăguț, 2016; Mountrakis et al., 2011; Pal and Mather, 2005), et implémenté dans plusieurs bibliothèques libres d'algorithmes. Chaque algorithme a ses propres avantages et limites, leurs performances dépendent des jeux de données étudiés :

- La **régression linéaire multiple (MLR, multi-linear regression)** est très souvent utilisée dans la littérature pour estimer des paramètres à partir d'images satellitaires. C'est une approche paramétrique, qui suppose qu'une combinaison de relations linéaire relie les variables explicatives à la variable expliquée, mais qui est régulièrement utilisée de la même manière que les algorithmes non paramétriques. Cette méthode a l'avantage d'être très simple à appliquer, sans paramètres à fixer,

mais le pré-requis d'une relation linéaire peut être un facteur limitant pour la précision des résultats.

- L'algorithme **Random Forest (RF)** est de plus en plus utilisé en régression. Son fonctionnement est simple, il repose sur la construction d'arbres de décision binaire qui sont ensuite combinés pour obtenir une règle de décision. Chaque arbre est construit sur un échantillonnage aléatoire des variables explicatives. Plusieurs paramètres de l'algorithme peuvent être fixés, comme le nombre d'arbre, le nombre de variables explicatives pour chaque arbre, ou la profondeur des arbres. Parmi les avantages du RF, on peut citer sa simplicité et le fait de pouvoir extraire, en sortie du modèle, les variables explicatives les plus utilisées. Il est aussi cité pour son efficacité à gérer des grands nombres d'échantillons grâce à la parallélisation de la construction des arbres. Sa principale limite est le fait de nécessiter un grand jeu de données pour pouvoir construire des arbres indépendants et éviter une trop grande spécialisation du modèle sur l'échantillon fourni (sur-apprentissage).
- L'algorithme **Support Vector Regression (SVR)** est un peu moins utilisé pour les problèmes de régression. On trouve en revanche beaucoup de littérature utilisant le Support Vector Machine (SVM), son équivalent pour les problèmes de classification binaire ou multi-classes. Cet algorithme nécessite aussi une paramétrisation. Son fonctionnement est plus compliqué que le MLR ou le RF, et dans le cas d'un noyau gaussien on ne peut pas directement extraire la contribution de chaque variable explicative. Son principal avantage est sa capacité à être utilisé efficacement sur des petits jeux de données comme ceux que l'on a souvent dans le cas de l'estimation des paramètres de structure forestière.

Le tableau II.7 présente le paramétrage des algorithmes. La description des paramètres peut être trouvée sur la page de description des algorithmes implémentés dans scikit-learn (SVR⁷, RF⁸) de Python (Pedregosa et al., 2011). La description et l'analyse de sensibilité des paramètres du SVR est présentée plus loin dans cette section.

Tableau II.7 : Paramètres des algorithmes de régression.

	MLR	RF	SVR
Paramètres fixés	Pas de paramètres	Nbr max primitives par branche = sqrt(N) Les autres paramètres ont des valeurs par défaut indiquées dans la description de la fonction	Kernel : RBF
Paramètres optimisés		Nombre d'arbres : [50,100] Profondeur des arbres : [10,NA]	Paramètre de coût : $2^{[0,1,2,3,4,5,6,7,8]}$ Gamma : [0.001,0.005,0.01,0.05,0.1,0.5,1,5] Epsilon : [0.001,0.005,0.01,0.05,0.1,0.5,1]

2.8.1 Comparaison des algorithmes

La Figure II.21 présentait l'effet de la réduction de dimension selon les trois algorithmes de régression. Ici la figure II.22 rassemble les résultats des trois algorithmes MLR, RF et SVR en sélection Forward, qui donne les meilleurs résultats en termes de sélection des primitives. On constate une amélioration claire du RMSE relatif avec le SVR pour l'âge et la densité d'arbres (7 à 10 % de RMSE relatif), et dans une moindre mesure pour les quatre autres paramètres AGB, BA, DBH et hauteur (4 à 5 %). Le SVR est d'une part plus performant que le MLR pour travailler sur des données reliées non linéairement, et d'autre part il est plus performant

⁷ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

⁸ <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>

que le RF pour travailler sur des petits jeux de données. Pour la suite de l'étude, nous utiliserons le SVR qui donne de meilleurs résultats sur ce jeu de données, mais nous recommandons de tester le RF lorsqu'un plus large échantillonnage est disponible. Le MLR est beaucoup plus simple et rapide à mettre en place mais n'est cependant pas capable de retrouver aussi bien les relations entre primitives satellitaire et paramètres forestier que les algorithmes de régression non linéaires.

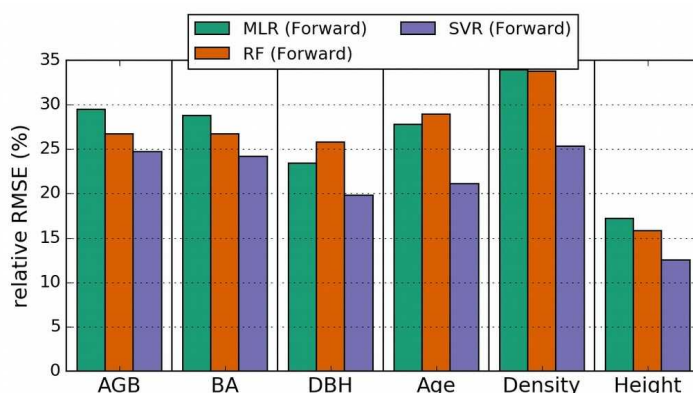


Figure II.22 : Comparaison des performances des algorithmes de régression MLR, RF et SVR avec une sélection Forward sur les 6 paramètres forestiers. Les algorithmes RF et SVR ont fait l'objet d'une optimisation des paramètres.

Nous avons également évalué le temps de calcul nécessaire à chacun des algorithmes pour réaliser une sélection Forward et fournir le meilleur modèle. Avec 31 primitives, il y a donc 31 étapes contenant chacune de 1 à 31 tests de combinaisons, ce qui fait un total de 496 processus. Le MLR n'a pas d'optimisation des paramètres et est calculé en ~30 secondes. Le RF a 4 combinaisons de paramètres à tester pour chaque processus, il est calculé en ~9h40. Enfin, le SVR a 504 combinaisons de paramètres à tester et est calculé en ~1h50. Ces temps calculés sont cependant spécifiques à la configuration de l'étude et de la machine (1 CPU, 2.9 GHz). Nous avons par ailleurs constaté que lorsqu'on a un échantillonnage plus important, supérieur à 1000-2000 échantillons, le RF peut être plus rapide que le SVR. Cette inversion de tendance s'explique par la capacité du RF à paralléliser la construction des arbres, le rendant plus efficace pour traiter des gros jeux de données bien qu'il soit plus lent sur notre « petit » échantillonnage de 83 points de mesure.

2.8.2. Paramétrisation du SVR

Le SVR a trois paramètres clés à optimiser. Une mauvaise gestion de ces paramètres peut dégrader les résultats ou provoquer un sur-apprentissage (« over-fitting ») qui empêcherait une bonne généralisation de la méthode. Le paramètre de coût (C) détermine à quel point le modèle va prendre en compte les échantillons mal estimés. Avec une haute valeur de C le modèle se complexifie pour s'adapter à tous les échantillons (pouvant provoquer du sur-apprentissage), tandis qu'un C faible va donner peu d'importance aux échantillons mal estimés au profit d'un modèle simple. Le coefficient Gamma (G) détermine à quel point un échantillon seul peut influencer le modèle. Comme avec C, une forte valeur de G peut provoquer du sur-apprentissage. Enfin, le paramètre Epsilon (E) définit le « tube » dans lequel les échantillons sont considérés comme étant bien estimés, il fonctionne comme un seuil.

La figure II.23 montre l'évolution des résultats des estimations (RMSE relatif en %) en faisant varier les paramètres C, G et E chacun à leur tour (la valeur des deux autres étant fixés par l'optimisation). Les statistiques sont issues d'une division des échantillons, avec 2/3 pour l'apprentissage (*Train*) et 1/3 pour la validation (*Test*), répétée 100 fois. L'écart type des

résultats est présenté autour des valeurs moyennes. Les graphiques montrent la courbe *Test* qui représente la validation indépendante, et la courbe *Train* qui représente la validation interne du modèle (validation sur les mêmes données qui ont servi à l'apprentissage). On constate tout d'abord que E donne les meilleurs résultats avec une valeur proche du minimum, mais c'est aussi là que les statistiques *Train* et *Test* sont les plus éloignées, ce qui indique un sur-apprentissage. Le RMSE relatif peut varier de 10 à 15 % selon la valeur de E. La variation du paramètre C engendre moins de variation du rRMSE (<10 %), le minimum de rRMSE des échantillons Test est obtenu en général avec C entre 10 et 50, le rRMSE tend ensuite à augmenter. Plus on augmente C, plus l'application sur les échantillons d'apprentissage (*Train*) donne un faible rRMSE. Les valeurs de C optimisées donnent donc des résultats plus encourageants pour la généralisation si le C est petit. Enfin, le paramètre G donne des meilleurs résultats quand il est proche de son minimum, entre 0.001 et 0.1. La variation de rRMSE provoquée par la variation de G est plus importante que pour C et E, jusqu'à 30 %. Comme pour C, si on augmente la valeur du paramètre G, les résultats *Test* sont moins bons, tandis que les résultats *Train* tendent vers le minimum d'erreur. Les valeurs de G optimisées ont des résultats très proches entre *Train* et *Test*, ce qui est intéressant pour la robustesse du modèle.

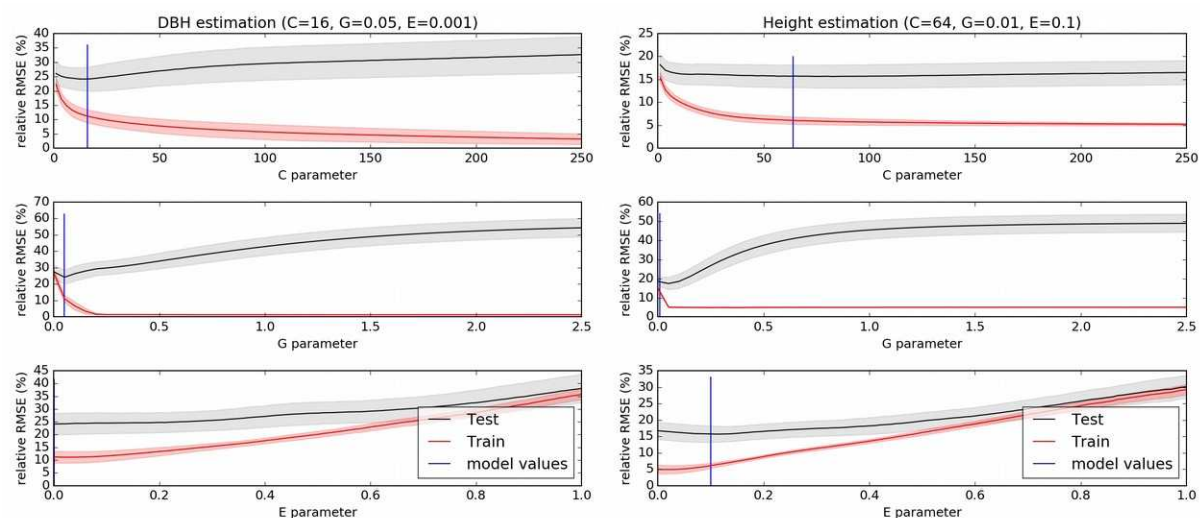


Figure II.23 : Évolution du RMSE relatif obtenu pour l'estimation du DBH et de la hauteur en faisant varier les paramètres C, G et E du SVR. Résultats obtenus avec 2/3 des échantillons pour l'apprentissage (*Train*) et 1/3 pour la validation (*Test*), sur 100 itérations. Chaque paramètre est analysé en fixant les deux autres paramètres (valeurs de l'optimisation).

La figure II.24 montre la sensibilité du SVR à la variation conjointe des paramètres C et G (le paramètre E étant fixé à 0.001) pour l'estimation du DBH. En observant l'évolution du RMSE relatif on voit que, quelle que soit la valeur de C, la variation de G a un effet fort sur la qualité des estimations. En revanche, le paramètre C a une importance pour $0.001 < G < 0.1$, mais au-delà la qualité dépend surtout de G.

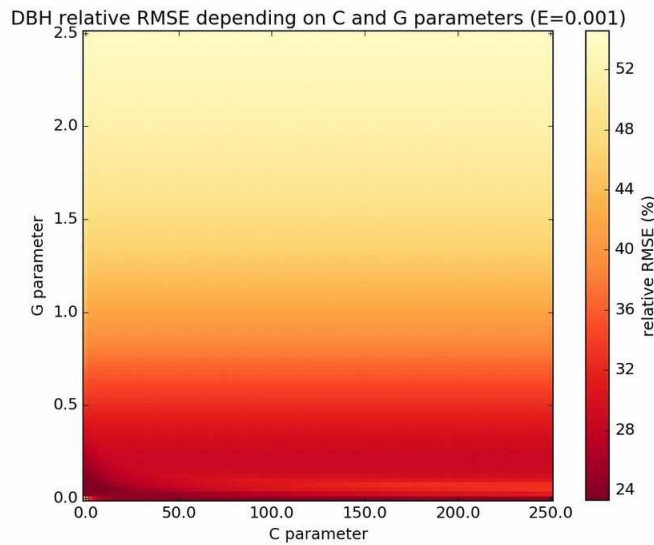


Figure II.24 : Effet de l'évolution conjointe des paramètres C et G du SVR sur le RMSE relatif pour l'estimation du DBH. Le paramètre E est fixé à 0.001.

En conclusion, nous avons gardé les valeurs indiquées dans le tableau II.7 pour l'optimisation des paramètres du SVR, même si l'analyse de sensibilité nous montre que tous les paramètres n'ont pas la même importance avec ce jeu de données. Le processus d'optimisation se fait comme suit :

- 20 divisions aléatoires des échantillons en apprentissage (2/3) et validation (1/3) avec vérification de la similarité des moyennes et variances (tests de Levene et Mann-Whitney). Les mêmes divisions sont réutilisées en cas de comparaison entre différentes méthodologies pour un même paramètre forestier.
- Pour chaque combinaison de paramètre (504 au total, tableau II.7) nous faisons l'apprentissage et l'application du modèle sur les 20 divisions en stockant les RMSE relatifs obtenus. La moyenne des RMSE relatifs des 20 divisions est utilisée pour évaluer les résultats de la combinaison de paramètres.
- Nous gardons au final la combinaison de paramètres C, G et E qui ont obtenu la meilleure moyenne de RMSE relatif

Cette validation croisée sur la base de 2/3 pour l'apprentissage et 1/3 pour la validation, répétée 20 fois, nous permet de sélectionner le meilleur paramétrage tout en réduisant le risque de sur-apprentissage.

2.8.3. Évaluation des modèles

À la suite de cette optimisation des paramètres avec ses propres méthodes d'évaluation, nous avons choisi d'évaluer les modèles obtenus en sortie du SVR avec une validation croisée « leave-one-out » (LOO). Chaque échantillon est estimé avec l'ensemble des échantillons moins lui-même, ce qui permet d'avoir une indépendance tout en gardant le maximum d'échantillons pour l'apprentissage. Un autre avantage est qu'il n'y a pas le facteur aléatoire d'une division des échantillons, donc une seule estimation de chaque échantillon est nécessaire, sans intervalle de confiance. Les statistiques de validation extraites ensuite sont :

- Le coefficient de détermination (r^2) entre 0 et 1, qui donne une indication sur la variance expliquée par le modèle sur le paramètre estimé. $r^2 = 1$ signifie que la relation linéaire est parfaite entre les valeurs mesurées et prédites.

- L'erreur quadratique moyenne (RMSE) dans la même unité que le paramètre forestier estimé, qui mesure une erreur moyenne en donnant plus d'importance aux échantillons très mal estimés.
- Le RMSE relatif en pourcentage, qui est le RMSE divisé par la moyenne des échantillons de référence pour le paramètre forestier estimé.
- L'erreur absolue moyenne (MAE), qui comme le RMSE mesure une erreur associée au modèle mais qui donne le même poids à tous les échantillons.
- Le MAE relatif, en pourcentage, est le MAE divisé par la moyenne des échantillons de référence pour le paramètre forestier estimé.
- Le biais moyen, qui est la moyenne (sans valeur absolue) des erreurs.

Afin d'évaluer la significativité des statistiques obtenues avec le LOO, nous avons réalisé des tests sur des divisions aléatoires des échantillons d'apprentissage et de validation, avec une évolution du ratio de 10 à 90 % pour l'apprentissage. La figure II.25 montre les RMSE relatifs obtenus avec ces ratios pour les 6 paramètres forestiers estimés. On peut voir que la qualité des estimations sur les échantillons de validation (*Test*, ligne rouge) s'améliore (le RMSE relatif baisse) quand on augmente le ratio apprentissage/validation. L'écart type augmente en raison du plus faible nombre d'échantillons pour la validation lorsque le ratio augmente. Nous avons superposé à ces tests les résultats obtenus en validation LOO pour un nombre d'échantillons correspondant au ratio apprentissage (de 10 à 100 %, 500 divisions aléatoires). On peut voir que la courbe des résultats avec une validation LOO suit parfaitement celle d'une validation avec division apprentissage/validation. Ce résultat montre que l'on garde un maximum d'indépendance tout en maximisant le nombre d'échantillons d'apprentissage avec la validation LOO, et ainsi confirme sa bonne représentativité.

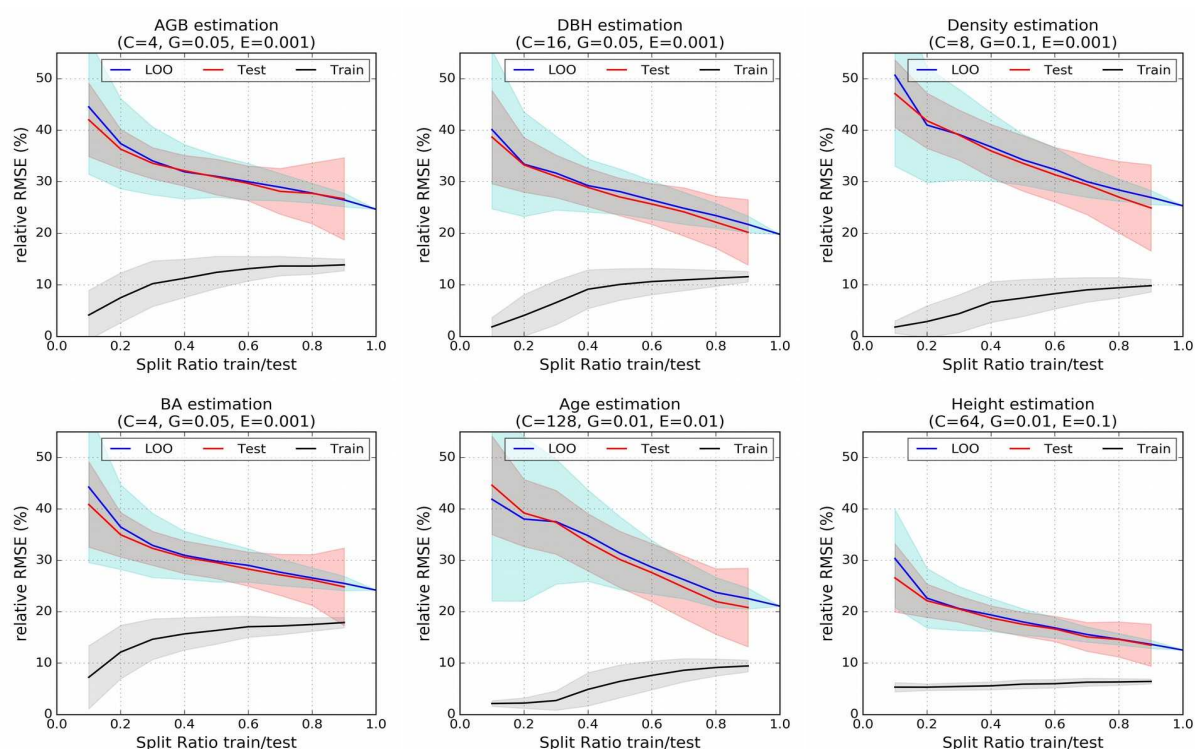


Figure II.25 : Évolution des performances (RMSE relatif) pour les 6 paramètres forestiers en fonction du ratio apprentissage(train)/validation(test). La moyenne (ligne) et l'écart type (zones pleines autour) sont calculés sur 500 divisions aléatoires. La validation LOO est calculée sur la même échelle de 10 à 100% des échantillons utilisés en entrée, avec 500 divisions aléatoires.

2.9. Résultats

Cette section présente les résultats des modèles obtenus avec la méthode de sélection des primitives Forward et une optimisation des paramètres du SVR décrite dans la section 2.8.2. La figure II.26 présente la sélection des primitives, indiquant les scores de RMSE relatifs (validation croisée LOO) associés aux différentes étapes de la sélection. Les types de primitives sont différenciés par des couleurs. Pour les paramètres AGB et BA, on constate à nouveau que l'amélioration avec la sélection Forward est limitée comparé à l'utilisation directe des 31 primitives. La progression du RMSE relatif est également assez lente entre 1 et 10 primitives sélectionnées. Pour les quatre autres paramètres forestiers, cette progression est plus forte, avec une réduction du RMSE relatif jusqu'à ~20 % pour l'âge entre 1 et 10 primitives sélectionnées. Ces modèles profitent aussi plus de la sélection Forward comparé au résultat avec 31 primitives. Cette amélioration est moins visible sur le profil de la hauteur, car les résultats sont déjà très bons mais la réduction en proportion est aussi importante que sur le DBH, l'âge et la densité d'arbres.

Concernant les types de primitives sélectionnées, nous pouvons voir une forte alternance des types de primitives sélectionnées en Forward. Il y a systématiquement 3 types différents dans les 5 premières primitives utilisées, voir dans les 3 premières pour les paramètres AGB, densité et hauteur. Enfin, si on regarde les courbes de RMSE relatifs, on voit qu'il serait possible d'enlever plusieurs primitives tout en ayant un résultat presque équivalent au meilleur pour la plupart des paramètres. Malgré cela, il est préférable de garder une méthode automatique et d'éviter les interventions manuelles dans la chaîne de traitement.

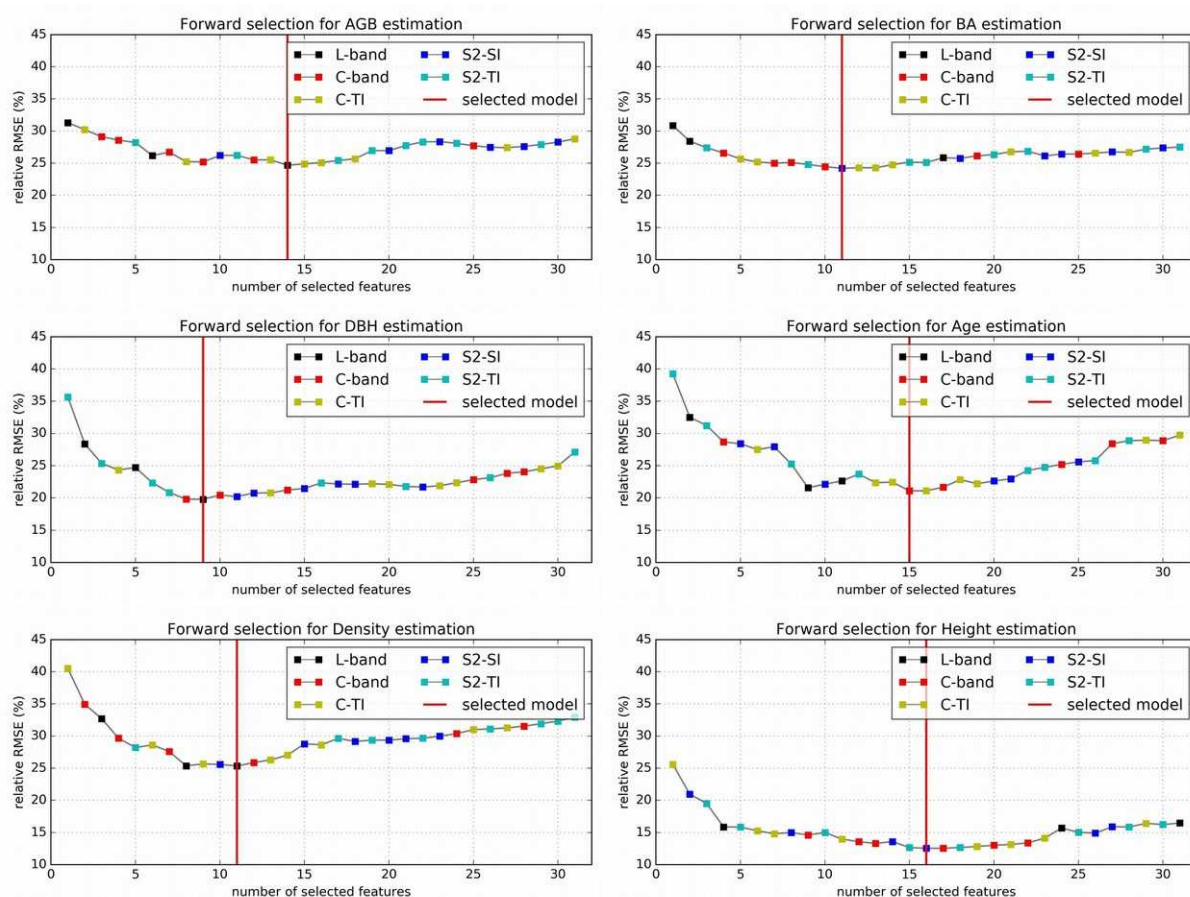


Figure II.26 : Profils de sélection Forward pour l'estimation des paramètres forestiers avec un SVR. L'appartenance des primitives aux 5 différents types est indiquée par les couleurs des carrés.

Les primitives retenues dans chaque modèle sont affichées dans le tableau II.8, rangées par types de primitives. Le chiffre entre parenthèses indique l'ordre dans la sélection Forward. Les statistiques de validation sont indiquées dans la colonne de droite. On constate que les 5 types (voir tableau II.6 : L-band, C-band, C-TI, S2-SI et S2-TI) sont utilisés dans le modèle final à l'exception du DBH qui n'utilise pas les indices spectraux de Sentinel-2. De 9 à 16 primitives sont sélectionnées selon les paramètres, les 5 à 10 premières ont une plus grande importance pour le modèle final (voir figure II.26). Les primitives issues de Sentinel-1 (C et C-TI) sont les plus fréquemment sélectionnées, ainsi que les coefficients de rétrodiffusion en bande L relativement au nombre initial (3 primitives en bande L, contre 6 à 8 pour les autres types de primitives, tableau II.6). Les primitives L-band semblent importantes pour les modèles. Pour les paramètres AGB et BA, ce résultat est attendu car la bande L montre des corrélations linéaires fortes avec ces paramètres. Pour les 4 paramètres DBH, âge, densité et hauteur, au moins une primitive L-band est sélectionnée dans les 5 premières alors que les relations linéaires avec ces paramètres sont assez faibles (figure II.11). Les indices spectraux de Sentinel-2 sont également utilisés dans les modèles bien qu'ayant les plus faibles relations linéaires avec les paramètres forestiers.

On trouve dans ces sets de primitives une combinaison de variables utiles pour prédire quantitativement les paramètres forestiers, mais aussi d'autres variables qui ont une valeur plus qualitative pour distinguer les différents sous-groupes dans le jeu de données d'apprentissage. En effet, on a vu dans la description des données qu'à partir de 10 à 12 ans, lors de la première éclaircie, les peuplements peuvent avoir des relations allométriques qui évoluent avec la maturité (natural thinning) et les opérations sylvicoles (dépressages et éclaircies). Les images satellitaires suivent aussi ces différences structurales, ainsi que le changement dans la répartition de la biomasse (dans les feuilles, branches, troncs) selon la maturité des peuplements.

Tableau II.8 : Résultats statistiques et liste des primitives utilisées pour l'estimation des paramètres forestiers.
Les résultats sont obtenus avec un SVR et une sélection Forward sur les 31 primitives issues de ALOS2-PALSAR2, Sentinel-1 et Sentinel-2. Les numéros entre parenthèse indiquent l'ordre de sélection des primitives.

	ALOS2-PALSAR2	Sentinel-1		Sentinel-2		Statistiques Validation
AGB	L-HV(1) L-HH(6) L-Ratio(14)	C-VH-sum(12) C-VV-sum(4) C-VH-win(7) C-VV-win(9) C-Ratio-win(3)	C-VH-sum-Hcor(8) C-VH-win-Hcor(2) C-VV-win-hom(13)	S2-NDWI-sum(10)	S2-BI-sum-hom(5) S2-NDVI-sum-hom(11)	$r^2=0.82$ 17.2 tons/ha 25 %
BA	L-HV(1) L-HH(2)	C-VV-sum(8) C-VH-win(4) C-VV-win(7) C-Ratio-win(10)	C-VV-win-Hcor(5) C-VV-win-hom(6)	S2-BI-win(11)	S2-BI-sum-Hcor(9) S2-NDVI-win-Hcor(3)	$r^2=0.79$ 5.1 m ² /m ² 24 %
DBH	L-HV(5) L-HH(9) L-Ratio(2)	C-Ratio-win(8)	C-VV-sum-hom(4)		S2-BI-sum-Hcor(6) S2-BI-sum-hom(1) S2-BI-win-hom(3) S2-NDVI-win-Hcor(7)	$r^2=0.88$ 4.3 cm 20 %
Age	L-HV(9) L-HH(11) L-Ratio(2)	C-VV-sum(4) C-VV-win(15)	C-VV-sum-hom(14) C-VH-win-hom(6) C-VV-win-hom(13)	S2-BI-sum(7) S2-NDWI-sum(5) S2-BI-win(10)	S2-BI-sum-Hcor(12) S2-BI-sum-hom(1) S2-BI-win-hom(3) S2-NDVI-win-Hcor(8)	$r^2=0.89$ 4.8 ans 21 %
Density	L-HV(3) L-HH(8) L-Ratio(11)	C-VH-sum(4) C-VH-win(2) C-Ratio-win(7)	C-VH-sum-Hcor(9) C-VH-sum-hom(6) C-VV-sum-hom(1)	S2-NDWI-win(10)	S2-BI-sum-hom(5)	$r^2=0.85$ 211 trees/ha 25 %
Height	L-HV(4)	C-VV-sum(13) C-VV-win(12) C-Ratio-win(9)	C-VV-sum-Hcor(11) C-VV-sum-hom(6) C-VH-win-hom(1) C-VV-win-hom(7)	S2-BI-sum(2) S2-NDVI-sum(14) S2-NDWI-sum(16) S2-NDWI-win(8)	S2-BI-sum-hom(3) S2-BI-win-hom(5) S2-NDVI-win-Hcor(10) S2-NDVI-win-hom(15)	$r^2=0.94$ 1.7 m 13 %

Nous avons vérifié si les primitives sélectionnées par types étaient statistiquement similaires entre les différents paramètres forestiers. Le tableau II.9 reprend le tableau II.8 en indiquant le nombre de primitives sélectionnées par types pour chaque paramètre forestier. Nous avons appliqué un test de Khi2 sur ce tableau pour tester s'il y avait des différences significatives dans la distribution des primitives. Le résultat est de 8.68 avec un $p_0 = 0.563$, on ne peut donc pas rejeter l'hypothèse d'équivalence des paramètres pour la distribution des primitives avec ce test.

Pour compléter ce test, nous avons réalisé des nouveaux tests de Khi2 pour comparer la distribution des primitives de chaque paramètre forestier, à une distribution théorique fixée :

$$N_{théorique} = \frac{N_{type} \times N_{modèle}}{N_{total}} \quad (9)$$

Avec $N_{théorique}$ le nombre théorique de primitives sélectionné pour un type, N_{type} le nombre maximum de primitives du type, $N_{modèle}$ le nombre de primitives sélectionnées par le modèle (tous types) et $N_{total}=31$ le nombre initial de primitives en entrée de la sélection. Cette distribution théorique suppose que les primitives sont sélectionnées équitablement dans les 5 types L, C, C-TI, S2-SI et S2-TI. Les résultats sont présentés dans le tableau II.10. Les hypothèses d'équivalence des distributions sont globalement confirmées, seul le DBH a un p_0 presque significatif. En effet, pour le paramètre DBH on peut voir dans les tableaux II.8 et II.9 que les primitives L et S2-TI sont majoritairement sélectionnées comparé à C, C-TI et S2-SI.

Tableau II.9 : Tableau de contingence pour la fréquence de sélection des différents types de primitives satellitaires pour les 6 paramètres forestiers.

	ALOS2-PALSAR2	Sentinel-1		Sentinel-2		Total
	L	C	C-TI	S2-SI	S2-TI	
AGB	3	5	3	1	2	14
BA	2	4	2	1	2	11
DBH	3	1	1	0	4	9
Age	3	2	3	3	4	15
Density	3	3	3	1	1	11
Height	1	3	4	4	4	16
Total	15	18	16	10	12	

Tableau II.10 : Tests du Khi2 pour chaque paramètre forestier entre la distribution des primitives sélectionnées et une distribution théorique déterminée (équation 9).

	AGB	BA	DBH	Age	Density	Height
Khi2	5.84	3.56	9.23	2.14	5.67	0.47
p_0	0.212	0.469	0.056	0.711	0.225	0.976

2.9.1. Hauteur moyenne (Height)

La figure II.27 présente la validation des estimations de hauteur moyenne. La mesure ou l'estimation de la hauteur des forêts est assez récurrente dans la littérature car plusieurs données de télédétection très précises permettent de fournir ces estimations, notamment le Lidar aéroporté et la photogrammétrie. Les résultats d'estimation avec notre méthode sont très bons, c'est le paramètre qui obtient les plus faibles erreurs parmi les six paramètres étudiés sur ce site d'étude. Le RMSE de 1.7 m (12 % de RMSE relatif) est comparable avec les erreurs de mesure sur le terrain. On ne constate pas de saturation et très peu valeurs aberrantes (« outliers »). Les faibles valeurs, en dessous de 5 à 6 m de hauteur, sont un peu plus dispersées en relatif.

Sur ce jeu de données, une valeur de hauteur dominante a été dérivé des mesures de hauteur (n-1 arbres les plus hauts sur n ares=0.01 ha). Les deux paramètres de hauteur sont très reliés,

car nous travaillons sur des plantations avec des arbres tous du même âge. L'estimation de la hauteur dominante donne des résultats quasi-identiques à celui de la hauteur moyenne.

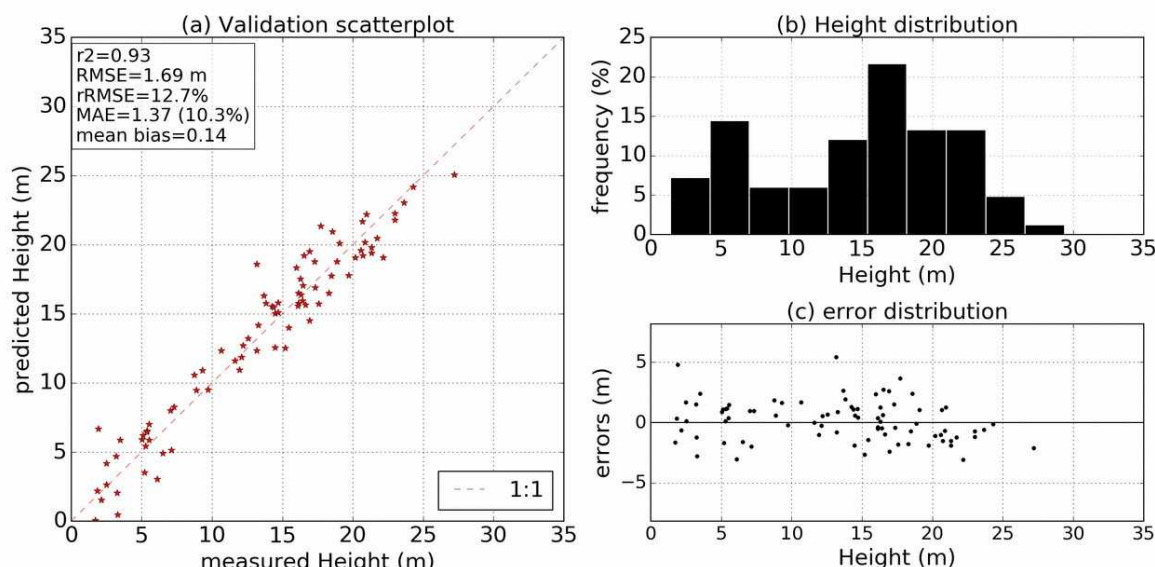


Figure II.27 : Graphiques de validation des résultats d'estimation de la hauteur moyenne (Height).

2.9.2. Diamètre moyen à 130 cm / mean DBH

La figure II.28 présente la validation des estimations du diamètre moyen à 130cm (DBH). Le résultat est très bon, avec un r^2 de 0.88, et une erreur absolue moyenne (MAE) de 13.8 % en relatif. Plusieurs outliers dégradent le RMSE, à 20 % en relatif. On distingue en particulier deux outliers à ~12 cm (sur-estimé) et à ~41 cm (sous-estimé). L'analyse des relations entre le DBH et les primitives ne permet pas de mettre en évidence une particularité de ces deux échantillons. Concernant les relations entre paramètres forestiers, le premier échantillon à ~12 cm a un DBH un peu faible par rapport à l'âge du peuplement, le deuxième à ~40 cm est un peuplement âgé (50 ans) mais avec peu d'arbres et peu de biomasse (AGB~50 tonnes/ha). Mises à part ces quelques particularités, les outliers ne sont pas faciles à caractériser, il s'agit plus d'un profil général que d'une anomalie en particulier. Ainsi, bien que l'estimation du DBH soit très bonne, il faudra faire attention à des erreurs potentiellement fortes localement.

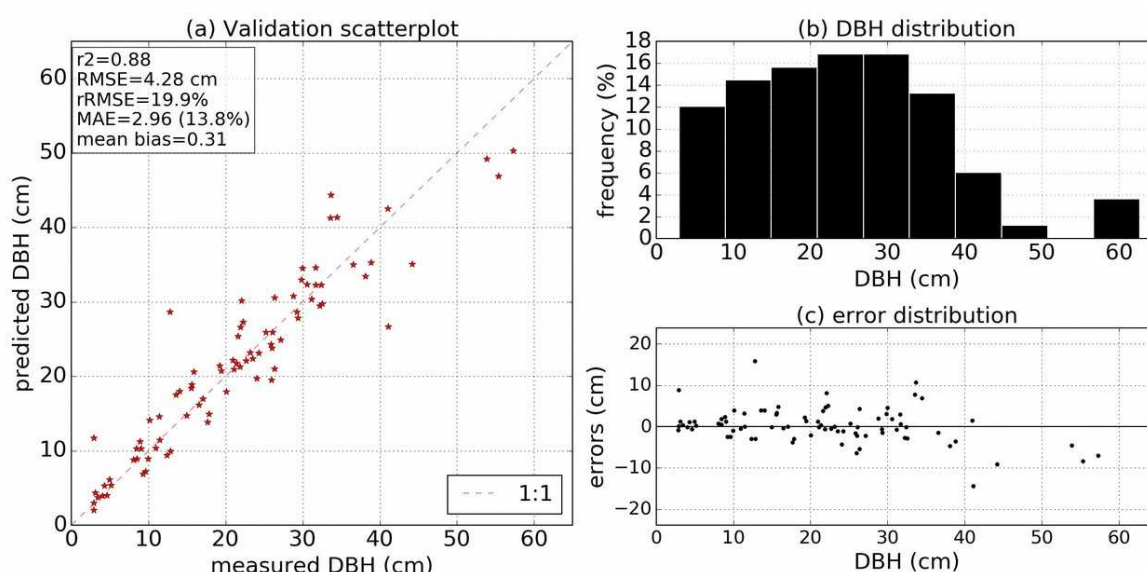


Figure II.28 : Graphiques de validation des résultats d'estimation du diamètre moyen à 130 cm (DBH).

2.9.3. Âge des peuplements

La figure II.29 présente la validation des estimations de l'âge des peuplements. Ce paramètre est spécifique aux plantations, car dans les forêts gérées par régénération, l'âge des arbres est variable et un âge moyen aurait peu de sens en termes de gestion ou d'écologie comparé aux paramètres de structure forestière. L'âge est fortement relié au DBH sur ces forêts, et les résultats d'estimations ressemblent à ceux du DBH. L'estimation est globalement bonne, nous pouvons voir cependant quelques outliers dans les faibles valeurs en dessous de 5 ans, et entre 25 et 30 ans.

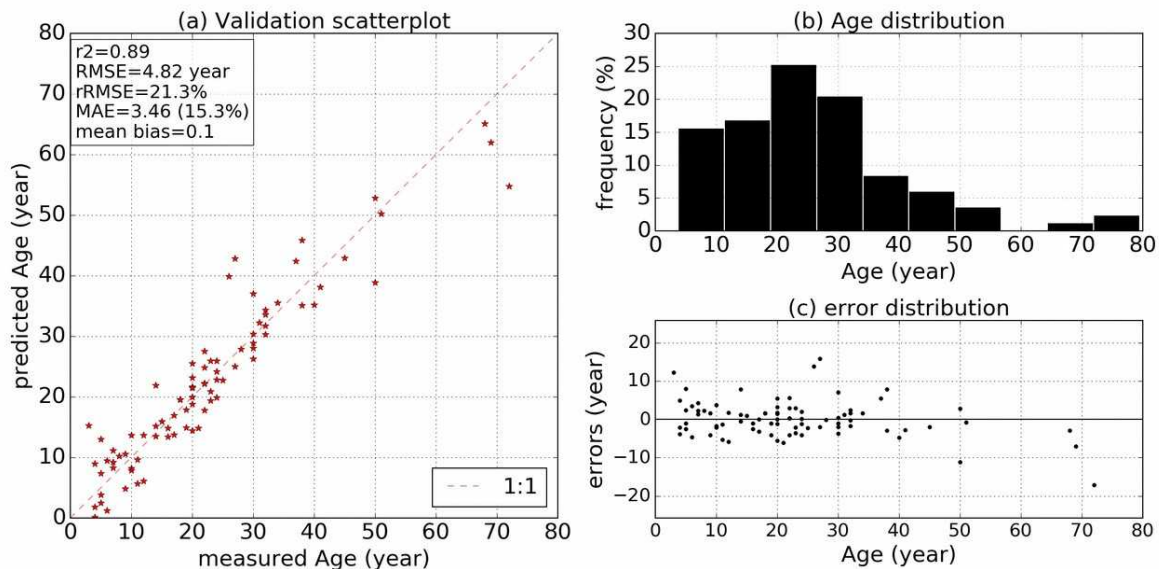


Figure II.29 : Graphiques de validation des résultats d'estimation de l'âge.

2.9.4. Densité d'arbres

La figure II.30 présente la validation des estimations de la densité d'arbres par hectare. Deux types de peuplements sont présents dans la zone : les plantations en rang et les semis. Les semis ont généralement des densités plus grandes pour les jeunes peuplements, et les peuplements matures sont plus hétérogènes en raison des éclaircies plus fortes pour ce type de boisement. Les plantations sont globalement plus homogènes, avec des densités plus fortes pour les peuplements intermédiaires et matures. L'estimation de la densité sur ce site d'étude donne de bons résultats, bien que le scatterplot soit plus dispersé que pour le DBH, l'âge et la hauteur. La figure II.30.b montre une distribution avec deux pics de fréquences autour de 550 et de 1400 arbres par hectares. Les résultats sont similaires en termes de r^2 entre les deux groupes. La valeur extrême à 2600 arbres par hectare est sous-estimée, il s'agit d'un reboisement par semi en rangs assez larges où les rangées sont très marquées. La plupart des autres fortes densités sont entre 1300 et 1500 arbres par hectares, le modèle estime cette valeur extrême (2600 arbres/ha) à la limite haute de ces valeurs (1600 arbres/ha). On note également un outlier à ~1000 arbres/ha qui est estimé à 500 arbres par hectares. Cet échantillon ne ressort pas particulièrement lors de l'analyse des primitives utilisées par le modèle, mais c'est le même échantillon que l'outlier de la figure II.28 avec un DBH à ~12 cm. Les autres échantillons autour de 1000 arbres/ha sont très variés sur photographie aérienne, et la visualisation des échantillons mesurés autour de 500 arbres/ha (valeur à laquelle l'outlier a été estimé) montre qu'en effet l'apparence est la même sur des images à 50 cm de résolution spatiale, tout comme avec les images Sentinel-2 à 10 m de résolution.

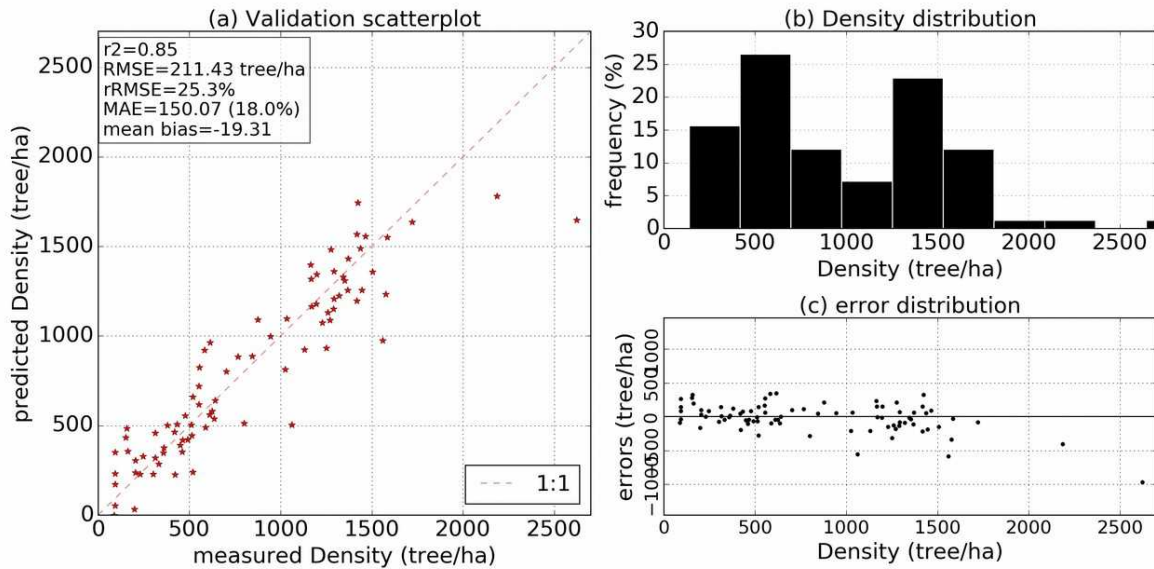


Figure II.30 : Graphiques de validation des résultats d'estimation de la densité d'arbres (Density).

2.9.5. Surface terrière / Basal Area (BA)

La figure II.31 présente la validation des estimations de surface terrière (BA). Nous pouvons voir une sur-estimation des valeurs intermédiaires (15 à 25 m²/ha). Les fortes valeurs sont plutôt bien prédites, mais légèrement sous-estimées. Les deux paramètres AGB et BA sont très proches, nous obtenons des résultats similaires (figure II.32). La plupart des mesures terrain permettent d'obtenir la surface terrière (mesures par arbres à la placette ou méthode avec chaîne relascopique), mais l'estimation de la biomasse n'est pas toujours possible. Sur ces autres sites où l'estimation d'une biomasse aérienne de référence à la placette n'est pas faisable, le fait que la surface terrière soit un indicateur assez proche du volume global et de la biomasse aérienne nous permettra d'extrapoler sur les capacités de notre méthode pour produire des cartes de biomasse lorsque la donnée BA est disponible.

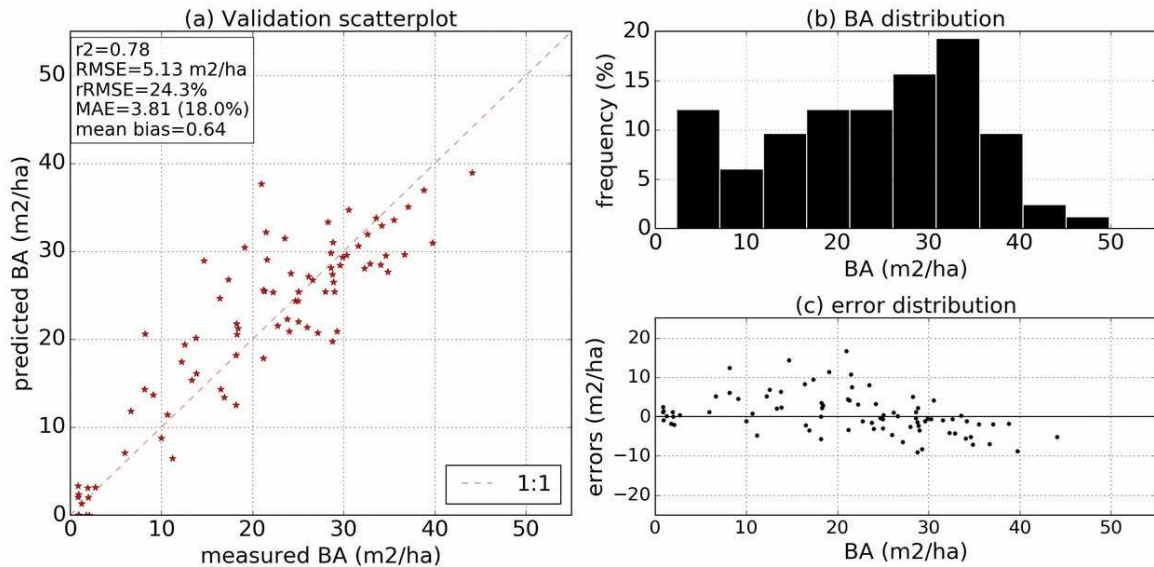


Figure II.31 : Graphiques de validation des résultats d'estimation de la surface terrière (BA).

2.9.6. Aboveground Biomass (AGB)

La figure II.32 présente la validation des estimations de l'AGB. Les résultats sont globalement bons, mais on constate de larges sur-estimations entre 50 et 70 tonnes/ha. Cette plage de valeurs correspond à des peuplements très divers en termes de relations entre les paramètres forestiers. Parmi ces valeurs d'AGB, on trouve des peuplements âgés de 15 à 50 ans, des densités de 100 à 1500 arbres/ha, des hauteurs de 9 à 22 m, etc. C'est autour de 15-20 ans qu'interviennent en général les premières éclaircies, à partir de là les peuplements peuvent avoir une très grande variabilité de paramètres forestiers. En revanche, bien que cette variabilité s'observe encore pour les fortes biomasses (voir figure II.3 les corrélations entre paramètres forestiers), le modèle produit d'assez bonnes estimations pour les biomasses entre 80 et 140 tonnes/ha, avec une légère sous-estimation.

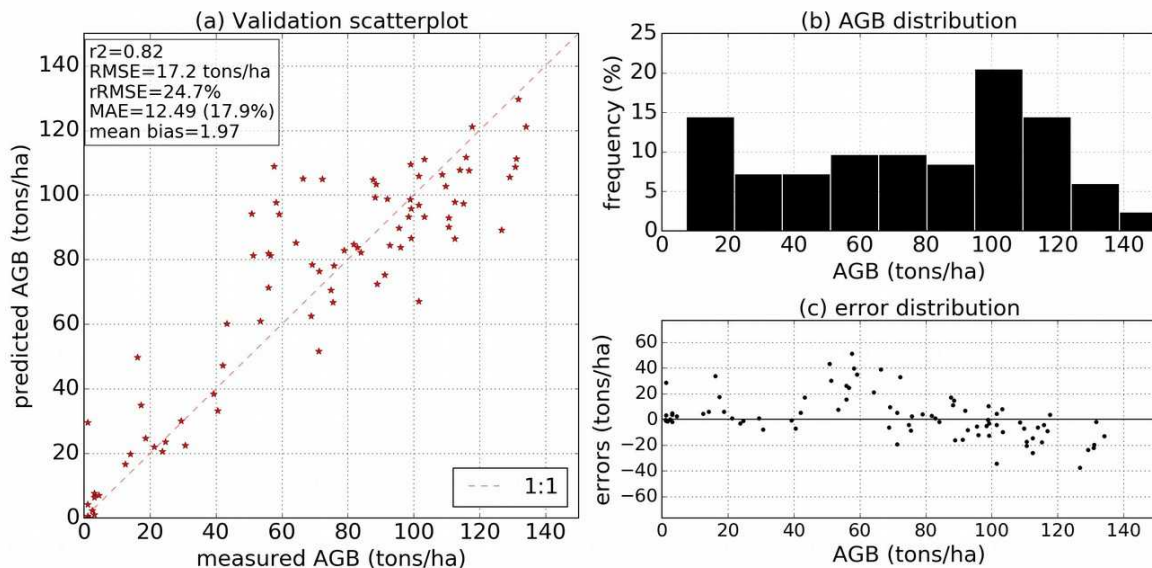


Figure II.32 : Graphiques de validation des résultats d'estimation de la biomasse aérienne (AGB).

La figure II.33 montre deux peuplements mesurés avec une même biomasse (AGB), mais très différents sur les autres paramètres. Dans les deux cas, l'AGB est fortement surestimé. L'observation des photographies, des relations entre paramètres forestiers et des primitives satellitaires ne permet pas de mettre en évidence un type de peuplement particulier qui serait mal estimé par le modèle. Étant donné que ces différences de gestion des peuplements sont encore présentes pour les fortes valeurs de biomasse, mais que les estimations sont meilleures, on peut penser que cette plage de valeurs ~50-70 tonnes/ha est une zone charnière où il est difficile pour le modèle de bien estimer l'AGB en raison de la trop grande variabilité des valeurs des primitives satellitaires pour une biomasse similaire. Concernant les très faibles valeurs, on peut voir dans la figure II.33.a une biomasse aérienne estimée à presque 30 tonnes/ha, l'observation des photographies aérienne nous permet de constater qu'il s'agit d'un peuplement où plusieurs vieux arbres n'ont pas été coupés tandis que le reste vient d'être planté (~4 ans). Cela illustre encore la grande variabilité des profils de peuplement que l'on peut trouver, même dans des forêts gérées de manière intensive, et ainsi l'importance d'avoir à la fois des données à haute résolution pour l'estimation mais également des jeux de données de référence terrain qui reflètent cette diversité des gestions.



Figure II.33 : Photographies aériennes des zones de mesures dans deux peuplements de pin maritime. Les cercles jaunes représentent la zone de mesure, les zones illustrées font 200×170 m. (a) AGB = 51 tonnes/ha (estimée à 81 tonnes/ha), âge = 51 ans, densité = 91 arbres/ha ; (b) AGB = 56 tonnes/ha (estimée à 82 tonnes/ha), âge = 14 ans, densité = 1370 arbres/ha.

2.10. Analyse de l'apport des types de données satellitaires

L'objectif de cette section est de quantifier l'amélioration des résultats en utilisant l'ensemble des primitives satellitaires décrites, par rapport à un seul type de données. Dans un deuxième temps, nous évaluons aussi l'apport d'une meilleure résolution spatiale (avec SPOT-6), par rapport aux données Sentinel-2, pour les indices de texture spatiale.

La littérature existante sur l'utilisation des données SAR en bande L et C, des données optiques et des indices de texture spatiale, suggère que la bande L donne le plus d'informations sur le volume des forêts, et ainsi sur certains paramètres de structure tels que la surface terrière (BA) et le DBH. Nous avons testé l'utilisation de chaque type de primitive un à un pour analyser les performances de chacun. La figure II.34 montre les RMSE relatifs obtenus avec la bande L, la bande C, les indices de textures issues de la bande C (C-TI), les indices spectraux (S2-SI), et les indices de textures issues des indices spectraux (S2-TI). Nous avons ajouté les performances en combinant toutes les primitives issues de Sentinel-1 (14), puis de Sentinel-2 (14), ainsi que les résultats en combinant l'ensemble des 31 primitives.

On constate tout d'abord que la combinaison de l'ensemble des 31 primitives en entrée de la méthode donne systématiquement les meilleurs résultats, avec une amélioration du RMSE relatif de 4 à 10 % selon les paramètres forestiers par rapport à l'utilisation de chaque type de données seuls. Les indices de texture en bande C (C-TI) ont presque systématiquement les meilleurs résultats parmi les 5 types seuls, y compris pour l'estimation de l'AGB où on aurait pu penser que le coefficient de rétrodiffusion en bande L (ALOS) serait plus performant. En effet, on a vu figure II.11 que les primitives en bande L avaient les plus fortes relations linéaires avec les paramètres AGB et BA, mais la combinaison dans un SVR des différents indices de texture en bande C permet d'avoir des résultats légèrement meilleurs pour l'estimation de ces deux paramètres sur ce jeu de données.

Pour l'estimation des paramètres DBH, âge, densité (Density) et hauteur (Height), l'ensemble des indices de texture en bande C montraient des fortes relations linéaires avec ces paramètres donc ces bons résultats d'estimation paraissent logiques. Les indices spectraux issus de Sentinel-2 (S2-SI) ont des résultats en général moins bons que les textures spatiales issues de ces indices (S2-TI). La combinaison des deux types issus de Sentinel-2 améliore la qualité des

résultats, tandis que pour Sentinel-1, l'ajout des coefficients de rétrodiffusion n'améliore que très peu comparé à l'utilisation des textures uniquement.

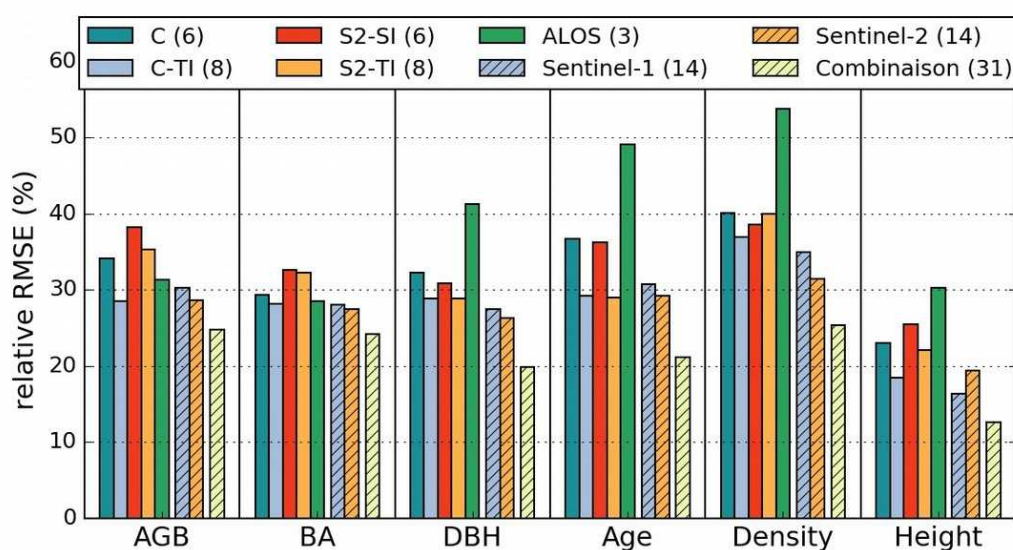


Figure II.34 : Comparaison des RMSE relatifs obtenus pour l'estimation des 6 paramètres forestiers avec chaque différent type de primitives, les données Sentinel-1 et Sentinel-2, et la combinaison des 31 primitives. Utilisation d'un algorithme SVR avec une sélection Forward.

Les résultats de la figure II.34 suggèrent donc qu'il y a des synergies entre types de données que le SVR est capable d'exploiter pour améliorer l'estimation des paramètres forestiers. Nous avons donc testé l'ensemble des couples de types de données pour voir quelles étaient les synergies les plus évidentes.

La figure II.35 présente les résultats des couples de types de données dans des tableaux symétriques autour de la diagonale qui contient les résultats des types seuls. Pour les paramètres AGB et BA, la meilleure combinaison est L + C-TI qui obtenaient les deux meilleurs résultats seuls. On retrouve le même schéma pour l'estimation de l'âge : la combinaison des deux meilleurs types seuls produit le meilleur résultat (C-T + S2-TI). Pour les autres paramètres forestiers DBH, densité et hauteur, les résultats des combinaisons sont plus surprenants. En effet, on trouve dans la meilleure combinaison le type de donnée qui donne le meilleur résultat utilisé seul (les indices de texture C-TI ou S2-TI), mais c'est la combinaison avec la bande L qui donne les résultats alors que c'est le type de donnée qui donne les moins bons résultats lorsqu'il est utilisé seul pour l'estimation des paramètres DBH, densité et hauteur. Cela pourrait s'expliquer par le fait que les primitives issues de la bande L sont moins corrélées avec les données Sentinel : l'algorithme est capable d'intégrer cette donnée de manière complémentaire aux autres pour améliorer le modèle bien que la bande L ne soit pas la plus performante.

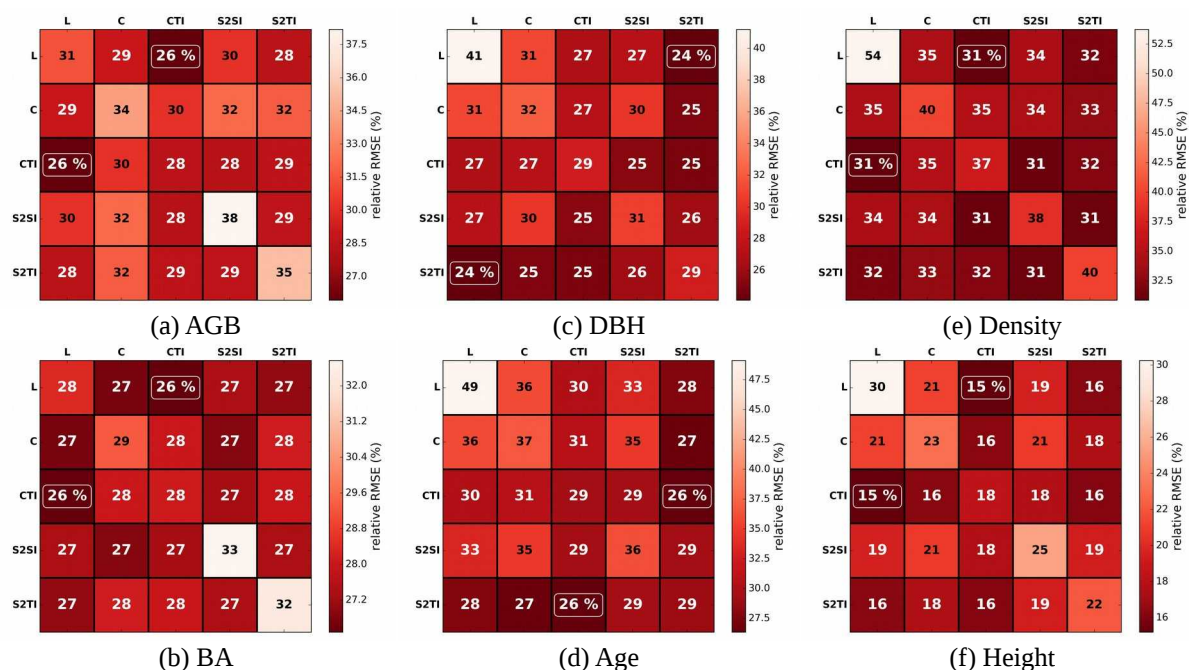


Figure II.35 : RMSE relatifs en tableaux croisés des combinaisons entre types de données satellitaires. Les meilleures combinaisons sont encadrées en blanc.

Cette analyse sur les différents types de données satellitaires illustre l'intérêt d'utiliser toutes ces données en combinaison pour favoriser les synergies et améliorer l'estimation des paramètres forestiers. La méthodologie issue de l'ensemble de ces tests permet d'avoir un set de primitives large et de sélectionner automatiquement ensuite celles qui se combinent le mieux dans un SVR pour l'estimation de chaque paramètre de structure. Ces synergies peuvent changer selon le type de forêts, c'est pourquoi il est intéressant de garder en entrée l'ensemble des primitives issues des mosaïques ALOS, de Sentinel-1 et de Sentinel-2, même si certaines apparaissent moins importantes avec ce jeu de données terrain sur du pin maritime dans les Landes.

Ajout de textures à très haute résolution spatiale

Une couverture annuelle SPOT-6/7 est acquise sur la France métropolitaine depuis 2013 et disponibles en libre accès pour les organismes de service public. Ces images ont une meilleure résolution spatiale que les images Sentinel, avec 6 m en mode multi-spectral et 1.5 m en mode panchromatique. Nous avons ajouté des primitives de texture spatiale issues de SPOT-6 pour tester l'adaptabilité de la chaîne de traitement, et la possibilité d'améliorer les résultats avec une meilleure résolution des images optiques. Les métriques d'homogénéité et de corrélation de Haralick ont été extraites des indices NDVI, BI et de l'image PAN. Ces 6 primitives issues de SPOT-6 ne sont ajoutées qu'à titre expérimental dans notre chaîne, car ce type d'image ne remplit pas totalement nos critères de libre accès et de reproductibilité et les données peuvent ne pas être comparables d'une tuile à l'autre ou d'une année à l'autre. Les images sont acquises à des dates différentes et avec des angles de visées qui peuvent varier, aucun algorithme d'harmonisation des mosaïques n'est appliqué sur ces données. La figure II.36 montre l'exemple de ces hétérogénéités d'acquisition sur la mosaïque de la totalité des Landes, ainsi qu'un zoom sur la bordure entre deux images panchromatiques sur un autre site d'étude dans la forêt d'Orléans.

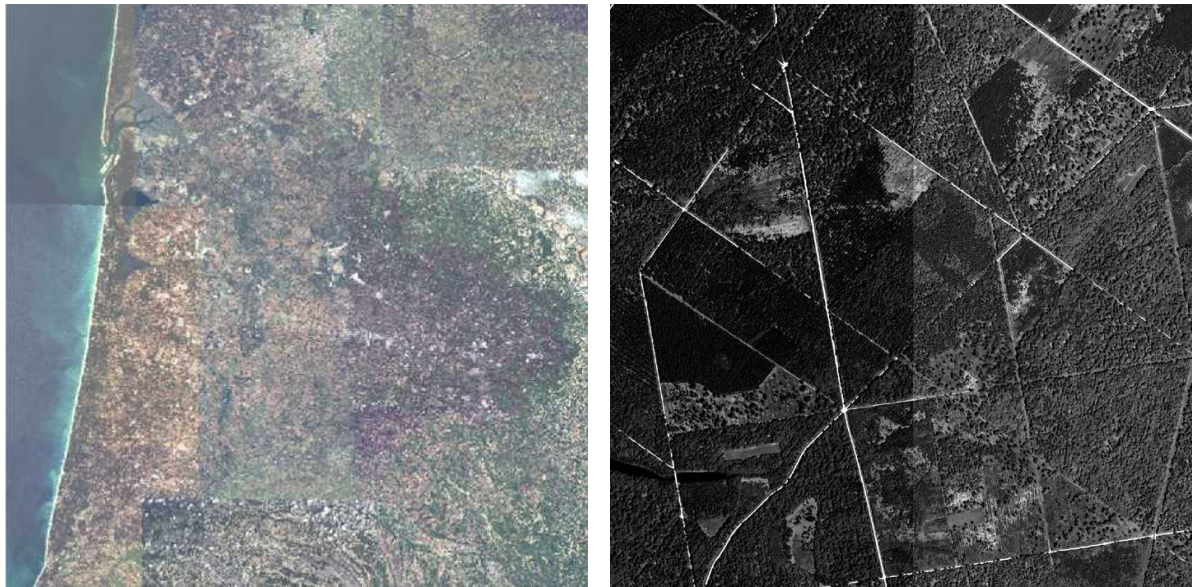


Figure II.36 : Exemple des problèmes radiométriques pour les mosaïques SPOT. (a) mosaïque couleur sur Landes de Gascogne, (b) limite de tuiles PAN (étirements identiques) sur la forêt d'Orléans.

Les résultats de l'ajout des indices de texture issus de SPOT-6 sont présentés dans la figure II.37 et le tableau II.11. Les différences ne sont pas assez prononcées pour conclure à une amélioration avec SPOT-6. Le DBH présente la meilleure amélioration, avec un gain de 2.5 % de RMSE relatif. L'ajout de cette donnée avec une meilleure résolution spatiale n'a donc pas d'effet vraiment significatif sur ce jeu de données, mais ce procédé montre qu'on peut facilement intégrer de nouvelles primitives. Cette option est intéressante lorsqu'on s'intéresse à une zone en particulier et que l'on a des données autres que les données en libre accès ALOS et Sentinel, potentiellement plus précises et pouvant améliorer les résultats (Lidar, THRS, hyper-spectral, etc.). L'ajout d'informations exogènes, sur la topographie ou le type de sol par exemple, est aussi une piste intéressante et qui peut être explorée avec la méthode que nous proposons.

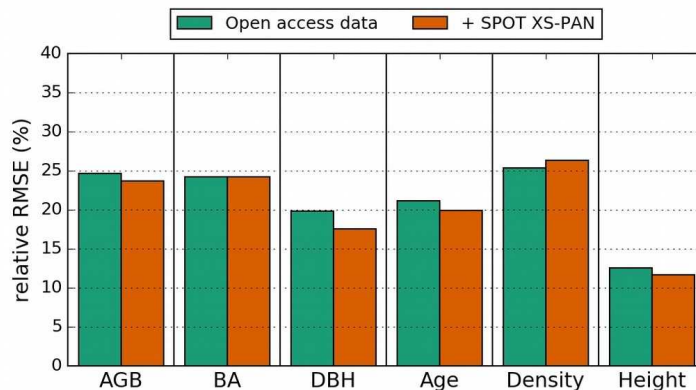


Figure II.37 : Comparaison des RMSE relatifs obtenus avec un SVR et une sélection Forward, sur les 31 primitives (ALOS2-PALSAR2, Sentinel-1 et Sentinel-2) ou avec l'ajout des textures spatiales à meilleure résolution (SPOT-6 multi-spectral et panchromatique).

Le tableau II.11 reprend les résultats obtenus avec les meilleurs types de primitives (première ligne), la meilleure combinaison de deux types (deuxième ligne), l'ensemble des primitives en libre accès (troisième ligne) et enfin l'ajout des données SPOT-6 à ces primitives (quatrième ligne). On voit bien la progression nette lorsqu'on combine deux types de primitives par rapport à un seul. La combinaison de l'ensemble des primitives améliore encore les résultats de manière importante pour le DBH, l'âge et la densité. L'amélioration est un peu

moins marquée pour les paramètres AGB, BA et hauteur moyenne. Enfin, l'ajout des données SPOT ne paraît pas être déterminant pour la qualité des estimations.

Tableau II.11 : Statistiques de validation (RMSE relatif) pour les différentes configurations.

	RMSE relatif (SVR/Forward)					
	AGB	BA	DBH	Age	Densité	Hauteur
Meilleur(s) type(s)	31% L-band	28% L-band/C-TI	29% C-TI/S2-TI	29% C-TI/S2-TI	37% C-TI	18% C-TI
Meilleure combinaison de 2 types	26% L-band + C-TI	26% L-band + C-TI	24% L-band + S2-TI	26% C-TI + S2-TI	31% L-band + C-TI	15% L-band + C-TI
Combinaison 5 types	25%	24%	20%	21%	25%	13%
Ajout SPOT-6	24%	24%	18%	20%	26%	12%

2.11. Conclusion

Nous avons décrit la construction d'une méthodologie et l'analyse des résultats sur un site d'étude avec des peuplements de pin maritime dans la forêt des Landes de Gascogne. Les données issues des mosaïques ALOS2-PALSAR2, de Sentinel-1 et de Sentinel-2 peuvent être utilisées en combinaison pour estimer les principaux paramètres forestiers qui permettent un suivi pour la gestion des forêts et les stocks de carbone.

Notre objectif ici était de tester les différentes approches utilisées dans la littérature pour aboutir à une méthode unifiée qui réponde aux critères que nous avons défini :

- Générique, c'est-à-dire qu'il est possible théoriquement d'appliquer la chaîne de traitement sur n'importe quelle zone et n'importe quelle forêt. Le modèle produit en fin de chaîne est spécifique au jeu de données en entrée, mais la méthode appliquée est identique.
- Automatique et modulaire : il est possible de modifier certaines étapes (primitives supplémentaires, algorithme de régression) mais les étapes de sélection des primitives et de paramétrisation sont automatiques.
- Utilisation de données satellitaires à haute résolution spatiale, en libre accès et acquises régulièrement sur l'ensemble du globe. La chaîne est également basée sur des bibliothèques « open source » (OTB, GDAL, bibliothèques Python).

Les différents tests et analyses réalisés sur ce jeu de données nous ont permis de sélectionner 31 primitives satellitaires qui sont utiles pour l'estimation des paramètres forestiers. Le coefficient de rétrodiffusion radar en bande L et les indices de texture spatiale issues de Sentinel-1 ou Sentinel-2 sont particulièrement complémentaires. L'ensemble des types de primitives est utile pour obtenir les meilleurs résultats. Nous recommandons l'utilisation du SVR et d'une sélection des primitives de type Forward pour des raisons d'efficacité sur les petits jeux de données de référence, ce qui est souvent le cas pour l'estimation des paramètres forestiers.

Chapitre III : Généralisation de la méthode

Chapitre III : Généralisation de la méthode.....	61
3.1. Application à une même essence.....	61
3.1.1. Présentation des données.....	61
3.1.2. Estimation des paramètres forestiers.....	65
3.2. Application à plusieurs essences forestières.....	74
3.2.1. Présentation des nouveaux sites d'étude.....	74
3.2.2. Analyse des jeux de données.....	76
3.2.3. Application de la méthode.....	79
3.2.4. Impact de la nomenclature des types de forêts.....	83
3.3. Conclusion.....	86

Nous avons construit une méthode efficace pour estimer les paramètres forestiers sur un site d'étude avec des plantations de pin maritime dans les Landes de Gascogne. L'objectif ensuite est de tester l'application de cette méthode sur d'autres sites et d'autres types de forêts.

La première question est de savoir si un modèle construit pour une essence, et sur un site donné, peut s'appliquer sur la même essence mais sur un autre site (avec potentiellement sol, climat et pratiques de sylviculture différentes). En effet, lors de la spatialisation pour créer des cartes de paramètres forestiers, un modèle appris sur une forêt en particulier va être appliqué aux forêts environnantes ou éloignées mais supposées être le même type de forêt.

Dans un deuxième temps, il est nécessaire de tester si notre méthode construite sur un type de forêts bien particulier peut s'appliquer à d'autres types de forêts, notamment des feuillus et des forêts non plantées. En fonction des résultats de ces deux premières analyses, nous allons mieux connaître les paramètres qui définissent le type de forêt présente au sol (du point de vue télédétection) et définir un système de nomenclature à plusieurs niveaux qui permettrait, d'une part, de mieux généraliser et spatialiser les modèles, et d'autre part, de savoir quel modèle appliquer selon la connaissance et la précision de l'information sur le type de forêt au sol.

3.1. Application à une même essence

3.1.1. Présentation des données

Nous disposons de mesures terrain sur un autre site de pin maritime dans les Landes. Il s'agit d'une forêt publique (domaniale) dans la zone côtière au nord du premier site d'étude. Les données nous ont été acquises et fournies par l'ONF. Nous distinguerons par la suite les deux sites des Landes par les appellations Landes-INRA (le premier site d'étude) et Landes-ONF. La figure III.1 montre la localisation de ce second site d'étude.

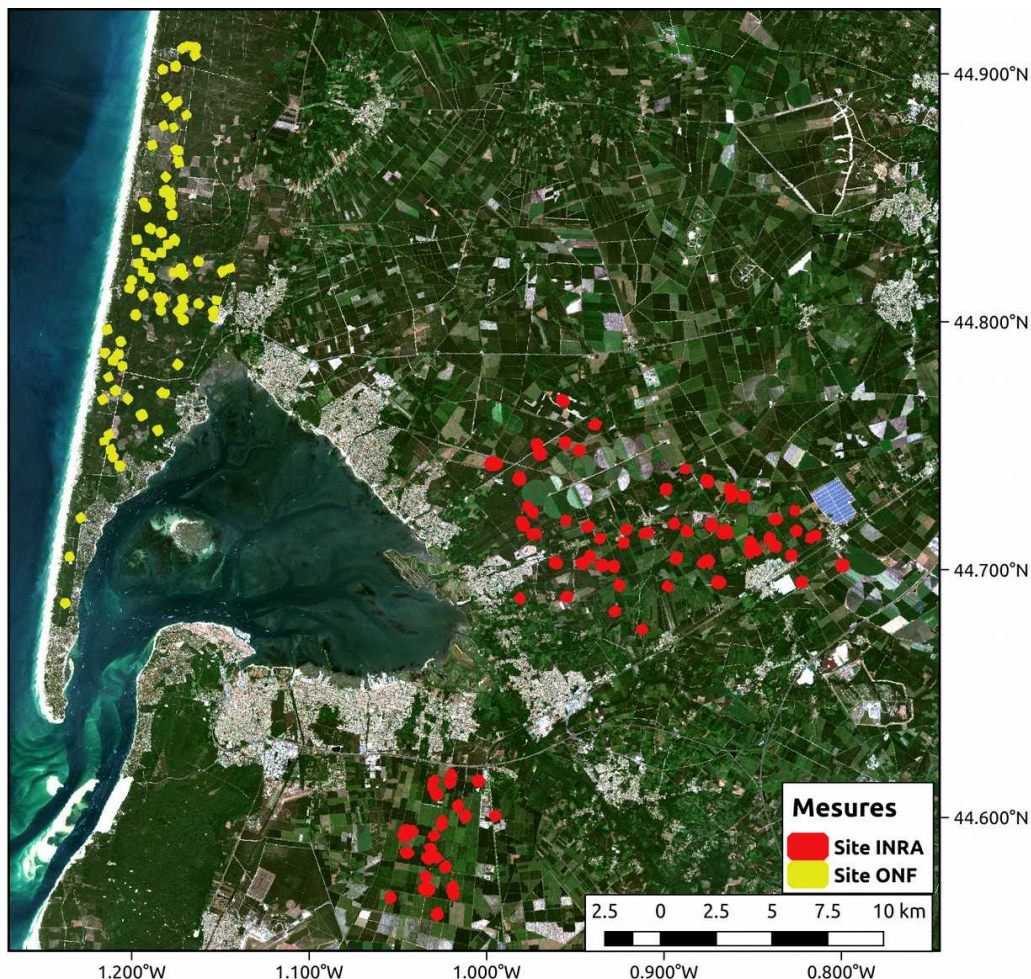


Figure III.1 : Localisation des sites d'étude Landes-INRA et Landes-ONF.

Les arbres ont été mesurés et comptés dans 83 placettes de 15 m de rayon (0.07 ha). Le seuil de pré-comptage est de 10 cm, ce qui veut dire que les arbres avec un diamètre (DBH) inférieur à 10 cm ne sont pas comptabilisés dans les mesures. Nous avons compilé ces mesures à la placette pour obtenir la surface terrière (BA), le DBH moyen et dominant (DBH moyen des $n-1$ arbres les plus gros sur n arbres = 0.01 ha), et la densité d'arbres par hectare. La hauteur dominante est directement donnée à la placette par l'ONF. Le climat et l'essence forestière sont donc les mêmes que sur le site Landes-INRA, à la différence près de la présence de quelques chênes dans les parcelles du site ONF. Le site Landes-ONF est situé plus proche de la côte, dans la zone dunaire, ce qui implique des variations d'angle d'incidence (IA) local pour les données radar. La figure III.2 montre la distribution de l'angle d'incidence local d'une image Sentinel-1 en orbite descendant (la même que celle utilisée pour le site Landes-INRA). Pour rappel, le site Landes-INRA a un angle d'incidence aux alentours de 33°. On observe dans la figure III.2 que la topographie fait varier l'angle d'incidence local entre 25 et 45 degrés. On a vu sur le précédent site d'étude que les coefficients de rétrodiffusion en bande L et C pouvaient avoir une relation linéaire forte avec les paramètres forestiers, ces relations pourraient être perturbées par la topographie dans la zone dunaire.

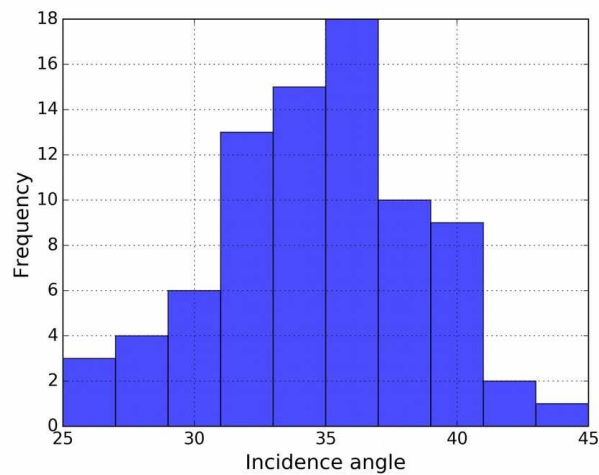
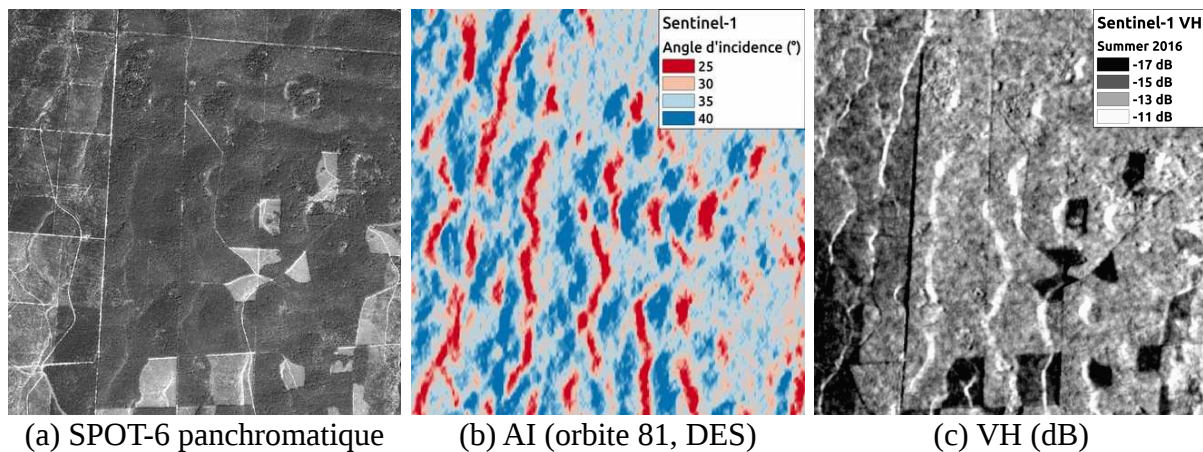


Figure III.2 : Distribution des valeurs d'angle d'incidence local (par pixel) d'une image Sentinel-1 en orbite descendante (008) sur les placettes du site Landes-ONF.

La figure III.3 montre localement l'effet de l'angle d'incidence local sur les valeurs Sentinel-1. Certaines variations de pente sont légèrement visibles sur SPOT-6 (1.5 m de résolution spatiale) avec généralement une augmentation de la réflectance globale (le sol plus brillant est davantage visible). L'angle d'incidence local peut varier rapidement entre 25 et 40°, et on voit sur l'image de droite (Sentinel-1 VH, été 2016) que ces pentes ont un effet nettement visible sur le coefficient de rétrodiffusion (plus de signal provenant des pentes à 25° qui sont face au radar).



(a) SPOT-6 panchromatique (b) AI (orbite 81, DES) (c) VH (dB)

Figure III.3 : Illustrations des angles d'incidence Sentinel-1 sur une même empreise dans la zone dunaire.

Nous avons quatre paramètres forestiers en commun entre les deux sites : surface terrière (BA), DBH, densité d'arbres (Density), et hauteur dominante (Hdom). Les distributions des valeurs sont assez proches entre les deux sites, hormis pour la densité d'arbres. La figure III.4 et le tableau III.1 montrent la distribution et les statistiques descriptives des paramètres forestiers pour les deux sites.

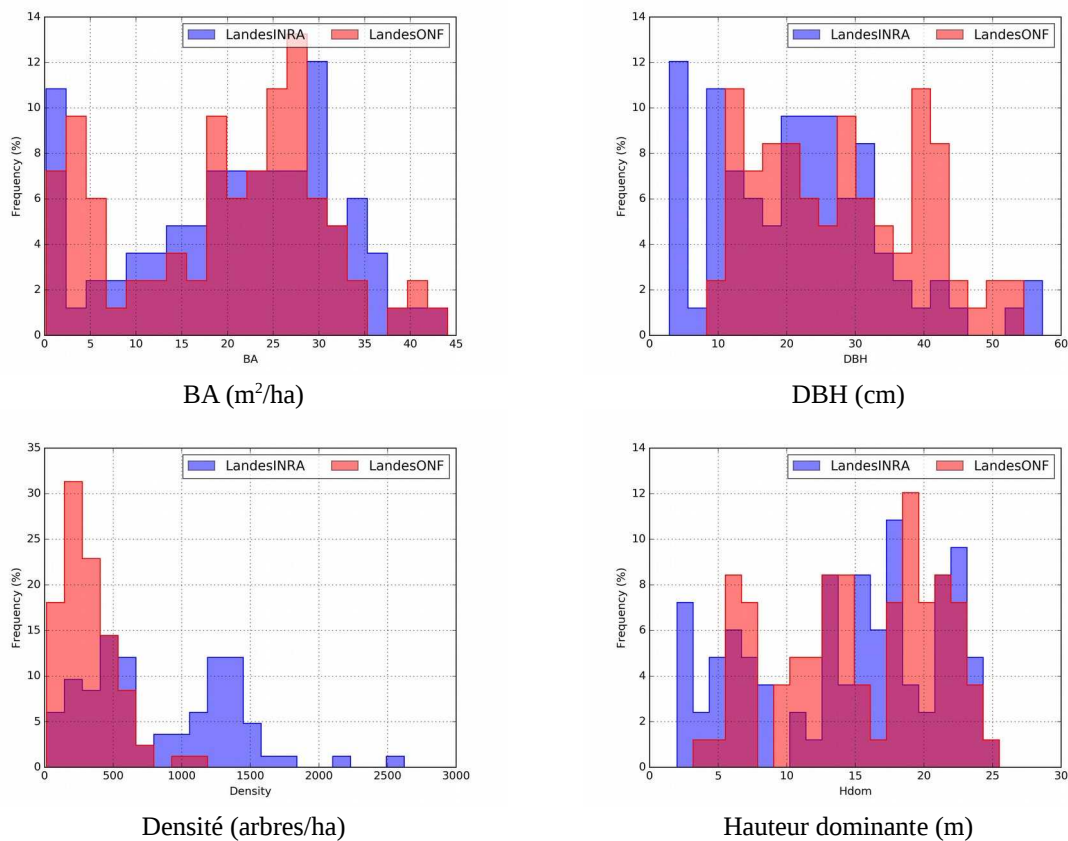


Figure III.4 : Distributions (fréquence en %) des paramètres forestiers sur les deux sites d'étude des Landes.

Les densités sont moins fortes dans les forêts de l'ONF échantillonnées, et la moyenne de DBH légèrement plus élevée. Le site Landes-INRA présente plus de DBH très faibles, correspondant aux fortes densités (peuplement très jeune), alors que le site Landes-ONF présente plus de DBH autour de 40 cm, correspondant à des peuplements presque matures.

Tableau III.1 : Statistiques descriptives des quatre paramètres forestiers sur les deux sites d'étude des Landes. Moy. : moyenne ; E. T. : écart type.

	BA (m ² /ha)		DBH (m)		Densité (arbres/ha)		Hauteur dominante (m)	
	Moy.	E. T.	Moy.	E. T.	Moy.	E. T.	Moy.	E. T.
Landes-INRA	21.16	10.94	21.52	12.36	835.42	534.31	14.36	6.74
Landes-ONF	19.56	11.1	28.11	11.68	319.08	197.31	15.31	5.69

Comme dans le cas du site Landes-INRA, nous avons digitalisé des polygones d'environ 1 hectare autour des placettes de mesure. Nous avons extrait les valeurs des 31 primitives satellitaires sur ces polygones pour regarder les relations entre primitives et paramètres forestiers. La figure III.5 montre les coefficients de détermination obtenus sur les deux sites avec les paramètres BA, DBH, Density et Hdom. Les primitives ont des relations linéaires avec les paramètres forestiers beaucoup plus faibles sur les peuplements de pin maritime du site Landes-ONF que sur ceux du site Landes-INRA. Les r^2 plafonnent entre 0.3 et 0.35. La bande L et les textures Sentinel-2 ont les meilleures corrélations avec BA. Le coefficient de rétrodiffusion et les textures de Sentinel-1 en bande C ont les meilleures corrélations avec le DBH. Les corrélations avec la densité sont très faibles, au maximum $r^2=0.2$ pour les primitives de Sentinel-2. Enfin, les corrélations avec la hauteur dominante sont mieux distribuées entre les types de primitives, mais restent faibles avec des r^2 de ~ 0.2 pour la bande L et les textures Sentinel-1, à $r^2 \sim 0.3$ pour Sentinel-1.

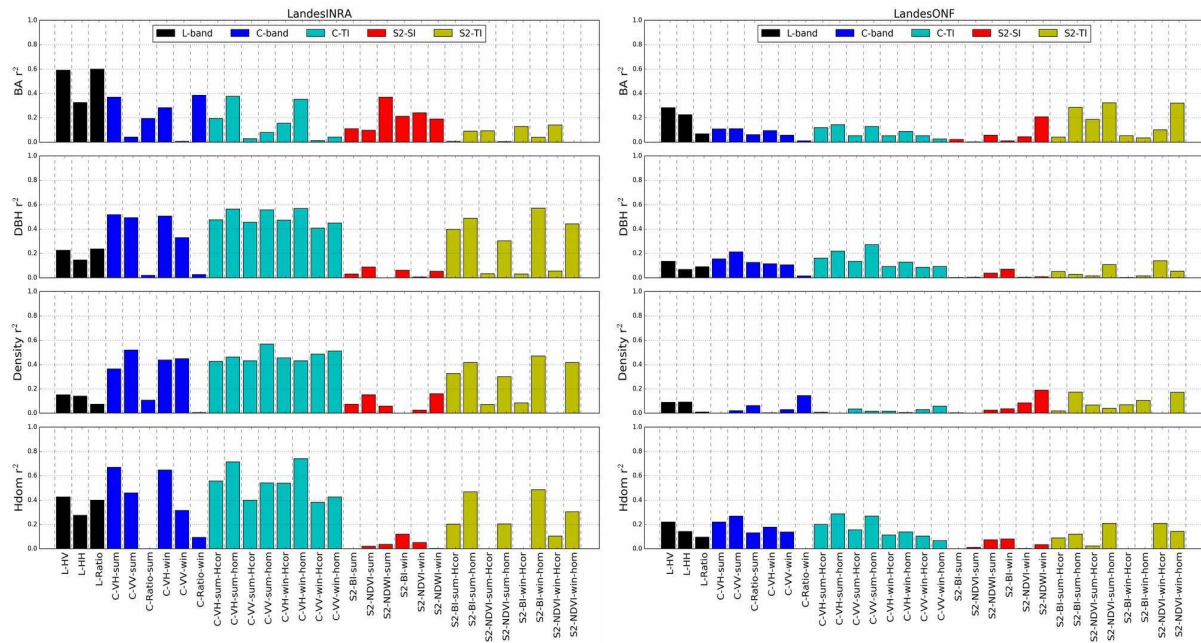


Figure III.5 : Coefficients de détermination r^2 entre les primitives et les paramètres forestiers sur les deux sites d'étude des Landes, calculés sur les polygones.

3.1.2. Estimation des paramètres forestiers

Pour analyser l'applicabilité de notre méthode à ce nouveau site de pin maritime, nous avons procédé avec quatre approches : (1) application de la méthode aux nouvelles données, (2) application des modèles obtenus sur le site Landes-INRA aux nouvelles données sur site Landes-ONF, (3) recherche des meilleurs estimations possibles en apprenant sur le site Landes-INRA et en appliquant sur le site Landes-ONF (ce qui implique une petite transformation de la méthode), et (4) application de la méthode aux deux sites ensemble, réunis comme un seul jeu de données.

1) Application de la méthode au site Landes-ONF

Dans un premier temps, nous avons donc appliqué la méthode dans son ensemble sur les nouvelles données du site ONF. Ce cas représente la meilleure des situations possibles : nous développons un modèle spécifique à partir de données disponibles in situ.

Étant donné le type de forêts, similaire au premier site (gestion uniforme à la parcelle), nous avons également digitalisé des polygones sur ce nouveau site de pin maritime Landes-ONF. Les polygones ont une taille moyenne de 1 ha et ont été photo-interprétés sur la base de photographies aériennes à 50 cm de résolution spatiale. Nous avons à nouveau vérifié que cette base plus large pour l'extraction des primitives satellitaires permettait d'obtenir de meilleures estimations. Comparé à l'utilisation des surfaces réellement mesurées (0.07 ha), l'élargissement à 1 ha améliore les r^2 d'environ 0.1 pour l'estimation de chacun des paramètres forestiers. Le RMSE relatif baisse de 3 % pour le paramètre DBH, et jusqu'à 5 % pour le paramètre BA. Ainsi, le bénéfice de travailler sur des unités de surface plus larges est confirmé sur ce deuxième site de forêts mono-spécifiques et équiennes où il est possible de digitaliser de plus grandes zones.

Le tableau III.2 montre les statistiques de validation sur ces données et les compare aux résultats précédemment obtenus sur le site Landes-INRA. Malgré les faibles corrélations observés dans la Figure III.4 pour le site ONF, on arrive à obtenir des estimations cohérentes à partir de nos 31 primitives. Les résultats d'estimation du paramètre BA sont similaires à ceux

obtenus sur le site INRA. Les trois autres paramètres forestiers sont moins bien estimés, mais là où les meilleurs r^2 entre les primitives et les paramètres forestiers vont de 0.2 à 0.3, on obtient des r^2 de 0.44 à 0.69 pour la validation des estimations de ces paramètres DBH, Density et Hdom.

Tableau III.2 : Statistiques de validation sur les sites Landes-INRA et Landes-ONF au niveau polygone (SVR, sélection Forward). La méthode est appliquée indépendamment sur chaque site.

	BA (m ² /ha)			DBH (m)			Densité (arbres/ha)			Hauteur dominante (m)		
	r^2	RMSE	(%)	r^2	RMSE	(%)	r^2	RMSE	(%)	r^2	RMSE	(%)
Landes-INRA	0.79	5.12	24.21	0.88	4.26	19.80	0.85	211.73	25.34	0.94	1.67	12.54
Landes-ONF	0.80	5.03	25.73	0.54	8.06	28.67	0.44	148.01	46.39	0.69	3.20	20.90

La figure III.6 présente les scatterplots de validation en différenciant les échantillons par leurs angles d'incidence local : les résultats sont moins bons pour le site Landes-ONF, en partie à cause d'une sous-estimation ou saturation pour les hautes valeurs des paramètres forestiers. C'est surtout le cas pour la hauteur dominante et le DBH. Cette saturation n'est pas visible sur l'estimation du paramètre BA, et la qualité des estimations de la densité d'arbres est similaire à ce qu'on trouvait pour le premier site Landes-INRA dans cet intervalle de valeurs. On peut également noter que l'estimation du DBH dominant est meilleure que celle du DBH moyen. Le paramètre DBHdom correspond plus à une valeur sur un ensemble de pixels, ce qui est plus cohérent avec notre méthode d'extraction des valeurs et pourrait donc expliquer les meilleurs résultats de l'estimation du DBH dominant.

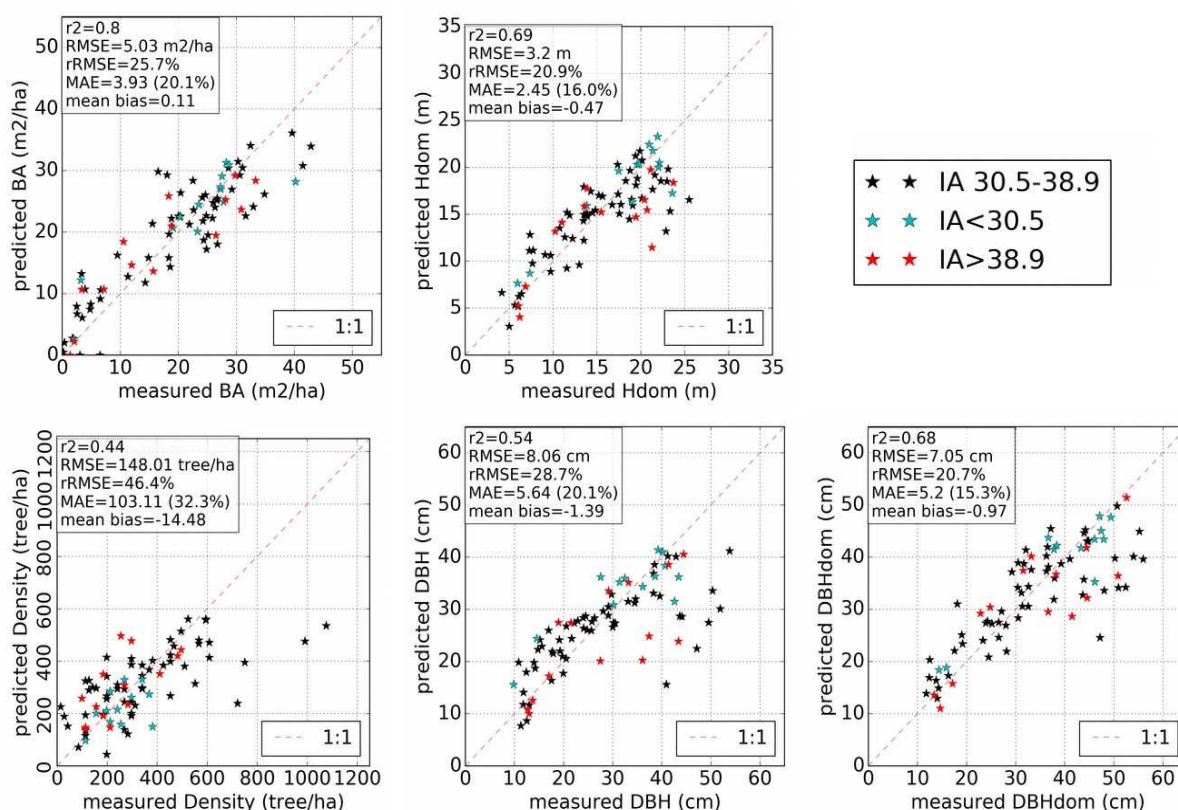


Figure III.6 : Graphiques de validation pour l'estimation des paramètres forestiers du site Landes-ONF, coloration selon les classes d'angle d'incidence local (IA).

On peut voir sur les figures que les échantillons mal estimés sont répartis entre les différents angles d'incidences, ces valeurs aberrantes ne correspondent donc pas seulement à des peuplements situés en pente (i.e. IA < 30°) ou ombre (IA > 38.9). De plus, nous avons testé de

calibrer à nouveau les modèles seulement sur les échantillons compris dans l'intervalle ± 1 écart type autour de la moyenne, soit 56 échantillons sur les 83. Les résultats de validation sont similaires entre l'utilisation de l'ensemble des échantillons ou le filtrage des échantillons en zone de pente ($\pm 2\%$ de RMSE relatif et ± 0.05 de r^2). Ainsi l'angle d'incidence local ne semble pas être directement la cause des moins bonnes statistiques. En revanche, il pourrait avoir un effet négatif sur l'établissement des relations statistiques entre les primitives SAR et les paramètres forestiers (intensité radar faussée par l'angle d'incidence).

Lorsqu'on regarde les types de primitives utilisés selon le site (tableau III.3), nous observons que la méthode sélectionne des primitives dans tous les types pour l'estimation des paramètres sur le site Landes-ONF. Mais il y a des différences dans les types de primitives sélectionnées. La bande L est nettement moins utilisée, et par contre les indices spectraux en optique sont plus utilisés. La bande C est aussi légèrement moins sélectionnée. Ainsi, le modèle issu des données Landes-INRA repose plus sur des éléments de structure de la forêt (bande C et S2-TI), tandis que le modèle issu des données ONF s'appuie plus sur des données biophysiques ou biochimiques (indices et texture optiques).

Tableau III.3 : Répartition des types de primitives sélectionnées dans les méthodes appliquées aux sites Landes-INRA et Landes-ONF. Le pourcentage est le nombre de primitive pour chaque type divisé par le nombre total de primitives sélectionnées.

		Répartition des types de primitives (%)				
		L	C	C-TI	S2-SI	S2-TI
Landes-ONF	BA	8%	17%	17%	33%	25%
	DBH	13%	20%	40%	13%	13%
	Density	12%	6%	18%	29%	35%
	Hdom	10%	10%	30%	10%	40%
	Moyenne	11%	13%	26%	22%	28%
Landes-INRA	BA	18%	36%	18%	9%	18%
	DBH	33%	11%	11%	0%	44%
	Density	27%	27%	27%	9%	9%
	Hdom	6%	19%	25%	25%	25%
	Moyenne	21%	23%	20%	11%	24%

2) Application des modèles du site Landes-INRA au site Landes-ONF

Les modèles issus de l'application de la méthode sur le site Landes-INRA ont été utilisés pour estimer les paramètres forestiers sur le site Landes-ONF. Ce cas représente la situation où un modèle créé à partir des données d'un autre site, est appliqué, en aveugle et sans modifications, sur un nouveau site.

Les résultats de validation montrent une très mauvaise qualité des estimations, avec des $r^2 < 0.1$ pour les quatre paramètres forestiers. Nous avons essayé d'analyser cette défaillance du transfert des modèles en observant les relations entre paramètres forestiers pour les deux sites. La figure III.7 montre ces inter-corrélations en séparant les deux sites.

Le DBH et la hauteur dominante (Hdom) ont une relation similaire entre les deux sites, il n'y a donc pas de différence de structure des arbres. Concernant la densité d'arbres par hectare (Density vs DBH ; Density vs Hdom), on peut voir qu'elle plafonne autour de ~600 arbres/ha pour le site ONF. Sur le site INRA, il y a une partie avec des fortes densités (1000 à 1500 arbres/ha), des hauteurs inférieures à 12 m, et un DBH < 20 cm (qui coïncide avec la première éclaircie). Et ensuite, on voit des densités inférieures à 1000 arbres/ha avec la même

dynamique de DBH et Hdom que le site ONF. Nous avons ainsi le même type d'arbres, mais avec une phase de croissance qui diffère entre les deux groupes de peuplements.

La surface terrière est une donnée complexe dans laquelle la densité d'arbres à son rôle à jouer. Dans le graphique Density vs BA, lorsqu'on regarde les densités inférieures à 600 arbres/ha et BA entre 10 et 40 m²/ha, nous retrouvons les comportements communs des deux peuplements (DBH > 20 cm, Hdom > 12 m). Pour le même BA < 10 m²/ha, il y a des arbres de l'ONF avec une densité inférieure à ceux de l'INRA, mais un DBH et une hauteur supérieure (DBH entre 10-20 cm et Hdom de 5 à 12 m par rapport à DBH entre 0 et 10 cm et Hdom < 5 m pour l'INRA). Finalement, pour BA > 10 m²/ha et densité > 1000 arbres/ha, nous retrouvons les placettes de l'INRA avec DBH entre 10 et 20 cm et Hdom entre 5 et 12 m. La forte densité de ces peuplements donne ainsi des fortes valeurs de BA.

Ces différences peuvent provenir du fait que le sol est moins fertile dans la zone dunaire comparé au site INRA, mais aussi du mode de gestion qui est principalement en semi pour le site ONF, alors qu'on trouve aussi des plantations sur le site INRA. Les plantations nécessitent des opérations d'éclaircies moins fortes, ce qui explique le fait de trouver des parcelles avec une forte valeur de surface terrière et une forte densité d'arbres par hectare.

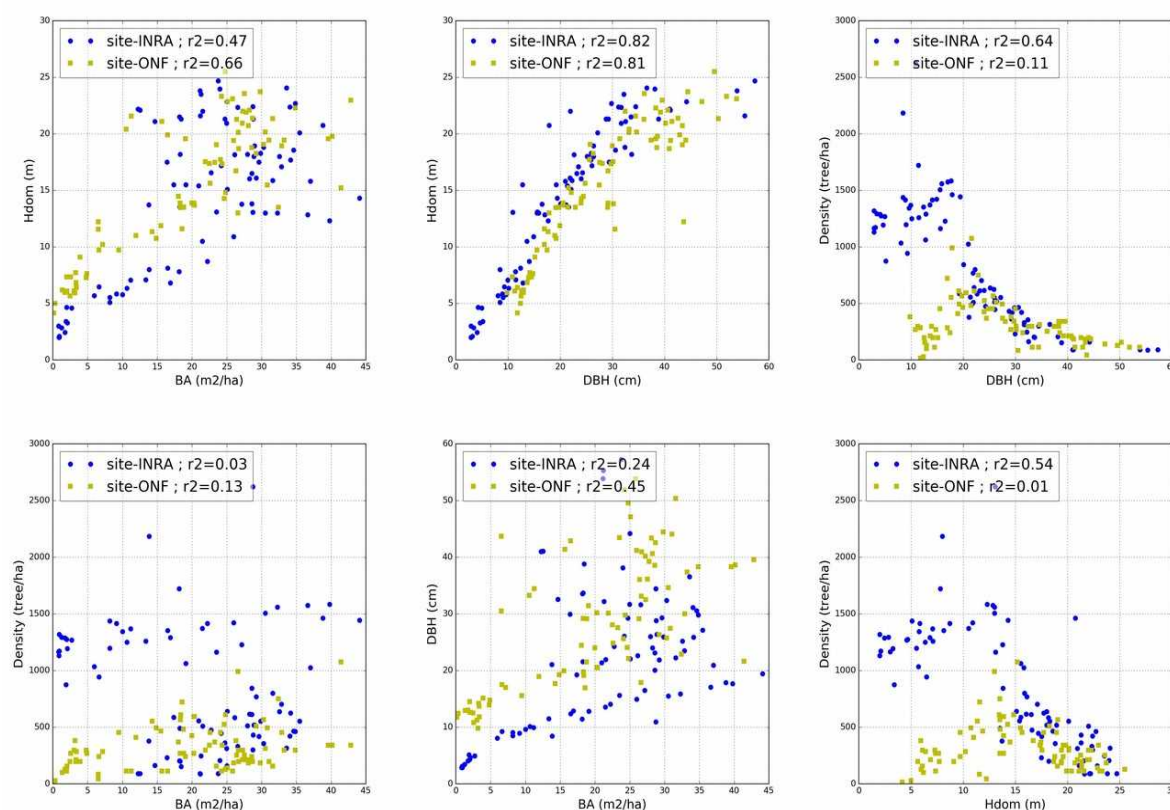


Figure III.7 : Inter-corrélations entre paramètres forestiers pour les placettes de pin maritime du site Landes-INRA et du site Landes-ONF.

Nous avons également analysé la distribution des valeurs des primitives par placettes, ainsi que les relations entre les paramètres forestiers et les primitives sur les deux sites. Nous avons choisi 6 primitives représentatives des différents types de données, et régulièrement sélectionnées dans les modèles : L-HV, C-VH, C-VH-sum-hom, C-VV-sum-hom, S2-BI-sum et S2-BI-sum-hom. La figure III.8 présente les corrélations entre les 6 primitives et la surface terrière (BA), la figure III.9 présente les mêmes corrélations avec la hauteur dominante (Hdom).

Plusieurs types de différences sont observables sur ces figures :

- Les valeurs de la primitive L-HV du site ONF sont globalement plus élevées pour des mêmes valeurs de surface terrière ou de hauteur, la différence est plus marquée sur les $BA < 15 \text{ m}^2/\text{ha}$. Ce phénomène est encore plus marqué avec L-HH (non montré ici). Nous avons vu que pour $BA < 15 \text{ m}^2/\text{ha}$, les peuplements du site ONF étaient composés par arbres plus grands et moins nombreux comparés avec une forte densité et des arbres plus petits sur le site INRA. La bande L étant sensible au tronc, le signal est donc plus fort pour le site ONF. Par ailleurs, la densité est inférieure et donc le sol plus visible, il peut y avoir un retour sol ou une interaction sol-tronc plus élevée aussi. Le même phénomène est observé pour S2-BI-sum qui représente l'intensité globale de la réflectance dans le visible en été. La différence entre les deux sites tend à se réduire pour les fortes valeurs de hauteur et de surface terrière. Les densités d'arbres dans les jeunes peuplements sont plus faibles sur la zone dunaire (site ONF), la réflectance du sol ou du sous-bois est donc plus importante dans les pixels Sentinel-2. De plus, le type de sol dans cette zone est du sable fin (forte réflectance), recouvert parfois par un sous-bois important de genêts (fleurs jaunes avec forte réflectance).
- Ensuite, quand on regarde les valeurs pour les primitives Sentinel-1 (coefficient de rétrodiffusion et indice de texture), les différences sont moins marquées pour la hauteur et la surface terrière. On observe cependant que les relations entre paramètres forestiers et données Sentinel-1 sont plus dispersées dans le cas du site ONF, particulièrement pour la hauteur. On retrouve toujours le deuxième nuage de points avec les fortes densités du site INRA. Les relations avec C-VV sont légèrement différentes (figure non montrée ici), avec une baisse globale du signal dans le cas du site ONF comparé au site INRA. Ce phénomène s'explique encore par les propriétés du sol et du sous-bois : plus de sable et une plus forte couverture (genêts principalement) au sol dans la zone dunaire.
- Enfin, on peut voir une plus forte différence dans les relations entre les paramètres forestiers et un indice de texture en optique (S2-BI-sum-hom, homogénéité sur l'indice de brillance en été). Il est difficile de parler de tendance avec cette dispersion, mais il semble que les valeurs de l'indice baissent avec l'augmentation de la hauteur et de la surface terrière dans le cas du site INRA. Cette relation pourrait être expliquée par le fait que les jeunes peuplements sont très homogènes et couvrants, tandis que les peuplements matures sont constitués d'arbres plus dispersés, sélectionnés selon les formes du tronc et donc potentiellement répartis de manière plus hétérogène dans les parcelles. La relation entre S2-BI-sum-hom et les paramètres forestiers semble inversée dans le cas du site ONF dans la zone dunaire. En effet, lorsqu'on regarde les images à haute résolution des peuplements matures, on constate qu'ils sont plus couvrants et plus homogènes que ceux du site INRA dans les terres. Cette différence pourrait s'expliquer par une valorisation différente dans la filière bois : une sélection des tiges de haute qualité dans les terres pour avoir quelques éléments de bois d'œuvre, tandis que les parcelles de l'ONF pourraient être davantage destinées au bois-industrie ou au bois-énergie.

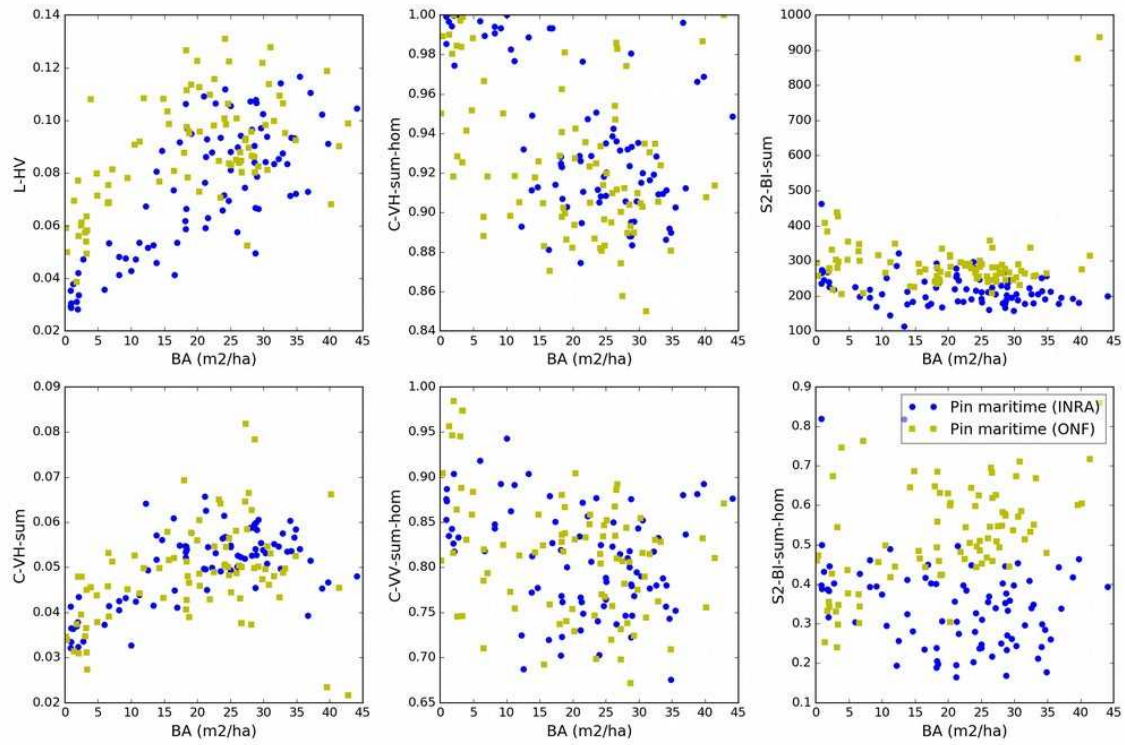


Figure III.8 : Corrélations entre les primitives et la surface terrière pour les polygones de pin maritime du site Landes-INRA et du site Landes-ONF.

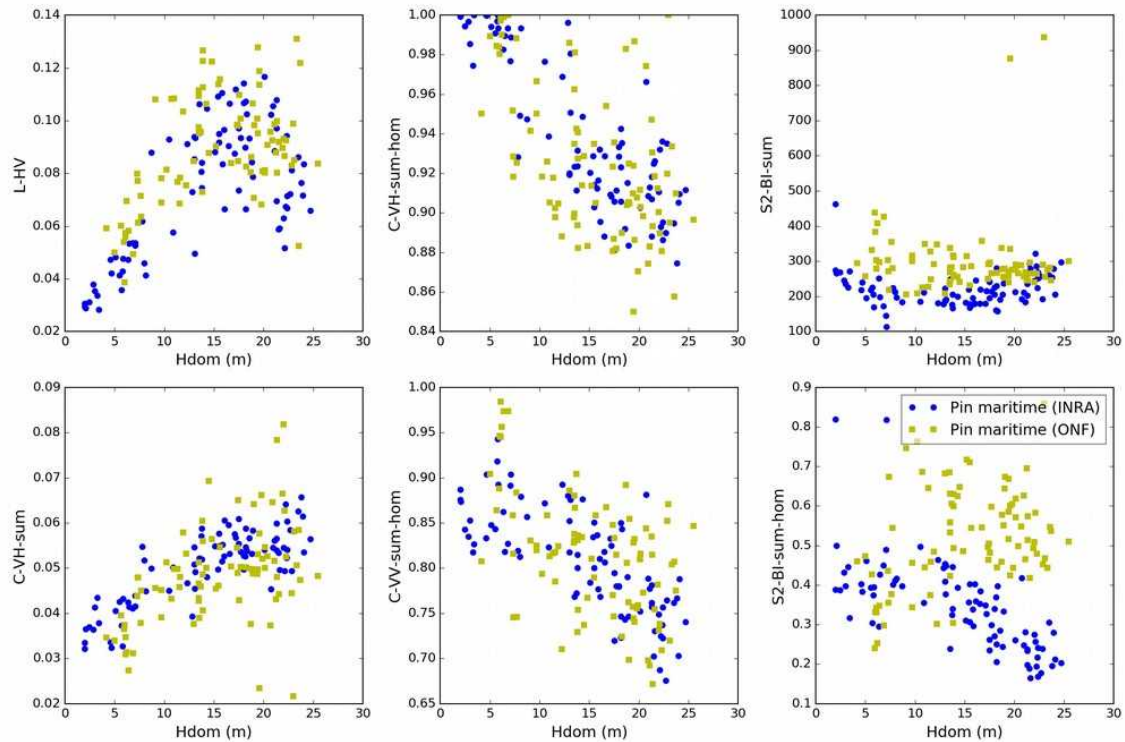


Figure III.9 : Corrélations entre les primitives et la hauteur dominante pour les polygones de pin maritime du site Landes-INRA et du site Landes-ONF.

Pour compléter ces observations, la figure III.10 montre les histogrammes de distribution de ces mêmes 6 primitives satellitaires avec l'ajout du S2-BI-win-hom (homogénéité de l'index de brillance en hiver). On voit en effet des différences sur L-HV, S2-BI-sum et S2-BI-sum-hom. Les plus basses valeurs observées sur le site INRA ne sont pas présentes sur le site ONF, et inversement pour les fortes valeurs pour ces trois primitives satellitaires. L'homogénéité de

l'index de brillance en hiver (S2-BI-win-hom) ne présente pas de grosses différences contrairement aux écarts observés en été (S2-BI-sum-hom), ce qui renforce notre hypothèse de l'influence du sous-bois, notamment la floraison du genêt dans le cas des Landes-ONF. Dans le cas des primitives Sentinel-1, les dynamiques de valeurs sont plus similaires entre les trois primitives.

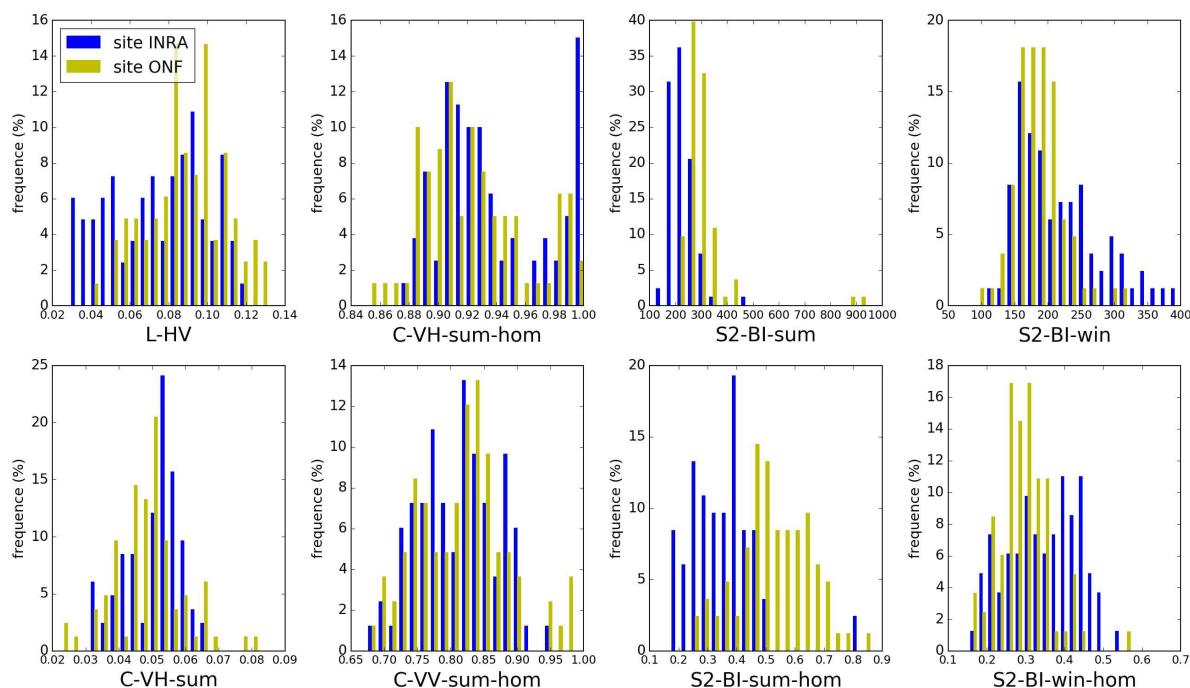


Figure III.10 : Distribution des valeurs des primitives satellitaires au niveau polygone (moyennes) sur les sites Landes-INRA et Landes-ONF.

Nous confirmons par les analyses de cette section ce que nous avons déjà détecté dans la section précédente : les deux sites ont des pratiques de sylviculture différentes qui impactent la densité, la gestion du sous-bois et l'homogénéité des peuplements matures. Les données satellitaires les mieux adaptés à décrire chaque site sont légèrement différentes. D'où la nécessité de garder l'ensemble des primitives et une méthode automatique pour notre chaîne de traitement.

3) Meilleurs résultats possibles du site Landes-INRA vers le site Landes-ONF

Les analyses précédentes montrent que les deux sites n'ont pas les mêmes patterns de relations entre primitives satellitaires et paramètres forestiers, ce qui explique la difficulté du transfert de modèle entre les deux ensembles de peuplements de pin maritime.

Pour terminer ces tests sur la transposition de la méthode, nous avons légèrement modifié la méthode de paramétrisation du SVR et de sélection des primitives satellitaires. L'objectif est de voir quels sont les meilleurs résultats qu'il serait possible d'avoir avec notre méthode en apprenant le modèle sur le site Landes-INRA pour l'appliquer sur le site Landes-ONF. De cette manière, nous pouvons constater si le problème de transposition est causé plus par des différences réelles entre les deux sites ou s'il vient de la méthode utilisée (problème de sur-apprentissage par exemple).

Les échantillons d'apprentissage sont constitués de l'ensemble des échantillons du site Landes-INRA. Les échantillons de validation sont constitués de l'ensemble des échantillons du site Landes-ONF, ils servent à sélectionner les paramètres du SVR ainsi que les primitives à ajouter dans la méthode de sélection Forward. De cette manière, nous mettons en place le

modèle le plus performant possible pour estimer les paramètres du site Landes-ONF avec les échantillons du site Landes-INRA dans le cadre de notre méthodologie. Les résultats sont présentés dans le tableau III.5, nous avons appliqué la méthode à l'ensemble des échantillons Landes-ONF dans un premier temps (lignes « Total »), puis seulement aux échantillons contenus dans l'intervalle +/- 1 écart type autour de la moyenne des angles d'incidence (lignes « IA-std ») pour évaluer l'impact de l'angle d'incidence local. Les statistiques présentées pour le site Landes-INRA sont issues d'une validation croisée leave-one-out à partir de la configuration sélectionnée par la méthode.

Tableau III.4 : Statistiques de validation de la méthode avec apprentissage sur les références du site Landes-INRA et validation sur le site Landes-ONF pour l'optimisation des paramètres du SVR et la sélection des primitives satellitaires.

		BA (m ² /ha)		DBH (cm)		Densité (arbres/ha)		Hdom (m)	
		r ²	RRMSE	r ²	RRMSE	r ²	RRMSE	r ²	RRMSE
Total	Landes-INRA	0.55	35%	0.71	32%	0.48	47%	0.86	18%
	Landes-ONF	0.58	38%	0.26	36%	0.13	60%	0.38	29%
IA-std	Landes-INRA	0.66	30%	0.72	31%	0.63	40%	0.86	18%
	Landes-ONF	0.60	37%	0.42	34%	0.20	66%	0.39	29%

Ces résultats montrent qu'il est possible d'améliorer la généralisation d'un site à l'autre en modifiant la méthode pour une optimisation adaptée des paramètres du SVR et de la sélection des primitives. En comparaison de lorsqu'on applique au site Landes-ONF les modèles issus de la méthode optimisée sur le site Landes-INRA (section précédente), les coefficients de détermination r² passent de 0.0-0.08 à 0.13-0.58 avec cette nouvelle approche.

Ces résultats sont censés représenter la meilleure généralisation possible avec notre méthode, cependant ils sont de bien moins bonne qualité que ceux obtenus lorsqu'on applique la méthode indépendamment sur chaque site, ou directement sur les deux sites réunis (tableau III.5, paragraphe suivant). Les primitives sélectionnées dans cette approche d'un site à l'autre ressemblent à celles sélectionnées dans le cas d'une application directe de la méthode au site Landes-ONF, ce qui explique la perte de performance pour les données Landes-INRA dans cette configuration.

4) Application de la méthode aux deux sites réunis

Nous avons vu dans le Chapitre II (section 2.8.3) que l'augmentation du nombre d'échantillons d'apprentissage est théoriquement bénéfique pour la précision des estimations des paramètres forestiers. Avec ce deuxième site, la quantité d'échantillon est doublée pour ces peuplements de pins maritimes. Si on considère les courbes montrées dans la figure II.27 (évolution des performances selon le nombre d'échantillons d'apprentissage et de validation), les performances pourraient être améliorées de 5 à 10 % de RMSE relatif en extrapolant les courbes de la validation croisée LOO.

Ainsi, nous avons rassemblé les deux jeux de données en un seul puis nous avons à nouveau appliqué la méthode sur ces données. Les résultats de validation sont présentés dans le tableau III.4 avec le résultat global et la différenciation des statistiques sur chacun de deux sites. En comparant les résultats des tableaux III.4 et III.5, nous observons que l'estimation des deux sites réunis (partie 4) donne de meilleurs résultats que l'estimation d'un site sur l'autre (parties 2 et 3), mais légèrement moins bons que l'estimation sur chaque site séparément (partie 1) (notamment pour les paramètres DBH et Density). Les résultats pour les paramètres BA et Hdom sont assez similaires à ceux obtenus sur les deux sites séparément. Deux interprétations de ce résultat sont possibles : i) à partir d'un certain seuil, l'augmentation du nombre

d'échantillons n'améliore pas significativement les résultats, ou ii) les deux sites, bien qu'avec la même essence forestière et dans la même zone géographique, ne peuvent finalement pas être assimilés à un seul type de forêts par rapport aux relations entre primitives satellitaires et paramètres forestiers. Les différences constatées dans les relations linéaires entre primitives et paramètres forestiers dans la figure III.8 à 10 suggèrent plutôt que la deuxième interprétation serait la bonne. Ainsi, le SVR ne bénéficie pas de l'augmentation du nombre d'échantillons et les deux sites sont à considérer comme des sous-groupes de ce type de forêts et non un seul et même type de forêt.

Il est important de noter que la combinaison des deux sites dans une même chaîne de traitement ne détériore que faiblement la qualité des estimations (par rapport aux estimations individuelles) et peut-être utilisée quand la connaissance du terrain n'est pas exhaustive et ne permet pas de séparer plus précisément les types de forêts.

Tableau III.5 : Statistiques de validation des estimations à partir d'un jeu de données rassemblant les deux sites. Les lignes Landes-INRA et Landes-ONF correspondent aux statistiques différenciées obtenus par site avec le modèle combinant les deux sites.

	BA (m ² /ha)			DBH (m)			Densité (arbres/ha)			Hdom (m)		
	r ²	RMSE	(%)	r ²	RMSE	(%)	r ²	RMSE	(%)	r ²	RMSE	(%)
Total 2 sites	0.77	5.34	26.03	0.66	7.30	30.46	0.78	225.51	39.61	0.80	2.78	19.04
Landes-INRA	0.75	5.5	25.37	0.77	5.91	28.15	0.79	252.92	31.56	0.90	2.14	14.84
Landes-ONF	0.79	5.18	26.74	0.49	8.46	31.41	0.14	194.26	57.60	0.67	3.30	22.32

Conclusions sur cette expérience

Cette série de test illustre les difficultés à travailler sur des forêts tempérés dans les zones où les conditions environnementales et les types de gestion forestière peuvent changer rapidement d'un peuplement à l'autre. Nous avons travaillé sur deux sites proches géographiquement avec des peuplements de pin maritime. Le site Landes-INRA est situé dans l'intérieur de la forêt des Landes, tandis que le site Landes-ONF est situé dans la zone dunaire proche du littoral. Ainsi, la topographie, le type de sol et la gestion peuvent être différents entre les deux sites d'étude.

L'application de la méthode mise au point dans le Chapitre II donne des résultats très bons sur le site Landes-INRA à l'intérieur des terres, et légèrement moins bons sur le site Landes-ONF de la zone dunaire. L'estimation de la surface terrière, généralement fortement lié à la biomasse des forêts, est assez précise sur les deux sites avec $r^2 \sim 0.8$ et 25 % de RMSE relatif. Cependant, sur ces jeux de données il est difficile de transférer les modèles d'un site à l'autre, alors que l'essence forestière et le climat sont les mêmes. Ainsi, le type de sol, la gestion différente et la topographie (dans le sens où elle déforme les relations entre paramètres forestiers et données satellitaires) nous forcent à considérer ces deux jeux de données comme étant des types de forêts différents. Néanmoins, les forêts sont suffisamment proches pour que la combinaison des deux sites dans la méthode donne des résultats qui restent stables pour l'estimation de la surface terrière et la hauteur notamment, comparé à l'application de la méthode sur chaque site individuellement. Cela suggère que même si les modèles sont difficilement généralisables entre ces deux sites, un modèle SVR construit sur les deux jeux de données ensemble sera capable de donner des résultats presque aussi bons que sur chaque site différencié. Il est cependant probable que cette solution ne soit pas utilisable sur l'ensemble des types de forêts, mais que certains regroupements soient possibles. Une solution de nomenclature à plusieurs niveaux serait possible pour l'estimation des paramètres forestiers sur de grande surfaces lorsqu'on est pas sûr avec précision des types de forêts au sol.

3.2. Application à plusieurs essences forestières

3.2.1. Présentation des nouveaux sites d'étude

Deux nouveaux sites d'étude sont ajoutés aux sites de pin maritime des Landes. La figure III.11 présente la localisation des sites et des placettes de mesures. Comme dans le cas du site Landes-ONF, ces deux nouveaux jeux de données ont été acquis et fournis par l'ONF.

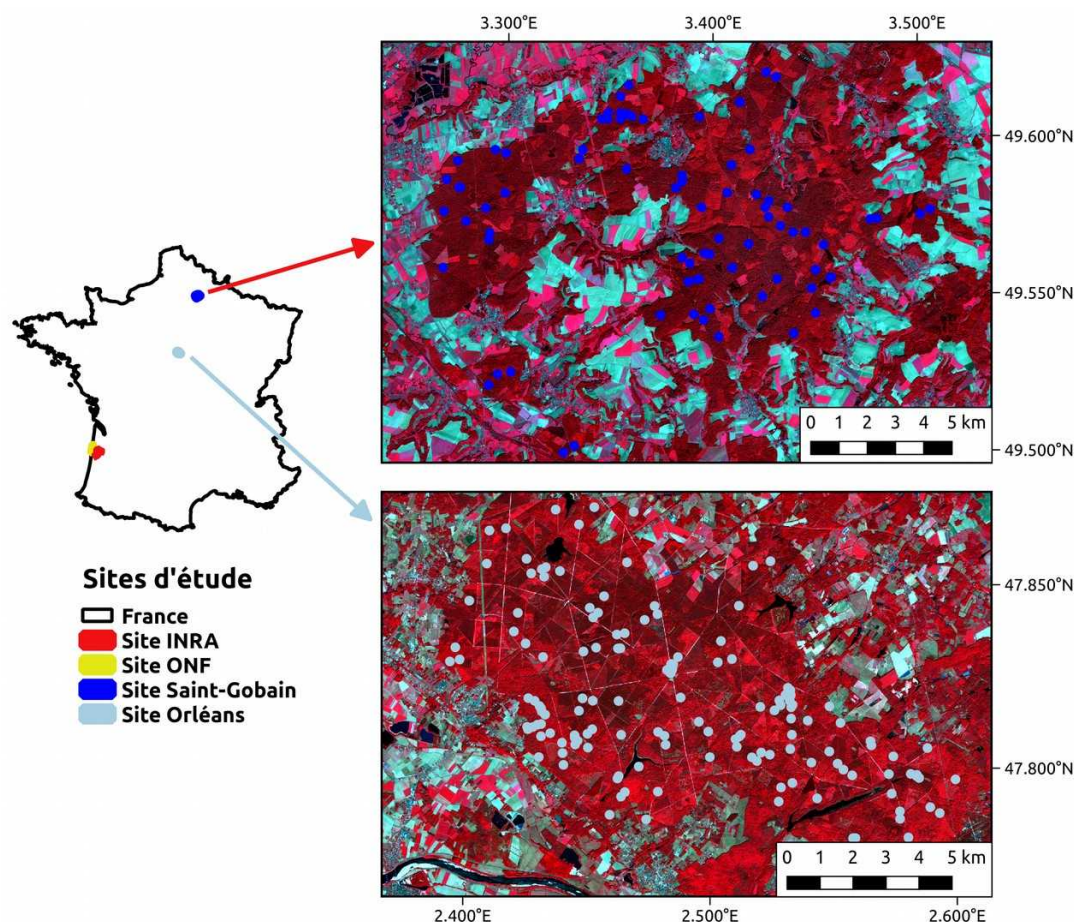


Figure III.11 : Localisation des quatre sites d'étude et visualisation des nouveaux sites dans les forêts d'Orléans et de Saint Gobain sur des images optiques en fausse couleur (PIR-R-V). Les points colorés représentent les placettes de mesure.

Nous avons calculé à la placette le DBH moyen et dominant, la densité d'arbres, la surface terrière (BA). La hauteur dominante nous a été donnée par l'ONF qui l'a calculé à l'aide de données Lidar et de mesures sur un sous-échantillon. Les mesures ont été faites sur des placettes de 15 à 18 mètres de rayon (0.07 à 0.1 ha), avec un seuil de pré-comptage de 17 cm (les arbres avec un DBH inférieur à 17 cm ne sont pas comptabilisés). Ce seuil de pré-comptage peut avoir une incidence forte sur les valeurs des paramètres forestiers compilés à la placette (Annexe A).

Forêt de Saint Gobain

La forêt de Saint Gobain (identifiée StGobain par la suite) est située à ~150 km au Nord Est de Paris (Figure III.10). Le sol est riche et il y a un peu de relief. Les essences majoritaires sont le chêne, le hêtre et le charme, mais on recense 17 essences présentes dans les 180 placettes de mesure, dont 11 qui sont majoritaires en surface terrière dans au moins une placette. Il s'agit d'une forêt diversifiée, avec des arbres d'âges et de diamètres différents à l'intérieur de chaque placette.

Afin de pouvoir concentrer nos modèles sur certains types de forêts, nous avons trié les placettes par essence ou par groupe d'essences. Les chênes sessile et pédonculés ont été regroupés en une seule essence de chêne. Les chênes verts, chênes rouge et chêne liège, si présent, ne sont pas intégrés à la classe « chêne » car trop différents d'un point de vue de structure forestière et des arbres. Trois groupes peuvent être extraits avec suffisamment de placettes pour être statistiquement représentatifs (>20 placettes). Il s'agit de 78 placettes, dont 28 placettes ont plus de 66 % de la surface terrière en chêne, 25 placettes sont des mélanges de chênes et de charmes (en surface terrière, plus de 25 % de chaque essence et moins de 66 % de chêne), et 25 placettes ont plus de 66 % de la surface terrière en hêtre. Chacune des trois classes a une variabilité des paramètres forestiers suffisante pour être intégrées dans des modèles statistiques, la classe chêne-charme a une variabilité un peu plus faible (tableau III.6). Les 102 placettes restantes ne peuvent pas rentrer dans des groupes statistiquement représentatifs. En dehors des trois essences principales (chêne, hêtre et charme) on trouve notamment le frêne, le châtaigner, le tilleul, l'érable, l'aulne, le bouleau et le merisier.

Forêt de Les Bordes (Orléans)

Le deuxième site est situé dans la forêt d'Orléans, à ~100 km au Sud de Paris (Figure III.11). Sur cette zone, la topographie est plate, et les sols sont plus pauvres et plus acides que dans la forêt de StGobain. D'après les chercheurs de l'ONF qui nous ont fourni les données, les forêts du site d'Orléans sont moins développées que celles de StGobain, avec une gestion parfois en taillis pour les feuillus, ce qui peut amener à une sous-estimation des paramètres forestiers lors des mesures avec un seuil de pré-comptage de 17 cm de DBH. Les essences majoritaires sont le chêne et le pin sylvestre. On trouve une dizaine d'essences forestières présentes sur les placettes (8 essences identifiées, plus une classe de feuillus et une classe de conifères non identifiés). De la même manière que pour StGobain, nous avons extrait deux classes d'essences représentants 88 placettes : 25 placettes ont plus de 66 % de pin sylvestre en surface terrière et 63 placettes ont plus de 66 % de chênes avec moins de 10 % de conifères. La dynamique des valeurs de paramètres forestiers est à peu près similaire au site de StGobain, mais les hauteurs dominantes sont globalement plus faibles. Les 115 placettes restantes ne peuvent pas rentrer dans des groupes statistiquement représentatifs.

Surface représentative pour l'extraction des primitives

Nous avons précédemment utilisé des polygones d'une surface minimale d'un hectare autour des placettes de mesure, afin de calculer des valeurs de primitives satellitaires sur la base de la moyenne des pixels (voir section 2.6). Cette procédure était possible en raison de la gestion homogène des parcelles de pin maritime. Les forêts des sites d'étude de St-Gobain et d'Orléans sont beaucoup plus hétérogènes que les peuplements de pin maritime des deux sites des Landes. L'observation des placettes sur des images aérienne optique à 0.5 m de résolution spatiale montre des hétérogénéités très rapidement autour des placettes, avec en plus le mélange d'essences (plus difficilement photo-interprétable) qui peut introduire davantage de variabilité. Ainsi, les placettes ne sont vraiment représentatives que de leur propre surface. L'élargissement de cette surface, pour extraire les primitives satellitaires, risquerait d'introduire des erreurs quant aux paramètres forestiers mesurés par rapport à la surface à l'extérieur des placettes. En conséquence, pour les sites d'étude de St-Gobain et d'Orléans, nous avons extrait les valeurs de primitives satellitaires sur les surfaces des placettes uniquement.

3.2.2. Analyse des jeux de données

Distribution des paramètres forestiers

Tableau III.6 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers pour l'ensemble des classes de forêts sur les 4 sites d'étude.

Plots	Classes de forêts	Surface terrière BA (m ² /ha)		DBH (cm)		Hauteur dom. (m)		Densité (arbre/ha)	
		Moy.	E. T.	Moy.	E. T.	Moy.	E. T.	Moy.	E. T.
83	Pin maritime (Landes-INRA)	21.2	10.9	21.5	12.4	14.4	6.7	835	538
83	Pin maritime (Landes-ONF)	19.6	11.1	28.1	11.7	14.3	5.7	319	199
63	Chêne (Orléans)	17.4	5.9	37.2	12.4	22.1	3.6	174	89
25	Pin Sylvestre (Orléans)	22.6	6.9	34.4	8.2	20.3	3.9	238	135
115	Autres (Orléans)	20.3	6.8	33.8	8.3	20.8	3.1	228	97
28	Chêne (StGobain)	22.9	9.3	38.1	7.0	26.9	3.5	166	69
25	Hêtre (StGobain)	18.4	9.4	37.1	9.2	27.1	6.3	149	73
25	Chêne-charme (StGobain)	20.3	6.2	33.5	3.5	27.2	1.7	201	64
102	Autres (StGobain)	19.3	8.7	34.7	7.5	26.0	3.9	188	80

Les différentes essences forestières ont des surfaces terrières (BA) moyennes similaires, avec une variance un peu plus faible pour les trois classes du site d'Orléans ainsi que pour la classe mixte chêne-charme de StGobain. Il y a plus de différences pour les autres paramètres forestiers. Le DBH moyen est plus élevé sur les sites d'Orléans et de StGobain, et sa variance globalement plus faible. On observe la même dynamique pour la hauteur dominante, ces deux paramètres ont donc une représentativité statistique moins bonne pour la construction des modèles de régression. La densité d'arbres est plus faible sur les sites d'Orléans et de StGobain.

La figure III.12 présente les distributions du DBH moyen, de la hauteur dominante et de la surface terrière sur les différentes classes de forêts. La dynamique de ces deux paramètres est meilleure sur les pins maritimes (plus large répartition des valeurs). Les classes des sites d'Orléans et de StGobain ont plus de forts DBH (dans la plage 30-50 cm), et les mesures du site StGobain ont des hauteurs dominantes principalement réparties sur les fortes valeurs (dans la plage 20-35 m). Les distributions du paramètre BA ont des plages de valeurs plus similaires entre les classes de forêt.

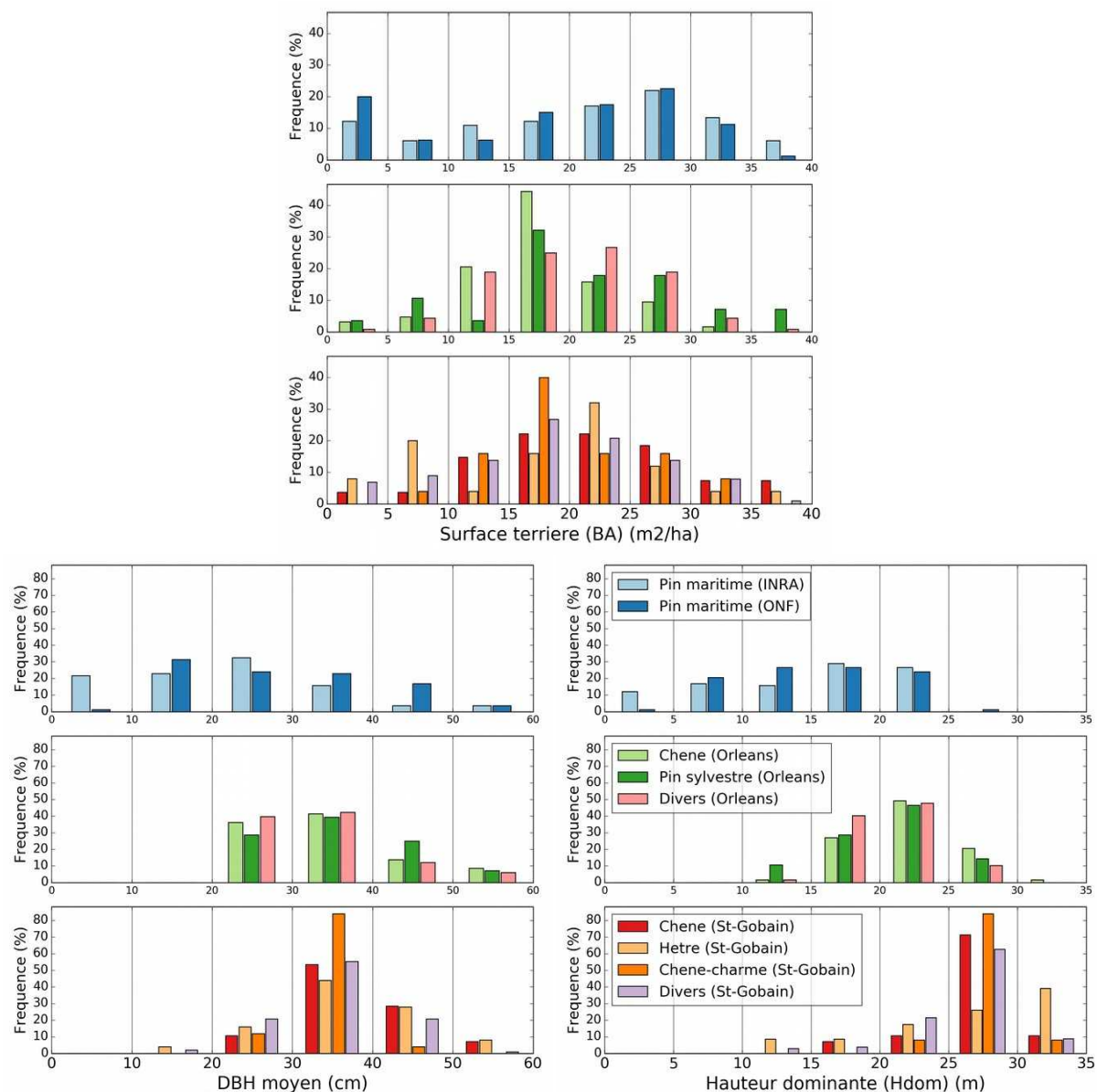


Figure III.12 : Histogrammes des distributions du DBH moyen, de la hauteur dominante et de la surface terrière des placettes de référence selon les sites et les classes de forêt.

Les données de mesure des sites de l'ONF nous ont été fournies au niveau arbre. Nous avons calculé la variabilité des diamètres des arbres à l'intérieur de chaque placette, la figure III.13 présente la dynamique de ces valeurs pour chacune de nos classes de forêt. Cette variabilité des diamètres peut être reliée à l'hétérogénéité de la forêt. On observe que les pins maritimes ont dans l'ensemble une faible variabilité du DBH, car ce sont des peuplements équiennes (tous les arbres ont le même âge, par conséquent les diamètres sont assez similaires). La variation des diamètres des arbres est plus forte dans le site de StGobain, encore plus pour la classe chêne, reflétant une plus grande diversité des âges et en plus des essences. Sur le site d'Orléans, les pins sylvestres sont assez réguliers avec une faible variabilité du DBH. Les chênes d'Orléans ont une variabilité plus importante, mais sont surtout très irréguliers avec dans la même classe des placettes très homogènes et d'autres qui ont des écarts types de 25-30 cm pour le diamètre des arbres. Cette hétérogénéité dans les placettes et entre les placettes risque de compliquer la construction d'un modèle statistique pour l'estimation des paramètres forestiers sur ces chênes d'Orléans.

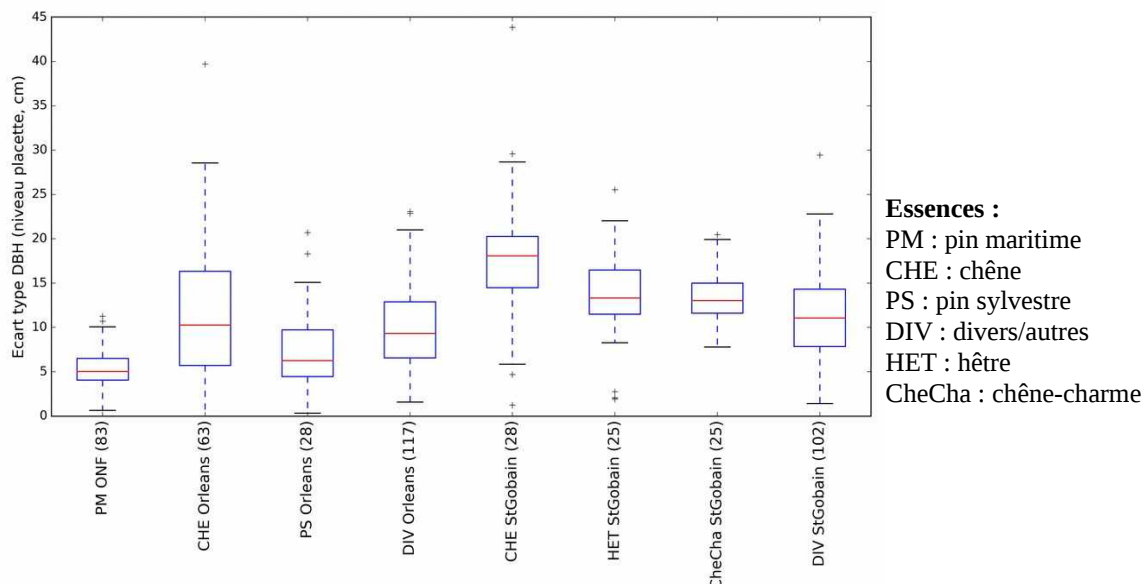


Figure III.13 : Variabilité intra-placettes des mesures de DBH des arbres. Les boîtes à moustaches montrent la médiane et les quartiles de chaque classe de forêt.

Relations entre paramètres forestiers et données satellitaires

Les sites StGobain et Orléans présentent des relations très variables entre les paramètres forestiers et les 31 primitives satellitaires. La figure III.14 présente les relations linéaires (r^2) entre les paramètres forestiers et les données satellitaires sur les classes chêne et pin sylvestre du site d'Orléans. Les mesures sur les placettes de chêne sont très peu corrélées linéairement avec les primitives (le r^2 ne dépasse pas 0.25). Les mesures sur les placettes de pin sylvestre sont beaucoup plus corrélées avec les primitives, mais avec un pattern encore différent de ceux observés sur les deux sites de pin maritimes (les indices spectraux montrent des relations significatives avec le DBH, la densité et la hauteur dominante). Concernant le site StGobain, avec uniquement des feuillus, les relations sont très faibles comme dans le cas des chênes sur Orléans. Peu de primitives ont des relations linéaires significatives avec les paramètres forestiers, et ces primitives peuvent être différentes pour chaque essence.

Ainsi, ces nouvelles observations montrent encore la nécessité d'utiliser une approche avec des régressions multiples et non linéaires, afin de combiner l'ensemble des données satellitaires pour caractériser les forêts.

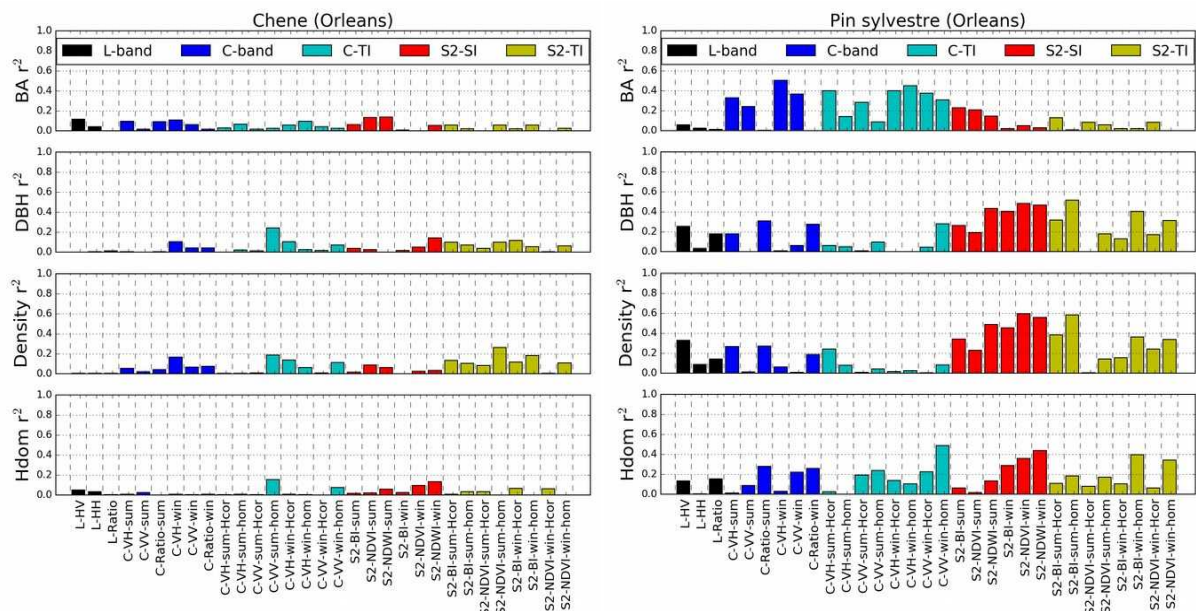


Figure III.14 : Coefficients de détermination r^2 entre les paramètres forestiers et les données satellitaires pour le site d'Orléans (classes chêne et pin sylvestre).

3.2.3. Application de la méthode

Nous avons appliqué la chaîne de traitement à ces 7 nouvelles classes de forêt. Dans un premier temps, la méthode est appliquée individuellement sur chaque classe de forêt, avec un SVR et une méthode de sélection Forward sur les 31 primitives issues des données satellitaires. Les résultats de ces 7 classes, ainsi qu'un rappel des résultats obtenus avec les deux sites de pin maritime, sont présentés dans le tableau III.7.

Tableau III.7 : Statistiques de validation des estimations de BA, DBH, Hdom et Densité pour l'ensemble des classes de forêts des 4 sites. Le pourcentage représente le RMSE relatif.

Classes de forêts	Surface terrière BA (m²/ha)			DBH (cm)			Hauteur dom. (m)			Densité (arbre/ha)		
	r^2	RMSE	%	r^2	RMSE	%	r^2	RMSE	%	r^2	RMSE	%
Pin maritime (Landes-INRA)	0.79	5.12	24 %	0.88	4.26	20 %	0.94	1.67	13 %	0.85	211.73	25 %
Pin maritime (Landes-ONF)	0.80	5.03	26 %	0.54	8.06	29 %	0.69	3.2	21 %	0.44	148.01	46 %
Chêne (Orléans)	0.43	4.5	26 %	0.61	7.76	21 %	0.52	2.49	11 %	0.62	54.97	32 %
Pin Sylvestre (Orléans)	0.77	4.13	20 %	0.74	4.67	13 %	0.85	1.59	8 %	0.74	68.10	29 %
Autres (Orléans)	0.47	4.93	24 %	0.42	6.32	19 %	0.36	2.47	12 %	0.57	63.36	28 %
Chêne (StGobain)	0.64	5.88	26 %	0.73	3.66	10 %	0.86	1.34	5 %	0.54	46.30	28 %
Hêtre (StGobain)	0.52	5.58	36 %	0.67	5.35	14 %	0.80	2.99	11 %	0.58	48.50	33 %
Chêne-charme (StGobain)	0.80	2.82	14 %	0.82	1.52	5 %	0.59	1.09	4 %	0.74	32.47	16 %
Autres (StGobain)	0.31	7.21	37 %	0.23	6.53	19 %	0.60	2.50	10 %	0.32	65.87	35 %

Les estimations sont globalement bonnes, même si on observe quelques disparités de qualité entre les différentes forêts et les différents paramètres forestiers. Les meilleurs résultats en termes de r^2 sont obtenus avec le premier site d'étude sur les pins maritimes dans les Landes. En effet, ce site est le plus adapté à ces méthodes de part la régularité des peuplements, la variance des paramètres mesurés, et l'absence totale de relief. Les résultats de validation des placettes qui font partie des classes discriminées sont meilleurs comparé aux autres (classes « Autres »), soulignant l'importance de traiter séparément les différents types de forêts lorsque c'est possible. Les estimations sont globalement meilleures sur les chênes de St-Gobain que sur les chênes d'Orléans. Cette différence pourrait s'expliquer d'une part par une hétérogénéité

plus forte sur Orléans (gestions à la fois en taillis et en futaie, contre futaie uniquement pour St-Gobain), et le seuil de pré-comptage à 17 cm qui peut impacter fortement les valeurs mesurées dans les taillis et introduire du bruit pour les relations avec les données satellitaires (voir Annexe A).

La figure III.15 présente le scatterplot de validation de la surface terrière pour les 7 classes de forêts identifiées. Les résultats sont globalement bons avec 25 % de RMSE relatif et un r^2 de 0.73. On observe une légère sous-estimation des valeurs au-delà de 35 m²/ha. Les placettes de la classe chêne-charme (points jaunes) sont particulièrement bien estimées. Les placettes de chêne du site Orléans ont le moins bon résultat si on regarde le r^2 , mais les points sont surtout concentrés sur une petite plage de données entre 15 et 25 m²/ha et seul un outlier sort du nuage de points de l'ensemble des placettes.

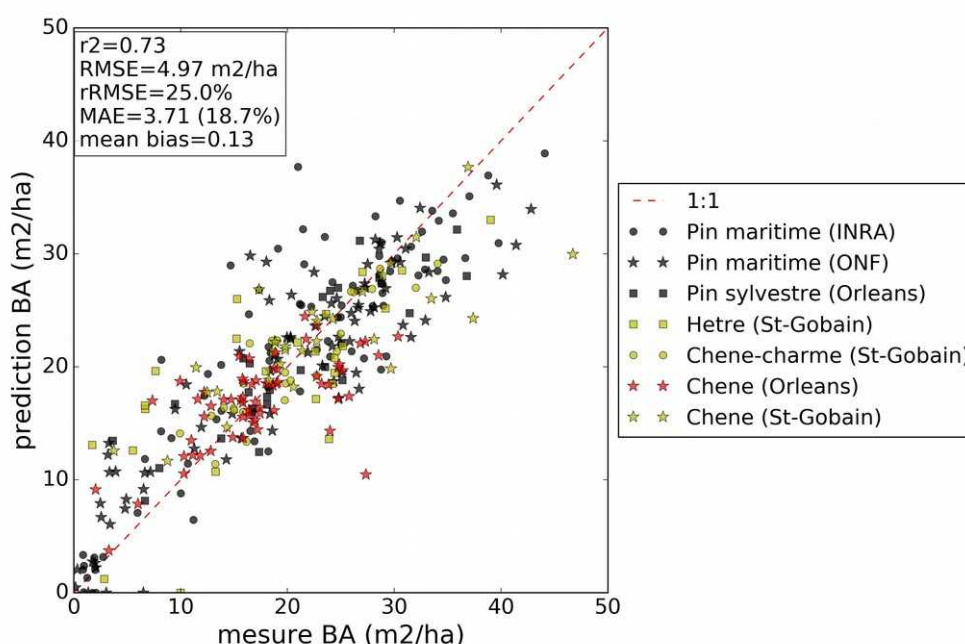


Figure III.15 : Graphique de validation des estimations de surface terrière (BA) pour les 7 classes de forêts identifiées.

La figure III.16 montre le scatterplot de validation des estimations de DBH moyen. Les estimations sont globalement très bonnes, avec un RMSE relatif en dessous de 20 %. Les estimations sont plus mauvaises pour les classes pin maritime (ONF) et chêne (Orléans) entre 35 et 50 cm de DBH. Les classes pin maritime (INRA) et la classe chêne-charme obtiennent les meilleurs résultats. La bonne corrélation pour la classe chêne-charme ($r^2=0.82$) est d'autant plus intéressante du fait que les valeurs de DBH sont réparties sur une très petite dynamique. Le modèle est capable de différencier des DBH assez proches sur cette classe de forêt.

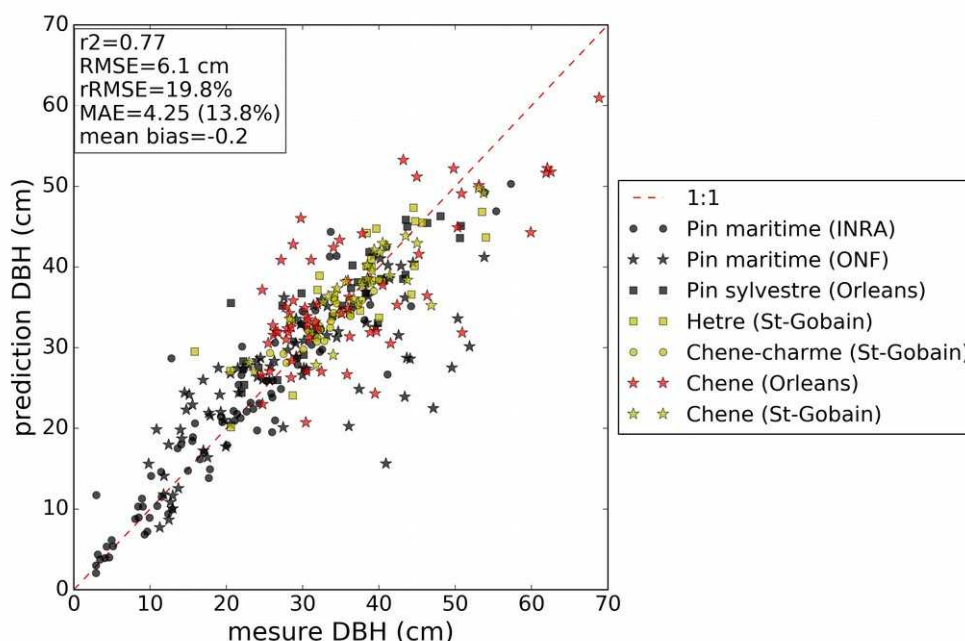


Figure III.16 : Graphique de validation des estimations de DBH moyen pour les 7 classes de forêts identifiées.

La figure III.17 montre le scatterplot de validation des estimations de hauteur dominante. On obtient les meilleurs résultats sur ce paramètre avec un RMSE relatif de 12 % et un r^2 de 0.9 sur ces 7 classes de forêt réunies. Le RMSE va de 1.09 m pour la classe chêne-charme à 3.2 m pour la classe pin maritime (ONF). Aucune saturation n'est visible, y compris pour les plus fortes valeurs de la classe hêtre jusqu'à 35 m de hauteur dominante.

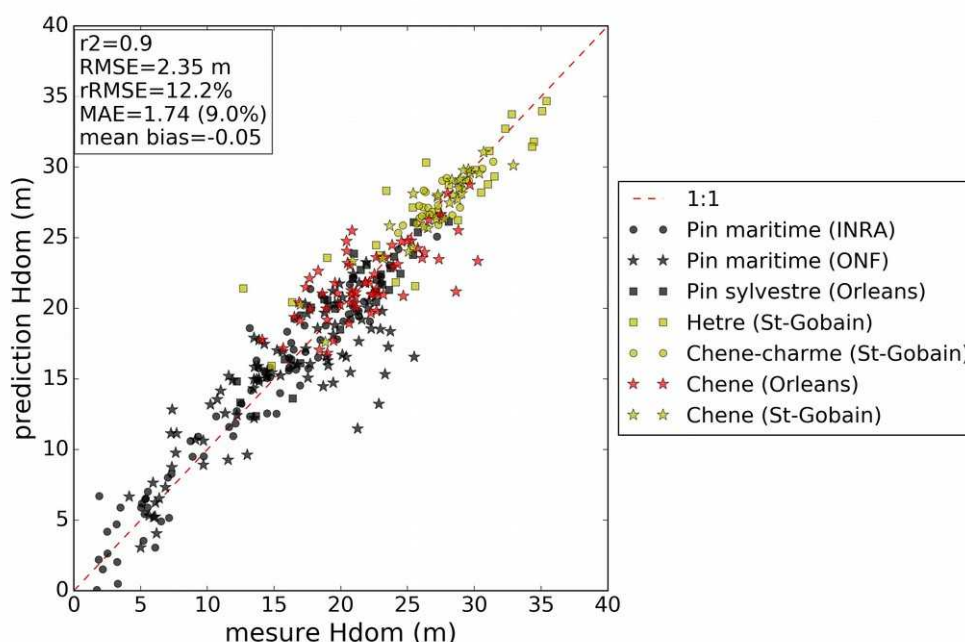


Figure III.17 : Graphique de validation des estimations de hauteur dominante (Hdom) pour les 7 classes de forêts identifiées.

La figure III.18 montre le scatterplot de validation des estimations de la densité d'arbres. Le graphique est repositionné sur la plage de valeurs 0-800 arbres/ha pour une meilleure visibilité des mesures sur les sites Orléans et StGobain. La densité est globalement moins bien estimée

que les autres paramètres forestiers, avec des RMSE relatifs souvent autour de 30 %. On arrive cependant à différencier les plus faibles et les plus fortes densités sur les sites de feuillus, même si la plage de valeurs est petite. Il serait intéressant d'avoir d'autres mesures de densité d'arbre autour de 700-800 arbres par hectare pour ce type de classe de feuillus afin de pouvoir suivre les forêts en reboisement (après des incidents de type tempête, incendie ou dépérissement).

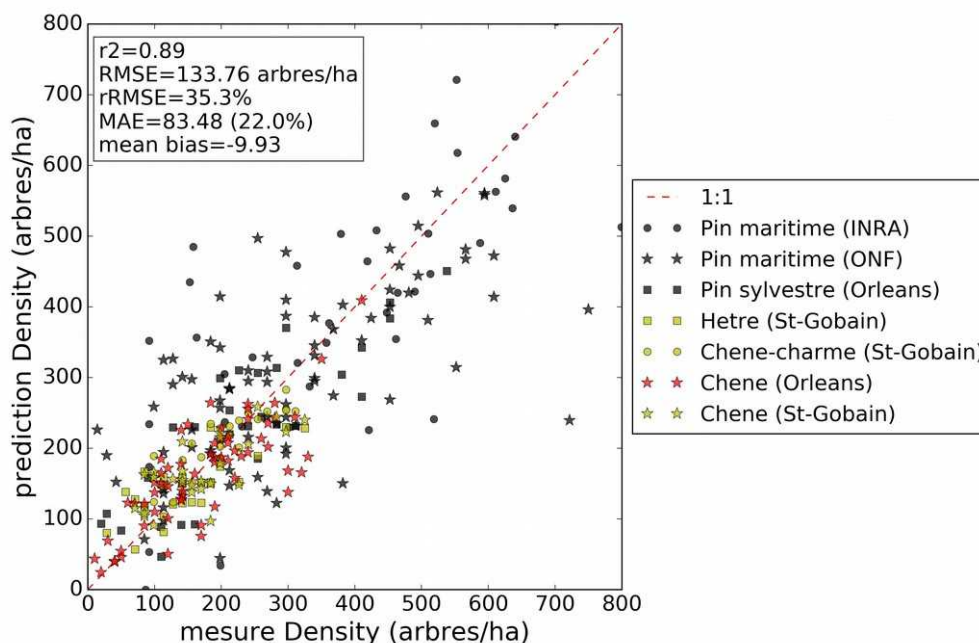


Figure III.18 : Graphique de validation des estimations de densité d'arbres pour les 7 classes de forêts identifiées. Le graphique est repositionné sur la plage de valeurs 0-800 pour une meilleure visibilité des mesures sur les sites Orléans et StGobain.

Le tableau III.8 présente la répartition par types des primitives sélectionnées par la méthode Forward pour l'estimation des paramètres forestiers. Il n'y a pas de différences significatives avec le test de Khi2 par rapport à une distribution théorique (équation 9, section 2.9.1), mais on peut observer quelques grandes tendances. Les coefficients de rétrodiffusion en bande L semblent globalement moins utilisés dans les classes de forêts plus hétérogènes (âges et diamètres des arbres plus variables dans chaque placette), mais continuent à être utilisés pour les estimations de hauteur dominante. Les indices de texture en bande C et en optique restent beaucoup sélectionnés dans les modèles. Les textures nous renseignent sur l'organisation spatiale de la forêt (densité, dispersion des arbres, contraste arbres-sol). Le choix de ces primitives par l'algorithme indique un lien très fort entre l'organisation spatiale et les paramètres dendrométriques propres des forêts gérées par l'homme, par rapport aux forêts naturelles. Enfin, on observe toujours une utilisation de plusieurs types de primitives ensembles mêmes s'ils ont des relations linéaires quasi-nulles avec les paramètres forestiers, cela souligne encore les synergies des informations apportées par les différents types de données satellitaires.

Tableau III.8 : Répartition des types de primitives sélectionnées pour chaque classe de forêt et chaque paramètre forestier. Les pourcentages sont calculés par rapport au nombre de primitives sélectionnées par la méthode Forward.

			Répartition des types de primitives (%)					Nb. sélectionnées
			L	C	C-TI	S2-SI	S2-TI	
Landes-INRA		BA	18%	36%	18%	9%	18%	11
		DBH	33%	11%	11%	0%	44%	9
		Density	27%	27%	27%	9%	9%	11
		Hdom	6%	19%	25%	25%	25%	16
Landes-ONF		BA	8%	17%	17%	33%	25%	12
		DBH	13%	20%	40%	13%	13%	15
		Density	12%	6%	18%	29%	35%	17
		Hdom	10%	10%	30%	10%	40%	10
Orléans	Chêne	BA	13%	38%	13%	13%	13%	8
		DBH	10 %	20%	10%	10%	40%	10
		Density	0%	33%	0%	0%	33%	3
		Hdom	27%	9%	27%	27%	18%	11
	Pin sylvestre	BA	0%	33%	44%	0%	22%	9
		DBH	0%	36%	18%	9%	36%	11
		Density	0%	20%	20%	20%	40%	5
		Hdom	22%	22%	22%	11%	22%	9
StGobain	Chêne	BA	0%	10%	30%	20%	40%	10
		DBH	0%	33%	33%	0%	33%	3
		Density	0%	0%	100%	0%	0%	3
		Hdom	17%	0%	50%	17%	17 %	6
	Hêtre	BA	0%	13%	13%	25%	50%	8
		DBH	0%	25%	38%	13%	25%	8
		Density	0%	0%	14%	29%	57%	7
		Hdom	7%	13%	20%	13%	47%	15
	Chêne-charme	BA	11%	11%	11%	22%	44%	9
		DBH	0%	43%	29%	14%	14%	7
		Density	0%	20%	40%	30%	10%	10
		Hdom	0%	30%	50%	10%	10%	10

3.2.4. Impact de la nomenclature des types de forêts

Exemple chêne

Les chênes des sites d'Orléans et de StGobain sont sur des sols très différents (sol plus pauvre sur Orléans), en conséquence, comme nous l'avons vu pour les deux sites de pin maritime dans les Landes, les forêts n'ont pas les mêmes caractéristiques bien qu'il s'agisse de la même essence. Nous avons testé la construction de modèles par paramètre forestiers pour ces deux classes de chênes réunies et avons comparé les résultats à ceux obtenus avec des modèles différenciés selon le site. Le tableau III.9 présente cette comparaison, les lignes « 2 sites » montrent les résultats obtenus lorsqu'on fait un modèle unique pour les chênes des deux sites ensemble. Les estimations de BA et Hdom sont légèrement dégradées pour les deux sites lorsqu'on procède avec un modèle unique, de manière plus marquée sur Orléans. Dans le cas du DBH et de la densité, les résultats sur les chênes d'Orléans sont plutôt similaires entre le modèle Orléans seul et les chênes des deux sites ensembles. En revanche les résultats sont plus fortement dégradés sur StGobain lorsqu'on mélange les chênes des deux sites pour l'apprentissage des modèles, les r^2 obtenus sur le DBH et la densité chutent respectivement de

0.73 à 0.17 et de 0.54 à 0.12. Nous savons que cette forêt de StGobain est plutôt gérée en futaie, tandis qu'on trouve des taillis et des futaies mélangées sur Orléans, cela pourrait expliquer le fait que StGobain bénéficie plus d'être isolé en tant que classe unique pour ses chênes.

Tableau III.9 : Statistiques de validation des estimations de BA, DBH, Hdom et Densité pour l'ensemble des classes de forêts des 4 sites. Le pourcentage représente le RMSE relatif à chaque classe.

Site (chêne)	Modèle (chênes)	Surface terrière BA (m ² /ha)			DBH (cm)			Hauteur dom. Hdom (m)			Densité Density (arbre/ha)		
		r ²	RMSE	%	r ²	RMSE	%	r ²	RMSE	%	r ²	RMSE	%
Orléans	Orléans seul	0.43	4.50	26 %	0.61	7.76	21 %	0.52	2.49	11 %	0.62	54.97	32 %
	2 sites	0.21	5.46	31 %	0.66	7.45	20 %	0.36	2.92	13 %	0.56	58.95	34 %
StGobain	StGobain seul	0.64	5.88	26 %	0.73	3.66	10 %	0.86	1.34	5 %	0.54	46.30	28 %
	2 sites	0.55	6.46	28 %	0.17	6.39	17 %	0.67	2.16	8 %	0.12	65.87	40 %

Toutes les essences

Parmi les 7 classes de forêt identifiées (en enlevant les classes « Autre »), nous avons testé l'impact des regroupements sur la qualité des modèles pour l'estimation des paramètres forestiers. Le tableau III.10 présente les résultats (statistiques sur l'ensemble des placettes des 7 classes, soit 322 échantillons) obtenus en procédant avec un modèle par classe, puis un seul modèle pour les 7 classes. Cette première comparaison confirme l'intérêt de discriminer les essences ou les types de forêts. Les résultats restent encore intéressants pour certaines applications, mais on observe une dégradation des résultats marquée, particulièrement pour la surface terrière. Nous avons comparé les scatterplots de validation des deux approches, on constate que pour la surface terrière la dégradation des statistiques provient à la fois d'une plus grande dispersion des points et d'une sous-estimation des fortes valeurs (saturation). Pour le DBH, il s'agit surtout d'une plus grande saturation. Pour la hauteur et la densité d'arbre, il s'agit d'une plus grande dispersion des points.

Tableau III.10 : Statistiques de validation pour les placettes des 7 classes de forêts (hors classes « Autres ») selon une approche avec un modèle par classe ou un modèle unique construit pour l'ensemble des classes.

Classe	Surface terrière (m ² /ha)			DBH (cm)			Hauteur dominante (m)		
	r ²	RMSE	%	r ²	RMSE	%	r ²	RMSE	%
Modèles par classes	0.73	4.97	25 %	0.77	6.1	20 %	0.90	2.35	12 %
Modèle unique	0.48	6.92	35 %	0.60	8.2	27 %	0.77	3.52	18 %

Séparation entre feuillus et conifères

Si on procède à des regroupements moins brutaux, par exemple discriminer seulement les feuillus et les conifères, les résultats sont bons pour les conifères mais les estimations des feuillus restent largement moins bonnes que lors d'une estimation classe par classe. Les tableaux III.11 et III.12 montrent ces résultats avec les regroupements pour les feuillus et les conifères. On voit donc que le fait de rassembler les 3 classes de conifères ensemble dans un modèle ne détériore que très peu la qualité des estimations. Ce sont trois classes de pin gérés de manière régulière, avec très peu de relief. Concernant 4 classes de feuillus, on constate une dégradation plus importante de la qualité des estimations. Les résultats par classe sont initialement déjà légèrement moins bons que pour les conifères, et les résultats avec un modèle unique pour les feuillus donne des r² entre 0.25 et 0.5, donc une variabilité des valeurs qui est assez faiblement expliquée par notre modèle de régression.

Tableau III.11 : Statistiques de validation pour les placettes des 3 classes de conifères selon une approche avec un modèle par classe ou un modèle construit pour l'ensemble placettes de conifères.

Classe	BA (m ² /ha)			DBH (cm)			Hauteur dominante (m)		
	r ²	RMSE	%	r ²	RMSE	%	r ²	RMSE	%
Modèle par classe de conifère	0.79	4.96	24 %	0.76	6.23	24 %	0.85	2.44	16 %
Modèle unique pour les conifères	0.70	5.86	29 %	0.71	6.78	26 %	0.77	3.04	19 %

Tableau III.12 : Statistiques de validation pour les placettes des 4 classes de feuillus (hors classe « Autre ») selon une approche avec un modèle par classe ou un modèle construit pour l'ensemble des 4 classes feuillus.

Classe	BA (m ² /ha)			DBH (cm)			Hauteur dominante (m)		
	r ²	RMSE	%	r ²	RMSE	%	r ²	RMSE	%
Modèle par classe de feuillus	0.58	5.0	26 %	0.64	5.92	16 %	0.78	2.22	9 %
Modèle unique pour les feuillus	0.24	6.76	35 %	0.48	7.17	20 %	0.53	3.19	13 %

La figure III.19 illustre la différence de résultats pour les classes de feuillus entre (a) un modèle par classe et (b) un modèle unique. On peut clairement voir que le modèle n'arrive pas à trouver de relation claire dans le cas d'un modèle unique, les points ne sont pas forcément plus dispersés mais le modèle n'est pas capable pas différencier les valeurs de surface terrière.

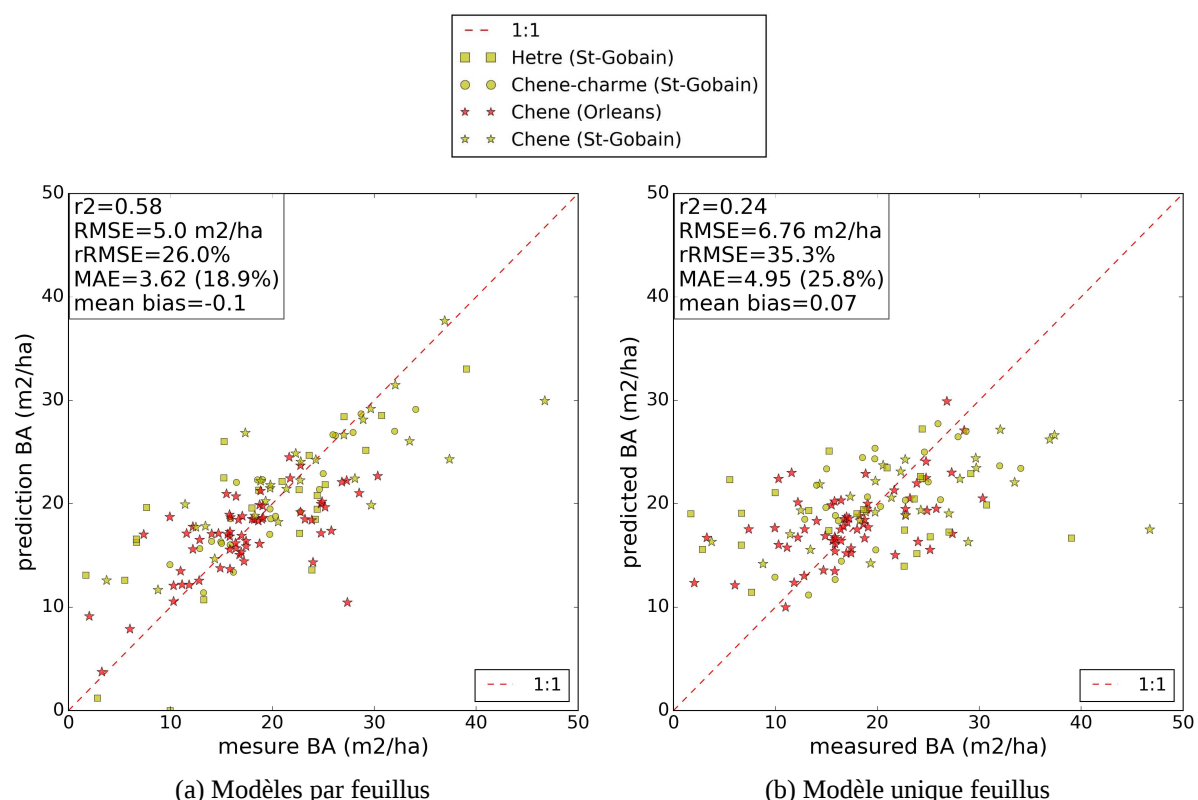


Figure III.19 : Graphiques de validation des estimations de surface terrière (BA) des feuillus selon une approche avec des modèles différenciés par classe (a), ou avec un modèle unique pour les 4 classes (b).

Nous n'avons pas pu aller jusqu'au bout et tester toutes les combinaisons possibles afin d'avoir les classes les plus larges avec les meilleurs résultats possibles, mais ces premiers résultats soulignent déjà l'importance du choix de la nomenclature.

Ici, notre approche est centrée sur des essences en particulier, mais d'autres approches pourraient améliorer les résultats ou permettre d'intégrer plus de placettes dans les classes. À ce stade, des discussions avec des gestionnaires forestiers et/ou des experts sont nécessaires pour améliorer notre définition des classes de forêt : les combinaisons d'essences qui

pourraient être similaires d'un point de vue sylviculture, les différences de structure des arbres qui peuvent être plus ou moins marquées selon les essences, etc.

Un approche statistique pourrait aussi être mise en place, avec par exemple des méthodes automatiques de regroupement (« clustering ») des placettes basées sur les relations entre paramètres forestiers, et les caractéristiques qualitatives (essence forestière, topographie, sol, climat, etc.).

3.3. Conclusion

La méthode mise au point précédemment dans le Chapitre II a été appliquée à d'autres forêts avec différentes essences, modes de gestion et climats. Les résultats montrent que la méthode est transférable à des types de forêts très différents, y compris des forêts beaucoup plus naturelles que la forêt des Landes qui comprend surtout des peuplements équiennes et mono-spécifiques. En revanche, les résultats en termes de transfert de modèle montrent qu'il est difficile d'apprendre sur un type de forêts et d'appliquer sur d'autres si les peuplements ne sont pas similaires, que ce soit par la gestion ou par les essences forestières présentes.

Les résultats de validation montrent encore une fois que le paramètre de hauteur est estimé de manière plus précise que les autres paramètres, y compris pour des valeurs les plus fortes jusqu'à 35 m de hauteur dominante. On observe quelques différences dans les types de primitives sélectionnées pour les modèles de régression. La bande L est moins utilisée pour les nouvelles forêts plus hétérogènes que nous avons intégré aux tests. Les indices de texture en optique ou en bande C sont encore très utilisés, indiquant ainsi une forte relation entre l'organisation spatiale et les paramètres dendrométriques. Sur la plupart des forêts de ces nouveaux jeux de données, les relations linéaires entre les primitives et les paramètres forestiers individuellement sont très faibles, mais le SVR est capable de les utiliser en combinaison (3 à 4 types sur 5) pour obtenir de bonnes estimations des paramètres forestiers. Les 31 primitives sélectionnées parmi les données en libre accès permettent d'obtenir des résultats déjà satisfaisants sur ces différents types de forêts, même si d'autres données plus précises pourraient éventuellement améliorer encore ces estimations.

Nos analyses montrent que les estimations sont meilleures quand des modèles spécifiques à différents types de forêt sont développés, par rapport à un modèle unique. L'idée n'est pas de développer une pléthore de modèles pour application à des types de forêt trop spécifiques (car il n'y aurait pas non plus les mesures terrain nécessaires pour calibrer chaque modèle), mais de rester générique. Ainsi, notre étude porte sur les choix de regroupement et leur impact sur la qualité des estimations. Nos résultats montrent que les conifères sont assez robustes au regroupement, tandis que les feuillus sont plus sensibles. D'autres tests et recherches sont nécessaires afin de déterminer une nomenclature hiérarchique (classes emboîtées) qui permette d'obtenir des résultats satisfaisants sur l'ensemble des forêts.

Chapitre IV : Cartographie des forêts

Chapitre IV : Cartographie des forêts.....	87
4.1. Cartes des types de forêts.....	87
4.1.1. Données existantes.....	87
4.1.2. Production de cartes des forêts.....	89
4.2. Spatialisation des modèles.....	91
4.2.1. Méthodes de spatialisation.....	91
4.2.2. Évaluation des différents approches.....	92
4.3. Analyse des cartes de paramètres forestiers.....	95
Site Orléans.....	95
Site Saint-Gobain.....	98
Sites Landes.....	98
4.4. Conclusion.....	100

La méthode mise en place a permis d'estimer les paramètres forestiers sur des petites zones au sein des peuplements forestiers (placettes ou polygones). Nous voulons à présent démontrer la capacité des données satellitaires à spatialiser ces estimations sur des zones plus larges.

4.1. Cartes des types de forêts

4.1.1. Données existantes

Nous avons montré dans le Chapitre II la possibilité d'estimer avec notre méthode les paramètres forestiers sur des placettes et des peuplements forestiers bien identifiés. La généralisation de la méthode pour construire des cartes à haute résolution nécessite d'avoir des informations sur l'occupation du sol afin de pouvoir appliquer les modèles sur les forêts uniquement, et si possible de distinguer plusieurs classes de forêts au sol quand on dispose de modèles spécifiques à ces classes de forêt.

Dans cette section, nous recensons succinctement des produits disponibles (cartes d'occupation du sol pour les forêts) qui pourraient répondre à nos besoins en termes de résolution spatiale et temporelle, d'accessibilité des données, et de qualité.

Masques des forêts

Plusieurs produits permettant de faire des cartes de forêt/non-forêt sont faits à partir de données de télédétection ou d'observations terrain. Le tableau IV.1 présente une liste des produits à haute résolution spatiale que nous avons évalués pour nos propres besoins, cette liste n'est pas exhaustive.

Tableau IV.1 : Produits disponibles pour la cartographie des forêts à haute résolution spatiale.

Fournisseur	Produit	Informations	Fréquence de mise à jour	Zone	Résolution spatiale
JAXA	forest/non-forest map	masque des forêts	Annuel 2007-2010 et >2015	Global	Raster ~25 m
Copernicus	Tree cover density	couverture forestière (%)	2012 et 2015	Europe	Raster ~20 m
Théia OSO	OCS	feuillus/conifères	Annuel à partir de 2016	France	Raster 10 à 20 m
IGN	BD forêt	essences forestières	Jusqu'à 15 ans	France	Vecteur 1:10 000

Différents éléments sont à prendre en compte pour l'utilisation de ce type de produits sur la cartographie des forêts :

- Plusieurs de ces produits ont une unité de cartographie minimale (« Minimum Map Unit », MMU) supérieure ou égale à 0.5 ha (~ 50 pixels à 10 m de résolution spatiale). C'est le cas ici des produits « Tree cover density » de Copernicus et de la BD forêt de l'IGN. Cette MMU ne permet pas de bien délimiter les petites forêts dispersées ou très différentes sur des surfaces réduites, que l'on trouve beaucoup en France. De plus, on trouve souvent au sein des peuplements forestiers des prairies, des habitats dispersés ou d'autres éléments qui ne sont pas des forêts et qui ont une taille inférieure à la MMU, donc comptabilisés comme forêt.
- L'information doit être récente car l'occupation du sol peut changer rapidement ainsi que la gestion des forêts.
- Stabilité inter-annuelle : on constate sur certains produits annuels (comme ici le masque des forêts de la JAXA) une trop grande instabilité des zones classées en forêt entre les années, qui ne correspond pas à la réalité.
- La nomenclature et la qualité de la carte des forêts. Certaines cartes fournissent des données peu précises au niveau local, même si la statistique sur une plus grande surface reste conforme à la réalité.

Après considération de ces différents éléments pour évaluer les produits que nous pourrions utiliser, la carte d'occupation du sol OCS-OSO paraît être la plus adaptée à nos besoins : mise à jour annuelle, bonne résolution spatiale (10 à 20 m), stabilité inter-annuelle et informations sur la qualité de la carte. La nomenclature actuelle ne différencie que les feuillus et les conifères sans plus de détail sur les essences, mais la méthode de production avec la chaîne de traitement « iota² »⁹ pourrait permettre à terme un élargissement de la nomenclature des forêts.

Cartes des types/essences des forêts

Nous avons donc utilisé la carte OCS-OSO comme masque des forêts pour l'application des modèles. Cependant, nous avons montré dans le Chapitre III que l'utilisation de modèles différenciés pour les essences forestières améliorerait fortement les résultats. Nous avons donc cherché à obtenir une cartographie plus fine des essences forestières sur le masque des forêts. La BD forêt de l'IGN a une nomenclature précise qui correspond à nos besoins, mais la comparaison de cette cartographie avec les informations d'essences forestières sur les placettes de mesures nous a révélé de nombreuses incohérences sur les sites Orléans et StGobain. Ces incohérences peuvent provenir de changements intervenus depuis la production (> 15 ans sur certains départements), ou être liées à la méthode elle-même (MMU, définition des classes, etc.). Des produits plus fiables pour nos besoins (qualité et/ou mise à jour) pourraient être disponibles localement : Plans Simples de Gestion (PSG) sur les forêts privées, plans d'aménagements sur les forêts publiques, ou des cartes produites occasionnellement sur des zones spécifiques. Afin de proposer une méthode généralisable à n'importe quelle zone, nous avons choisi de produire notre propre cartographie des essences forestières sur les sites d'étude.

⁹ <https://iota2.readthedocs.io/en/latest/index.html>

4.1.2. Production de cartes des forêts

Nous présentons dans cette section les caractéristiques et la méthode de production de la carte des essences forestières sur les sites d'études. L'objectif est uniquement d'avoir un support pour appliquer ensuite nos modèles. Nous ne présenterons pas ici de comparaisons ou de validation indépendante des méthodes et des résultats de classification obtenus pour les cartes d'essences.

Méthode de classification

Nous avons procédé en utilisant une méthode de classification par apprentissage similaire à celle utilisée pour produire la carte OCS-OSO. Le tableau IV.2 résume les grandes lignes de la méthode.

Tableau IV.2 : Informations sur la méthode de classification pour la cartographie des essences forestières.

Images	Primitives	Références	Algorithme	Post-traitement
Sentinel-1 2015-2017	Moyennes mensuelles	Essences des placettes de mesure, nomenclature identique aux modèles	Random Forest (OTB) <i>profondeur : 25</i> <i>nombre arbres : 100</i> <i>éch. min. nœuds : 10</i> <i>100 % éch. apprentissage</i>	Suppression des pixels isolés (GDAL sieve) ou vote majoritaire sur des segments issus d'une segmentation <i>mean-shift</i>
Sentinel-2 2015-2017	Images sans nuages sur le site d'étude. NDVI, NDWI et BI.			

Cette méthode a été utilisée pour cartographier les types de forêts sur les sites StGobain et Orléans. L'ensemble des pixels des placettes de référence ont été utilisés pour l'apprentissage afin d'obtenir la meilleure classification possible. Les résultats sont satisfaisants pour nos besoins. Concernant les deux sites d'étude des Landes, le produit OCS-OSO était suffisant pour discriminer les feuillus et les conifères, sachant que 90 % des conifères de cette région sont des peuplements de pin maritime.

Cartes de classification

Les statistiques issues de la comparaison de ces cartes avec les références ont été calculées. Elles sont présentées à titre indicatif, mais il ne s'agit pas d'une validation indépendante.

- Orléans : 93 % de bonne classification, kappa = 0.88
- StGobain : 84 % de bonne classification, kappa = 0.77

Les pixels sont donc classés selon leur vraisemblance avec les pixels de référence (les placettes) sur la base des séries temporelle Sentinel-1 et Sentinel-2. Les figures IV.1 et IV.2 présentent les cartes d'essences obtenues sur les sites Orléans et StGobain. Ces cartes nous permettront d'appliquer les modèles de chaque classe sur les pixels considérés comme de la même classe par le classifieur.

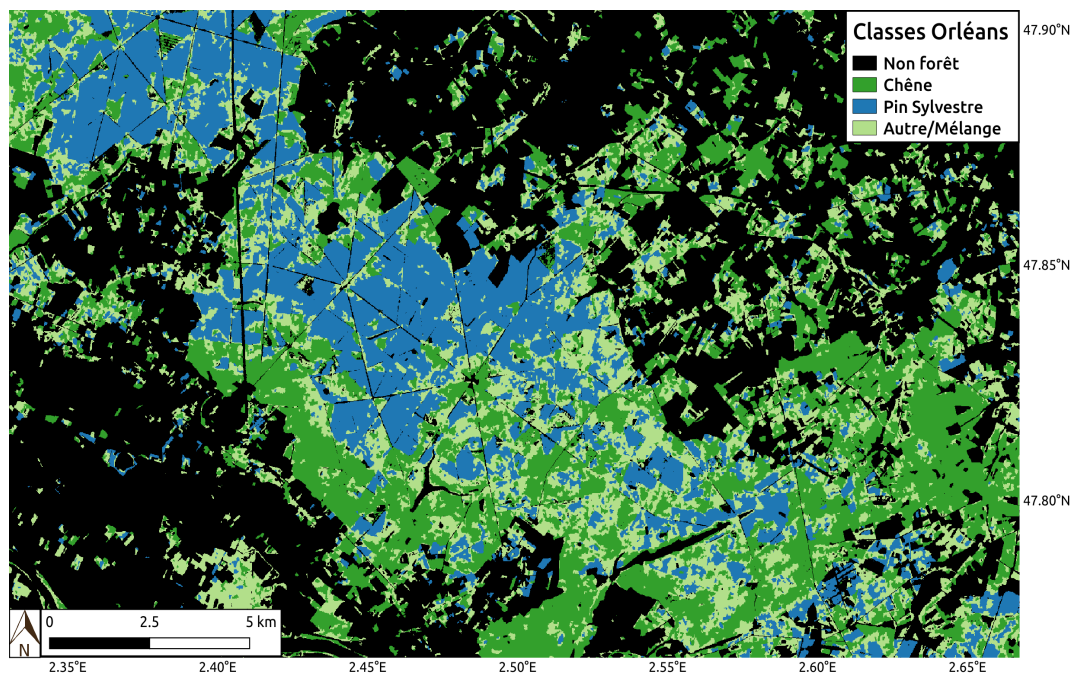


Figure IV.1 : Classification des types de forêt sur le site d'Orléans.

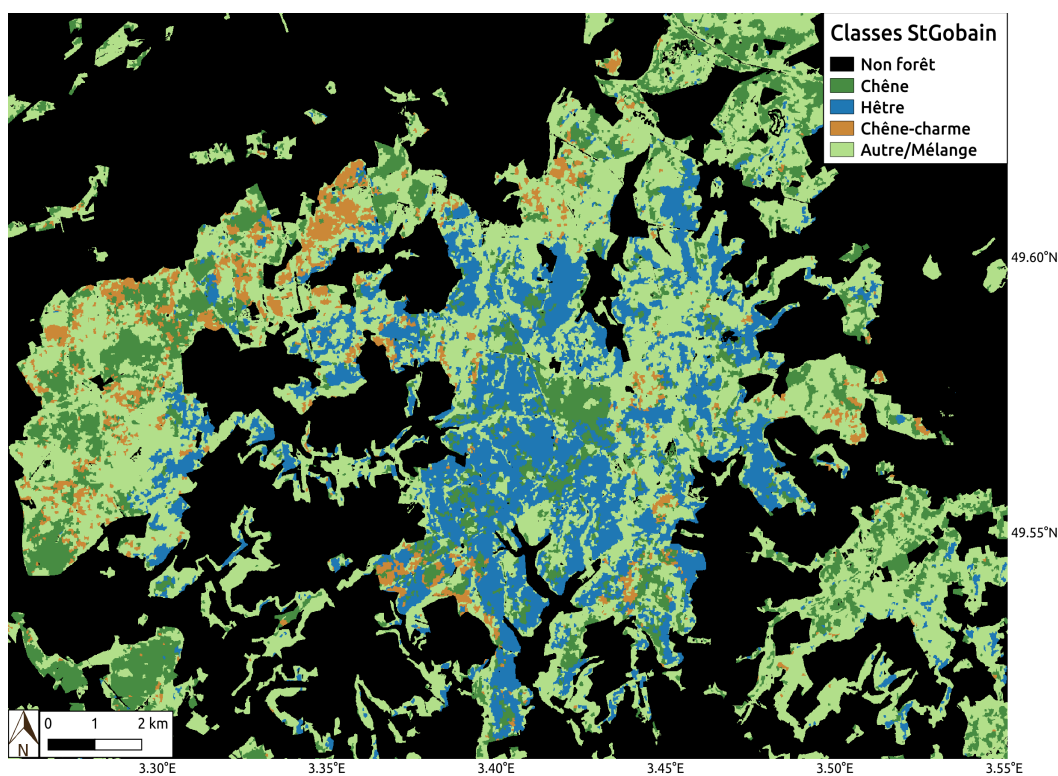


Figure IV.2 : Classification des types de forêt sur le site de StGobain.

La carte de classification du site Orléans (figure IV.1) paraît cohérente avec la réalité. On compte 8185 ha de chêne, 6228 ha de pin sylvestre, et 6585 ha de forêts mélangées (classe « Autre ») qui se trouvent principalement aux frontières entre les peuplements de chênes et ceux de pin sylvestre. La carte de classification du site StGobain (figure IV.2) est plus hétérogène dans la répartition des essences. Ces forêts sont plus mélangées, et les classes utilisées dans la nomenclature sont physiquement beaucoup plus proches entre elles (différents types de feuillus) que celles du site Orléans (feuillus/conifères). On compte 4880 ha de chêne, 3140 ha de hêtre (principalement sur la partie Est de la forêt), 838 ha de mélange chêne-charme, et enfin 13 694 ha de forêts mélangées (classe « Autre »).

4.2. Spatialisation des modèles

4.2.1. Méthodes de spatialisation

L'apprentissage des modèles avec les données satellitaires s'est fait sur des polygones, c'est-à-dire sur des moyennes d'ensembles de pixels. Nous avons vu par ailleurs, dans le cas des peuplements homogènes de pin maritime dans les Landes, que la méthode produit de meilleures estimations à partir de polygones plus larges que les placettes de mesure. Cependant, cette approche peut poser des problèmes pour ensuite étendre les modèles sur les pixels d'une zone étendue, avec une application sur les pixels individuels et non sur des ensembles de pixels. En effet, les valeurs des pixels peuvent avoir une variabilité plus importante que celle des polygones utilisés en apprentissage. De plus, l'algorithme SVR utilisé nécessite la normalisation des valeurs en entrée, cette normalisation peut fournir des « mauvaises » valeurs si les moyennes et écarts types des pixels de la zone sont trop différents des moyennes et écarts types des polygones utilisés dans l'apprentissage.

Pour cette raison, nous avons testé quatre différentes options pour l'application des modèles, en prenant en compte l'aspect pixel/objet et la normalisation des données en entrée :

Approche pixel

Dans cette approche, nous avons appliqué les modèles directement aux pixels de la zone à spatialiser. Pour la normalisation des données en entrée du SVR nous avons étudié deux options :

- Normalisation des pixels de la carte selon les valeurs des polygones de l'apprentissage. Cette configuration est appelée par la suite « approche pixel avec normalisation modèle ».
- Normalisation des pixels de la carte selon les valeurs de l'ensemble des pixels appartenant à la classe cible. Cette configuration est appelée par la suite « approche pixel avec normalisation classe ».

Approche objet (segmentation)

Une segmentation a été produite avec la méthode mean-shift (Comaniciu and Meer, 2002) en utilisant l'application OTB. Les données d'entrée sont les 31 primitives utilisées pour les modèles, ces primitives sont normalisées et rééchantillonnées entre 0 et 1 pour la segmentation. Le choix des paramètres a été fait pour être applicable sur toutes les zones sans nouvelle optimisation : *rayon spatial* = 3 pixels, *rayon spectral* = 0.01, et *taille minimum* = 16 pixels. L'utilisation d'une segmentation pour travailler sur des objets au lieu de pixels peut permettre à la fois d'améliorer la qualité des estimations (moyennes de pixels), et d'améliorer l'aspect visuel des cartes. Pour la normalisation des données en entrée du SVR nous avons étudié les mêmes options que dans l'approche pixel :

- Utilisation d'une segmentation et normalisation des pixels selon les valeurs des polygones d'apprentissage : « approche objet normalisation modèle ».
- Utilisation d'une segmentation et normalisations des pixels selon les valeurs de l'ensemble des pixels appartenant à la classe cible : « approche objet normalisation classe ».

4.2.2. Évaluation des différents approches

Pour évaluer la qualité des cartes produites lors des quatre différentes configurations, nous avons mis en place trois indicateurs : 1) la distribution des valeurs prédites sur l'ensemble des pixels de forêt, 2) la validation des estimations sur les placettes de référence, et 3) une évaluation visuelle des cartes (hétérogénéité, artefacts, etc.). Il est cependant difficile de vraiment évaluer la qualité de la spatialisation des modèles sans validation indépendante.

Distribution des valeurs

La figure IV.3 montre la distribution des valeurs de paramètres forestiers estimés sur le site d'Orléans. On voit une correspondance très forte entre les différentes configurations, mais aussi avec les valeurs de référence. Ce résultat est très prometteur concernant la capacité du SVR à généraliser les estimations des paramètres forestiers à l'ensemble des pixels d'une carte, pourvu que les références soient représentatives de la réalité.

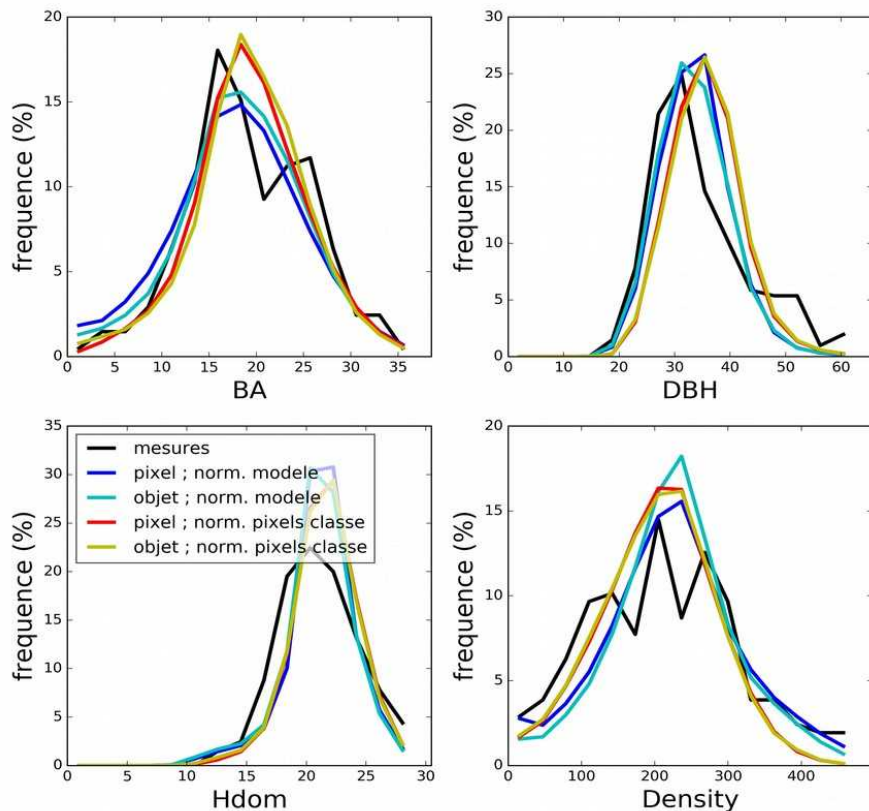


Figure IV.3 : Distribution des valeurs de paramètres forestiers prédits sur l'ensemble des pixels de forêt du site d'Orléans (2.1 millions de pixels). Les valeurs des références sont superposées sur les mêmes intervalles. BA (surface terrière, m²/ha) ; DBH (diamètre à 130cm, cm) ; Hdom (hauteur dominante, m) ; Density (densité d'arbres, arbres/ha)

Lorsque nous regardons ces mêmes distributions pour le site Landes-ONF (figure IV.4), les résultats sont légèrement moins bons pour le cas où on normalise avec les valeurs du modèle : dans le cas de BA il y a un pic de fréquence à 17 m²/ha, et dans le cas de la densité un pic de fréquence à 300 arbres/ha. Cette sur-représentation de certaines valeurs indique une déficience du modèle de régression qui n'arrive pas à reproduire certaines dynamiques. Cette différence, par rapport aux résultats sur Orléans, peut être due à la présence d'autres types de forêts dans la forêt des Landes parmi les pixels du masque, non représentés par les mesures terrain. Nous observons une légère tendance à avoir des distributions plus larges dans le cas de la normalisation sur l'ensemble des pixel ou objets de la classe cible.

Ainsi, l'utilisation d'images segmentées par rapport à l'application directement au pixel n'a pas d'impact visible sur la distribution des valeurs prédites. Il n'y a pas de grandes différences entre les distributions de valeurs obtenues avec les options « pixel » ou « objet ». En revanche la normalisation semble être un facteur plus important pour obtenir une meilleure représentativité des valeurs.

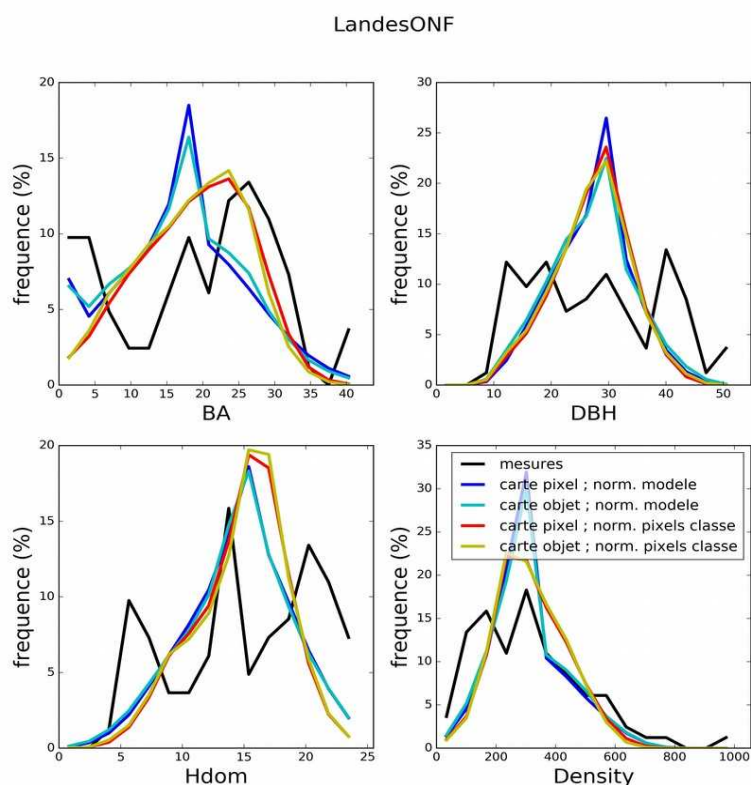


Figure IV.4 : Distribution des valeurs de paramètres forestiers prédits sur l'ensemble des pixels de forêt du site Landes-ONF (2.16 millions de pixels). Les valeurs des références sont superposées sur les mêmes intervalles. BA (surface terrière, m²/ha) ; DBH (diamètre à 130cm, cm) ; Hdom (hauteur dominante, m) ; Density (densité d'arbres, arbres/ha)

Validation des estimations sur de placettes de référence

Nous avons ensuite calculé les statistiques de validation des valeurs prédites sur les surfaces des placettes de mesure. La figure IV.5 montre les résultats en termes de RMSE relatif pour les deux mêmes sites Orléans et Landes-ONF. Pour le site Orléans, les résultats sont encore très similaires entre configurations. Pour le site Landes-ONF, les résultats de validation sont meilleurs avec la normalisation sur les valeurs du modèle, particulièrement pour le paramètre BA. L'utilisation d'une segmentation n'apporte pas d'amélioration par rapport à une application au pixel. Encore une fois, la différence de performances entre les deux configurations pour la normalisation (normalisation par rapport aux valeurs des polygones d'apprentissage ou des valeurs des pixels/objets de la classe cible) pourrait s'expliquer par la présence de types de forêts autres que le pin maritime sur le masque forêt du site Landes-ONF. Les pixels de ces autres forêts (îlots chênes par exemple) déforment la dynamique des valeurs de pixels par rapport aux pins maritimes seuls. Ceci démontre l'importance d'avoir une carte précise et à haute résolution des types de forêt au sol. Les types de forêts présents dans les placettes du site Orléans sont plus diversifiés, améliorant par conséquent la capacité du modèle à généraliser sur les autres forêts.

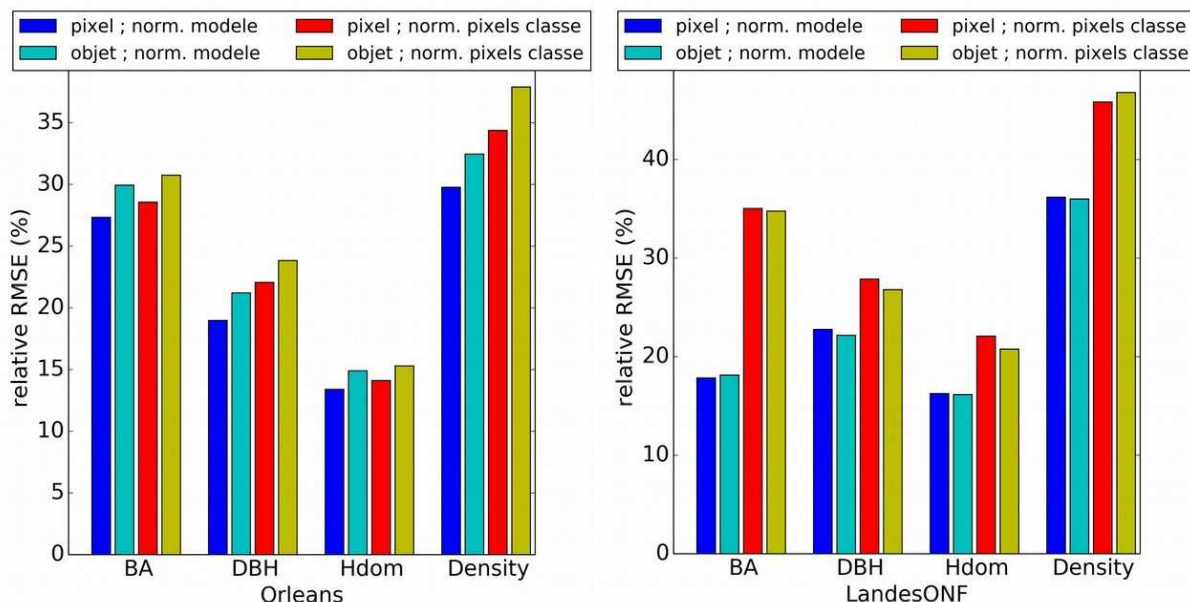


Figure IV.5 : Statistiques de validation (RMSE relatif) des cartes de prédictions sur les pixels des placettes de référence. BA (surface terrière, m²/ha) ; DBH (diamètre à 130cm, cm) ; Hdom (hauteur dominante, m) ; Density (densité d'arbres, arbres/ha)

Aspect visuel des cartes

La segmentation n'apporte donc pas de bénéfice visible dans notre évaluation comparée à l'application directement au niveau pixel. En revanche, l'aspect visuel des cartes est parfois grandement amélioré avec l'application sur les objets issus de la segmentation. La figure IV.6 présente un exemple de comparaison entre l'application pixel et l'application objet sur un zoom du site Orléans, pour l'estimation de la surface terrière. La variabilité de la surface terrière selon les peuplements est bien visible dans les deux cas, mais la carte pixel a un aspect beaucoup plus bruité. Cette approche objet pourrait donc être une option valable pour produire des cartes qui ont vocation à être visualisées, en plus de produire des statistiques de zone. La dynamique des valeurs et la variabilité locale sont conservées, mais le bruit est largement réduit. La qualité des estimations en objet semble similaire au pixel, mais cela reste à valider avec d'autres données indépendantes.

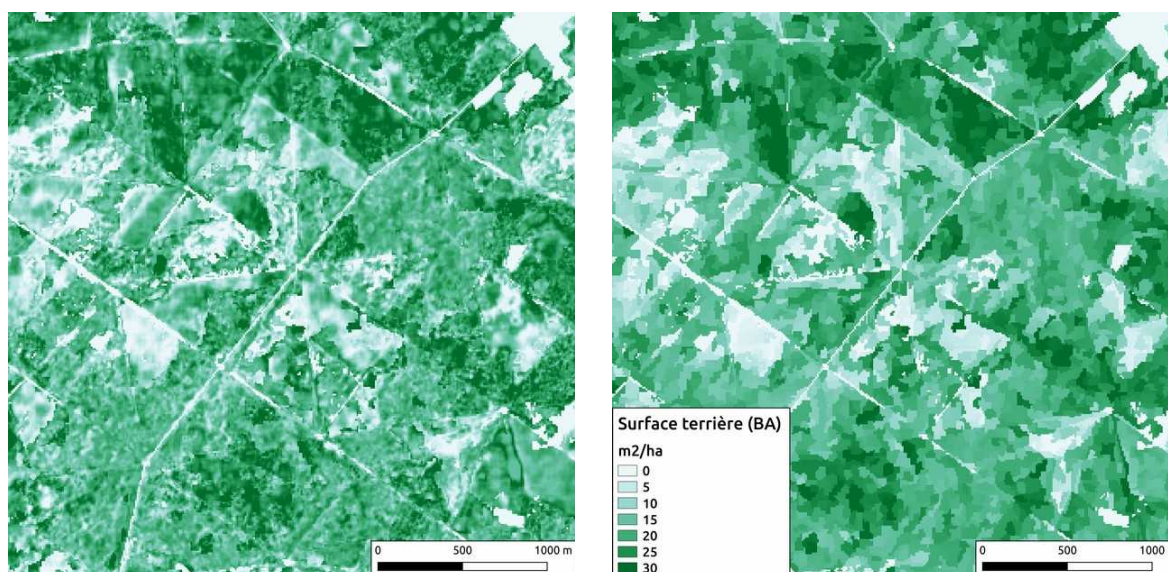


Figure IV.6 : Illustrations des cartes de surface terrière (BA) produites sur le site Orléans avec les méthodes pixel et objet. La normalisation pour le SVR est faite sur les valeurs d'apprentissage.

En plus de ces trois critères d'évaluation, une carte de confiance pourrait être fournie avec les cartes de paramètres forestiers. Cette carte de confiance peut être basée sur le modèle lui-même, ou bien directement avec une valeur par pixel qui indique si les données satellitaires sont comprises dans les dynamiques de valeurs utilisées en apprentissage. Par exemple comptabiliser le nombre de primitives qui sont éloignées de plus de deux écarts types de la moyenne des valeurs d'apprentissage, 1 moins ce ratio représente la confiance qu'on peut accorder à l'application du modèle sur cette série de valeurs du pixel.

4.3. Analyse des cartes de paramètres forestiers

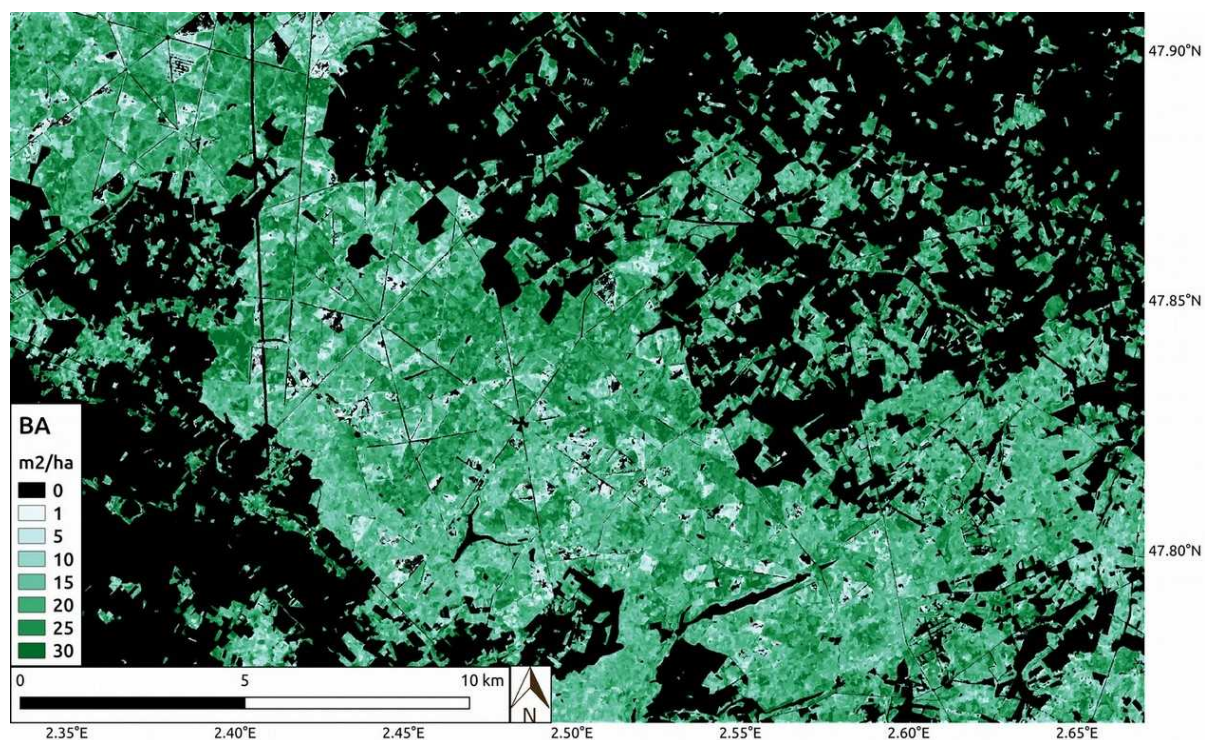
Site Orléans

Les cartes des paramètres BA, DBH, densité d'arbres et hauteur dominante sont présentées en figures IV.7, IV.8, IV.9 et IV.10. Le tableau IV.3 reprend les statistiques descriptives des paramètres estimés pour les classes chêne, pin sylvestre et autre/mélange. Les 3 classes de forêt ont des statistiques relativement similaires, avec juste une surface terrière estimées légèrement plus basse pour les chênes (16 m²/ha contre ~20 pour les autres), et une densité d'arbre en moyenne plus élevée sur les pins sylvestres. La variabilité de hauteur dominante est plus importante sur les pins sylvestres (E. T. = 4.4 m pour une moyenne = 20.5 m), qui reflète en effet la plus grande diversité des stades de développement pour cette classe observée sur les images satellitaires et les cartes de prédictions.

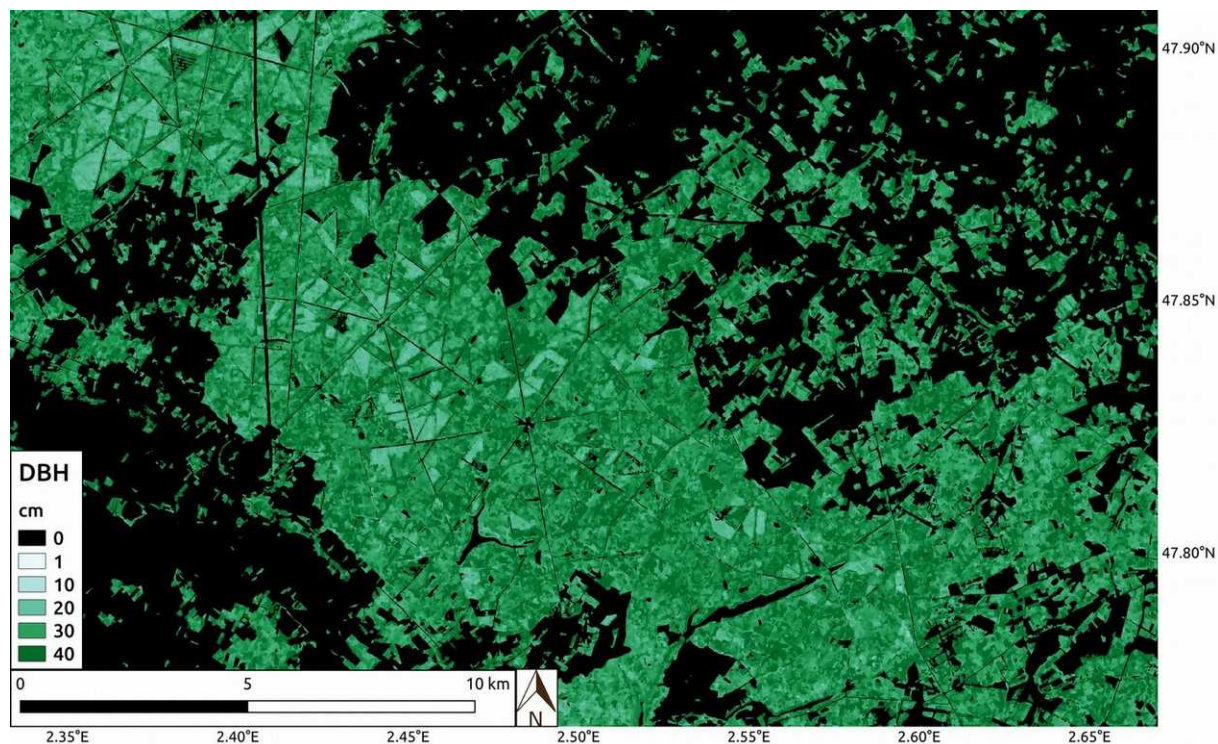
Tableau IV.3 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers estimés sur l'ensemble des pixels de forêt du site Orléans.

Orléans	BA (m ² /ha)		DBH (cm)		Densité (arbre/ha)		Hdom (m)	
	moy.	E. T.	moy.	E. T.	moy.	E. T.	moy.	E. T.
Chêne	16.0	5.0	33.9	7.2	199	70	21.4	2.3
Pin sylvestre	20.5	7.6	33.1	6.5	288	88	20.5	4.4
Autre/mélange	20.2	6.5	33.6	5.2	206	70	21.3	1.7

On voit sur la carte de prédiction de la surface terrière (figure IV.7) que les plus fortes valeurs se situent plutôt sur les peuplements de pin sylvestre (cf. carte de classification des essences figure IV.1), mais la variabilité entre peuplements y est aussi la plus forte.



Sur la carte des prédictions de DBH (figure IV.8), les différences dans l'ensemble du site sont moins marquées que pour la surface terrière. On distingue surtout plusieurs zones/parcelles avec un DBH plus faible et très homogène spatialement.



La plupart des zones avec un plus faible DBH sont caractérisées également, sur la carte de densité (figure IV.9), par une plus forte densité d'arbres par hectare. Il s'agit de jeunes peuplements de pin sylvestre. Les zones classées en chêne ont globalement une plus faible densité d'arbres par hectare que les peuplements de pin sylvestre.

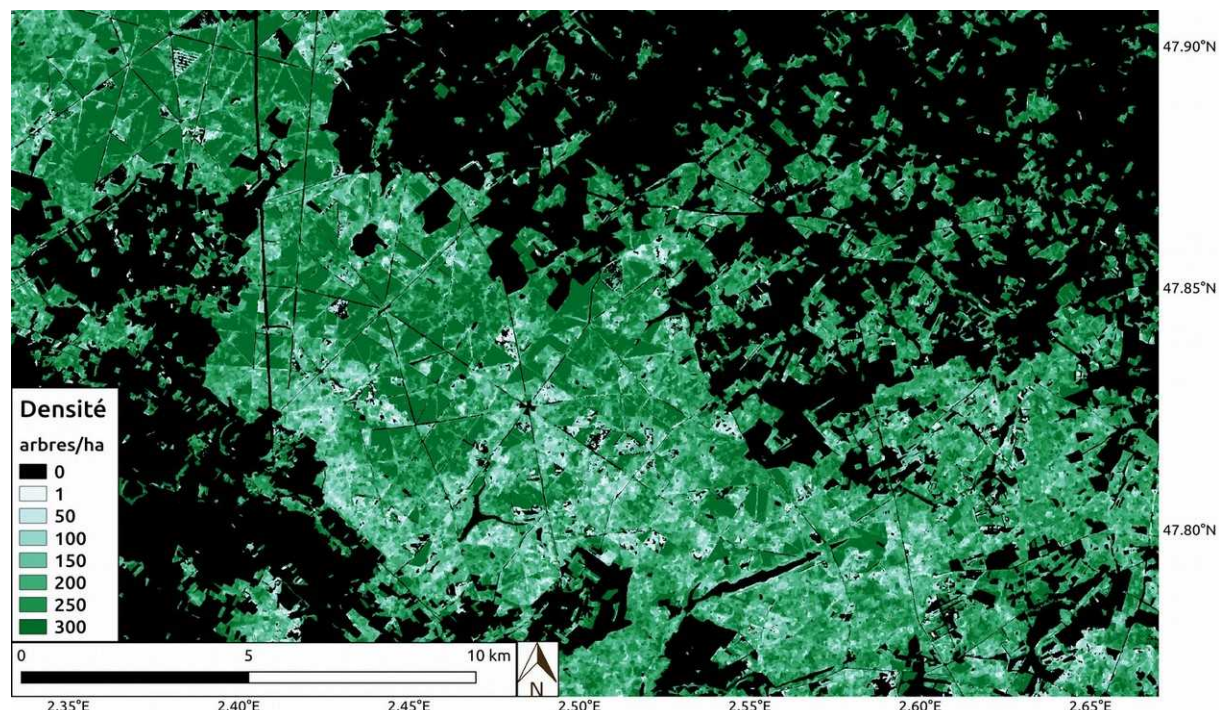


Figure IV.9 : Carte de prédiction de la densité d'arbres (arbres/ha) sur le site Orléans.

Enfin, la carte de prédiction des hauteurs dominantes (figure IV.10) ressemble plus à la carte de DBH. Les différences entre essences sont moins marquées, mais on distingue clairement les jeunes peuplements de pin sylvestre qui ont une hauteur plus faible et assez homogène par parcelle.

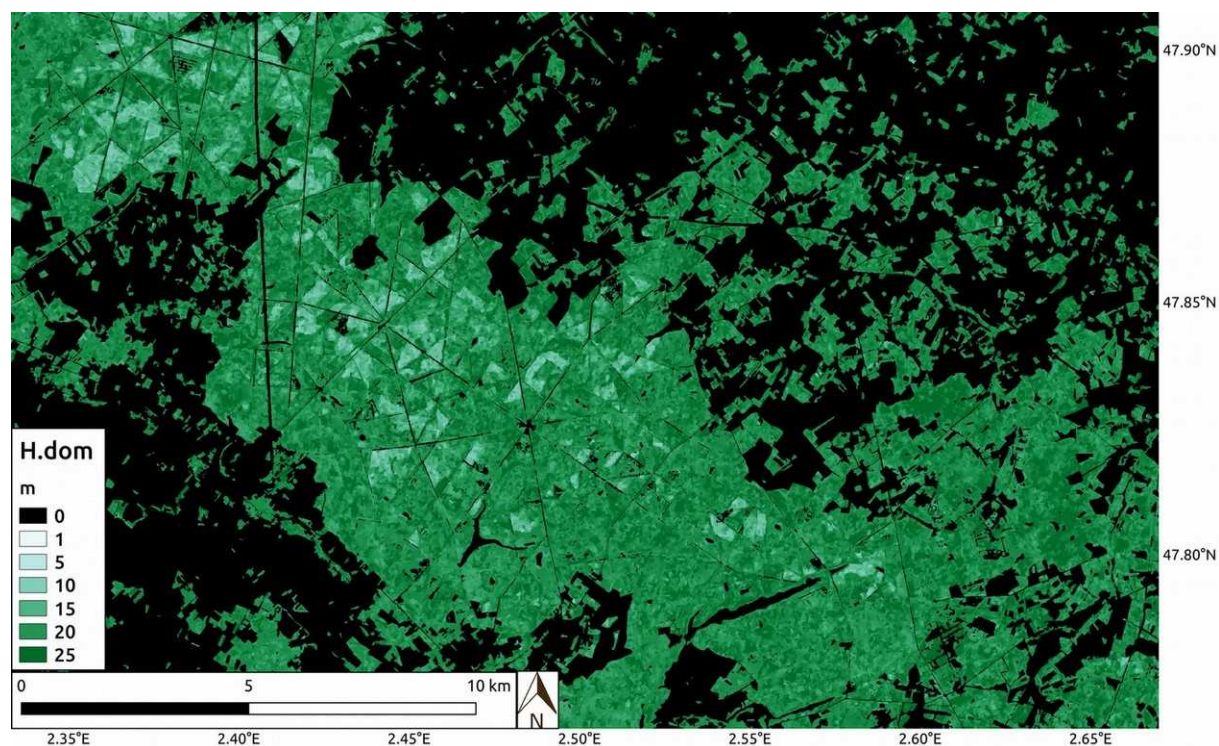


Figure IV.10 : Carte de prédiction de la hauteur dominante (m) sur le site Orléans.

Site Saint-Gobain

De la même manière, les cartes de prédiction des paramètres forestiers ont été produites sur le site StGobain. Le tableau IV.4 présente les statistiques descriptives pour les 4 classes de forêt. Les statistiques moyennes sont encore similaires entre les classes, on note une plus faible surface terrière moyenne sur la classe mixte (autre/mélange), et une hauteur moyenne du hêtre plus forte. Les différences sont plus visibles lorsqu'on compare les écarts-types entre les classes, les paramètres prédits sur la classe hêtre ont globalement des écarts types plus forts, tandis que les écarts-types pour la classe chêne-charme sont plus faibles.

Tableau IV.4 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers estimés sur l'ensemble des pixels de forêt du site StGobain.

StGobain	BA (m ² /ha)		DBH (cm)		Densité (arbre/ha)		Hdom (m)	
	moy.	E. T.	moy.	E. T.	moy.	E. T.	moy.	E. T.
Chêne	21.6	6.0	36.6	5.6	165	48	25.8	3.6
Hêtre	20.7	9.7	36.8	8.1	155	54	28.0	7.9
Chêne-charme	19.9	6.1	32.9	2.8	188	46	25.9	1.7
Autre/mélange	15.6	6.0	34.8	5.4	188	43	24.7	4.8

Sur ces deux sites Orléans et StGobain, nous avons pu calculer un DBH dominant (DBHdom) à partir des mesures au niveau arbre. L'estimation du DBHdom à l'aide de la méthode mise en place donne des résultats de validation très proches de ceux du DBH moyen, mais la variabilité du DBHdom est plus importante entre les différentes classes (par exemple 51.5 cm de moyenne pour la classe chêne contre 40.3 cm pour la classe autre/mélange). Ainsi, d'autres paramètres forestiers que ceux classiquement disponibles (BA, DBH, Densité et Hauteur dominante présentés en commun pour tous les sites) peuvent être utilisés en entrée de la chaîne et spatialisés sur des cartes à haute résolution. Ces paramètres, tels que le DBH dominant (important pour les forestiers), la fraction de couvert ou encore la taille de houppier sont disponibles sur certains sites de mesure et pourraient avoir un intérêt et un « sens physique » à être spatialisés avec la méthode que nous avons utilisée ici.

Sites Landes

Le tableau IV.5 présente les statistiques descriptives des paramètres estimés sur les pixels de conifères des sites Landes INRA et ONF, représentant principalement du pin maritime. La biomasse moyenne est plus faible sur la zone dunaire (site ONF) avec ~50 tonnes/ha contre ~65 tonnes/ha de moyenne dans le site INRA. Cette différence n'est pas visible dans les estimations de DBH et Hdom, c'est la densité d'arbres qui est estimée beaucoup plus grande sur le site INRA.

Les valeurs d'AGB présentées ont été calculées avec la relation $AGB = a \times BA \times Hdom$, déterminée à partir des données de référence du site INRA, avec $a = 0.1886$. De cette manière, une valeur de paramètre AGB peut être spatialisé aussi sur le site ONF à partir des estimations de BA et Hdom. Le procédé de calcul de l'AGB à partir des valeurs estimées d'autres paramètres semble robuste sur ce site et ces peuplements de pin maritime. En comparaison, l'AGB estimée et spatialisée directement à partir des données satellitaires sur le site INRA, où nous avons des valeurs AGB de référence, donne des statistiques similaire avec 64.1 tonnes/ha en moyenne et un écart type de 33.6. Ces valeurs sont très proches des statistiques indiquées dans le tableau IV.4 (moy. = 64.8 et E. T. = 37.9 tonnes/ha).

Tableau IV.5 : Statistiques descriptives (moyenne et écart type) des paramètres forestiers estimés sur l'ensemble des pixels de conifères des sites Landes INRA et ONF (pin maritime).

Landes	BA (m ² /ha)		DBH (cm)		Densité (arbre/ha)		Hdom (m)		AGB * (tonnes/ha)	
	moy.	E. T.	moy.	E. T.	moy.	E. T.	moy.	E. T.	moy.	E. T.
INRA	20.0	8.0	25.2	11.4	750	336	15.9	5.8	64.8	37.9
ONF	16.9	8.8	27.0	7.3	307	128	14.6	4.4	50.2	32.7

* AGB calculé à partir des estimations de BA et Hdom ($AGB = 0.1886 \times BA \times Hdom$)

La figure IV.11 présente la carte de prédiction d'AGB sur le site Landes-INRA. La spatialisation est utile pour localiser les zones avec une forte biomasse, mais également pour obtenir des statistiques d'ensemble. On compte ~1.2 million de tonnes de bois sec sur l'ensemble du site INRA (~18.7 milliers d'hectares de pin maritime), et ~1.1 million de tonnes sur le site ONF de la zone dunaire (~21.6 milliers d'hectares de pin maritime).

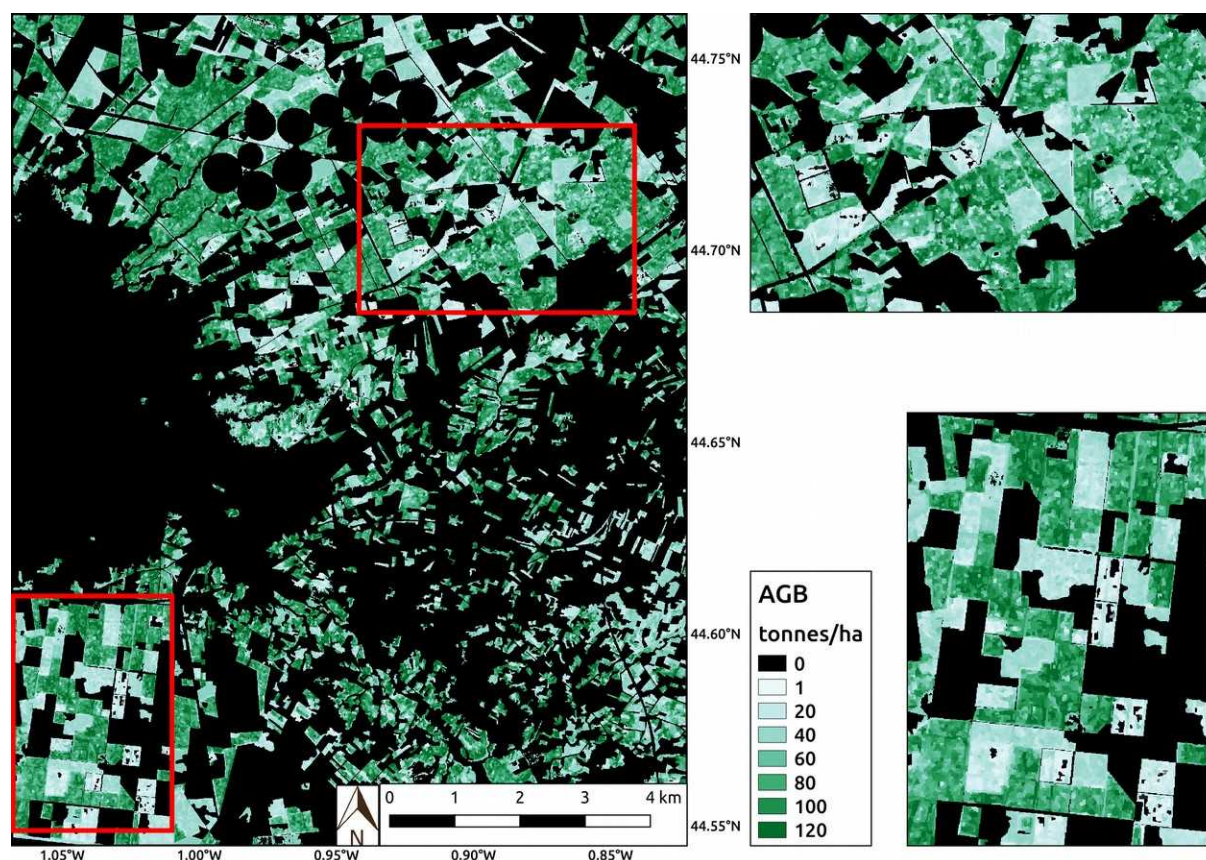


Figure IV.11 : Carte de prédiction de la biomasse aérienne de bois (AGB) pour les pins maritimes du site Landes-INRA.

La superposition des cartes de biomasse sur une carte de DBH nous permet d'aller plus loin dans l'analyse de la composition des sites. Le tableau IV.6 présente la répartition de la biomasse par tranche de DBH. Une grande partie de la biomasse est comprise dans les forêts avec un DBH entre 20 et 40 cm (65 % sur le site INRA, 85 % sur le site ONF). Les modes de gestion de ces types de forêts mono-spécifiques et équiennes sont souvent bien connu et harmonisés, en règle générale on opère une coupe rase sur les peuplements de pin maritime quand ils atteignent 40 à 50 cm de DBH. Ces cartes de paramètres forestiers peuvent donc nous renseigner sur la « pérennité » de la biomasse. Ici par exemple sur le site INRA, les 260 mille tonnes de bois sec du compartiment 0-20 cm peuvent être considérées comme potentiellement en expansion (ces forêts vont stocker plus de carbone), tandis que les ~165

mille tonnes de bois sec dans les forêts au-dessus de 40 cm de DBH seront potentiellement récoltées à courte échéance, le bilan carbone dépend alors de l'utilisation qui est faite du bois.

Tableau IV.6 : Répartition de la biomasse aérienne (AGB) par tranche de DBH moyen.

Site	DBH moyen (cm)				
	< 20 cm	20 - 40 cm	40 - 60 cm	60 - 80 cm	> 80 cm
INRA (tonnes)	260 431 tonnes	792 071 tonnes	144 820 tonnes	19 872 tonnes	88 tonnes
ONF (tonnes)	91 472 tonnes	922 425 tonnes	71 869 tonnes	-	-

4.4. Conclusion

Ces tests préliminaires pour la spatialisation de la méthode montrent que notre méthode est capable de produire des cartes à haute résolution des paramètres forestiers. Nous remarquons cependant certains défauts lors de la spatialisation des prédictions : i) nous pouvons rencontrer des problèmes de représentativité des valeurs d'apprentissage, révélés notamment par des problèmes de normalisation ; ii) avec un approche pixel, le bruit observé sur les images en entrée (notamment sur les ratios de bandes radar et les indices de texture) peut se répercuter sur les cartes produites, sans forcément représenter une réalité des forêts ; iii) les erreurs de classification des types de forêt (essences dans notre cas) ont une forte influence sur la qualité de la spatialisation.

Avec nos jeux de données de référence, il n'est pour l'instant pas possible de trancher clairement sur l'approche à utiliser pour la normalisation des primitives (images en entrée de l'application) et l'application pixel ou objet. Concernant l'identification des classes de forêts sur les images, avant l'application des modèles différenciés par classe, nous avons utilisé une méthode simple de classification par apprentissage avec les séries temporelles Sentinel-1 et Sentinel-2. Nous n'avons pas fait de tests comparatifs et de validation des classifications, mais plusieurs études montrent que ce type de méthode permet d'obtenir efficacement des cartes de l'occupation du sol et des essences forestières (Fassnacht et al., 2016; Immitzer et al., 2016; Inglada et al., 2017; Pelletier et al., 2016; Sheeren et al., 2016). Des recherches supplémentaires sur ce sujet permettraient d'améliorer cette étape de notre méthodologie.

Les cartes présentées n'ont pas été validées de manière indépendante (il n'y a pas de validation croisée leave-one-out, l'ensemble des mesures est utilisée pour calibrer le modèle qui est appliqué à tout l'image). La problématique de la validation des cartes est une question importante pour l'amélioration de la méthode et l'utilisation future de ce type de produit. D'autres cartes issues de données plus précises (par exemple Lidar ou photogrammétrie), ou des cartographies issues de plans de gestion pourraient être utilisées localement pour faire des inter-comparaisons. Sur des zones plus larges, il serait possible de travailler en collaboration avec l'inventaire forestier pour faire des statistiques comparatives. Ces comparaisons seraient possibles à l'échelle du département ou de grands peuplement, il faut une densité de points d'inventaire suffisamment importante pour obtenir des statistiques fiables.

Chapitre V : Conclusion

5.1. Conclusions

L'objectif général de la thèse vise à évaluer le potentiel de la télédétection spatiale pour produire des cartes de paramètres forestiers et/ou de biomasse avec une résolution spatiale décamétrique et mises à jour régulièrement, puis de mettre en place les bases d'une méthodologie automatique et générique pour la production de ces cartes. Nous avons décliné ces objectifs en plusieurs questions méthodologiques : 1) les images optiques et radar disponibles en libre accès mondialement peuvent-elles permettre d'estimer les paramètres forestiers avec une précision acceptable ($< 25\text{-}30\%$ d'erreurs), et leur combinaison permet-elle d'améliorer les résultats ? 2) Quels algorithmes et quelles méthodes de calibration utiliser pour faire ces estimations ? 3) Comment construire une méthode générique capable d'être utilisée pour les différents types de forêts ? Et 4) Comment spatialiser les estimations pour produire des cartes ?

Dans le chapitre II, nous avons montré que les données issues des séries temporelles Sentinel-1 et Sentinel-2, avec les mosaïques annuelles ALOS2-PALSAR2, étaient capable d'estimer les paramètres forestiers (AGB, BA, DBH, densité, hauteur et âge) sur des forêts équiennes et mono-spécifiques de pin maritime dans les Landes de Gascogne. Les résultats de validation en termes d'erreur relative vont de 10% pour la hauteur à 25% pour l'AGB. Ces erreurs sont comparables à celles observées dans les études avec des données 3D telles que le Lidar ou la photogrammétrie. Nous avons trouvé sur ces forêts, et sur d'autres forêts plus mélangées, que le SVR avec une méthode de sélection des primitives par itérations croissantes (méthode dite « stepwise » en forward) était plus performant que d'autres algorithmes (MLR et RF) et que d'autres méthodes de réduction de dimension (ACP, méthode de sélection backward, ou aucune sélection). Ces tests nous ont permis de mettre en évidence les synergies entre les données optiques, radar multi-fréquence et les indices de texture spatiale. Tous les types de données ensemble contribuent à produire les meilleurs résultats possibles avec la méthode proposée.

Dans le Chapitre III nous avons appliqué la méthode, mise au point dans le chapitre II, à différentes forêts sur 3 nouveaux sites d'étude. Nous avons montré la capacité de notre approche à produire de bonnes estimations sur des forêts plus complexes que les pins maritimes du premier site des Landes. Les résultats de validation en termes d'erreur relative vont de 5% à 30% selon les différentes essences pour les paramètres BA, DBH, densité et hauteur dominante. Les résultats sont meilleurs lorsque les essences sont différenciées les unes des autres pour la construction des modèles. Ces erreurs sur différents types de forêts sont du même ordre de grandeur que celles observées dans des études avec des données 3D de type Lidar et photogrammétrie (Holopainen et al., 2015; Pearse et al., 2018; Rahlf et al., 2014; Vastaranta et al., 2013; X. Yu et al., 2015). En plus de la problématique des modèles par essence, nous avons mis en évidence les difficultés posées pour une même essence d'arbre mais avec des gestions sylvicoles et du sous-bois/sol différentes. La transposition d'un modèle n'est pas possible directement d'un site à l'autre, en revanche la jointure des deux sites dans un même modèle peut permettre d'avoir des résultats presque équivalents au traitement individuel de chaque site. Pour aller plus loin à propos de la nomenclature, nous avons testé plusieurs rapprochements d'essences dans l'ensemble des 4 sites d'étude. Il apparaît que les conifères de nos jeux de données peuvent être traités ensemble avec de bons résultats d'estimation, tandis que les mêmes rapprochements pour les feuillus détériorent fortement la

qualité des estimations. Une approche qui pourrait être envisagée est de créer une nomenclature hiérarchique, avec plusieurs niveaux de classes imbriquées. Ceci permettrait de fournir les meilleurs résultats possibles lorsque l'information d'essence est disponible, et de remonter dans la hiérarchie à un niveau supérieur (regroupement d'essences avec caractéristiques similaires) lorsque cette information n'est pas disponible.

Par ailleurs, nous avons observé que les primitives utilisées pour chaque essence (ou classe de forêt) étaient différentes. La solution proposée de sélection automatique des primitives permet de spécialiser les modèles sur chaque forêt et paramètre forestier, comme faisant partie d'une chaîne de traitement de bout en bout avec la sélection des primitives, la paramétrisation et la calibration du modèle de régression SVR. Cette sélection des primitives améliore significativement les résultats (jusqu'à 10 % d'erreur relative en moins), et le fait de réduire le set de primitives en entrée de la sélection, sur la base de nos jeux de données, risquerait de trop spécialiser la méthode sur ces mêmes jeux de données. La contrepartie est que les primitives sélectionnées peuvent être différentes entre les classes de forêts, et ainsi accentuer le risque de mauvaises estimation lorsqu'on applique un modèle sur un type de forêt qui n'est pas rencontré dans l'apprentissage du modèle.

Dans le Chapitre IV, nous avons testé plusieurs approches pour la production des cartes : comment appliquer les modèles à une zone plus large autour de la zone des références. Avec nos données et un SVR, l'utilisation d'une approche pixel ou objet n'est pas déterminante (résultats très proches) pour la qualité des estimations. De même, la normalisation (nécessaire pour le SVR) des valeurs de pixels, pour l'application à l'ensemble des images, peut se faire : soit selon les valeurs utilisées en calibration des modèles (valeurs moyennées par polygones), soit avec les valeurs de l'ensemble des pixels forêt de l'image (ou la classe de forêt concernée). En l'état, avec les données des 4 sites d'étude et la méthode proposée, ce choix de normalisation n'a pas montré d'incidence significative sur la spatialisat ion. L'utilisation d'une approche objet (segmentation mean-shift) a principalement pour impact de réduire les artefacts sur les images, et de produire ainsi une cartographie plus « propre » des paramètres forestiers pour une meilleure visualisation. Malgré cette absence d'impact de l'approche pixel/objet et de la normalisation sur les valeurs de calibration, il serait recommandé, pour de futures utilisations de la méthode sur de nouveaux jeux de données, de continuer à tester toutes les approches. Au-delà de ces tests méthodologiques, nous avons aussi abordé la question des cartes de classification des forêts (par exemple, par essences) nécessaires à l'application des modèles différenciés. Les cartes librement disponibles ne répondent pas à nos besoins (pas assez de résolution spatiale, de classes, trop anciennes ou pas de bonne qualité). Plusieurs études ont montré que les méthodes de classification par apprentissage, avec des données de type Sentinel-1 ou Sentinel-2, peuvent permettre d'obtenir efficacement des cartes de l'occupation du sol et des essences forestières (Fassnacht et al., 2016; Immitzer et al., 2016; Sheeren et al., 2016). Nous avons utilisé les séries temporelles Sentinel-1 et Sentinel-2, avec un algorithme Random Forest, pour obtenir rapidement des cartes des essences.

Enfin, et de manière plus générale, l'une des complexités de la thèse porte sur l'obtention et le rassemblement de données de référence terrain. Grace aux équipes de l'ISPA-INRA et de l'ONF, nous avons pu travailler sur des jeux de données de très bonne qualité bien adaptés à une analyse par télédétection. L'utilisation d'image satellites et de modèles statistiques nécessite des données géolocalisées, au sein de petites zones relativement homogènes (0.5 à 1 ha), et couvrant un ensemble de valeurs des paramètres forestiers suffisamment large pour extraire des relations statistiques avec les images satellites. Nous avons également vu que les

informations sur le type de gestion, le sous-bois ou les conditions environnementales peuvent être nécessaires afin de discriminer efficacement des grands ensembles de forêts en entrée des modèles. L'accès aux jeux de données de mesures terrain est difficile auprès des organismes, et il est encore plus difficile d'obtenir des données qui répondent à tous les critères listés ci-dessus. Il y a pourtant de nombreuses mesures réalisées chaque année dans toutes les forêts en France. Les données des organismes forestiers nationaux ne sont pas tout le temps diffusées car non centralisées ou non numérisées, les données accessibles de l'inventaire forestier national ne sont pas précisément localisées pour des raisons de confidentialité, et les données des acteurs locaux sont souvent sur peu de forêts, sont hétérogènes dans la méthode, et sont peu numérisées. L'échange entre ces acteurs, de domaines différents et pas forcément proches du monde de la télédétection, devient vite complexe et difficile. Il n'est pas simple de rassembler ces jeux de données qui permettent l'entraînement et la validation des modèles issus de la télédétection pour la production de cartes de paramètres forestiers. L'une des réussites de la thèse est d'avoir commencé à réunir les acteurs des forêts et les faire parler ensemble. Au-delà des échanges avec des laboratoires de recherche publique, nous avons pu initier des travaux d'échanges scientifiques et de transmission de données avec le Centre National de la Propriété Forestière (CNPF), la recherche et développement de l'ONF, le laboratoire de l'inventaire forestier (LIF, IGN), ainsi que des entreprises travaillant sur les forêts (Terranis, Open Forêt), pour établir les bases de futures transmissions de compétences en télédétection spatiale. La mutualisation des savoirs, des méthodes et des données est une des clés pour aller vers une connaissance plus précise et un meilleur suivi local et national des forêts.

Le développement opérationnel de la télédétection spatiale des ressources forestières, telle que nous en avons exposé les bases, avec des données globales et libre accès, pourrait donner des outils supplémentaires aux différents acteurs autour des forêts. Pour l'inventaire forestier national, une cartographie haute résolution de ces paramètres, en amont des campagnes de mesures, peut aider à l'échantillonnage du territoire : différencier les zones avec des paramètres de structure homogènes (où moins de points de mesure sont nécessaires), ou au contraire très variable (où il faut donc densifier les points de mesure). Et en aval, à l'aide des mesures terrain compatibles, il serait possible de spatialiser plus finement les statistiques d'inventaire. De même, auprès des collectivités locales, des entreprises et des forestiers, des cartes haute résolution de la ressource forestière peuvent constituer un outil supplémentaire à ceux existants pour aider à planifier les campagnes de mesures et opérations sylvicoles sur les peuplements, et caractériser les ressources en bois dans le contexte des paysages, des trames vertes et de la biodiversité.

5.2. Perspectives

La perspective globale est donc de pouvoir appliquer ce type de méthode à l'échelle régionale ou nationale, voir dans d'autres pays. Cela présente des défis en termes de passage à l'échelle (quantité de données et traitements), mais aussi pour consolider la méthodologie. Plusieurs actions et axes d'analyses peuvent être cités :

i) Quelle est l'influence de la topographie sur les données satellitaires ? Les données optiques, les indices de texture, et particulièrement les données radar, sont affectés par les pentes. Il existe des techniques de correction radiométrique (Shepherd and Dymond, 2003; Small, 2011; Ulander, 1996) qui peuvent être appliquées dans certaines conditions. Il s'agit de mieux connaître leur domaine d'applicabilité et le résultat de leur application, et ainsi définir

les seuils de pente en dessous desquels les données satellitaires sont utilisables pour établir des modèles statistiques.

ii) La robustesse des modèles est déterminante pour le suivi temporel des forêts. Il faut tester la stabilité des estimations sur plusieurs années, et déterminer à partir de quand le rapport entre l'erreur du modèle et l'augmentation de la biomasse permet de constater des évolutions. De même, après une tempête et le déblayage des arbres impactés par exemple, peut-on constater la baisse de biomasse du peuplement ?

iii) Des indicateurs de confiance seraient très utiles pour une bonne utilisation des cartes de paramètres forestiers. En plus des statistiques de validation des modèles, des cartes de confiance peuvent être produites. Par exemple, il est possible de calculer un indice à partir des écarts entre les valeurs des pixels (vecteur, si on considère plusieurs primitives) où le modèle est appliqué et les valeurs utilisées en apprentissage du modèle (par exemple, écart à la moyenne en rapport de l'écart type).

iv) Enfin, des recherches doivent être faites sur la nomenclature en entrée des modèles. Les classes de forêt peuvent être définies selon les essences d'arbres, les groupes d'essences, les types de structure et de gestion, ou encore le sol et les conditions environnementales. Un travail de test (avec un plus large spectre de forêts) et d'expertise est nécessaire pour définir les classes de forêts qui permettront d'avoir des ensembles cohérents et pour construire de meilleurs modèles statistiques.

Bibliographie

- Abdullahi, S., Kugler, F., Pretzsch, H., 2016. Prediction of stem volume in complex temperate forest stands using TanDEM-X SAR data. *Remote Sens. Environ.* 174, 197–211. <https://doi.org/10.1016/j.rse.2015.12.012>
- Ahern, F.J., Leckie, D.J., Drieman, J.A., 1993. Seasonal changes in relative C-band backscatter of northern forest cover types. *IEEE Trans. Geosci. Remote Sens.* 31, 668–680. <https://doi.org/10.1109/36.225533>
- Andersen, H.-E., Barrett, T., Winterberger, K., Strunk, J., Temesgen, H., 2009. Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using airborne LiDAR and field plot data in a model-assisted sampling design. *IUFRO Div 4 Symp. Extending For. Inventory Monit.*
- Attarchi, S., Gloaguen, R., 2014. Improving the Estimation of Above Ground Biomass Using Dual Polarimetric PALSAR and ETM+ Data in the Hyrcanian Mountain Forest (Iran). *Remote Sens.* 6, 3693–3715. <https://doi.org/10.3390/rs6053693>
- Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., Coomes, D., Piboule, A., Vincent, G., 2019. A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests. *Remote Sens.* 11, 1086. <https://doi.org/10.3390/rs11091086>
- Austin, J.M., Mackey, B.G., Van Niel, K.P., 2003. Estimating forest biomass using satellite radar: an exploratory study in a temperate Australian Eucalyptus forest. *For. Ecol. Manag.* 176, 575–583. [https://doi.org/10.1016/S0378-1127\(02\)00314-6](https://doi.org/10.1016/S0378-1127(02)00314-6)
- Baghdadi, N., le Maire, G., Fayad, I., Bailly, J.S., Nouvellon, Y., Lemos, C., Hakamada, R., 2014. Testing Different Methods of Forest Height and Aboveground Biomass Estimations From ICESat/GLAS Data in Eucalyptus Plantations in Brazil. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7, 290–299. <https://doi.org/10.1109/JSTARS.2013.2261978>
- Baghdadi, N., Maire, G.L., Bailly, J., Osé, K., Nouvellon, Y., Zribi, M., Lemos, C., Hakamada, R., 2015. Evaluation of ALOS/PALSAR L-Band Data for the Estimation of Eucalyptus Plantations Aboveground Biomass in Brazil. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8, 3802–3811. <https://doi.org/10.1109/JSTARS.2014.2353661>
- Bastin, J.-F., Barbier, N., Couteron, P., Adams, B., Shapiro, A., Bogaert, J., Cannière, C.D., 2014. Aboveground biomass mapping of African forest mosaics using canopy texture analysis: toward a regional approach. *Ecol. Appl.* 24, 1984–2001. <https://doi.org/10.1890/13-1574.1>
- Beaudoin, A., Le Toan, T., Goze, S., Nezry, E., Lopes, A., Mougin, E., Hsu, C.C., Han, H.C., Kong, J.A., Shin, R.T., 1994. Retrieval of forest biomass from SAR data. *Int. J. Remote Sens.* 15, 2777–2796. <https://doi.org/10.1080/01431169408954284>
- Beguet, B., Guyon, D., Boukir, S., Chehata, N., 2014. Automated retrieval of forest structure variables based on multi-scale texture analysis of VHR satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 96, 164–178. <https://doi.org/10.1016/j.isprsjprs.2014.07.008>
- Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Berninger, A., Lohberger, S., Stängel, M., Siegert, F., 2018. SAR-Based Estimation of Above-Ground Biomass and Its Changes in Tropical Forests of Kalimantan Using L- and C-Band. *Remote Sens.* 10, 831. <https://doi.org/10.3390/rs10060831>

- Berninger, A., Lohberger, S., Zhang, D., Siegert, F., 2019. Canopy Height and Above-Ground Biomass Retrieval in Tropical Forests Using Multi-Pass X- and C-Band Pol-InSAR Data. *Remote Sens.* 11, 2105. <https://doi.org/10.3390/rs11182105>
- Bolton, D.K., White, J.C., Wulder, M.A., Coops, N.C., Hermosilla, T., Yuan, X., 2018. Updating stand-level forest inventories using airborne laser scanning and Landsat time series data. *Int. J. Appl. Earth Obs. Geoinformation* 66, 174–183. <https://doi.org/10.1016/j.jag.2017.11.016>
- Boukir, S., Orny, C., Chehata, N., Guyon, D., Wigneron, J.-P., 2013. Détection de changements structurels sur des images satellite haute résolution. Application en milieu forestier. *Trait. Signal* 30, 401–429. <https://doi.org/10.3166/ts.30.401-429>
- Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., Asner, G.P., 2018. An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR. *Remote Sens. Environ.* 206, 156–173. <https://doi.org/10.1016/j.rse.2017.12.030>
- Bouvier, M., Durrieu, S., Fournier, R.A., Renaud, J.-P., 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sens. Environ.* 156, 322–334. <https://doi.org/10.1016/j.rse.2014.10.004>
- Brede, B., Verrelst, J., Gastellu-Etchegorry, J.-P., Clevers, J.G.P.W., Goudzwaard, L., den Ouden, J., Verbesselt, J., Herold, M., 2020. Assessment of Workflow Feature Selection on Forest LAI Prediction with Sentinel-2A MSI, Landsat 7 ETM+ and Landsat 8 OLI. *Remote Sens.* 12, 915. <https://doi.org/10.3390/rs12060915>
- Bruniquel, J., Lopes, A., 1997. Multi-variate optimal speckle reduction in SAR imagery. *Int. J. Remote Sens.* 18, 603–627. <https://doi.org/10.1080/014311697218962>
- Carreiras, J.M.B., Vasconcelos, M.J., Lucas, R.M., 2012. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sens. Environ.* 121, 426–442. <https://doi.org/10.1016/j.rse.2012.02.012>
- Castel, T., 2002. Retrieval biomass of a large Venezuelan pine plantation using JERS-1 SAR data. Analysis of forest structure impact on radar signature. *Remote Sens. Environ.* 79, 30–41. [https://doi.org/10.1016/S0034-4257\(01\)00236-X](https://doi.org/10.1016/S0034-4257(01)00236-X)
- Castillo, J.A.A., Apan, A.A., Maraseni, T.N., Salmo, S.G., 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS J. Photogramm. Remote Sens.* 134, 70–85. <https://doi.org/10.1016/j.isprsjprs.2017.10.016>
- Castillo-Santiago, M., Ricker, M., Jong, B., 2010. Estimation of tropical forest structure from SPOT5 satellite images. *Int. J. Remote Sens. - INT J REMOTE SENS* 31, 2767–2782. <https://doi.org/10.1080/01431160903095460>
- Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B., Yamakura, T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145, 87–99. <https://doi.org/10.1007/s00442-005-0100-x>
- Chen, L., Ren, C., Zhang, B., Wang, Z., Xi, Y., 2018. Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. *Forests* 9, 582. <https://doi.org/10.3390/f9100582>
- Chen, L., Wang, Y., Ren, C., Zhang, B., Wang, Z., 2019. Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data. *Remote Sens.* 11, 414. <https://doi.org/10.3390/rs11040414>
- Chi, H., Sun, G., Huang, J., Guo, Z., Ni, W., Fu, A., 2015. National Forest Aboveground Biomass Mapping from ICESat/GLAS Data and MODIS Imagery in China. *Remote Sens.* 7, 5534–5564. <https://doi.org/10.3390/rs70505534>

- Chi, H., Sun, G., Huang, J., Li, R., Ren, X., Ni, W., Fu, A., 2017. Estimation of Forest Aboveground Biomass in Changbai Mountain Region Using ICESat/GLAS and Landsat/TM Data. *Remote Sens.* 9, 707. <https://doi.org/10.3390/rs9070707>
- Choi, C., Pardini, M., Papathanassiou, K., 2019. A Structure-Based Framework for the Combination of GEDI and Tandem-X Measurements Over Forest Scenarios, in: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. Presented at the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 4488–4490. <https://doi.org/10.1109/IGARSS.2019.8898130>
- Colin A., Thivolle-Cazat A., 2016. Disponibilités forestières pour l'énergie et les matériaux à l'horizon 2035, Expertises. IGN, FCBA, ADEME.
- Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 603–619. <https://doi.org/10.1109/34.1000236>
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297. <https://doi.org/10.1007/BF00994018>
- Couteron, P., Pelissier, R., Nicolini, E.A., Paget, D., 2005. Predicting tropical forest stand structure parameters from Fourier transform of very high-resolution remotely sensed canopy images. *J. Appl. Ecol.* 42, 1121–1128. <https://doi.org/10.1111/j.1365-2664.2005.01097.x>
- Currie, W.S., Bergen, K.M., 2008. Temperate Forest, in: *Encyclopedia of Ecology*. Elsevier, pp. 3494–3503. <https://doi.org/10.1016/B978-008045405-4.00704-7>
- d'Oliveira, M.V.N., Reutebuch, S.E., McGaughey, R.J., Andersen, H.-E., 2012. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sens. Environ.* 124, 479–491. <https://doi.org/10.1016/j.rse.2012.05.014>
- Dassot, M., Constant, T., Fournier, M., 2011. The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges. *Ann. For. Sci.* 68, 959–974. <https://doi.org/10.1007/s13595-011-0102-2>
- Davi, H., Soudani, K., Deckx, T., Dufrene, E., Dantec, V.L., François, C., 2006. Estimation of forest leaf area index from SPOT imagery using NDVI distribution over forest stands. *Int. J. Remote Sens.* 27, 885–902. <https://doi.org/10.1080/01431160500227896>
- Debastiani, A.B., Sanquetta, C.R., Corte, A.P.D., Pinto, N.S., Rex, F.E., 2019. Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest. *Ann. For. Res.* 0. <https://doi.org/10.15287/afr.2018.1267>
- Deleuze, C., Morneau, F., Renaud, J.P., Vivien, Y., Rivoire, M., Santenoise, P., Longuetaud, F., Mothe, F., Hervé, J.C., Vallet, P., 2014. Estimer le volume total d'un arbre, quelles que soient l'essence, la taille, la sylviculture, la station 12.
- Deo, R., Russell, M., Domke, G., Andersen, H.-E., Cohen, W., Woodall, C., 2017. Evaluating Site-Specific and Generic Spatial Models of Aboveground Forest Biomass Based on Landsat Time-Series and LiDAR Strip Samples in the Eastern USA. *Remote Sens.* 9, 598. <https://doi.org/10.3390/rs9060598>
- Dixon, R.K., Brown, S., Houghton, R.A., Solomon, A.M., Trexler, M.C., Wisniewski, J., 1994. Carbon Pools and Flux of Global Forest Ecosystems. *Sci. New Ser.* 263, 185–190.
- Dobson, M.C., Ulaby, F.T., Pierce, L.E., Sharik, T.L., Lin, C., Sarabandi, K., 1995. Estimation of Forest Biophysical Characteristics in Northern Michigan with SIR-C/X-SAR. *IEEE Trans. Geosci. Remote Sens.* 33, 19.
- Donager, J.J., Sankey, T.T., Sankey, J.B., Sanchez Meador, A.J., Springer, A.E., Bailey, J.D., 2018. Examining Forest Structure With Terrestrial Lidar: Suggestions and Novel

- Techniques Based on Comparisons Between Scanners and Forest Treatments. *Earth Space Sci.* 5, 753–776. <https://doi.org/10.1029/2018EA000417>
- Dong, J., Kaufmann, R.K., Myneni, R.B., Tucker, C.J., Kauppi, P.E., Liski, J., Buermann, W., Alexeyev, V., Hughes, M.K., 2003. Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sens. Environ.* 84, 393–410. [https://doi.org/10.1016/S0034-4257\(02\)00130-X](https://doi.org/10.1016/S0034-4257(02)00130-X)
- Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines, in: *Proceedings of the 9th International Conference on Neural Information Processing Systems, NIPS'96*. MIT Press, Denver, Colorado, pp. 155–161.
- Dube, T., Mutanga, O., 2015a. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS J. Photogramm. Remote Sens.* 101, 36–46. <https://doi.org/10.1016/j.isprsjprs.2014.11.001>
- Dube, T., Mutanga, O., 2015b. Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS J. Photogramm. Remote Sens.* 108, 12–32. <https://doi.org/10.1016/j.isprsjprs.2015.06.002>
- Dube, T., Sibanda, M., Shoko, C., Mutanga, O., 2017. Stand-volume estimation from multi-source data for coppiced and high forest *Eucalyptus* spp. silvicultural systems in KwaZulu-Natal, South Africa. *ISPRS J. Photogramm. Remote Sens.* 132, 162–169. <https://doi.org/10.1016/j.isprsjprs.2017.09.001>
- Duncanson, L.I., Niemann, K.O., Wulder, M.A., 2010. Integration of GLAS and Landsat TM data for aboveground biomass estimation. *Can. J. Remote Sens.* 36, 129–141. <https://doi.org/10.5589/m10-037>
- FAO, 2018. LA SITUATION DES FORETS DU MONDE 2018 (SOFO): les forets au service du developpement durable. FOOD & AGRICULTURE ORG, Place of publication not identified.
- FAO, 2015. Global forest resources assessment 2015.
- Fassnacht, F.E., Lati, H., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* 24.
- Fassnacht, K.S., Gower, S.T., MacKenzie, M.D., Nordheim, E.V., Lillesand, T.M., 1997. Estimating the leaf area index of North Central Wisconsin forests using the landsat thematic mapper. *Remote Sens. Environ.* 61, 229–245. [https://doi.org/10.1016/S0034-4257\(97\)00005-9](https://doi.org/10.1016/S0034-4257(97)00005-9)
- Flamenco, H.N., Gonzalez-Benecke, C.A., Wightman, M.G., 2019. Long-term effects of vegetation management on biomass stock of four coniferous species in the Pacific Northwest United States. *For. Ecol. Manag.* 432, 276–285. <https://doi.org/10.1016/j.foreco.2018.09.033>
- Franklin, S., Maudie, A.J., Lavigne, M., 2001. Using spatial co-occurrence texture to increase forest structure and species composition classification accuracy. *Photogramm. Eng. Remote Sens.* 67, 849–855.
- Freitas, S.R., Mello, M.C.S., Cruz, C.B.M., 2005. Relationships between forest structure and vegetation indices in Atlantic Rainforest. *For. Ecol. Manag.* 218, 353–362. <https://doi.org/10.1016/j.foreco.2005.08.036>
- Frison, P.-L., Fruneau, B., Kmiha, S., Soudani, K., Dufrêne, E., Toan, T.L., Koleček, T., Villard, L., Mougin, E., Rudant, J.-P., 2018. Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. *Remote Sens.* 10, 2049. <https://doi.org/10.3390/rs10122049>

- Gao, T., Zhu, J., Deng, S., Zheng, X., Zhang, J., Shang, G., Huang, L., 2016. Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history. *Int. J. Appl. Earth Obs. Geoinformation* 52, 155–165. <https://doi.org/10.1016/j.jag.2016.06.004>
- Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data Softw. Stand.* 2. <https://doi.org/10.1186/s40965-017-0031-6>
- Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VEN μ S, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* 114, 1747–1755. <https://doi.org/10.1016/j.rse.2010.03.002>
- Hagolle, O., Huc, M., Villa Pascual, D., Dedieu, G., 2015. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VEN μ S and Sentinel-2 Images. *Remote Sens.* 7, 2668–2691. <https://doi.org/10.3390/rs70302668>
- Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. *Remote Sens. Environ.* 185, 221–232. <https://doi.org/10.1016/j.rse.2016.02.023>
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* SMC-3, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Harding, D.J., Carabajal, C.C., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophys. Res. Lett.* 32. [https://doi.org/10.1029/2005GL023471@10.1002/\(ISSN\)1944-8007.ICESAT1](https://doi.org/10.1029/2005GL023471@10.1002/(ISSN)1944-8007.ICESAT1)
- Harrell, P.A., Kasischke, E.S., Bourgeau-Chavez, L.L., Haney, E.M., Christensen, N.L., 1997. Evaluation of approaches to estimating aboveground biomass in Southern pine forests using SIR-C data. *Remote Sens. Environ.* 59, 223–233. [https://doi.org/10.1016/S0034-4257\(96\)00155-1](https://doi.org/10.1016/S0034-4257(96)00155-1)
- Haywood, A., Stone, C., Jones, S., 2018. The Potential of Sentinel Satellites for Large Area Aboveground Forest Biomass Mapping, in: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. Presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Valencia, pp. 9030–9033. <https://doi.org/10.1109/IGARSS.2018.8517597>
- Ho Tong Minh, D., Le Toan, T., Rocca, F., Tebaldini, S., Villard, L., Réjou-Méchain, M., Phillips, O.L., Feldpausch, T.R., Dubois-Fernandez, P., Scipal, K., Chave, J., 2016. SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. *Remote Sens. Environ.* 175, 138–147. <https://doi.org/10.1016/j.rse.2015.12.037>
- Holopainen, M., Vastaranta, M., Karjalainen, M., Karila, K., Kaasalainen, S., Honkavaara, E., Hyypä, J., 2015. FOREST INVENTORY ATTRIBUTE ESTIMATION USING AIRBORNE LASER SCANNING, AERIAL STEREO IMAGERY, RADARGRAMMETRY AND INTERFEROMETRY—FINNISH EXPERIENCES OF THE 3D TECHNIQUES. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* II-3/W4, 63–69. <https://doi.org/10.5194/isprsannals-II-3-W4-63-2015>
- Imhoff, M.L., 1993. Radar backscatter/biomass saturation: observations and implications for global biomass assessment, in: *Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium*. Presented at the Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium, pp. 43–45 vol.1. <https://doi.org/10.1109/IGARSS.1993.322465>

- Immitzer, M., Vuolo, F., Atzberger, C., 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe 28.
- Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sens.* 9, 95. <https://doi.org/10.3390/rs9010095>
- Irulappa-Pillai-Vijayakumar, D.B., Renaud, J.-P., Morneau, F., McRoberts, R.E., Vega, C., 2019. Increasing Precision for French Forest Inventory Estimates using the k-NN Technique with Optical and Photogrammetric Data and Model-Assisted Estimators. *Remote Sens.* 11, 991. <https://doi.org/10.3390/rs11080991>
- Jayathunga, S., Owari, T., Tsuyuki, S., 2019. Digital Aerial Photogrammetry for Uneven-Aged Forest Management: Assessing the Potential to Reconstruct Canopy Structure and Estimate Living Biomass. *Remote Sens.* 11, 338. <https://doi.org/10.3390/rs11030338>
- Jochem, A., Hollaus, M., Rutzinger, M., Höfle, B., 2010. Estimation of Aboveground Biomass in Alpine Forests: A Semi-Empirical Approach Considering Canopy Transparency Derived from Airborne LiDAR Data. *Sensors* 11, 278–295. <https://doi.org/10.3390/s110100278>
- Joshi, N., Mitchard, E., Schumacher, J., Johannsen, V., Saatchi, S., Fensholt, R., 2015. L-Band SAR Backscatter Related to Forest Cover, Height and Aboveground Biomass at Multiple Spatial Scales across Denmark. *Remote Sens.* 7, 4442–4472. <https://doi.org/10.3390/rs70404442>
- Karila, K., Vastaranta, M., Karjalainen, M., Kaasalainen, S., 2015. Tandem-X interferometry in the prediction of forest inventory attributes in managed boreal forests. *Remote Sens. Environ.* 159, 259–268. <https://doi.org/10.1016/j.rse.2014.12.012>
- Kayitakire, F., Hamel, C., Defourny, P., 2006. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sens. Environ.* 102, 390–401. <https://doi.org/10.1016/j.rse.2006.02.022>
- Kempes, C.P., West, G.B., Crowell, K., Girvan, M., 2011. Predicting Maximum Tree Heights and Other Traits from Allometric Scaling and Resource Limitations. *PLoS ONE* 6, e20551. <https://doi.org/10.1371/journal.pone.0020551>
- Koch, B., 2013. Remote Sensing supporting national forest inventories NFA 15.
- Laurin, G.V., Balling, J., Corona, P., Mattioli, W., Papale, D., Puletti, N., Rizzo, M., Truckenbrodt, J., Urban, M., 2018. Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *J. Appl. Remote Sens.* 12, 1. <https://doi.org/10.1117/1.JRS.12.016008>
- le Maire, G., Marsden, C., Nouvellon, Y., Grinand, C., Hakamada, R., Stape, J.-L., Laclau, J.-P., 2011. MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass. *Remote Sens. Environ.* 115, 2613–2625. <https://doi.org/10.1016/j.rse.2011.05.017>
- le Maire, G., le François, C., Soudani, K., Davi, H., Dantec, V.L., Saugier, B., Dufrêne, E., 2006. Forest leaf area index determination: A multiyear satellite-independent method based on within-stand normalized difference vegetation index spatial variability. *J. Geophys. Res. Biogeosciences* 111. <https://doi.org/10.1029/2005JG000122>
- Le Quéré, C., Andrew, R.M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P.A., Korsbakken, J.I., Peters, G.P., Canadell, J.G., Arneth, A., Arora, V.K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L.P., Ciais, P., Doney, S.C., Gkritzalis, T., Goll, D.S., Harris, I., Haverd, V., Hoffman, F.M., Hoppema, M., Houghton, R.A., Hurtt, G., Ilyina, T., Jain, A.K., Johannessen, T., Jones, C.D., Kato, E., Keeling, R.F., Goldewijk, K.K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger,

- J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F.N., van der Laan-Luijkx, I.T., van der Werf, G.R., Viovy, N., Walker, A.P., Wiltshire, A.J., Wright, R., Zaehle, S., Zheng, B., 2018. Global Carbon Budget 2018. *Earth Syst. Sci. Data* 10, 2141–2194. <https://doi.org/10.5194/essd-10-2141-2018>
- Le Toan, T., 2007. SAR data for forestry and agriculture.
- Le Toan, T., Beaudoin, A., Riom, J., Guyon, D., 1992. Relating forest biomass to SAR data. *IEEE Trans. Geosci. Remote Sens.* 30, 403–411. <https://doi.org/10.1109/36.134089>
- Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., Ulander, L., 2011. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sens. Environ.* 115, 2850–2860. <https://doi.org/10.1016/j.rse.2011.03.020>
- Lee, S.-K., Fatoyinbo, T., Marselis, S.M., Qi, W., Hancock, S., Armston, J., Dubayah, R., 2019. Spaceborne Data Fusion for Large-Scale Forest Parameter Estimation: GEDI Lidar Tandem-X INSAR Missions, in: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. Presented at the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 4491–4494. <https://doi.org/10.1109/IGARSS.2019.8899854>
- Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo, F.D.B., Hunter, M.O., Oliveira, R. de, 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophys. Res. Lett.* 32. <https://doi.org/10.1029/2005GL023971>
- Li, A., Huang, C., Sun, G., Shi, H., Toney, C., Zhu, Z., Rollins, M.G., Goward, S.N., Masek, J.G., 2011. Modeling the height of young forests regenerating from recent disturbances in Mississippi using Landsat and ICESat data. *Remote Sens. Environ.* 115, 1837–1849. <https://doi.org/10.1016/j.rse.2011.03.001>
- Lisein, J., Bonnet, S., Lejeune, P., Deseilligny, M., 2014. Modélisation de la canopée forestière par photogrammétrie depuis des images acquises par drone. *Rev. Française Photogramm. Teledetection* 45–54.
- Liu, K., Wang, J., Zeng, W., Song, J., 2017. Comparison and Evaluation of Three Methods for Estimating Forest above Ground Biomass Using TM and GLAS Data. *Remote Sens.* 9, 341. <https://doi.org/10.3390/rs9040341>
- Liu, Y., Gong, W., Xing, Y., Hu, X., Gong, J., 2019. Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS J. Photogramm. Remote Sens.* 151, 277–289. <https://doi.org/10.1016/j.isprsjprs.2019.03.016>
- Lu, D., 2005. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *Int. J. Remote Sens.* 26, 2509–2525. <https://doi.org/10.1080/01431160500142145>
- Luckman, A., Baker, J., Honzák, M., Lucas, R., 1998. Tropical Forest Biomass Density Estimation Using JERS-1 SAR: Seasonal Variation, Confidence Limits, and Application to Image Mosaics. *Remote Sens. Environ.* 63, 126–139. [https://doi.org/10.1016/S0034-4257\(97\)00133-8](https://doi.org/10.1016/S0034-4257(97)00133-8)
- Luyssaert, S., Schulze, E.-D., Börner, A., Knohl, A., Hessenmöller, D., Law, B.E., Ciais, P., Grace, J., 2008. Old-growth forests as global carbon sinks. *Nature* 455, 213–215. <https://doi.org/10.1038/nature07276>
- Macedo, F.L., Sousa, A.M.O., Gonçalves, A.C., Marques da Silva, J.R., Mesquita, P.A., Rodrigues, R.A.F., 2018. Above-ground biomass estimation for *Quercus rotundifolia* using vegetation indices derived from high spatial resolution satellite images. *Eur. J. Remote Sens.* 51, 932–944. <https://doi.org/10.1080/22797254.2018.1521250>

- Marsden, C., le Maire, G., Stape, J.-L., Seen, D.L., Roupsard, O., Cabral, O., Epron, D., Lima, A.M.N., Nouvellon, Y., 2010. Relating MODIS vegetation index time-series with structure, light absorption and stem production of fast-growing Eucalyptus plantations. *For. Ecol. Manag.* 259, 1741–1753. <https://doi.org/10.1016/j.foreco.2009.07.039>
- Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S.J., 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sens. Environ.* 209, 90–106. <https://doi.org/10.1016/j.rse.2017.12.020>
- Mcroberts, R., Tomppo, E., 2007. Remote sensing support for national forest inventories. *Remote Sens. Environ.* 110, 412–419. <https://doi.org/10.1016/j.rse.2006.09.034>
- Melnikova, I., Awaya, Y., Saitoh, T.M., Muraoka, H., Sasai, T., 2018. Estimation of Leaf Area Index in a Mountain Forest of Central Japan with a 30-m Spatial Resolution Based on Landsat Operational Land Imager Imagery: An Application of a Simple Model for Seasonal Monitoring. *Remote Sens.* 10, 179. <https://doi.org/10.3390/rs10020179>
- Meng, S., Pang, Y., Zhang, Z., Jia, W., Li, Z., 2016. Mapping Aboveground Biomass using Texture Indices from Aerial Photos in a Temperate Forest of Northeastern China. *Remote Sens.* 8, 230. <https://doi.org/10.3390/rs8030230>
- Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., Seifert-Granzin, J., 2014. Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sens. Environ.* 155, 109–119. <https://doi.org/10.1016/j.rse.2014.01.029>
- Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., Gourlet-Fleury, S., 2015. Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sens. Environ.* 159, 307–317. <https://doi.org/10.1016/j.rse.2014.12.019>
- Mitchard, E.T.A., Saatchi, S.S., Lewis, S.L., Feldpausch, T.R., Woodhouse, I.H., Sonké, B., Rowland, C., Meir, P., 2011. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sens. Environ.* 115, 2861–2873. <https://doi.org/10.1016/j.rse.2010.02.022>
- Mitchard, E.T.A., Saatchi, S.S., White, L.J.T., Abernethy, K.A., Jeffery, K.J., Lewis, S.L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H., Meir, P., 2012. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences* 9, 179–191. <https://doi.org/10.5194/bg-9-179-2012>
- Mitchard, E.T.A., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N.S., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T.R., Meir, P., 2009. Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophys. Res. Lett.* 36. <https://doi.org/10.1029/2009GL040692>
- Motte, E., Zribi, M., Fanise, P., Egido, A., Darrozes, J., Al-Yaari, A., Baghdadi, N., Baup, F., Dayau, S., Fieuzal, R., Frison, P.-L., Guyon, D., Wigneron, J.-P., 2016. GLORI: A GNSS-R Dual Polarization Airborne Instrument for Land Surface Monitoring. *Sensors* 16, 732. <https://doi.org/10.3390/s16050732>
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* 66, 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* 25, 3999–4014. <https://doi.org/10.1080/01431160310001654923>
- Nelson, R., Ranson, K.J., Sun, G., Kimes, D.S., Kharuk, V., Montesano, P., 2009. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sens. Environ.* 113, 691–701. <https://doi.org/10.1016/j.rse.2008.11.010>

- Ni, W., Woodcock, C.E., Jupp, D.L.B., 1999. Variance in Bidirectional Reflectance over Discontinuous Plant Canopies. *Remote Sens. Environ.* 69, 1–15. [https://doi.org/10.1016/S0034-4257\(98\)00125-4](https://doi.org/10.1016/S0034-4257(98)00125-4)
- Nicolas, H., Dusseux, P., 2016. La télédétection et la gestion du pâturage.
- Noordermeer, L., Bollandsås, O.M., Ørka, H.O., Næsset, E., Gobakken, T., 2019. Comparing the accuracies of forest attributes predicted from airborne laser scanning and digital aerial photogrammetry in operational forest inventories. *Remote Sens. Environ.* 226, 26–37. <https://doi.org/10.1016/j.rse.2019.03.027>
- Ozdemir, I., Karnieli, A., 2011. Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *Int. J. Appl. Earth Obs. Geoinformation* 13, 701–710. <https://doi.org/10.1016/j.jag.2011.05.006>
- Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* 26, 1007–1011. <https://doi.org/10.1080/01431160512331314083>
- Pearse, G.D., Dash, J.P., Persson, H.J., Watt, M.S., 2018. Comparison of high-density LiDAR and satellite photogrammetry for forest inventory. *ISPRS J. Photogramm. Remote Sens.* 142, 257–267. <https://doi.org/10.1016/j.isprsjprs.2018.06.006>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 187, 156–168. <https://doi.org/10.1016/j.rse.2016.10.010>
- Persson, H., Fransson, J.E.S., 2014. Forest Variable Estimation Using Radargrammetric Processing of TerraSAR-X Images in Boreal Forests. *Remote Sens.* 6, 2084–2107. <https://doi.org/10.3390/rs6032084>
- Persson, H.J., 2016. Estimation of Boreal Forest Attributes from Very High Resolution Pléiades Data. *Remote Sens.* 8, 736. <https://doi.org/10.3390/rs8090736>
- Piermattei, L., Karel, W., Wang, D., Wieser, M., Mokroš, M., Surový, P., Koreň, M., Tomašík, J., Pfeifer, N., Hollaus, M., 2019. Terrestrial Structure from Motion Photogrammetry for Deriving Forest Inventory Data. *Remote Sens.* 11, 950. <https://doi.org/10.3390/rs11080950>
- Proisy, C., 1999. Apport des données radar à synthèse d'ouverture pour l'étude de la dynamique des écosystèmes forestiers. Université Toulouse 3 Paul Sabatier.
- Proisy, C., Couteron, P., Fromard, F., 2007. Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. *Remote Sens. Environ.* 109, 379–392. <https://doi.org/10.1016/j.rse.2007.01.009>
- Qi, W., Dubayah, R.O., 2016. Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping. *Remote Sens. Environ.* 187, 253–266. <https://doi.org/10.1016/j.rse.2016.10.018>
- Qi, W., Lee, S.-K., Hancock, S., Luthcke, S., Tang, H., Armston, J., Dubayah, R., 2019a. Improved forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. *Remote Sens. Environ.* 221, 621–634. <https://doi.org/10.1016/j.rse.2018.11.035>
- Qi, W., Saarela, S., Armston, J., Ståhl, G., Dubayah, R., 2019b. Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sens. Environ.* 232, 111283. <https://doi.org/10.1016/j.rse.2019.111283>
- Quegan, S., Yu, J.J., 2001. Filtering of multichannel SAR images. *IEEE Trans. Geosci. Remote Sens.* 39, 2373–2379. <https://doi.org/10.1109/36.964973>

- Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., Astrup, R., 2014. Comparison of four types of 3D data for timber volume estimation. *Remote Sens. Environ.* 155, 325–333. <https://doi.org/10.1016/j.rse.2014.08.036>
- Rizzoli, P., Martone, M., Gonzalez, C., Wecklich, C., Borla Tridon, D., Bräutigam, B., Bachmann, M., Schulze, D., Fritz, T., Huber, M., Wessel, B., Krieger, G., Zink, M., Moreira, A., 2017. Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS J. Photogramm. Remote Sens.* 132, 119–139. <https://doi.org/10.1016/j.isprsjprs.2017.08.008>
- Rosenqvist, A., Shimada, M., Milne, A.K., 2007. The ALOS Kyoto & carbon initiative. Presented at the International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3614–3617. <https://doi.org/10.1109/IGARSS.2007.4423628>
- Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., Morel, A., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci.* 108, 9899–9904. <https://doi.org/10.1073/pnas.1019576108>
- Santoro, M., Cartus, O., 2018. Research Pathways of Forest Above-Ground Biomass Estimation Based on SAR Backscatter and Interferometric SAR Observations. *Remote Sens.* 10, 608. <https://doi.org/10.3390/rs10040608>
- Santoro, M., Cartus, O., Fransson, J.E.S., Wegmüller, U., 2019. Complementarity of X-, C-, and L-band SAR Backscatter Observations to Retrieve Forest Stem Volume in Boreal Forest. *Remote Sens.* 11, 1563. <https://doi.org/10.3390/rs11131563>
- Santoro, M., Eriksson, L., Fransson, J., 2015. Reviewing ALOS PALSAR Backscatter Observations for Stem Volume Retrieval in Swedish Forest. *Remote Sens.* 7, 4290–4317. <https://doi.org/10.3390/rs70404290>
- Sarker, L.R., Nichol, J.E., 2011. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ.* 115, 968–977. <https://doi.org/10.1016/j.rse.2010.11.010>
- Sarker, Md.L.R., Nichol, J., Ahmad, B., Busu, I., Rahman, A.A., 2012. Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation. *ISPRS J. Photogramm. Remote Sens.* 69, 146–166. <https://doi.org/10.1016/j.isprsjprs.2012.03.002>
- Shaiek, O., Loustau, D., Trichet, P., Meredieu, C., Bachtobji, B., Garchi, S., EL Aouni, M.H., 2011. Generalized biomass equations for the main aboveground biomass components of maritime pine across contrasting environments. *Ann. For. Sci.* 68, 443–452. <https://doi.org/10.1007/s13595-011-0044-8>
- Sheeren, D., Fauvel, M., Josipovic, V., Lopes, M., Planque, C., Willm, J., Dejoux, J.-F., 2016. Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series 29.
- Shepherd, J.D., Dymond, J.R., 2003. Correcting satellite imagery for the variance of reflectance and illumination with topography. *Int. J. Remote Sens.* 24, 3503. <https://doi.org/10.1080/01431160210154029>
- Singh, M., Evans, D., Friess, D., Tan, B., Nin, C., 2015. Mapping Above-Ground Biomass in a Tropical Forest in Cambodia Using Canopy Textures Derived from Google Earth. *Remote Sens.* 7, 5057–5076. <https://doi.org/10.3390/rs70505057>
- Small, D., 2011. Flattening Gamma: Radiometric Terrain Correction for SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* 49, 3081–3093. <https://doi.org/10.1109/TGRS.2011.2120616>
- Song, C., Dickinson, M.B., Su, L., Zhang, S., Yaussey, D., 2010. Estimating average tree crown size using spatial information from Ikonos and QuickBird images: Across-

- sensor and across-site comparisons. *Remote Sens. Environ.* 114, 1099–1107. <https://doi.org/10.1016/j.rse.2009.12.022>
- Sun, G., Ranson, K., Kimes, D., Blair, J., Kovacs, K., 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sens. Environ.* 112, 107–117. <https://doi.org/10.1016/j.rse.2006.09.036>
- Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958. <https://doi.org/10.1021/ci034160g>
- Tomppo, E., 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. *Remote Sens. Environ.* 92, 1–20. <https://doi.org/10.1016/j.rse.2004.04.003>
- Trochta, J., Krůček, M., Vrška, T., Král, K., 2017. 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. *PLOS ONE* 12, e0176871. <https://doi.org/10.1371/journal.pone.0176871>
- Tuominen, S., Pekkarinen, A., 2005. Performance of different spectral and textural aerial photograph features in multi-source forest inventory. *Remote Sens. Environ.* 94, 256–268. <https://doi.org/10.1016/j.rse.2004.10.001>
- Turner, D.P., Cohen, W.B., Kennedy, R.E., Fassnacht, K.S., Briggs, J.M., 1999. Relationships between Leaf Area Index and Landsat TM Spectral Vegetation Indices across Three Temperate Zone Sites. *Remote Sens. Environ.* 70, 52–68. [https://doi.org/10.1016/S0034-4257\(99\)00057-7](https://doi.org/10.1016/S0034-4257(99)00057-7)
- Ulander, L.M.H., 1996. Radiometric slope correction of synthetic-aperture radar images. *IEEE Trans. Geosci. Remote Sens.* 34, 1115–1122. <https://doi.org/10.1109/36.536527>
- Urbazaev, M., Cremer, F., Migliavacca, M., Reichstein, M., Schmullius, C., Thiel, C., 2018. Potential of Multi-Temporal ALOS-2 PALSAR-2 ScanSAR Data for Vegetation Height Estimation in Tropical Forests of Mexico. *Remote Sens.* 10, 1277. <https://doi.org/10.3390/rs10081277>
- Vafaei, S., Soosani, J., Adeli, K., Fadaei, H., Naghavi, H., Pham, T., Tien Bui, D., 2018. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). *Remote Sens.* 10, 172. <https://doi.org/10.3390/rs10020172>
- Vaglio Laurin, G., Pirotti, F., Callegari, M., Chen, Q., Cuzzo, G., Lingua, E., Notarnicola, C., Papale, D., 2016. Potential of ALOS2 and NDVI to Estimate Forest Above-Ground Biomass, and Comparison with Lidar-Derived Estimates. *Remote Sens.* 9, 18. <https://doi.org/10.3390/rs9010018>
- Vastaranta, M., Holopainen, M., Karjalainen, M., Kankare, V., Hyypä, J., Kaasalainen, S., 2014a. TerraSAR-X Stereo Radargrammetry and Airborne Scanning LiDAR Height Metrics in Imputation of Forest Aboveground Biomass and Stem Volume. *IEEE Trans. Geosci. Remote Sens.* 52, 1197–1204. <https://doi.org/10.1109/TGRS.2013.2248370>
- Vastaranta, M., Niemi, M., Karjalainen, M., Peuhkurinen, J., Kankare, V., Hyypä, J., Holopainen, M., 2014b. Prediction of Forest Stand Attributes Using TerraSAR-X Stereo Imagery. *Remote Sens.* 6, 3227–3246. <https://doi.org/10.3390/rs6043227>
- Vastaranta, M., Wulder, M.A., White, J.C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., Holopainen, M., Hyypä, J., Hyypä, H., 2013. Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update. *Can. J. Remote Sens.* 39, 382–395. <https://doi.org/10.5589/m13-046>

- Wang, M., Sun, R., Xiao, Z., 2018. Estimation of Forest Canopy Height and Aboveground Biomass from Spaceborne LiDAR and Landsat Imageries in Maryland. *Remote Sens.* 10, 344. <https://doi.org/10.3390/rs10020344>
- Wessel, B., Huber, M., Wohlfart, C., Marschalk, U., Kosmann, D., Roth, A., 2018. Accuracy assessment of the global TanDEM-X Digital Elevation Model with GPS data. *ISPRS J. Photogramm. Remote Sens.* 139, 171–182. <https://doi.org/10.1016/j.isprsjprs.2018.02.017>
- Wijaya, A., Liesenberg, V., Susanti, A., Karyanto, O., Verchot, L.V., 2015. Estimation of Biomass Carbon Stocks over Peat Swamp Forests using Multi-Temporal and Multi-Polarizations SAR Data. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XL-7/W3, 551–556. <https://doi.org/10.5194/isprsarchives-XL-7-W3-551-2015>
- Wolter, P.T., Townsend, P.A., Sturtevant, B.R., 2009. Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data. *Remote Sens. Environ.* 113, 2019–2036. <https://doi.org/10.1016/j.rse.2009.05.009>
- Woodcock, C.E., Collins, J.B., Gopal, S., Jakabhazy, V.D., Li, X., Macomber, S., Ryherd, S., Judson Harward, V., Levitan, J., Wu, Y., Warbington, R., 1994. Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sens. Environ.* 50, 240–254. [https://doi.org/10.1016/0034-4257\(94\)90074-4](https://doi.org/10.1016/0034-4257(94)90074-4)
- Yu, X., Hyyppä, J., Karjalainen, M., Nurminen, K., Karila, K., Vastaranta, M., Kankare, V., Kaartinen, H., Holopainen, M., Honkavaara, E., Kukko, A., Jaakkola, A., Liang, X., Wang, Y., Hyyppä, H., Katoh, M., 2015. Comparison of Laser and Stereo Optical, SAR and InSAR Point Clouds from Air- and Space-Borne Sources in the Retrieval of Forest Inventory Attributes. *Remote Sens.* 7, 15933–15954. <https://doi.org/10.3390/rs71215809>
- Yu, Y., Saatchi, S., 2016. Sensitivity of L-Band SAR Backscatter to Aboveground Biomass of Global Forests. *Remote Sens.* 8, 522. <https://doi.org/10.3390/rs8060522>
- Yu, Y., Yang, X., Fan, W., 2015. Estimates of forest structure parameters from GLAS data and multi-angle imaging spectrometer data. *Int. J. Appl. Earth Obs. Geoinformation* 38, 65–71. <https://doi.org/10.1016/j.jag.2014.12.013>
- Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y., Myneni, R.B., 2014. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sens. Environ.* 151, 44–56. <https://doi.org/10.1016/j.rse.2014.01.025>
- Zhang, L., Shao, Z., Liu, J., Cheng, Q., 2019. Deep Learning Based Retrieval of Forest Aboveground Biomass from Combined LiDAR and Landsat 8 Data. *Remote Sens.* 11, 1459. <https://doi.org/10.3390/rs11121459>
- Zhang, X., Kondragunta, S., 2006. Estimating forest biomass in the USA using generalized allometric models and MODIS land products. *Geophys. Res. Lett.* 33. <https://doi.org/10.1029/2006GL025879>
- Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J., Yu, S., 2016. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *Int. J. Appl. Earth Obs. Geoinformation* 53, 1–15. <https://doi.org/10.1016/j.jag.2016.08.007>
- Zhu, X., Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS J. Photogramm. Remote Sens.* 102, 222–231. <https://doi.org/10.1016/j.isprsjprs.2014.08.014>
- Zribi, M., Guyon, D., Motte, E., Dayau, S., Wigneron, J.P., Baghdadi, N., Pierdicca, N., 2019. Performance of GNSS-R GLORI data for biomass estimation over the Landes forest. *Int. J. Appl. Earth Obs. Geoinformation* 74, 150–158. <https://doi.org/10.1016/j.jag.2018.09.010>

Annexes

Annexe A : Mesures et seuil de pré-comptage

Par manque de temps et de moyens, il n'a pas été possible durant la thèse de se rendre sur les sites d'étude pour observer les placettes de mesure. Cependant, nous avons pu organiser quelques journées de mesures dans des forêts aux alentours de Toulouse (Forêts privées aux alentours des communes de Rieumes en Haute-Garonne et de Saint-Antonin-Noble-Val en Tarn-et-Garonne). Les placettes mesurées ne sont pas suffisamment nombreuses pour constituer un groupe qui puisse être ajouté à nos modèles, mais ces mesures nous ont permis de constater les erreurs possibles avec de type de référence, ainsi que les biais liés au protocole, notamment le seuil de pré-comptage.

Les forêts échantillonnées contiennent principalement des chênes et des charmes, en gestion non régulière (mélanges d'âge et de classes de diamètres au sein d'une même placette). Notre protocole consiste à délimiter une zone carrée de 25×25 mètres, soit environ 1 pixel d'une image ALOS2-PALSAR2 et 4 pixels d'images Sentinel-1 et Sentinel-2. Ensuite, nous avons mesuré la circonférence des troncs (à 130 cm de hauteur) de tous les arbres d'une hauteur supérieure à 5 mètres (évaluation visuelle). Ces mesures ont ensuite été numérisées et transformées en valeurs surfaciques de diamètre moyen (DBH, cm), densité d'arbres (arbres/ha), et surface terrière (BA, m^2/ha) avec différents seuils de pré-comptage.

Après analyse des protocoles et des valeurs mesurées, on constate tout d'abord qu'il peut avoir des erreurs importantes pour les valeurs de densité d'arbre et de surface terrière : 1) lors de la délimitation du périmètre de mesure (qui constitue la référence de surface pour le calcul des valeurs à l'hectare ensuite) il est possible d'inclure trop d'arbres en limite de périmètre ou d'exclure des arbres qui devraient être comptés dans le périmètre ; 2) l'étape de mesure des arbres peut aussi faire l'objet d'« oublis » ou double comptages, surtout pour des fortes densités lorsqu'on a pas la possibilité de marquer les arbres. D'après ces mesures et nos discussions avec les professionnels, nous estimons qu'il peut y avoir jusqu'à 10-15% d'erreur sur les mesures de densité et de BA avec nos placettes de 25×25 m. Plus on augmente la surface échantillonnée par placette, plus on réduit ce risque d'erreurs (surtout par rapport aux erreurs de périmètre). Les variables de DBH et de hauteur d'arbre sont moins impactées par ce type d'erreur, car elles ne sont pas surfaciques et donc pas multipliées lors du passage à une unité surfacique plus grande (25×25 m, soit 0.0625 ha, vers 1 ha).

La deuxième analyse que l'on peut faire suite à ces mesures porte sur l'impact du seuil de pré-comptage sur les valeurs calculées à la placette. Le seuil de pré-comptage consiste à avoir une limite basse de DBH à partir de laquelle on mesure et on comptabilise les arbres. Ce seuil peut permettre d'une part de réduire le travail de mesure en considérant que les petits arbres sont négligeables dans la description d'une structure de forêt, et d'autre part, de cibler les mesures : pour un inventaire forestier qui vise à évaluer la valeur d'une forêt, seuls les gros arbres ont une réelle valeur sur le marché du bois. L'Inventaire Forestier National, réalisé par l'IGN, prend en compte les arbres à partir de 7.5 cm de DBH, mais ce seuil est souvent plus élevé dans les inventaires forestiers. Les mesures de nos sites d'Orléans et de Saint-Gobain ont été faites avec un seuil à 17 cm de DBH, 10 cm pour les plantations de pin maritimes de la zone dunaire des Landes, et il n'y a pas de seuil pour les pins maritimes du site de l'INRA. Pour les mesures que nous avons réalisés autour de Toulouse, ce seuil était de 5 cm afin de couvrir toute la dynamique de DBH. Nous avons calculé le DBH, la densité et la surface terrière en utilisant les seuils de 7.5 cm et de 17 cm de DBH afin d'évaluer l'impact sur les valeurs. La figure A.1 présente les graphiques de comparaisons entre les seuils 7.5 et 17 cm

pour les paramètres DBH, surface terrière et densité. On constate des biais importants pour les trois paramètres, et une absence totale de corrélation dans le cas de la densité. L'utilisation d'un seuil à 17 cm, comparé à 7.5 cm qui comptabilise la quasi-totalité des ligneux des placettes, amène une sous-estimation importante de la surface terrière (biais négatif de 7.8 m²/ha, figure A.1.a), une légère sur-estimation du DBH (biais positif de 7.1 cm, figure A.1.b), et un plafonnement de la densité d'arbres à 400-500 arbres au lieu de 1500-2000 arbres par hectare avec un seuil de 7.5 cm de DBH.

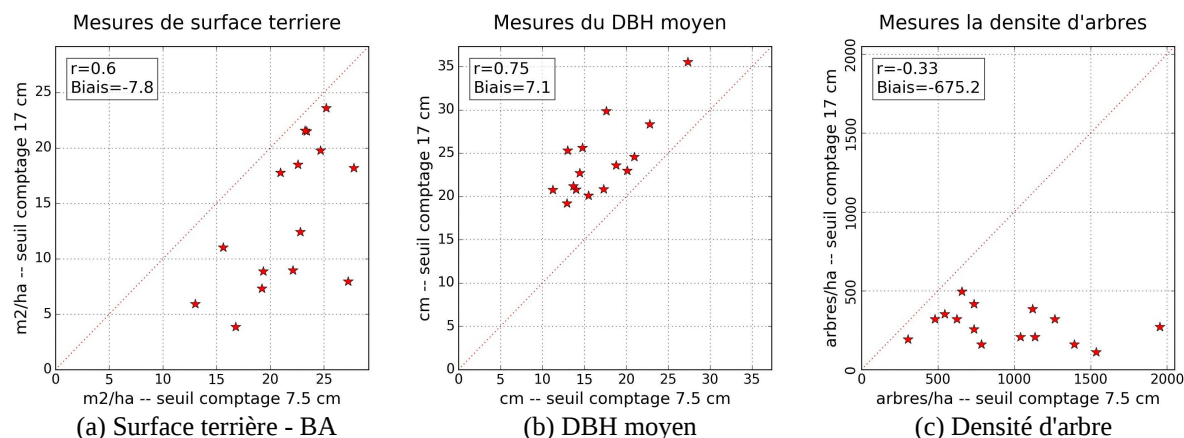


Figure A.1. Graphiques de comparaison des valeurs mesurées avec des seuils de pré-comptage à 7.5 et 17 cm de DBH pour les paramètres de surface terrière (a), DBH (b), et densité d'arbres (c).

Dans le cadre de la méthodologie mise au point dans la thèse, les erreurs et les biais de mesure peuvent avoir plusieurs impacts. Les arbres de faible circonférence ont peu d'importance dans le cadre de l'estimation de la ressource en bois, mais ils peuvent représenter un stock non négligeable de biomasse. Les types de signaux satellitaires que nous utilisons sont plus sensibles à la biomasse de manière générale : les arbres ayant un DBH entre 7.5 cm et le seuil utilisé (17 cm dans le cas des mesures ONF d'Orléans et de Saint-Gobain) ont un impact sur le coefficient de rétrodiffusion en bande L (ALOS-PALSAR), et sur la bande C et l'optique si les houppiers arrivent à la canopée ou si la couverture forestière est faible. En conséquence, avec un seuil de pré-comptage à 17 cm, il peut n'y avoir que de faibles différences de valeurs de surface terrière et densité d'arbre mesurées entre deux parcelles, alors qu'il y a une grande différence lorsque l'on prend en compte toutes les tiges à partir de 7.5 cm de DBH (voir figure A.1). Ces biais de mesure, ajoutés aux possibles erreurs, introduisent du bruit dans les données et peuvent détériorer les relations statistiques entre les paramètres forestiers mesurés et le signal satellitaire. Les résultats obtenus pour l'estimation des paramètres forestiers sur les chênes d'Orléans (section 3.2.3), moins bons que sur Saint-Gobain, pourraient être expliqués par ces biais de mesure. En effet, la variabilité du DBH est plus forte sur Orléans avec une partie de la gestion en taillis sous futaie, ce qui peut introduire des biais importants avec le seuil de pré-comptage de 17 cm.

Pour une application de la méthode d'estimation des paramètres forestiers par satellite à d'autres forêts et régions, il faudra faire attention à ces seuils de pré-comptage pour la combinaison des jeux de données et l'analyse des résultats.

Annexe B : Variation du signal radar avec l'humidité

Les variations temporelles observées sur les séries temporelles radar Sentinel-1 peuvent être gênantes pour la sélection des dates à intégrer dans la méthodologie. Nous nous sommes intéressés à l'origine de ces variations. L'hypothèse la plus évidente est que cette variation du signal radar est due aux variations d'humidité. Pour évaluer cette hypothèse, nous avons comparé le signal Sentinel-1 avec des sondes d'humidité sur le terrain, les produits satellitaires d'humidité, et les données de pluie. Dans le cadre du projet GLORI, des sondes MLX ont été utilisées pour mesurer l'humidité à 5 cm de profondeur dans une parcelle déboisée au milieu d'autres parcelles du site d'étude (sud-ouest du site). Nous avons sélectionné 3 parcelles boisées autour de la parcelle instrumentée pour faire nos tests, la figure B.1 montre ces 4 parcelles sur une image Spot-6 panchromatique. Il s'agit de peuplements adjacents étant soumis aux mêmes variations locales de climat. Ces parcelles ont des structures variées :

- id.1144 : déboisée (résidus de végétation basse).
- id.1104 : 1170 arbres/ha, 2 m de hauteur dominante, 1 tonne/ha de biomasse.
- id.1159 : 1163 arbres/ha, 10 m de hauteur dominante, 64 tonnes/ha de biomasse.
- id.1142 : 553 arbres/ha, 16 m de hauteur dominante, 131 tonnes/ha de biomasse.

Par ailleurs, nous avons téléchargé les cartes de SWI (Soil Water Index, livré à 0.1° de résolution spatiale) (Bauer-Marschallinger and Paulik, 2019) disponibles sur la plateforme Copernicus et les cartes de pluie fournies par la JAXA (produit GSMaP-rainfall, livré à 0.1° de résolution spatiale) (Kubota et al., 2007). Les séries temporelles SWI sont fortement corrélées linéairement à la mesure terrain sur un pas de temps journalier en 2016 ($r=0.91$), ce produit pourrait donc être utilisé comme proxy pour le suivi sur les autres parcelles en l'absence de mesures d'humidité sur le terrain.

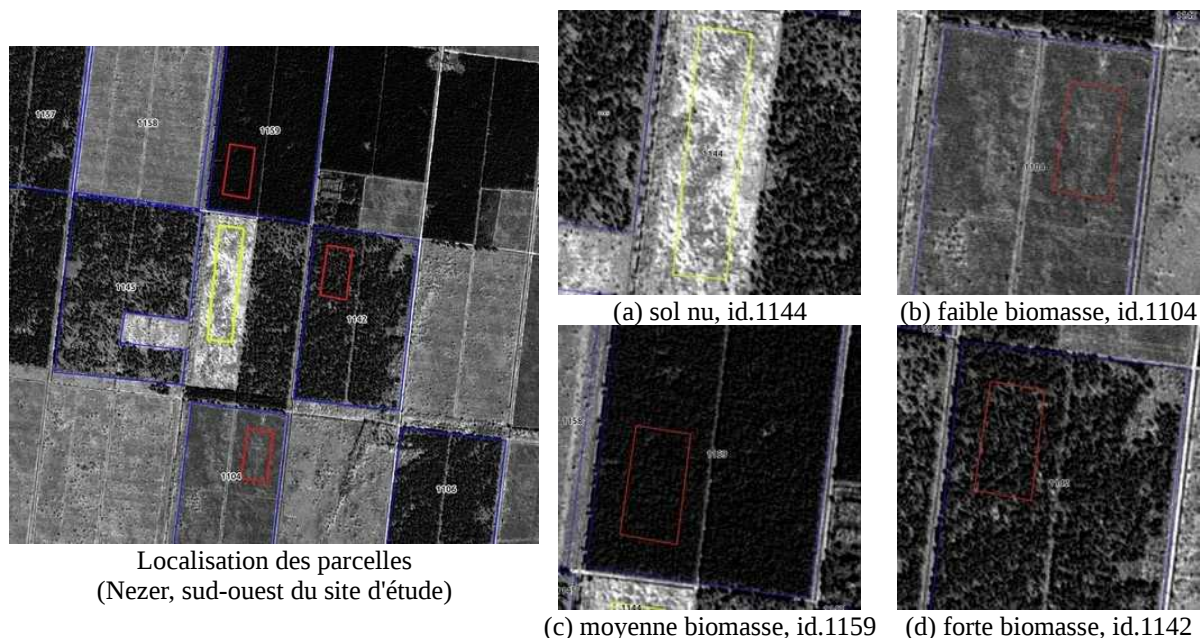


Figure B.1 : Illustration des parcelles utilisées pour étudier les variations d'humidité. La parcelle (a) est déforestée et instrumentée pour mesurer l'humidité du sol.

Nous avons comparé les séries temporelles Sentinel-1 avec les dates correspondantes de l'indice SWI sur l'ensemble de l'année 2016. Les données Sentinel-1 VV sont plus sensibles au sol et à l'humidité, tandis que les données VH sont plus sensibles au volume de la végétation. La figure B.2 montre la corrélation entre le signal VV et le produit SWI aux dates

d'acquisition de Sentinel-1, sur les pixels associés aux quatre parcelles étudiées. On constate une forte relation linéaire positive entre le SWI et le signal VV sur la parcelle déboisée ($r=0.84$), et une grande dynamique annuelle avec 6 dB de variation. Ensuite, la dynamique et la corrélation linéaire diminuent ($r\sim 0.65$) avec l'augmentation de la couverture arborée sur les parcelles 1104 et 1159. La parcelle 1142 (forte biomasse) a une dynamique réduite, mais la corrélation entre VV et SWI est plus forte ($r=0.75$) qu'avec les deux autres parcelles forestières. Cette parcelle a une plus faible densité d'arbres, ce qui pourrait expliquer une plus forte influence du sol et de l'humidité sur VV.

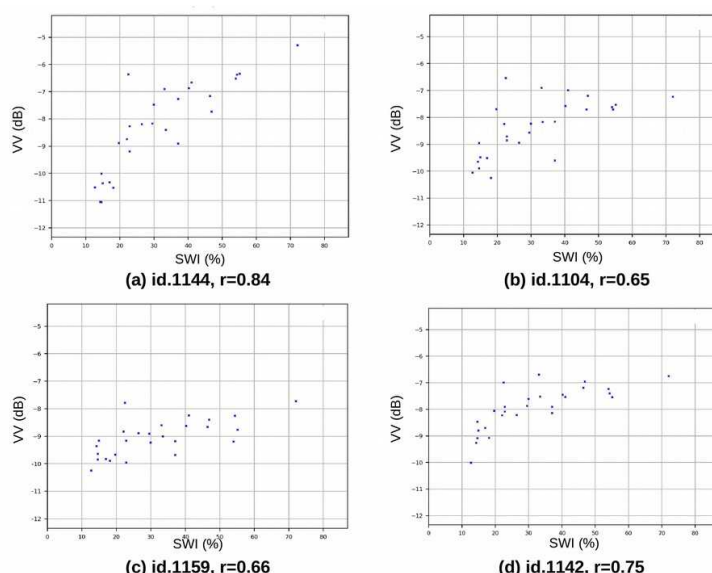


Figure B.2 : Corrélations entre les séries temporelles de SWI et de Sentinel-1 VV pour les quatre parcelles. (a) sol nu, (b) faible biomasse, (c) moyenne biomasse, (d) forte biomasse.

Si on regarde les mêmes corrélations mais avec le signal VH (figure B.3), la parcelle déboisée a encore la plus forte corrélation ($r=0.69$) et la plus grande dynamique annuelle (~ 5 dB). Par contre, la parcelle mature (1142) a une relation linéaire plus faible ($r=0.59$) et une dynamique plus faible (~ 2 dB). Cette corrélation reste significative, mais la variation temporelle d'humidité représentée par le SWI a moins d'impact que sur le signal VV.

Ainsi, à cet angle d'incidence (33°), pour le signal VV on constate plutôt une influence de la densité de couvert sur la relation avec l'humidité, tandis que pour VH on peut penser que la hauteur ou la biomasse ont plus d'influence (relations moins fortes avec l'humidité dans le cas de plus forte biomasses ou hauteurs).

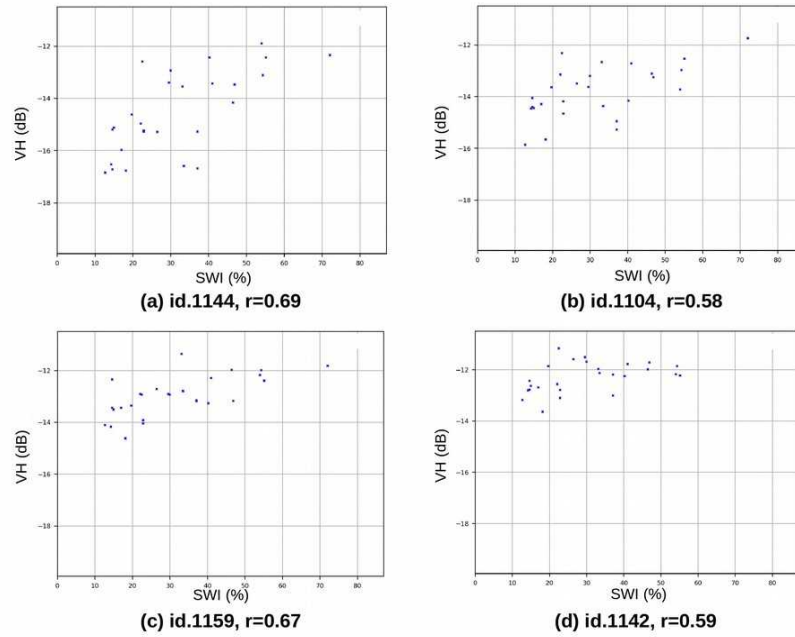


Figure B.3 : Corrélations entre les séries temporelles de SWI et de Sentinel-1 VH pour les quatre parcelles. (a) sol nu, (b) faible biomasse, (c) moyenne biomasse, (d) forte biomasse.

La figure B.4 montre un exemple de l'évolution temporelle de Sentinel-1, SWI, la sonde terrain et les pluies GSMaP sur la parcelle 1142 (553 arbres/hectares et 131 tonnes/ha). Cette représentation confirme visuellement les relations montrées dans les figures B.2 et B.3. Les données de pluies de GSMaP ne montrent pas de relation linéaire significative avec les données terrain ou le SWI au pas de temps journalier. L'utilisation d'un cumul sur 3 jours permet de voir une légère tendance ($r > 0.4$) avec la sonde terrain et le SWI. La hauteur de pluie fournie par GSMaP n'est pas suffisante pour décrire l'humidité du sol sur cette zone, mais la superposition des occurrences de pluie avec la variation du signal Sentinel-1, l'humidité mesurée et le SWI dans la figure B.4 permet de voir qu'une grande partie des larges variations du signal Sentinel-1 et de l'humidité sont cooccurentes avec des épisodes de pluie.

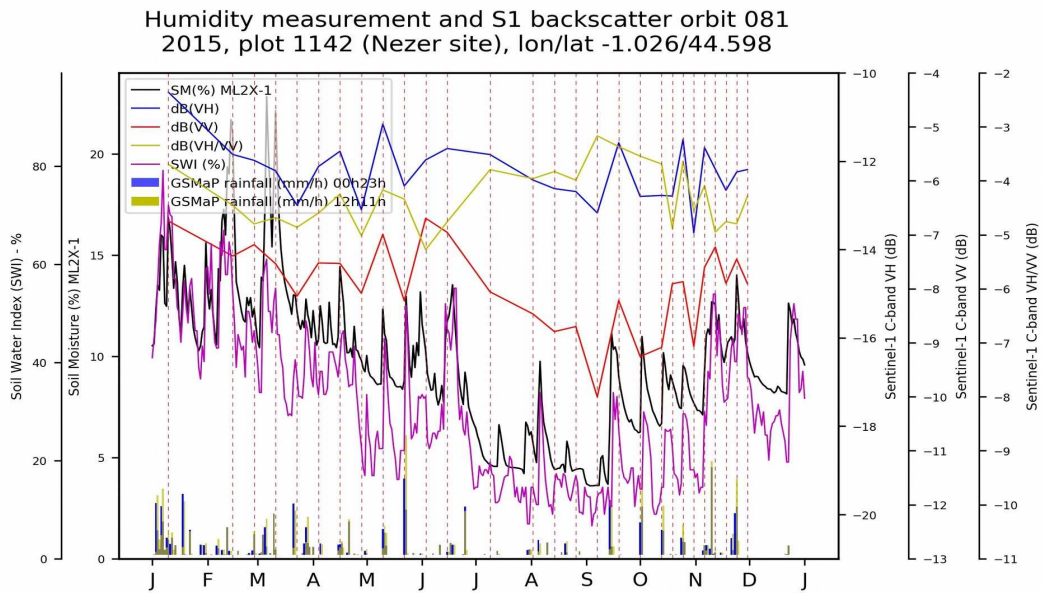


Figure B.4 : Profils temporels Sentinel-1 VH, VV et le ratio VH/VV sur la parcelle 1142 de pin maritime (553 arbres/ha et 131 tonnes/ha). Les profils temporels des produits SWI, GSMaP et de la sonde MLX (sur la parcelle déboisée 1144) sont superposés.

Annexe C : Communication scientifique

C.1. Journal international à comité de lecture

1. **Morin, D.**, Planells, M., Guyon, D., Villard, L., Mermoz, S., Bouvet, A., Thevenon, H., Dejoux, J.-F., Le Toan, T., Dedieu, G., 2019. Estimation and Mapping of Forest Structure Parameters from Open Access Satellite Images: Development of a Generic Method with a Study Case on Coniferous Plantation. *Remote Sensing* 11, 1275. <https://doi.org/10.3390/rs11111275>
2. Inglada, J., Vincent, A., Arias, M., Tardy, B., **Morin, D.**, Rodes, I., 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing* 9, 95. <https://doi.org/10.3390/rs9010095>
3. Valero, S., **Morin, D.**, Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., Koetz, B., 2016. Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions. *Remote Sensing* 8, 55. <https://doi.org/10.3390/rs8010055>
4. Battude, M., Al Bitar, A., **Morin, D.**, Cros, J., Huc, M., Marais Sicre, C., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment* 184, 668–681. <https://doi.org/10.1016/j.rse.2016.07.030>
5. Waldner, F., Lambert, M.-J., Li, W., Weiss, M., Demarez, V., **Morin, D.**, Marais-Sicre, C., Hagolle, O., Baret, F., Defourny, P., 2015. Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series. *Remote Sensing* 7, 10400–10424. <https://doi.org/10.3390/rs70810400>
6. Matton, N., Canto, G.S., Waldner, F., Valero, S., **Morin, D.**, Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P., 2015. An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series. *Remote Sensing* 7, 13208–13232. <https://doi.org/10.3390/rs71013208>
7. Li, W., Weiss, M., Waldner, F., Defourny, P., Demarez, V., **Morin, D.**, Hagolle, O., Baret, F., 2015. A Generic Algorithm to Estimate LAI, FAPAR and FCOVER Variables from SPOT4_HRVIR and Landsat Sensors: Evaluation of the Consistency and Comparison with Ground Measurements. *Remote Sensing* 7, 15494–15516. <https://doi.org/10.3390/rs71115494>
8. Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., **Morin, D.**, Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sensing* 7, 12356–12379. <https://doi.org/10.3390/rs70912356>
9. Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Matton, N., **Morin, D.**, Popescu, R., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Ahmad, I., Bégué, A., Wu, B., De Abelleira, D., Diarra, A., Dupuy, S., French, A., Ul Hassan Akhtar, I., Kussul, N., Lebourgeois, V., Le Page, M., Newby, T., Savin, I., Verón, S.R., Koetz, B., Defourny, P., 2015. Building a Data Set over 12 Globally Distributed Sites to Support the Development of Agriculture Monitoring Applications with Sentinel-2. *Remote Sensing* 7, 16062–16090. <https://doi.org/10.3390/rs71215815>

C.2. Conférence internationale à comité de lecture

1. D. Morin*, M. Planells & G. Dedieu, « *Estimation of forest parameters combining high resolution radar and optical spaceborne sensors* », **IGARSS 2019, Yokohama, Japan. Oral & proceedings.**
2. D. Morin*, M. Planells, S. Mermoz, A. Bouvet, L. Villard, T. Le Toan & Gérard Dedieu, « *Estimation of Forest Parameters by Combining Optical And Multi-frequency Radar Spaceborne Sensors* », **ESA Living planet symposium 2019, Milan, Italy. Poster.**
3. D. Morin*, M. Planells, D. Guyon, S. Mermoz, A. Bouvet, L. Villard, T. Le Toanb & G. Dedieu, « *Estimation of forest parameters combining multisensor high resolution remote sensing data* », **ForestSAT 2018, College Park, Maryland, USA. Oral.**
4. D. Morin, M. Planells, D. Guyon, L. Villard & G. Dedieu*, « *Estimation of forest parameters combining multisensor high resolution remote sensing data* », **IGARSS 2018, Valencia, Spain. Oral & proceedings.**
5. D. Morin*, M. Graham, M. Planells, S. Valero, J.F. Dejoux & G. Dedieu, « *High resolution optical and radar synergy for forest type classification at national scale* », **RAQRS'V (2017), Valencia, Spain. Oral.**
6. Delloye C.*, Weiss M., Baret F., Morin D., Defourny P., 2016. « *Accuracy Assessment of GAI Retrieval from SPOT5 Take According to Crop Type and Crop Development (BELCAM)* », **Living Planet Symposium 2016. Prague, Czech Republic. Poster & proceedings.**
7. Inglada J., Arias M.*, Tardy B., Morin D., Valero S., Hagolle O., Dedieu G., Sepulcre G., Bontemps S., Defourny P., 2015. « *Benchmarking of algorithms for crop type land-cover maps using Sentinel-2 image time series* », **IGARSS 2015. Milan, Italy. Oral & proceedings.**
8. Bontemps, S.*, Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Morin, D., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Defourny, P., Koetz, B., 2015. « *“Sentinel-2 for agriculture”: Supporting global agriculture monitoring* », **IGARSS 2015. Milan, Italy. Oral & proceedings.**
9. Valero, S., Morin, D.*, Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., 2015. « *Processing Sentinel-2 image time series for developing a real-time cropland mask* », **IGARSS 2015. Milan, Italy. Oral & proceedings.**
10. Battude M.*, Morin D., Simonneaux V., Al Bitar A., Cros J., Lhuissier L., Demarez V. « *Estimation of yield and water requirements of maize crops combining high spatial and temporal resolution images with a simple crop model, in the perspective of the Sentinel-2 mission* », **RAQRS'IV 2014, Valencia, Spain. Poster.**
11. Battude, M.*, Morin, D., Laurence, S., Valero, S., Sicre, C.M., Dejoux, J.-F., Huc, M., Hagolle, O., Cros, J., Inglada, J., **Weiss, M., Courault, D.** and Demarez, V., « *Estimation of phenological indicators and irrigated areas of crops with high temporal and spatial resolution images, in the perspective of the SENTINEL-2 mission* », **Global Vegetation Monitoring and Modeling (GV2M), 2014, Avignon, France. Oral.**
12. D. Morin*, M. Battude, S. Laurence, S. Valero, C. Marais Sicre, J.-F. Dejoux, M. Huc, Olivier Hagolle, J. Cros, J. Inglada, Marie Weiss, Dominique Courault, V. Demarez, « *Cartography of irrigated crops and estimation of biophysical variables with high temporal and spatial resolution images* », **Perspective Sentinel 2 workshop LPVE, 2014, Frascati, Italy. Oral.**

C.3. Conférence nationale

1. D. Morin*, M. Planells, G. Dedieu, « *Utilisation de l'imagerie spatiale multi-capteurs HR pour l'estimation des paramètres forestiers : Application aux parcelles de pin maritime des Landes* », **Séminaire R&D ADEME « Comment optimiser la contribution des forêts et de la filière bois à l'atténuation du changement climatique ? », 2017, Paris, France. Poster.**

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Auteur :

David MORIN

Titre :

Estimation et suivi de la ressource en bois en France par valorisation des séries multi-temporelles à haute résolution spatiale d'images optiques (Sentinel-2) et radar (Sentinel-1, ALOS-PALSAR).

Directeurs de thèse :

Milena PLANELLS et Gérard DEDIEU (HRD)

Date et lieu de soutenance :

6 février 2020

Salle de Conférence du CESBIO, Toulouse

Résumé :

L'estimation et le suivi de la ressource forestière sont des enjeux majeurs pour la gestion des territoires et les stratégies d'adaptation/mitigation face au changement climatique. Il y a actuellement un besoin fort d'une cartographie fine et régulière des ressources forestières pour améliorer ce suivi.

Nous avons évalué le potentiel des images satellitaires Sentinel-1, Sentinel-2 et ALOS-PALSAR pour l'estimation des paramètres forestiers en France. Une méthode automatique et générique a été mise en place et validée sur 4 sites d'étude (feuillus, conifères et forêts mixtes). L'utilisation des différents capteurs ensemble et d'algorithmes de régression multivarié non-linéaire permet d'obtenir des estimations avec des erreurs relatives de 5 à 30% pour la biomasse, la surface terrière, le DBH, la hauteur et la densité. Ces estimations peuvent ensuite être spatialisées à haute résolution sur des grands territoires avec les images satellitaires.

Mots clés :

Forêt ; Biomasse ; Télédétection ; Cartographie ; Optique ; Radar

École doctorale :

173 - SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace

Laboratoire :

Centre d'Études Spatiales de la BIOSphère (CESBIO)

Université de Toulouse, UMR 5126 (CNRS, UPS, CNES, IRD, INRA)

18 avenue Edouard Belin, bp 2801

31401 Toulouse cedex 9

Financements :

Centre national d'études spatiales (CNES)

Agence de l'environnement et de la maîtrise de l'énergie (ADEME)