https://lib.uliege.be https://matheo.uliege.be

Potentiel de l'utilisation de données Sentinel-1 pour caractériser les écosystèmes du Sud-Ouest de Madagascar

Auteur : Renard, Guillaume
Promoteur(s) : Bastin, Jean-François
Faculté : Gembloux Agro-Bio Tech (GxABT)
Diplôme : Master en bioingénieur : sciences et technologies de l'environnement, à finalité spécialisée
Année académique : 2021-2022
URI/URL : http://hdl.handle.net/2268.2/16106

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Potentiel de l'utilisation de données Sentinel-1 pour caractériser les écosystèmes du Sud-Ouest de Madagascar.

Guillaume Renard

TRAVAIL DE FIN D'ÉTUDES PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE MASTER BIOINGÉNIEUR EN SCIENCES ET TECHNOLOGIES DE L'ENVIRONNEMENT

Année académique 2021 - 2022

PROMOTEUR : Jean-François Bastin

Toute reproduction du présent document, par quelque procédé que ce soit, ne peut être réalisée qu'avec l'autorisation de l'auteur et de l'autorité académique de Gembloux Agro-Bio Tech.

Le présent document n'engage que son auteur.

Potentiel de l'utilisation de données Sentinel-1 pour caractériser les écosystèmes du Sud-Ouest de Madagascar.

Guillaume Renard

TRAVAIL DE FIN D'ÉTUDES PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE MASTER BIOINGÉNIEUR EN SCIENCES ET TECHNOLOGIES DE L'ENVIRONNEMENT

Année académique 2021 - 2022

PROMOTEUR : Jean-François Bastin

Remerciements

Je tiens tout d'abord à remercier pour ce travail mon promoteur, Jean-François Bastin, pour son encadrement, ses appuis logistiques et scientifiques, ses encouragements et échanges, qui m'ont permis la réalisation de ce Travail de Fin d'Etudes. Un grand merci également d'avoir trouvé des financements pour la mission de terrain.

Un grand merci à Felana et Olivia pour leur accueil à l'Université d'Antananarivo et pour leur aide pour la mise en place de la mission de terrain, et plus spécialement à Felana pour son accompagnement sur le terrain.

Merci à l'ARES pour l'octroi d'une bourse de voyage d'études.

Je tiens également à remercier Gilbert et Hasina pour leur travail et leur bonne humeur durant toute la durée de la mission, sans quoi cette dernière aurait été impossible.

Un grand merci à Arthur, sans qui cette épopée malgache aurait été beaucoup moins fun. Merci à Justine de nous avoir accompagné sur la RN7 et également de nous avoir fait découvrir la capitale sous un nouvel angle.

Un grand merci à Giulia, de m'avoir soutenu dans ce projet comme dans les autres, et pour tous les moments passées ensemble et à venir !

Un grand merci aux copains en tout genre : mouchtiques, cokotteurs, sterputes et AGéens qui ont rendu l'aventure gembloutoise inoubliable.

Enfin, je tiens à remercier mes parents, qui m'ont donné les clefs en main pour concrétiser mes projets et qui m'ont soutenu tout le long de mes études.

Abstract

Because of their endemism and their services to local communities, the dry forests and xerophitic thickets in southwestern Madagascar are ecosystems of major importance. Currently, human pressures and climate change are threatening them. The area is also characterised by the extreme poverty of its inhabitants, who rely greatly on these ecosystems. In order to preserve them and the ecosystemic services they provide, sustainable management is required. For this management, a detailed and spatially explicit assessment of vegetation is needed. Remote sensing is a key technology for large-scale characterisation of ecosystems. SAR technology in particular can potentially overcome some of the limitations of optical imagery in drylands. This study aims to investigate the potential of Sentinel-1 data to characterise structural attributes of this area. The use of Sentinel-1 and Sentinel-2 data of different seasons was studied to map woody cover (WC), above-ground woody biomass (AGB), canopy height (H95), as well as two height distribution metrics (Dim1H and Dim2H). Backscatter coefficients, spectral reflectances, derived indices, texture and temporal metrics were combined with reference data, i.e. field inventory data or photointerpretation data in Random Forest regressions to map these different variables. The objectives were to : (1) identify the factors affecting Sentinel-1 data's capacity to map response variables, (2) to compare the performance of Sentinel-1, Sentinel-2 and their joint use. The models were all cross-validated taking into account spatial autocorrelation and the photointerpretation models were also validated on field data. The comparison of performance was mainly carried out using the root mean squared error (RMSE). Sentinel-1's bands and indices were discovered to be crucial to its accuracy, and the adoption of multi-seasonal models seems relevant. Sentinel-1 outperformed Sentinel-2 in terms of accuracy [1] for WC (RMSE : 12,1% vs. 14,5%) and H95 (RMSE : 1,19 m vs. 1,39 m). For AGB, Sentinel-2 was more accurate (RMSE : 12,9 T.ha^{-1}) than Sentinel-1 (RMSE : 13,29 T.ha^{-1}). Additionally, the merged sensors improved the accuracies of most seasonalities for WC (RMSE : 11,8%) and AGB (RMSE : 12,7T.ha^{-1}). Different seasons are best depending on the sensor and response variable used. The season appears to have less of an impact on the merged S-1 and S-2 data, reflecting the complementarity of the sensors. The demonstrated ability to characterize these ecosystems using Sentinel-1 and merged Sentinel-1 and Sentinel-2 data is beneficial for conservation and sustainable management. The present study is a first step and further efforts are required to improve the characterisation of dry forests and xerophytic thickets in Madagascar.

1. The RMSE values presented here are those of the multi-season models

Résumé

En raison de leur endémisme et de leurs services aux populations locales, les forêts sèches et fourrés xérophiles du Sud-Ouest de Madagascar constituent des écosystèmes d'importance majeure. Ils sont aujourd'hui menacés par les pressions anthropiques et le changement climatique. Cette zone est également caractérisée par l'extrême pauvreté de ses habitants, qui dépendent grandement de ces écosystèmes. La mise en place d'une gestion durable de ces derniers s'impose comme nécessaire à leur conservation et au maintien des services écosystémiques qu'ils offrent. Cette gestion nécessite une évaluation quantitative précise et spatialement explicite de la végétation. La télédétection constitue une technologie clé pour la caractérisation à grande échelle des écosystèmes. La technologie SAR plus particulièrement, peut potentiellement surmonter certaines limitations de l'imagerie optique dans les zones arides. Cette étude vise ainsi à étudier le potentiel des données Sentinel-1 pour caractériser des attributs structurels de cette zone. L'utilisation de des données Sentinel-1 et Sentinel-2 de différentes saisons a été étudiée pour cartographier la couverture ligneuse (WC), la biomasse ligneuse aérienne (AGB), la hauteur de canopée (H95) ainsi que deux métriques de répartitions des hauteurs (Dim1H et Dim2H). Coefficients de rétrodiffusion, réflectances spectrales, indices dérivés, métriques de texture et temporelles ont été combinés avec des données de références, à savoir des données d'inventaires de terrain ou de photointerprétation, dans des régressions Random Forest afin de cartographier ces différentes variables. Le but était de : (1) identifier les facteurs influençant la capacité des données Sentinel-1 à cartographier les variables réponses, (2) comparer les performances de Sentinel-1, Sentinel-2 et de leur utilisation conjointe. Les modèles ont tous été validés en validation croisée en prenant en compte l'autocorrélation spatiale et les modèles de photointerprétation ont en outre été validés sur les données de terrain. La comparaison des performances a principalement été effectuée à l'aide de l'erreur quadratique moyenne (RMSE). Les bandes et indices dérivés se sont avérés importants pour la précision de Sentinel-1 et l'utilisation de modèles multi-saisons paraît pertinente. Sentinel-1 a obtenu une meilleure précision [2] pour le WC (RMSE : 12,1 %) et H95 (RMSE : 1,19 m) que Sentinel-2 (RMSE WC : 14,5 % ; RMSE H95 : 1,39 m). Pour l'AGB, Sentinel-2 s'est montré plus précis (RMSE : 12,9 $T.ha^{-1}$) que Sentinel-1 (RMSE : 13,29 $T.ha^{-1}$). La fusion des capteurs a en outre présenté des gains de précision pour la plupart des saisonnalités pour WC (RMSE : 11,8%) et AGB (RMSE : 12,7 $T.ha^{-1}$). Les saisons optimales varient selon le capteur et la variable réponse employés. L'influence de la saison semble moins importante pour les données S-1 et S-2 fusionnées, témoignant de la complémentarité des capteurs. La possibilité démontrée d'utiliser les données Sentinel-1 ainsi que les données Sentinel-1 et 2 fusionnées afin de caractériser ces écosystèmes se révèle utile dans le but de leur conservation et de leur gestion durable. L'étude ci-présente constitue une première approche et des efforts subséquents sont requis afin d'améliorer la caractérisation des forêts sèches et fourrés xérophiles malgaches.

2. Les valeurs des RMSE ici présentées sont celles des modèles multi-saisons

Table des matières

1 Introduction **1**
1.1 Contexte . . . 1
1.2 État de l'art . . . 1
1.3 Objectifs . . . 2

2 Matériel et méthodes **3**
2.1 Zone d'étude . . . 3
2.1.1 Localisation . . . 3
2.1.2 Topographie . . . 3
2.1.3 Pédologie . . . 3
2.1.4 Climat . . . 4
2.2 Composition floristique . . . 4
2.3 Ressources informatiques . . . 4
2.4 Design d'échantillonage . . . 5
2.4.1 Photo-interprétation . . . 5
2.4.2 Inventaires de terrain . . . 5
2.5 Données de référence . . . 6
2.5.1 Données de photo-interprétation . . . 6
2.5.2 Données de terrain . . . 7
2.5.3 Calcul métriques de terrain . . . 9
2.5.4 Extrapolation des données de terrain aux données de photointerprétation . . . 10
2.6 Variables prédictives . . . 11
2.6.1 Saisonnalité . . . 11
2.6.2 Imagerie radar à synthèse d'ouverture : Sentinel-1 . . . 12
2.6.3 Imagerie optique multispectrale : Sentinel-2 . . . 14
2.7 Calibration et évaluation des modèles . . . 16
2.7.1 Régression Random Forest . . . 16
2.7.2 Block CV . . . 17
2.7.3 Sélection et importance des variables . . . 17
2.8 Evaluation des modèles . . . 18

3 Résultats **20**
3.1 Sélection de variables . . . 20
3.2 Validation jeu de photointerprétation . . . 21
3.3 Influence variable réponse . . . 22
3.4 Influence prétraitements . . . 23
3.5 Comparaison satellites . . . 23
3.6 Influence saisonnalité . . . 24
3.7 Importance des variables . . . 25
3.7.1 Couverture ligneuse . . . 25
3.7.2 Biomasse aérienne . . . 26
3.7.3 Hauteur de canopée . . . 28
3.8 Cartes prédites . . . 29
3.9 Répartition spatiale des différences entre satellites . . . 30

4 Discussion **32**
4.1 Potentiel de photointerprétation 32
4.2 Comparaisons variables réponses 32
4.3 Facteurs d'influence des performances de Sentinel-1 33
4.4 Sentinel-1 versus Sentinel-2 35
4.5 Comparaison produits existants 37
4.6 Perspectives suplémentaires 37

5 Conclusion **38**

6 Contribution de l'étudiant **38**

7 Bibliographie **39**

8 Webographie **44**

Annexes **45**
A Visualisation RGB plateau calcaire 45
B Variables texturales Sentinel-1 45
C Comparaison masques nuages 47
D Importance des variables : WC photointerprété 48
E Illustration des blocs spatiaux 49
F Illustration des effets des prétraitements 50
G Valeurs observées vs. prédites : WC Sentinel-1 et 2 50
H Importance des variables : Variables S-1 et S-2 (sélection séparées) assemblées dans une RFR 51

Table des figures

1 Localisation, topographie et visualisation RGB de la zone d'étude 3
2 Diagramme ombrothermique, cartes des précipitations totales annuelles et cartes des sols dominants de la zone d'étude 4
3 Design d'échantillonnage pour (a) la photointerprétation et (b) les inventaires de terrain 6
4 Exemple parcelle de photointerprétation sur Collect Earth Online 7
5 Illustration du protocole point-intercept 8
6 Architecture des fourrés xérophiles (Randriamalala et al., 2022) 8
7 Corrélations entre classes de hauteurs et composantes principales 10
8 Corrélogramme des données de terrain et de photointerprétation 11
9 Séries mensuelles de NDVI issues de Modis 12
10 Etapes de Boruta (Hasan et al., 2020) 18
11 Vue d'ensemble de l'approche méthodologique. 19
12 Comparaison des RMSE des modèles avec ou sans sélection de variables Boruta pour les différentes variables réponses. L'écart interquartile correspond aux variations de RMSE selon les capteurs et saisonnalités employés. 20
13 Comparaison des RMSE des modèles de terrain validés en CV, des modèles de photointerprétation validés en CV et des modèles de photointerprétation validés sur les données de terrain en fonction des variables réponse. L'écart interquartile correspond aux variations de RMSE selon les capteurs et saisonnalités employés. 21
14 Comparaison des RSE des modèles de terrain validés en CV, des modèles de photointerprétation validés en CV et des modèles de photointerprétation validés sur les plots de terrain en fonction des variables réponse. L'écart interquartile correspond aux variations de RMSE selon les capteurs et saisonnalités employés. 22
15 Valeurs de RMSE des modèles de terrain et de photointerprétation pour le WC en fonction de la saison, du prétraitement. L'écart interquartile est dû aux différences des RMSE entre les saisonnalités employées. Les échelles de RMSE sont différentes pour les deux graphiques. 24
16 Valeurs de RMSE en fonction de la saison et du capteur des modèles de terrain pour les variables réponses WC, AGB et H95 validées en CV, et du modèle de photointerpétration WC validé sur les données de terrain. 25
17 Comparaison des scores Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour le WC. Les prédicteurs en noir sont des bandes, en rouge des indices dérivés, en bleu des métriques temporelles et en vert des indicateurs de texture. 27
18 Comparaison des scores Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour l'AGB. Les prédicteurs en noir sont des bandes, en rouge des indices dérivés, en bleu des métriques temporelles et en vert des indicateurs de texture. 27
19 Comparaison des scores Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour l'H95. Les prédicteurs en noir sont des bandes, en rouge des indices dérivés, en bleu des métriques temporelles et en vert des indicateurs de texture. 28
20 Cartes finales de WC, AGB et H95 29
20 Différences en fonction des trois capteurs pour les variables réponses WC (jeu de photointerprétation), AGB, et H95 avec les modèles multisaisons. Les différences positives indiquent des estimations plus élevées pour le jeu de données avec le capteur portant le premier nom (par exemple, Sentinel-1 dans la carte gauche), tandis que les différences négatives indiquent des valeurs plus élevées dans le produit portant le second nom (par exemple, Sentinel-2 dans ce cas). 31

21 Influence du sol sur la différence entre les valeurs de WC Sentinel-1 et Sentinel-2 (modèles multisaisonnaux). Les valeurs positives indiquent des valeurs plus élevées de Sentinel-1, tandis que des valeurs négatives indiquent des valeurs de Sentinel-2 plus élevées. 36
22 Visualisation RGB plus détaillée de la zone d'étude. 45
23 Visualisation des cartes réalisées avec les métriques de texture sélectionnées Sentinel-1 selon les tailles de fenêtres mobiles 3, 5, 7, 9. Comparaison des image RGB Google Earth et résumé des performances des modèles 46
24 a) Médiane de janvier 2021 sans filtre b) Médiane de janvier 2021 avec filtre s2cloudless c) Médiane de janvier 2021 avec filtre sen2cor 47
25 Comparaison des score Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour le WC de photointerprétation. 48
26 Une configuration des blocs spatiaux pour le jeu de données de terrain et de photointerprétation 49
27 Visualisation effets prétraitements de Mullissa et al. (2021) sur une image mono-date dans une zone de topographie importante. 50
28 Comparaison des valeurs observées et prédites pour les modèles Sentinel-1 (pre) et Sentinel-2 multisaisons 50
29 Importance des variables calculées par RF pour les variables sélectionnées par Boruta S-1 multi et S-2 multi séparement, assemblées dans une régression RF 51

1 Introduction

1.1 Contexte

Les forêts sèches et fourrés xérophiles du sud-ouest de Madagascar constituent des écosystèmes intéressants, possédant une végétation riche en espèces d'arbres et de fourrés endémiques (Seddon et al., 2000). Ces écosystèmes souffrent néanmoins de déforestation dont les principales causes sont l'agriculture sur brulis, la production de charbon et dans une moindre mesure le pâturage caprin (Randriamalala et al., 2016). Ces menaces sont en outre aggravées par le changement climatique, ce dernier affectant d'ores et déjà les ressources en eau, les évènements climatiques extrêmes et la productivité agricole à Madagascar (Nematchoua et al., 2018). Les dégradations de forêts naturelles, causées entre autres par les raisons évoquées ci-dessus, constituent une source importante de gaz à effets de serre (Pearson et al., 2017). Vieilledent et al. (2018) estime que la déforestation à Madagascar sur la période 2010-2014 a entraîné des émissions de carbone dans l'atmosphère de 40,2 Mt C. Il est également à noter que Madagascar a perdu 44 % de sa forêt naturelle depuis 1953 et que la forêt restante est hautement fragmentée, avec 46% de la forêt se trouvant à moins de 100 m de la lisière de la forêt (Vieilledent et al., 2018). Une gestion adaptée de ces écosystèmes s'impose dès lors comme nécessaire à leur survie, au support des populations locales et à la mitigation du changement climatique.

Concernant la situation économique de la zone d'étude, à savoir le sud du district de Tuléar II (cf. 2.1), la majorité la population est considérée comme pauvre et vit de l'agriculture (Neudert et al., 2015). Le climat semi-aride de la zone limitant les rendements agricoles, le développement de l'agriculture pluviale ne suffit pas à réduire la dépendance des ménages aux ressources forestières (Hanisch, 2015), tandis que l'élevage extensif, qui est moins impacté par la variabilité des pluies, augmente la résilience des populations face aux aléas climatiques (Hänke & Barkmann, 2017). Malheureusement, l'élevage de zébus est confronté à un phénomène de vols qui participe à l'insécurité de la zone (Francken & Minten, 2005). L'élevage caprin, moins sujet aux vols, a donc été identifié comme un moyen de gestion durable des fourrés xérophiles du Sud-Ouest de Madagascar, pouvant constituer une source importante de revenus pour les ménages et donc une alternative aux pratiques de production de charbon et de cultures sur brulis (Randriamalala & Hervé, 2022). Dans ce but, un Projet de Recherche et Développement (PRD), dont l'un des volets est l'étude du potentiel de cartographie de la zone par télédétection, est mis en place.

1.2 État de l'art

La télédétection constitue aujourd'hui une technologie clé pour l'étude à grande échelle de la végétation. Les approches basées sur des régressions pour dériver des variables continues de végétation ligneuses, telles que la couverture ligneuse et la biomasse aérienne, sont particulièrement prometteuses (Hansen et al., 2003; Bastin et al., 2017; Baumann et al., 2018). Néanmoins, les désaccords entre produits globaux (Hansen et al., 2003; Sexton et al., 2013) actuels dans les zones arides montre que nos connaissances de l'étendue de la végétation de ces écosystèmes sont limités (Bastin et al., 2017). Plus spécifiquement, les fourrés xérophiles du sud-ouest de Madagascar sont encore peu étudiés et suivis à l'heure actuelle alors qu'une évaluation quantitative précise et spatialement explicite est nécessaire à une gestion durable de ces écosystèmes.

La cartographie des zones arides est caractérisée par des challenges uniques, qui ne sont généralement pas rencontrés dans les zones plus humides. Les challenges majeurs sont le faible rapport signal/bruit de la végétation, la réflectance élevée du sol, la haute hétérogénéité spatiale entre la parcelle et l'échelle régionale, les saisons de croissances irrégulières dues aux pluies saisonnières irrégulières et aux sécheresses fréquentes (Smith et al., 2019).

L'utilisation des données Radar pour des écosystèmes arides a été identifiée comme prometteuse pour la cartographie de paramètres tels que la couverture ligneuse et la biomasse aérienne ligneuse étant donné qu'ils sont moins sensibles aux variations intra et interannuelles dans la strate herbacée et de la phénologie des feuilles (Archibald & Scholes, 2007). De plus, leur sensibilité aux paramètres structuraux de la végétation les rendent potentiellement puissants afin de cartographier la végétation ligneuse (Reiche et al., 2021). Concernant la biomasse aérienne (abrégée AGB), le potentiel des données radar à synthèse d'ouverture (SAR) pour la cartographie de l'AGB est connu depuis longtemps (Le Toan et al., 1992). En général, la rétrodiffusion des images acquises à des longueurs d'ondes courtes telles que la bande X ou la bande C, dont les images Sentinel-1 font partie, montrent de plus faibles corrélations à la biomasse que des images acquises dans des longueurs d'onde plus grandes telles que les bandes L ou P (Bouvet et al., 2018). En fait, la bande C pénètre à travers les feuilles et est dispersée/réfléchie par les petites branches, tandis que la bande L pénètre à travers les couches superficielles de la végétation et est dispersée par le tronc et les branches principales (Ghasemi et al., 2011). Les données SAR en bande L ont donc fait l'objet de diverses études afin de cartographier la biomasse aérienne ou ses variations (Mitchard et al., 2011; Ryan et al., 2012; Bouvet et al., 2018; McNicol et al., 2018). Toutefois, l'estimation d'AGB utilisant la bande C obtient de bons résultats dans des zones de plus faible biomasse étant donné que les petites branches sont les principaux éléments diffuseurs pour cette bande ; l'estimation sature usuellement vers 60 T.ha^{-1} (Ghasemi et al., 2011). De plus, les fréquences d'observations élevées combinées à la haute résolution spatiale de Sentinel-1 peuvent surmonter les faiblesses de la bande C pour la cartographie de l'AGB (Reiche et al., 2021), comme cela a été réalisé dans diverses études (Cartus et al., 2021; Argamosa et al., 2018). Concernant la couverture ligneuse (abrégé WC), les données SAR en bande C ont été utilisées dans diverses études (Mathieu et al., 2013; Main et al., 2016). D'autres paramètres structuraux tels que la hauteur de canopée ont également été étudiés à l'aide des données SAR en bande C (Bruggisser et al., 2021). L'utilisation de texture pour les données SAR fait également l'objet d'études qui mettent en avant leur apport aux données SAR (Mishra et al., 2019a; Nkeumoe Numbisi et al., 2018).

En outre, il existe un nombre croissant d'articles mettant en avant le potentiel de la fusion de données optiques et radar pour la cartographie de zones forestières, agricoles (Faye et al., 2020; Van Tricht et al., 2018), urbaines (Frantz et al., 2021), mais également de couverture et d'occupation du sol (Land-Use Land-Cover, LULC) (Carrasco et al., 2019; Ienco et al., 2019; Abera et al., 2022; Heckel et al., 2020). La diversité de domaines abordés par la fusion radar-optique met en avant son caractère transversal. Plus spécifiquement pour les zones arides, bien qu'elle n'ait pas encore été étudiée pour le Sud-Ouest de Madagascar, le potentiel de la fusion des données radar et optiques a été étudié pour la cartographie de l'AGB (Forkuor et al., 2020; Fremout et al., 2022; Laurin et al., 2018), le WC (Fremout et al., 2022; Baumann et al., 2018; Anchang et al., 2020) ou encore la hauteur de canopée (Kacic et al., 2021).

1.3 Objectifs

L'objectif général de ce travail est donc l'étude du potentiel de l'utilisation de données Sentinel-1 par rapport aux données optiques Sentinel-2 afin de cartographier les écosystèmes semi-arides du sud-ouest de Madagascar. La cartographie vise à quantifier différents paramètres structuraux, à savoir la couverture ligneuse (WC), la biomasse aérienne (AGB), la hauteur de canopée (H95) et deux métriques de structure des hauteurs(Dim1H et Dim2H). Le travail se concentre dans un premier temps nb sur les paramètres influençant la capacité des données Sentinel-1 à cartographier les différentes variables citées ci-dessus. Ces paramètres sont (1) la présence de prétraitements supplémentaires, (2) l'influence d'indicateurs dérivés, c'est-à-dire indices et indicateurs de texture, (3) l'influence de l'utilisation de métriques spatio-temporelles ainsi que (4) l'influence des saisons utilisées. Dans un second temps, le travail vise à comparer les performances des données Sentinel-1, Sentinel-2 et leur utilisation conjointe afin de cartographier les paramètres structuraux cités ci-dessus.

2 Matériel et méthodes

2.1 Zone d'étude

2.1.1 Localisation

La zone d'étude est délimitée au nord par le fleuve Fiherenana et par les limites administratives du district de Tuléar II. Elle est située dans le sud-ouest du pays, dans la région administrative d'Atsimo Andrefana. Elle est bordée à l'Ouest par le canal de Mozambique. Sa superficie est de 4820 km^2. La zone est également caractérisée par le fleuve Onilahy en son centre et le lac Tsimanampetsotsa au Sud-Ouest visibles à la Figure 1. Le lac fait partie du parc national de Tsimanampetsotsa. Parmi les autres caractéristiques notables, une bande de sols rougeâtres est visible traversant d'est en ouest la zone, appelée corridor d'Itombona (Carrière et al., 2018). L'Annexe A reprend une visualisation de cette zone.

2.1.2 Topographie

La topographie de la zone est illustrée à la Figure 1. Des pentes importantes sont observées entre la zone littorale et la zone de plateau adjacente ainsi qu'autour de fleuve Onilahy. Les altitudes maximales sont rencontrées au nord-ouest, coïncidant avec des pluviométries plus élevées. La topographie est tirée des données NASADEM à 30 m de résolution (NASA, 2020).

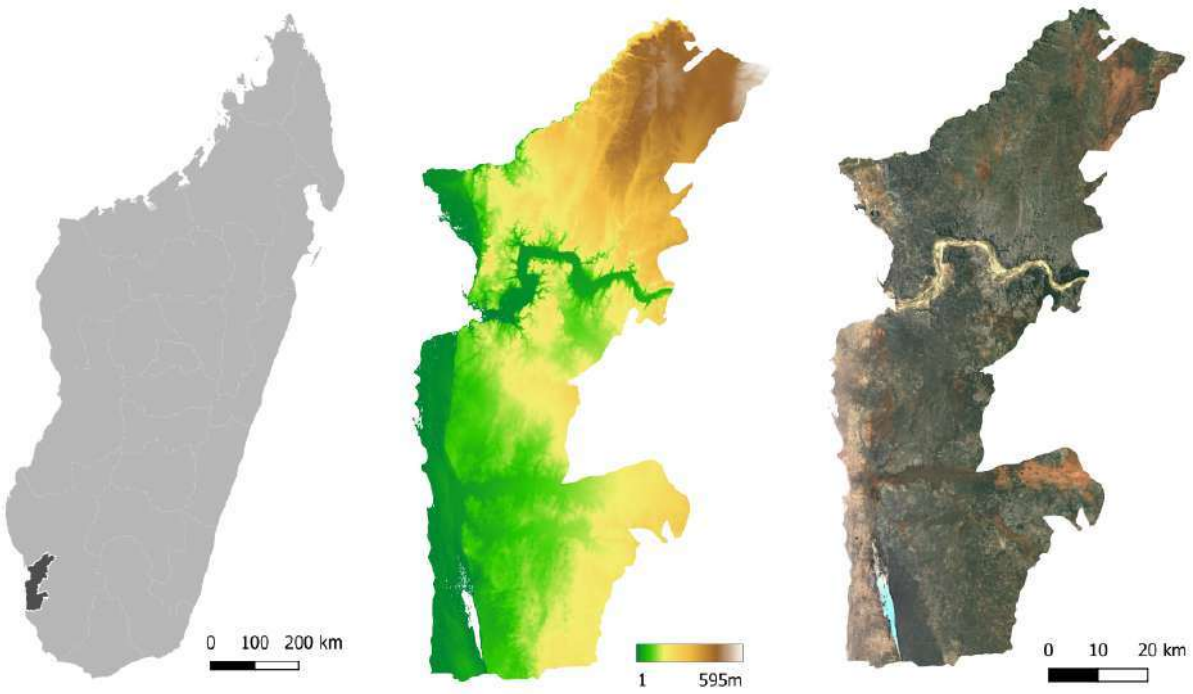


FIGURE 1 – Localisation, topographie et visualisation RGB de la zone d'étude

2.1.3 Pédologie

Les zones pédologiques (FAO) sont affichées à la Figure 1. Les inventaires de terrain ont principalement été réalisés sur les zones dominées par les régosols calcaires et les lithosols. Dans les zones dominées par les lithosols, les sols associés sont des cambisols calciques et des régosols calcaires. Ratovonamana et al. (2011) distingue les sols d'une partie de cette zone (parc de Tsimanampetsotsa) comme grès calcaires et sols ferrugineux, caractérisés par une teinte rougeâtre et les lie aux compositions végétales présentes. Bien que cette dernière classe de sol ne soit pas reprise explicitement par la

classification de la FAO, celle-ci se distingue sur la carte RGB (cf. Figure 1 ; Annexe A) pour le sud de la zone.

2.1.4 Climat

La zone d'étude est caractérisée par un climat semi-aride chaud avec une saison sèche s'étendant de mars à novembre et qui est accompagnée d'une baisse des températures qui atteint son minimum en juillet (cf. Figure 2). La pluviométrie annuelle est de 463 mm et la température moyenne de 24,1°C. La zone est caractérisée par un gradient de précipitations visible à la Figure 2. Les données climatiques utilisées sont issues des jeux de données WorldClim accédées via Google Earth Engine (Fick & Hijmans, 2017).

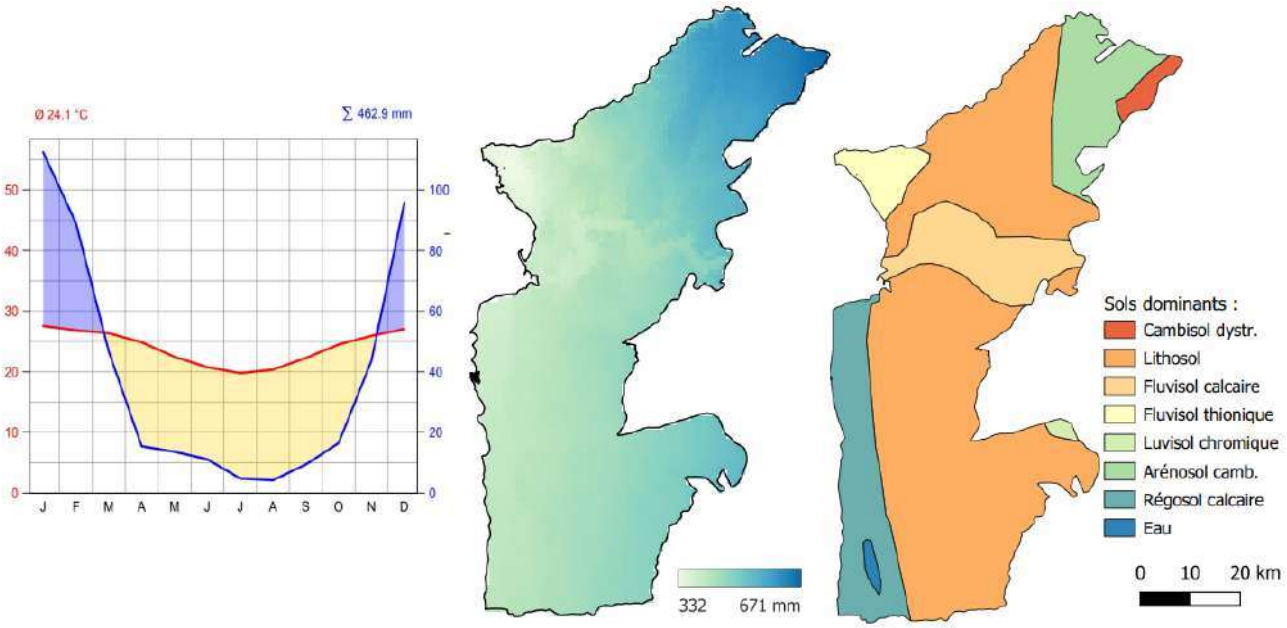


FIGURE 2 – Diagramme ombrothermique, cartes des précipitations totales annuelles et cartes des sols dominants de la zone d'étude

2.2 Composition floristique

Les types de formations forestières sont répartis selon un gradient est-ouest avec à l'Est des forêts sèches, puis en allant vers l'ouest, des formations de transitions, suivies des fourrés xérophiles (Ferry et al., 1998). Ces derniers sont divisés en hauts fourrés xérophiles, bas fourrés xérophiles sur sol sableux et bas fourrés xérophiles sur sol calcaire. Les forêts sèches et les formations de transition sont toutes deux caractérisées par la présence d'*Adansonia za* Baill. émergents et par la présence de trois strates. Les fourrés quant à eux sont caractérisés par deux strates de végétation plus ou moins confuses. Les fourrés sur sol sableux sont notamment caractérisés par la présence de *Commiphora lamii H. Perr.* Les bas fourrés xérophiles, quant à eux, reposent sur l'affleurement des grès calcaires. Les types de formation dépendent également fortement du type de sol.

2.3 Ressources informatiques

Google Earth Engine (abrégé GEE) a été utilisé pour l'extraction des données. GEE est une plateforme basée sur le cloud pour l'analyse géospatiale à l'échelle planétaire. Elle donne accès à des ressources informatiques de hautes performances pour le traitement de très grands jeux de données géospatiales. (Gorelick et al., 2017).

La récolte de données de photointerprétation a été réalisée sur la plateforme Collect Earth Online, qui est un outil web d'interprétation visuelle des terres. Il fait partie de la suite Open Foris, qui est un ensemble de logiciels open source développé par la FAO pour le monitoring environnemental (FAO, 2022). Il permet de collecter des données de références en utilisant des images satellites à haute résolution et des analyses de données à l'aide de Google Earth Engine. Plusieurs utilisateurs peuvent collecter de l'information simultanément.

Le traitement des données a été réalisé via RStudio. Les principales librairies utilisées pour mener à bien les divers traitements sont Boruta(Kursa & Rudnicki, 2010), RandomForest (RColorBrewer & Liaw, 2018) implémenté dans Caret (Kuhn, 2015)et blockCV (Valavi et al., 2018).

2.4 Design d'échantillonage

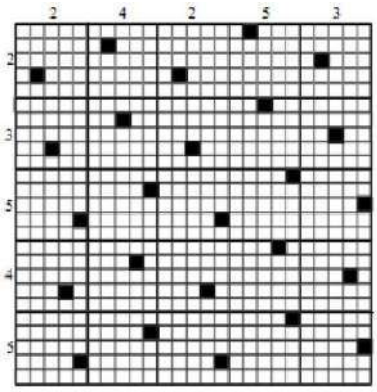
2.4.1 Photo-interprétation

La photointerprétation est réalisée sur la plateforme Collect Earth Online. Le design d'échantillonnage utilisé est un échantillonage stratifié systématique non-aligné (abrégé SSUS). Ce dernier utilise une grille systématique et place un échantillon aléatoirement dans chaque grille. Il combine les avantages d'un échantillonnage systématique, à savoir une bonne couverture spatiale et les avantages d'un échantillonnage aléatoire, à savoir la possibilité d'étude de propriétés statistiques (Banko, 1998). Un total de 505 placettes de 50x50 mètres disposées selon le design d'échantillonnage SSUS sont générées.

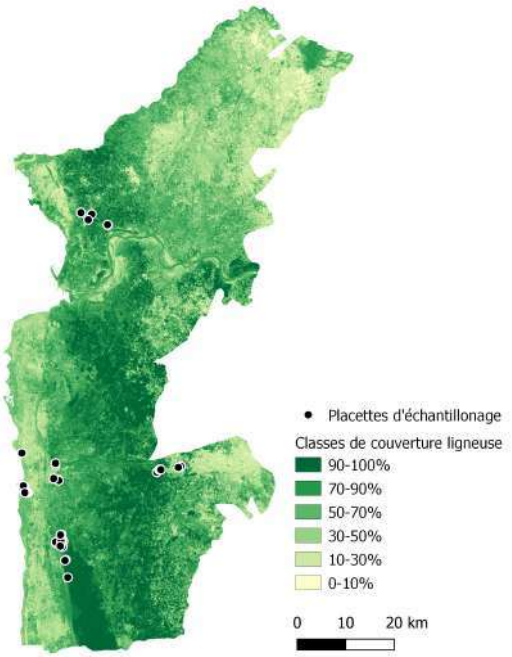
2.4.2 Inventaires de terrain

Afin de stratifier la zone d'étude en fonction de la couverture ligneuse évaluée par photointerprétation, les valeurs de couverture ligneuse des placettes photo-interprétées ont été utilisées comme données d'entraînement afin de produire des cartes de couverture de la végétation. Ces dernières ont été réalisées à l'aide de variables prédictives Sentinel-2, consistant en des bandes et des indices dérivés, qui ont été utilisés pour entraîner des modèles Random Forest dans GEE. Une carte quantifiant la couverture ligneuse a été produite et utilisée afin de stratifier la zone d'étude et d'échantillonner les différents niveaux de recouvrement tel qu'illustré à la Figure 3(b).

Les zones de travail ont ainsi été identifiées sur base de la carte 3(b) et par vérification visuelle des images satellites Bing Maps et/ou Google Earth de manière à couvrir les différents écosystèmes cités en 2.2 et les différents pourcentages de couverture ligneuse. La localisation des placettes d'inventaires a également été conditionnée par différentes contraintes logistiques et administratives résultant en la distribution spatiale -peu équilibrée- des placettes visible à la Figure 3(b).



((a)) Stratified Systematic Unaligned Sampling (SSUS)



((b)) Classes de couvertures ligneuses et placettes d'échantillonnage

FIGURE 3 – Design d'échantillonnage pour (a) la photointerprétation et (b) les inventaires de terrain

2.5 Données de référence

2.5.1 Données de photo-interprétation

Les données de photointerprétation consistent des valeurs de couverture ligneuses réparties sur 505 placettes de 50x50 mètres.

L'évaluation des WC a été réalisée sur Collect Earth Online (FAO, 2022). Chaque placette contenait 49 points de photo-interprétation disposés de manière systématique (voir Figure 4). Des données sur les parcelles de précipitations annuelles, de topographie, de texture et pH du sol ainsi que les évolutions annuelles des indices de végétation NDVI, NDWI et EVI étaient renseignées sur la plateforme Collect Earth Online afin de faciliter la photo-interprétation. À chaque point a été attribué une classe : soit élément ligneux, soit élément non ligneux, c'est-à-dire sol nu, couverture herbacée, etc. Le pourcentage de couverture ligneuse (WC), qui est défini comme la projection verticale des éléments d'une plante ligneuse au sol de chaque parcelle a ensuite été calculé à partir du nombre d'éléments ligneux en son sein.

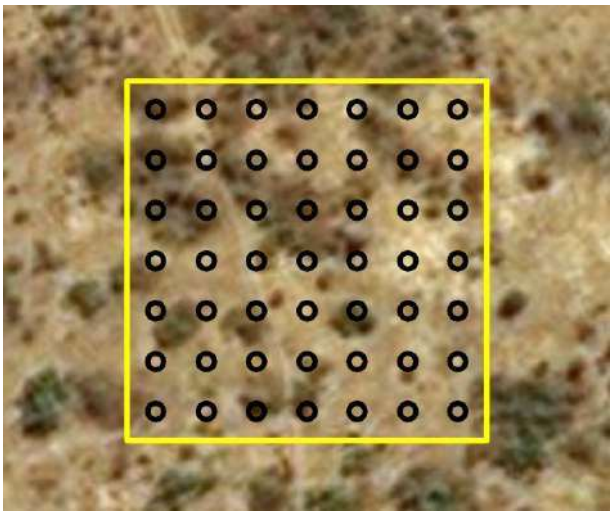


FIGURE 4 – Exemple parcelle de photointerprétation sur Collect Earth Online

2.5.2 Données de terrain

Les données de terrain ont été récoltées lors d'une campagne de récolte durant le mois de mai 2022. Elles ont été échantillonnées en suivant le design détaillé en 2.4.2. Les parcelles sont des carrés de 30x30 mètres au sol. Les coordonnées ont été récoltées en 16 points de la parcelle à l'aide d'un GNSS. Les 16 points étaient espacés de 10 mètres chacun et étaient marqués par des piquets afin de matérialiser la parcelle et les transects de l'échantillonnage point-intercept détaillés ci-après. La récolte de 16 positions GPS par parcelle a permis d'améliorer la précision de la géolocalisation des parcelles et de quadriller la parcelle afin de faciliter la mise en place des mesures. En effet, les 16 points GPS ont permis de calculer le centroïde des parcelles, en effectuant la moyenne des latitudes et longitudes et de les reconstituer avec précision.

Deux protocoles ont été réalisés pour chaque parcelle : un échantillonnage point-intercept et un inventaire en plein.

Protocole point-intercept Le protocole point-intercept est issu des protocoles du TERN (White et al., 2012). Il consiste en l'évaluation de la présence et en la description de la végétation en 54 points dans la parcelle répartis sur quatre transects. Deux transects sont dirigés vers le nord magnétique (nommés N1 et N2) et deux vers l'est (nommés E1 et E2), les axes sont espacés de dix mètres (cf. Figure 5). La hauteur, le nom vernaculaire et le type de végétation tel que défini par Randriamalala et al. (2022) (voir Figure 6) sont récoltés pour les plantes présentes.

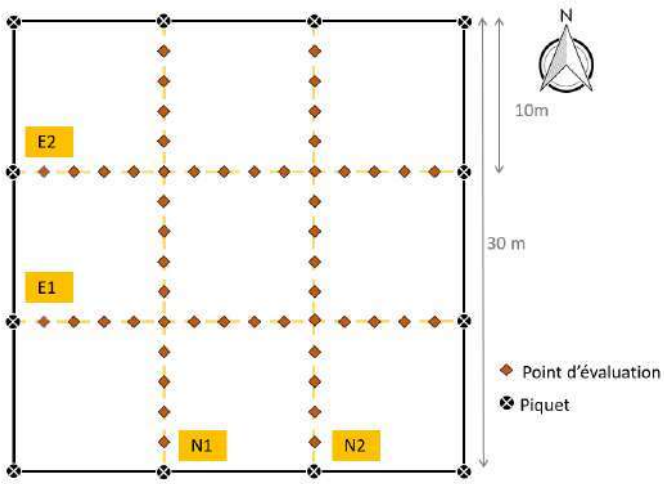


FIGURE 5 – Illustration du protocole point-intercept

Protocole d'inventaire en plein Le protocole d'inventaire en plein consiste en l'identification et en la mesure de différents paramètres de l'ensemble des individus dont la circonférence de la/les tiges basales sont supérieures à cinq centimètres. Les mesures sont effectuées en prenant en compte la diversité des fourrés xérophiles d'architecture du Sud-ouest malgache tels que défini par Randriamalala et al. (2022) qui sont illustrés à la Figure 6. Les mesures pour les trois catégories définies sont :

— **Type 1** : arbustes multitroncs avec une seule tige basale et les premières branches en dessous d'une hauteur de 1,3 m. Les informations récoltées pour ce type sont le diamètre de la tige basale mesurée à une hauteur de 0,3 m (D_{30}), la hauteur totale (H), le nombre de tiges au-delà de 1 m (N_c) et le nom vernaculaire de l'espèce

— **Type 2** : arbustes à plusieurs tiges basales. Les informations récoltées pour ce type sont le diamètre de la tige basale maximale mesuré à une hauteur de 0,3 m (D_{30max}), H, N_c et nom vernaculaire de l'espèce.

— **Type 3** : arbustes dont les premières branches dépassent une hauteur de 1,3 m. Ces arbustes peuvent être mono ou multitroncs. Les informations récoltées pour ce type sont : DBH, H et le nom vernaculaire de l'espèce.

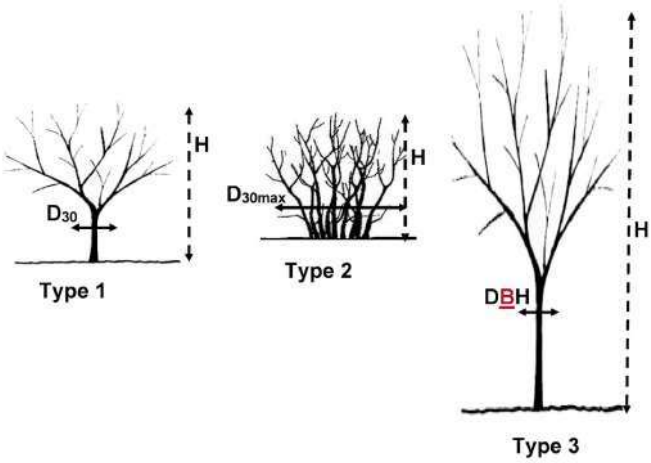


FIGURE 6 – Architecture des fourrés xérophiles (Randriamalala et al., 2022)

2.5.3 Calcul métriques de terrain

Les métriques à l'échelle des parcelles dérivées des données récoltées sont : la couverture ligneuse (WC), la biomasse aérienne (AGB) , la hauteur de canopée (H95) et deux métriques de répartition des hauteurs (Dim1H et Dim2H). Elles sont détaillées ci-dessous.

Couverture ligneuse

Le pourcentage de couverture ligneuse, définie comme le pourcentage de la surface du sol couverte lorsque les couronnes sont projetées verticalement, est calculé à partir des données récoltées avec le protocole point-intercept. Le pourcentage des points classés comme végétation ligneuse par parcelle constitue le WC.

Biomasse aérienne

Les équations allométriques de Randriamalala et al. (2022) ont été utilisées pour déterminer la biomasse à l'exception des individus de type 3 dont le DBH est supérieur à 20 cm et/ou du genre Andasonia pour lesquels l'équation de Vieilledent et al. (2012) est utilisée. Ces individus, qui étaient en réalité principalement des baobabs, sont donc considérés comme des arbres plutôt que des buissons/arbustes et sont en dehors de la gamme de validité des équations de Randriamalala et al. (2022). Les équations utilisées sont détaillées à la Table 1. Le genre des individus est utilisé afin de retrouver la densité de bois. Pour certains individus, la densité de bois était inconnue, cela arrivait lorsque la traduction entre nom vernaculaire et nom scientifique était impossible ou que la densité associée à un genre était inconnue. Dans ces cas-là, la moyenne de la densité de bois par parcelle a été calculée et appliquée aux individus de la parcelle dont la densité était inconnue.

Type	Equation
Type 1	$EXP(-2.629 + 2.103lnD_{30} + 1.043lnH + 1.464ln\rho - 1.537ln\rho lnN_c + 0.712lnD30ln\rho lnN_c)$
Type 2	$EXP(-0.702 + 1.958 * lnD_{30max} + 2.456 * ln - 2.378 * ln\rho * lnN_c - 0.653 * lnD_{30max} * lnH * ln + 1.49 * lnD30max * ln\rho * lnN_c)$
Type 3	$EXP(-1.3 + 3.019 * lnDBH - 0.828 * lnH + 1.741 * lnDBH * ln\rho 0.591 * lnDBH * lnH * ln\rho)$
Arbres	$EXP(-1,103 + 1,994 * lnDBH + 0,317 * lnH + 1,303 * ln\rho)$

Table 1 – Equations allométriques utilisées pour calculer l'AGB. Les paramètres D_{30}, H, DBH et N_c sont détaillés plus haut dans la section ;

Hauteur de canopée

La hauteur de canopée (H95) est extraite en calculant le percentile 95 des valeurs de hauteurs des parcelles d'inventaires en plein.

Métrique de répartition des hauteurs

Afin d'obtenir des métriques sur la répartition des hauteurs au sein de nos parcelles, les individus de chaque parcelle ont été répartis en classe de hauteur de un à neuf mètres, la dernière classe reprenant les individus de plus de neuf mètres. Une analyse en composantes principales a été réalisée sur ces classes des hauteurs. Les métriques de distribution des hauteurs sont les scores des parcelles sur les deux premiers axes de l'ACP. Il est observable à la Figure 7 que les tailles les plus corrélées avec la première dimension sont les individus possédant une taille entre trois et sept mètres. La seconde dimension quant à elle est surtout corrélée avec les individus entre un et trois mètres. Il faut toutefois garder à l'esprit que les deux premières composantes principales expliquent respectivement 41 et 17,4 % du jeu de données.

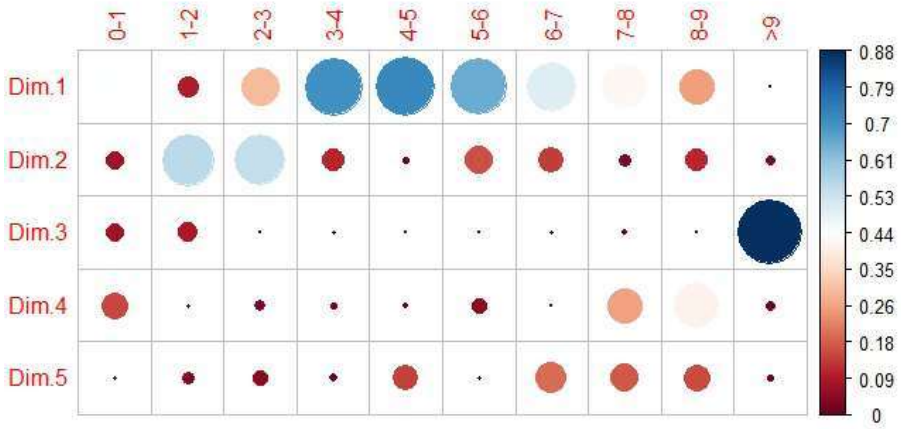


FIGURE 7 – Corrélations entre classes de hauteurs et composantes principales

2.5.4 Extrapolation des données de terrain aux données de photointerprétation

Les 41 placettes de terrain ont été utilisées afin d'évaluer le pourcentage de couverture ligneuse par photointerprétation à l'aide de Collect Earth Online. Les corrélations entre les différentes données de terrain et le WC de photointerprétation ont ensuite été calculées et sont observables à la Figure 8. Les corrélations élevées entre données de terrain et photointerprétation ont suggéré la possibilité d'extrapoler les valeurs récoltées sur le terrain pour les différentes métriques aux 505 plots de photointerprétation. Cette extrapolation possède l'avantage de multiplier par 12 la taille du jeu de données et de mieux répartir les parcelles au sein de la zone d'étude. Il faut toutefois garder à l'esprit que certaines des corrélations sont plus faibles et que les modèles à l'aide des données extrapolées sur ces métriques sont entachés d'une erreur.

La corrélation entre les données de couverture ligneuse de photointerprétation et de terrain étant de 0,90, l'hypothèse est formulée qu'il existe une erreur à la fois sur les valeurs de WC photointerprétation et à la fois sur les données de WC de terrain. Leur relation a donc été donc modélisée à l'aide d'une régression d'axe majeur (MAR) pour prendre en compte la répartition de l'erreur. Le package "lmodel2" sur Rstudio a été utilisé pour déterminer la relation (Legendre & Oksanen, 2018). La régression d'axe majeur entre les WC de terrain et de photointerprétation a été appliquée aux WC des 505 plots de photointerprétation. Les valeurs de WC obtenus à l'aide de cette relation, nommées dans les modèles "AdjustedWC", ont été considérées comme étant les plus proches de la vérité.

À partir de ces valeurs de WC ajustées, des régressions linéaires entre WC ajusté et log(AGB), H95, Dim1H et Dim2H ont été réalisées afin de pouvoir extrapoler ces métriques sur les 505 placettes. La régression linéaire a été réalisée entre WC ajusté et log(AGB) plutôt qu'AGB directement étant donné la corrélation plus importante. Les valeurs de log(AGB) extrapolées ont ensuite été retransformées en AGB. Les différentes relations sont reprises dans la Table 2.

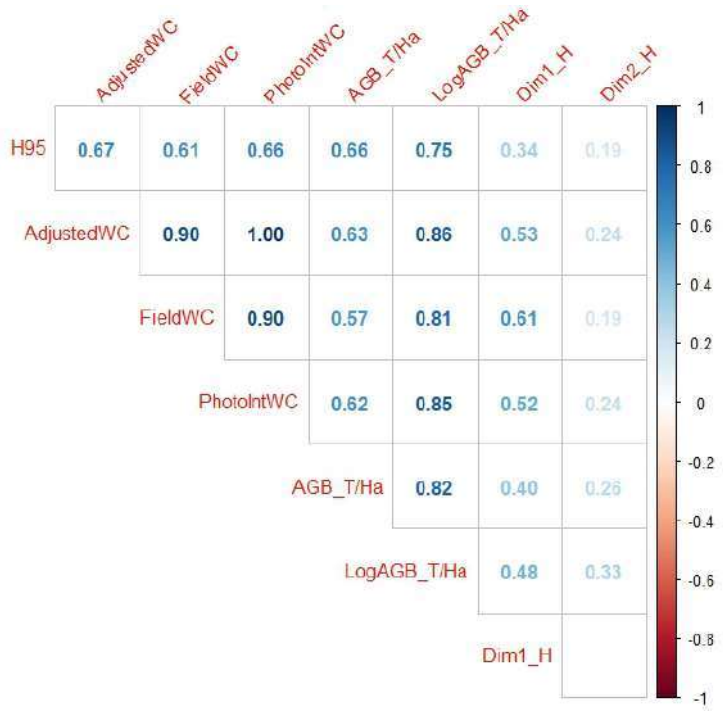


FIGURE 8 – Corrélogramme des données de terrain et de photointerprétation

Type	Equation
MAR	$WC_{ajuste} = -0.03034876 + 0.9811160 * WC_{photointerprete}$
OLS	$H95 = -1.1737 + 4.6497 * WC_{ajust}$
	$log(AGB) = -0.5122 + 2.3417 * WC_{ajuste}$
	$Dim1H = -2.3898 + 4.0570 * WC_{ajuste}$
	$Dim2H = -0.7157 + 1.2025 * WC_{ajuste}$

TABLE 2 – Equations reliant les variables réponses. Les types d'équation utilisés sont une régression d'axe majeur (MAR) et des régression linéaires au sens des moindres carrés ordinaires (OLS)

2.6 Variables prédictives

Afin de modéliser la couverture ligneuse, la biomasse aérienne, le percentile 95 de hauteur et les métriques de répartition des hauteurs, un ensemble de variables prédictives, composé de données Sentinel-1 et Sentinel-2, ont été préparées. Les sections suivantes donnent le détail de celles-ci.

2.6.1 Saisonnalité

Les données ont été acquises du 1er novembre 2020 jusqu'au 31 octobre 2021 inclus afin de fournir des données proches de la période de récolte de données de terrain tout en ne coupant pas la saison de croissance ni en ne couvrant la période de l'anomalie de Sentinel-1B qui s'est produite le 23 décembre 2021 (ESA, 2022). Les saisons considérées ici sont basées sur les valeurs de séries mensuelles de NDVI Modis accédées via Earth Map (FAO, 2022). Le but est de capter au mieux les variations phénologiques au sein de notre zone d'étude. Les saisons sont détaillées dans la Table 2.6.1.

En plus de ces quatre saisons de végétation, deux autres saisonnalités sont considérées pour la modélisation, à savoir la période annuelle couvrant toute l'année d'étude, ainsi que la combinaison des

quatre saisons de végétation, appelée multisaisons et abrégée "multi".

Saison	Abréviation	Période
Saison de croissance	Growing	Du 1er novembre au 31 janvier
Saison de végétation maximale	Grow	Du 1er février au 31 mars
Saison d'assèchement	Drying	Du 1er avril au 31 aout
Saison sèche	Dry	Du 1er septembre au 31 octobre

TABLE 3 – Saisons de végétation définies, abréviations et périodes associées

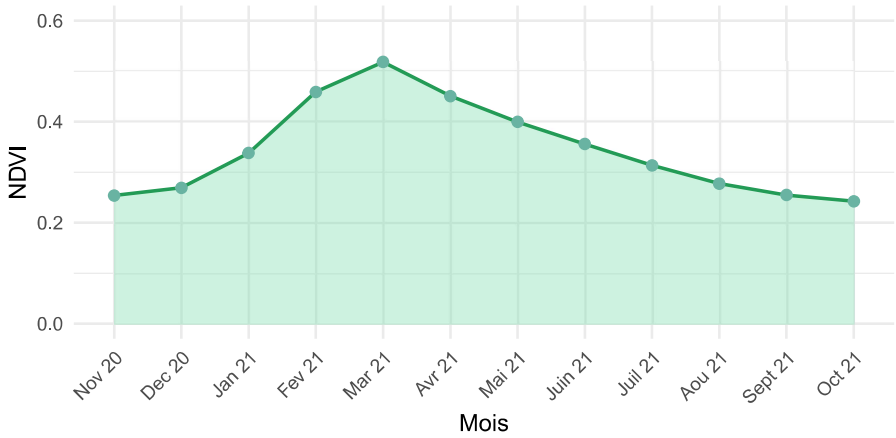


FIGURE 9 – Séries mensuelles de NDVI issues de Modis

2.6.2 Imagerie radar à synthèse d'ouverture : Sentinel-1

Les images Sentinel-1 sont des images radar à synthèse d'ouverture fonctionnant en bande C (fréquence centrale : 5,405 GHz). Elles ont été acquises en polarisations VV (transmission et réception verticales) et VH (transmission verticale et réception horizontale) dans le mode interférométrique (Interferometric Wide Swath, abrégé IW) et en orbites ascendantes et descendantes. Les données, accédées via GEE, sont les données "Ground Range Detected (GRD)", signifiant qu'elles ont subi un traitement multi-visée et été projetées au sol en utilisant la projection ellipsoïde WGS84. Les données S1 GRD disponibles sur GEE sont prétraitées par la boite à outils Sentinel afin d'appliquer le fichier d'orbites, enlever le bruit thermique, enlever le bruit de bord, effectuer la calibration radiométrique et effectuer la correction de terrain (Google, 2021).

2.6.2.1 Prétraitements supplémentaires Sentinel-1

Bien que les données Sentinel-1 GRD soient déjà prétraitées, des prétraitements supplémentaires sont recommandés pour les applications terrestres. Ces prétraitements recommandés sont issus de l'article de Mullissa et al. (2021) qui propose leur implémentation dans Earth Engine afin de rendre les données prêtes à l'analyse (Analysis-Ready Data, ARD) pour une grande gamme d'applications terrestres. Ces prétraitements sont :

— **Correction supplémentaire du bruit de bord**. Malgré les corrections implémentées dans la boite à outils Sentinel, des artefacts persistent dans certaines images. La correction supplémentaire apportée est basée sur un seuil d'angle d'incidence et sur la suppression de valeur de rétrodiffusions anormales parmi les pixels avoisinant la zone du bord (Mullissa et al., 2021)

— **Filtre speckle**. Les images SAR sont par nature affectées par le speckle ou chatoiement. C'est un bruit qui apparaît dans les images SAR sous forme de bruit granulaire, et qui est dû à l'interférence des ondes réfléchies par de nombreux diffuseurs élémentaires (Lee et al., 1994). Mullissa et al. (2021) propose différents filtres mono et multitemporels disponibles dans la littérature. Le filtre est choisi par une évaluation visuelle qualitative de la réduction du speckle dans des régions homogènes et de la préservation d'éléments subtils tels que des routes. Des paramètres additionnels peuvent être modifiés : le nombre d'images utilisées et la taille du kernel.

  Pour notre zone d'étude, l'évaluation visuelle a conduit au choix d'un filtre Lee Sigma amélioré (Lee et al., 2008) , utilisé dans un cadre multitemporel (Quegan & Yu, 2001), avec une taille de fenêtre de kernel égal à 15 et 15 images utilisées pour le filtre multitemporel.

— **Normalisation radiométrique du terrain**. Ce traitement est important afin de mitiger l'effet de la topographie sur les rétrodiffusions SAR. L'approche utilisée est une correction basée sur l'angle entre l'image SAR et la géométrie du terrain. Le modèle utilisé est optimisé pour la végétation ; il suppose que le terrain est un volume opaque de diffuseurs isotropes avec une densité de diffuseurs constante par unité de volume (Vollrath et al., 2020). Les zones affectées par les effets de repliements et d'ombrage peuvent être masquées. Celles-ci présentent une configuration géométrique spécifique qui permet leur identification sur base de comparaison de l'inclinaison de la pente et l'angle d'incidence. Néanmoins, la zone affectée est plus large qu'uniquement les pixels présentant cette configuration géométrique spécifique. Un tampon personnalisable est proposé afin de masquer les zones affectées passivement par les distorsions. La taille de tampon choisi est de 50 mètres afin d'identifier les zones affectées et donc peu fiables. Il est à noter que ces zones ne représentent qu'une très faible partie de notre zone d'étude.

Afin d'évaluer l'apport de ces prétraitements supplémentaires, des cartes ont été produites avec les données sans prétraitements supplémentaires, nommés "raw" dans les modèles, et d'autres, nommés "pre", avec.

2.6.2.2 Variables prédictives Sentinel-1

La Table 2.6.2 donne un aperçu des variables prédictives radar utilisées. Les variables prédictives sont calculées pour les orbites ascendantes et descendantes séparément, les médianes sont ensuite extraites pour chaque saison séparément. Afin de valoriser l'information des tendances intra-annuelles des données, des statistiques ont été réalisées sur la collection d'images SAR sur la période d'étude entière. En addition à ces métriques temporelles, deux indices sont calculés sur toutes les images VV et VH disponibles. Ils consistent en la somme et en la différence des polarisations. Les valeurs de rétrodiffusion étant exprimées en décibels, la différence (VH-VV) correspond à un quotient, la somme (VV+VH) correspond quant à elle à un produit (Laurin et al., 2018).

De plus, les 14 métriques de texture de la matrice de cooccurrence des niveaux de gris (GLCM) de Haralick et al. (1973) et quatre métriques additionnelles de Conners et al. (1984) ont été calculées pour les médianes des séries VV et VH par saisonnalité. Ces métriques de texture sont basées sur la construction et l'analyse d'une matrice de co-occurrences des niveaux de gris (GLCM), qui permet l'évaluation de l'agencement des pixels entre eux. Une valeur des gris, correspondant dans notre étude à une valeur entière enregistrée en 32 bits, a été attribuée à chaque pixel des bandes VV et VH des images. Dans GEE, deux paramètres sont ensuite définis : la taille de la fenêtre mobile (appelée également kernel) et le déplacement.

Les tailles de fenêtres mobiles 3, 5, 7 et 9 ont été testées en évaluant leurs performances pour cartographier le pourcentage de couverture ligneuse ajusté sur le jeu de 505 parcelles. Les performances étant similaires, l'évaluation visuelle des cartes a montré que la prise en compte d'éléments subtils était supérieure pour la taille de fenêtre mobile 3. Un extrait des cartes produites est présenté en Annexe B. Afin de réduire les corrélations au sein des variables texturales et de sélectionner celles portant le plus d'informations, la méthode proposée par King & Jackson (1999) a été utilisée. Une analyse en composantes principales a ainsi été réalisée sur les 18 variables pour chaque polarisation répartie sur les 505 parcelles de photointerprétation afin d'être représentative de la zone d'étude. Les variables les plus corrélées des quatre premières dimensions, qui expliquent plus de 90% de la variance des variables texturales ont été sélectionnées . Ces variables sélectionnées, identiques pour VV et VH, sont la dissimilarité (DISS), le cluster shade (SHADE) (Conners et al., 1984) et la corrélation (CORR) (Haralick et al., 1973). Les performances des nouvelles variables texturales ont été testées pour la couverture ligneuse de terrain à l'aide d'une comparaison avec les valeurs prédites par une régression RF et observées et par une évaluation visuelle des cartes obtenues. Ces deux évaluations indiquent que la réduction du nombre de variables n'impacterait à priori pas les performances du modèle.

Bandes, indices ou paramètres	Abréviation	Définition
Polarisation	VV	Transmission verticale-verticale
	VH	Transmission verticale-horizontale
Métriques temporelles	Pourcentiles 25 et 75, SD, moyenne, range	
Indices	VH-VV	Quotient
	VH+VV	Produit
Texture	DISS, SHADE, CORR 3x3	Dissimilarité, cluster shade et Sum average utilisés avec une fenêtre mobile de 3

Table 4 – Variables prédictives dérivées de Sentinel-1 par saisonnalité et prétraitement

2.6.3 Imagerie optique multispectrale : Sentinel-2

Les images Sentinel-2 sont des images optiques multispectrales à haute résolution avec des bandes allant de résolutions spatiales de 10 à 60m. Le produit level-2A, qui a été accédé via GEE, fournit des images de la réflectance en bas de l'atmosphère (*Bottom Of Atmosphere*, abrégé BOA) avec correction atmosphérique et une projection en WGS 84 (ESA). Les données ont été acquises aux dates similaires d'acquisition des données Sentinel-1, à savoir du 1er novembre 2020 jusqu'au 31 octobre.

2.6.3.1 Masque nuages

Afin de limiter l'influence des nuages sur les données, le pourcentage de pixels nuageux a été fixé à 10%. Les nuages résiduels ont ensuite été masqués. Il existe plusieurs masques dans la littérature pour Sentinel-2, qui ont fait l'objet de comparaisons dans diverses études. Parmi ceux-ci, les algorithmes s2cloudless (Zupanc, 2017) et sen2cor (Louis et al., 2016) ont été testés. Ils possèdent l'avantage d'être déjà implémentés dans Google Earth Engine et ont obtenu des précisions globales (abrégé PG) globalement satisfaisantes dans les études comparatives dont les résultats sont repris dans la Table 5. Il est à noter que sen2cor a fait l'objet de plus d'études que s2cloudless. Les deux algorithmes ont été comparés visuellement et sur GEE et sen2cor semble être plus adapté à la zone d'étude. En effet, sen2cloudless étant plus sujet aux erreurs dans les zones très lumineuses (Skakun et al., 2022), des zones de sable blanc sur les côtes sont considérées comme des nuages malgré différents paramétrages de celui-ci. L'Annexe C reprend une comparaison visuelle des deux masques.

Etude	PG sen2cor	PG s2cloudless
Tarrio et al. (2020)	78,7%	N.A.
Sanchez et al. (2020)	79%	52%
Baetens et al. (2019)	84%	N.A.
Skakun et al. (2022)	86%	89%

TABLE 5 – Précisions globales (PG) de sen2cor et s2cloudless issues de quatre études comparatives

2.6.3.2 Variables prédictives Sentinel-2

Les variables prédictives liées à Sentinel-2 sont reprises dans la Table 6. Dix bandes du capteur ont été extraites et déclinées pour les quatre saisons et l'année entière. Les indices spectraux NDVI, EVI et MSAVI2 ont été dérivés des bandes. Ces indices ont été choisis d'une part, en raison de leur popularité au sein de travaux similaires et d'autre part, en raison de leur prise en compte des paramètres environnementaux différents, à savoir les effets de sol et d'atmosphère. Le NDVI, ou indice de végétation par différence normalisée, est l'indice le plus communément utilisé. Il possède une réponse sensible à la végétation verte même pour des zones peu couvertes, il a également été démontré qu'il est lié à la structure de canopée et à la photosynthèse de canopée. Néanmoins, le NDVI est sensible aux effets de sol lumineux, à la couleur du sol, et à l'atmosphère (Xue & Su, 2017). Le "Modified Soil Adjusted Vegetation Index" (MSAVI2), ou indice modifié de végétation ajusté au sol, est un indice de végétation minimisant les effets dus à la réflectance du sol. L' "Enhanced Vegetation Index" (EVI), ou indice de végétation amélioré, est un indice corrigeant à la fois les effets de sol et atmosphériques. Les équations des indices [3] NDVI, MSAVI2 et EVI sont reprises dans la Table 6.

Analogiquement à Sentinel-1, des métriques temporelles et de textures ont été calculées pour Sentinel-2 sur le NDVI afin d'avoir des jeux similaires de variables prédictives pour les jeux de données radar et optique de manière à pouvoir les comparer au mieux. Le NDVI a été choisi étant donné qu'il est l'indicateur spectral le plus communément utilisé afin d'évaluer son potentiel pour la zone d'étude. Les métriques temporelles calculées sont identiques à celles calculées pour Sentinel-1. Concernant les variables texturales, celles-ci ont été sélectionnées en procédant de manière analogue à Sentinel-1. Les variables de texture de la matrice de cooccurrence des niveaux de gris (GLCM) ont été calculées pour les séries de NDVI. Les tailles des fenêtres mobiles, 3, 5, 7 et 9 ont été testées. L'évaluation des performances des modèles ne mettant pas en exergue de différences, l'évaluation visuelle des cartes entrainées avec ces les textures a mené au choix de la taille de fenêtre de 3. Afin de réduire le nombre de métriques de texture, et diminuer la corrélation des variables, une ACP a été réalisée et les métriques les plus corrélées avec les quatre premiers axes ont été sélectionnées.

3. L'abréviation NIR utilisée dans les formules de NDVI, MSAVI2 et EVI correspond à la bande NIR-1

Bandes, indices ou paramètres	Abréviation	Définition
Bandes Multispectrales	Blue	490 nm
	Green	560 nm
	Red	665 nm
	Red edge 1	750 nm
	Red edge 2	749 nm
	Redge edge 3	783 nm
	NIR-1	Proche infrarouge, 842 nm
	NIR-2	Proche infrarouge, 865 nm
	SWIR-1	Infrarouge à courtes longueurs d'ondes, 1610 nm
	SWIR-2	Infrarouge à courtes longueurs d'ondes, 2190 nm
Indices	NDVI	(NIR-RED)/(NIR1+RED)
	MSAVI2	$((2NIR+1)-((2NIR + 1)^2 - 8(NIR - RED))^{-2}) / 2$
	EVI	2.5((NIR-RED)/(NIR + 6RED - 7.5BLUE + 1))
Texture	SAVG, Corr, Imcorr2	Sum average, Correlation, Information Measure of Corr. 2 sur le NDVI
Métrique temporelle	Pourcentiles 25 et 75, SD, moyenne, range	

Table 6 – Variables prédictives dérivées de Sentinel-2 par saisonnalité

2.7 Calibration et évaluation des modèles

Les variables prédictives décrites ci-dessus sont utilisées afin de prédire les valeurs de couverture ligneuse, biomasse aérienne, hauteur de canopée et les deux métriques de répartition des hauteurs. Afin de pouvoir évaluer et comparer le pouvoir prédictif des données Sentinel-1 non-prétraités, des données Sentinel-1 prétraitées, des données Sentinel-2, de la fusion des données Sentinel-1 et Sentinel-2, désignées par la suite en tant que "capteurs", ainsi que des différentes saisonnalités définies en 2.6.1, un total de 120 modèles ont étés construits respectivement pour les jeux de données de 42 placettes de terrain et pour le jeu de données de 505 placettes de photointeprétation où les métriques ont été extrapolées. La Table 7 reprend les différentes configurations des régression RF réalisées. La méthodologie est reprise dans la Figure 11

Jeu de données	Variables réponse	Capteur	Saisonnalité
505 plots	AdjustedWC	s1 - pre	growing
42 plots	AGB	s1 - raw	grow
	H95	s2	drying
	Dim1H	s1s2	dry
	Dim2H		multi
			year

Table 7 – Ensemble des facteurs des différentes configurations des régressions Random Forest

2.7.1 Régression Random Forest

Les modèles prédictifs ont été construits à l'aide de l'algorithme Random Forest (RF). Cet algorithme effectue un apprentissage sur de multiples arbres de décisions aléatoires entraînés sur des sous-ensembles de données légèrement différents et agrège ensuite leurs prédictions en moyennant leurs résultats. RF a notamment réalisé d'excellentes performances lorsque le nombre de variables prédictives est supérieur au nombre d'observations (Biau & Scornet, 2016), ce qui est le cas du jeu de données de 42 placettes. Des études comparant des algorithmes d'apprentissage appliqués à la télédétection ont mis en avant les performances de Random Forest (Nagelkirk & Dahlin, 2020; Gan et al., 2022).

Le package 'randomForest' a été utilisé pour l'implémentation de RandomForest dans R(RColorBrewer & Liaw, 2018). Celui-ci a été utilisé au sein du package 'caret'(Kuhn, 2015). Ce dernier a été choisi pour sa facilité d'implémentation dans R, son utilisation fréquente dans la littérature ainsi que sa compatibilité avec le package 'blockCV'. Concernant les paramètres de RF, l'étude de leur influence a montré peu de variations des performances des modèles en fonction de leurs valeurs. Le nombre d'arbres (ntree) a donc été défini à 100, valeur à partir de laquelle le gain de précision est négligeable. Les valeurs par défaut des autres paramètres ont été conservées, à savoir, la taille minimale des noeuds terminaux (nodesize), fixée à cinq pour une régression RF, le nombre de prédicteurs aléatoire à chaque noeud (mtry), égal au nombre de variables prédictives divisé par 3 pour une régression RF. Le choix de ne pas chercher à optimiser au maximum les performances des modèles a été réalisé dans l'optique des objectifs de ce travail, qui cherche avant tout à comparer les facteurs influençant les performances de cartographie des données Sentinel.

2.7.2 Block CV

Les données écologiques sont souvent corrélées spatialement, c'est-à-dire que les observations proches les unes des autres (dans l'espace ou le temps) sont plus similaires que les observations distantes. L'autocorrélation spatiale prend généralement la forme de zones ou de gradients(Legendre, 1993). Afin d'éviter une surestimation des performances des modèles via une validation croisée conventionnelle, des blocs spatiaux ont été générés à l'aide du package "blockCV" (Valavi et al., 2018) afin de séparer les données de calibration et de validation. Les tailles des blocs générés sont de respectivement 10 et 2 km^2 pour les jeux de données de 505 et de 41 placettes. Pour chacun des modèles, les données ont été séparées en 10 et 5 folds pour les jeux de données de 505 et de 41 placettes. L'opération a ensuite été répétée 20 fois avec des configurations différentes des blocs spatiaux.

2.7.3 Sélection et importance des variables

Afin de pouvoir identifier l'ensemble des variables prédictives pertinentes pour les modèles et calculer l'importance des variables, l'algorithme de sélection de variables implémenté dans R 'Boruta' a été utilisé (Kursa & Rudnicki, 2010). Cet algorithme de sélection de variables est de type *all-relevant*. C'est-à-dire qu'il vise à identifier tous les attributs pertinents pour la modélisation, à la différence d'autres algorithmes de sélection, de type *minimal-optimal*, qui eux visent avant tout la sélection d'informations non redondantes et l'utilisation d'un nombre optimal de variables prédictives pour les performances des modèles (Kohavi & John, 1997). En effet, le but ici est d'identifier tous les attributs qui sont dans certaines circonstances pertinentes pour la cartographie. Trouver tous les attributs pertinents, au lieu uniquement des non redondants est particulièrement intéressant lorsque l'on s'intéresse à la compréhension des mécanismes reliés au sujet d'intérêt à savoir ici, le potentiel des données Sentinel-1 pour cartographier la zone d'étude.

Boruta est construit autour de Random Forest. Il évalue l'importance des variables en comparant les variables prédictives avec des "caractéristiques de l'ombre" ou "shadow features". Ces "shadow features" sont des copies du jeu de données initial dont les éléments sont réarrangés aléatoirement afin d'éliminer les corrélations avec la variable réponse. Un RF est exécuté sur le jeu de shadow features et le score Z est calculé. Le score Z des variables prédictives est ensuite calculé par RF puis les scores du jeu de base sont comparés avec le score Z maximum des shadow feature (abrégé MZSA [4]). Les variables dépassant le score de la MSZA sont comptabilisées comme un "hit". Le processus est alors répété et l'importance des variables est attribuée à l'aide du vecteur "hit". Un test statistique bilatéral est réalisé pour les variables dont l'importance est indéterminée. Les attributs dont l'importance est significativement plus basse que celle de la MZSA sont classés comme non importants et retirés du système tandis que les attributs significativement supérieurs à MZSA sont classés comme "importants". Les shadows

4. Maximum Z score among Shadows Atrribute

features sont ensuite retirées puis la procédure est répétée jusque qu'a ce que l'importance de toutes les variables est assignée ou que l'algorithme a atteint la limite d'itération maximale d'exécutions de RF. Les étapes de l'algorithme sont illustrées à la Figure 10 ci-dessous.

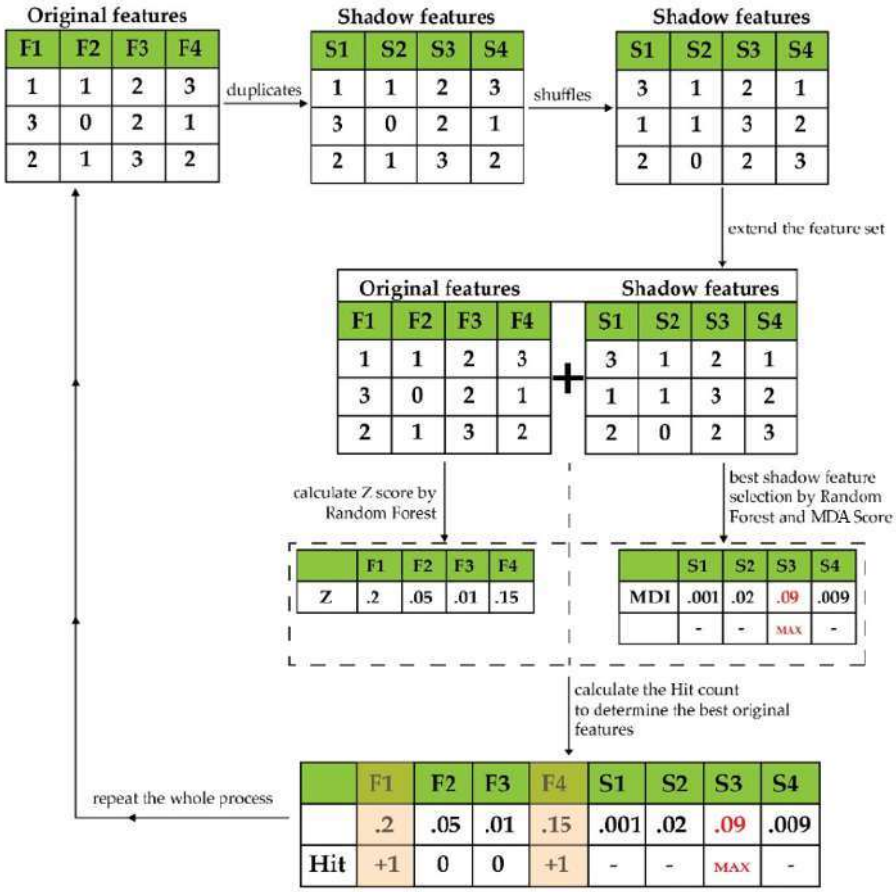


FIGURE 10 – Etapes de Boruta (Hasan et al., 2020)

2.8 Evaluation des modèles

Pour chaque modèle, les métriques suivantes ont été calculées : MAE, RSME, Biais, RSE .
La RMSE est utilisée afin d'entraîner et d'évaluer les modèles. Elle a l'avantage d'être de prendre davantage en compte l'impact des outliers. RMSE, biais et MAE sont utilisés afin d'évaluer les modèles pour une même variable réponse. Afin de comparer les différentes variables réponses, le RSE ("Relative Squared Error") est utilisé. Pour évaluer l'importance des variables, le score Z, calculé à partir dela diminution moyenne de la précision (Mean Decrease Accuracy, abrégé MDA), a été utilisé.

$$MAE = \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2}$$

$$RSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{3}$$

$$Biais = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{4}$$

où :

- y_i = la valeur actuelle
- $\hat{y}_i$ = la valeur observée
- $\overline{y}_i$ = la moyenne des valeurs observées
- n = le nombre d'observations

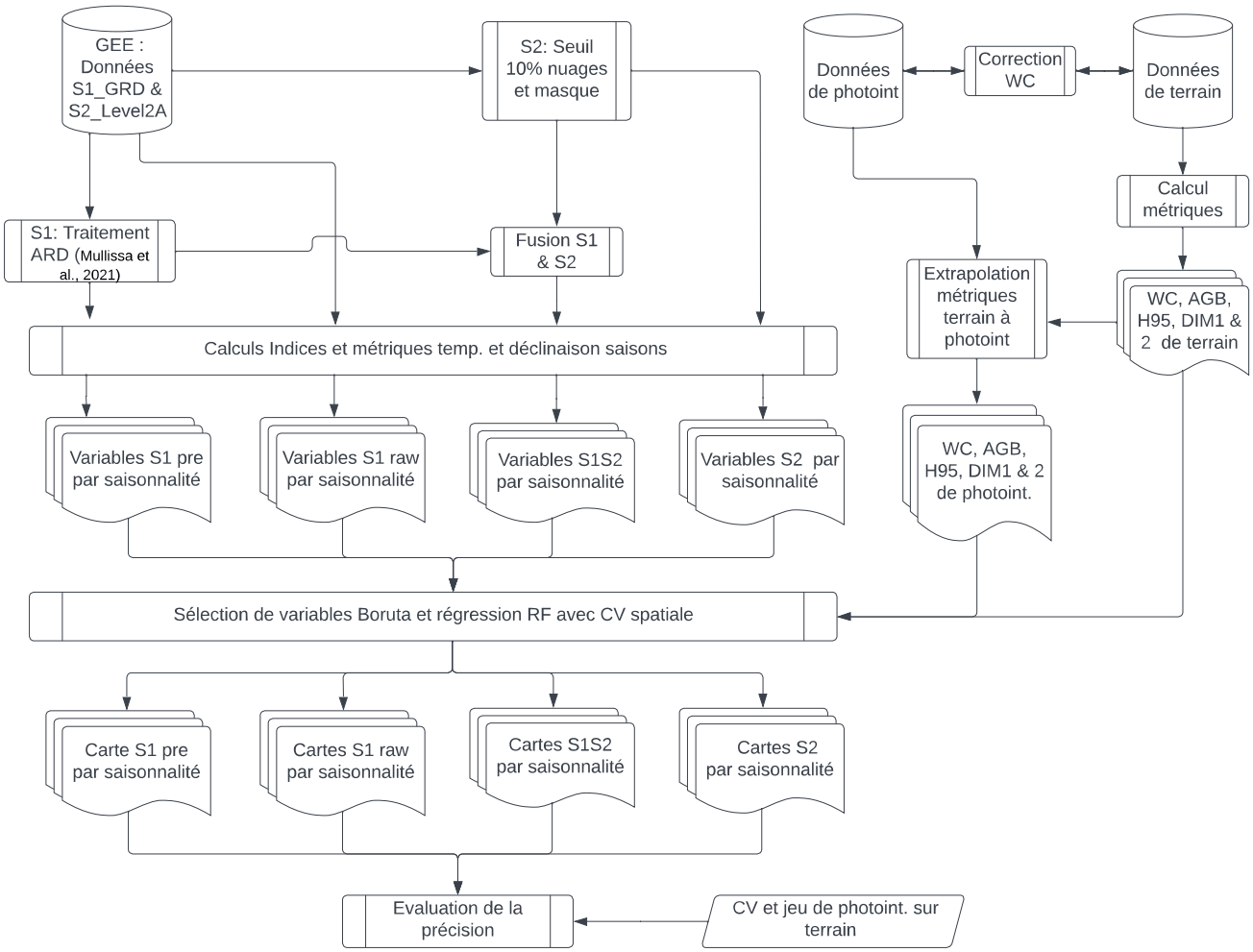


FIGURE 11 – Vue d'ensemble de l'approche méthodologique.

3 Résultats

Dans cette section sont présentés les résultats obtenus avec les modèles. La Table 8 présente un résumé des données afin d'avoir un point de comparaison avec les métriques d'évaluation des modèles. Etant donné le nombre important de modèles obtenus et de paramètres croisés entre eux, et afin de ne pas surcharger la section ci-présente. Certains résultats plus détaillés, tels que l'importance des variables, sont présentés uniquement pour les modèles jugés plus pertinents.

		WC [-]	AGB [T.ha^{-1}]	H [m]	Dim1H [-]	Dim2H [-]
	Moyenne	0.591	17.746	4.093	0.0105	-0.004
42	Médiane	0.644	10.690	3.60	-0.560	-0.3
	Ecart-type	0.266	18.353	1.899	2.072	1.350
	Moyenne	0.570	19.064	3.7078	0.149	0.037
505	Médiane	0.651	10.257	4.195	0.250	0.067
	Ecart-type	0.339	19.216	1.7852	1.152	0.341

Table 8 – Résumé des données de référence des deux jeux de données pour WC, AGB, H95, Dim1H et Dim2H

3.1 Sélection de variables

La Figure 13 compare les RMSE des modèles validés CV. Il est observable que pour chaque variable réponse, le RMSE diminue systématiquement après la sélection de variables Boruta. Cet effet est néanmoins moins marqué pour le jeu de photointerprétation. Les modèles avec sélection de variables sont donc utilisés pour la suite des résultats.

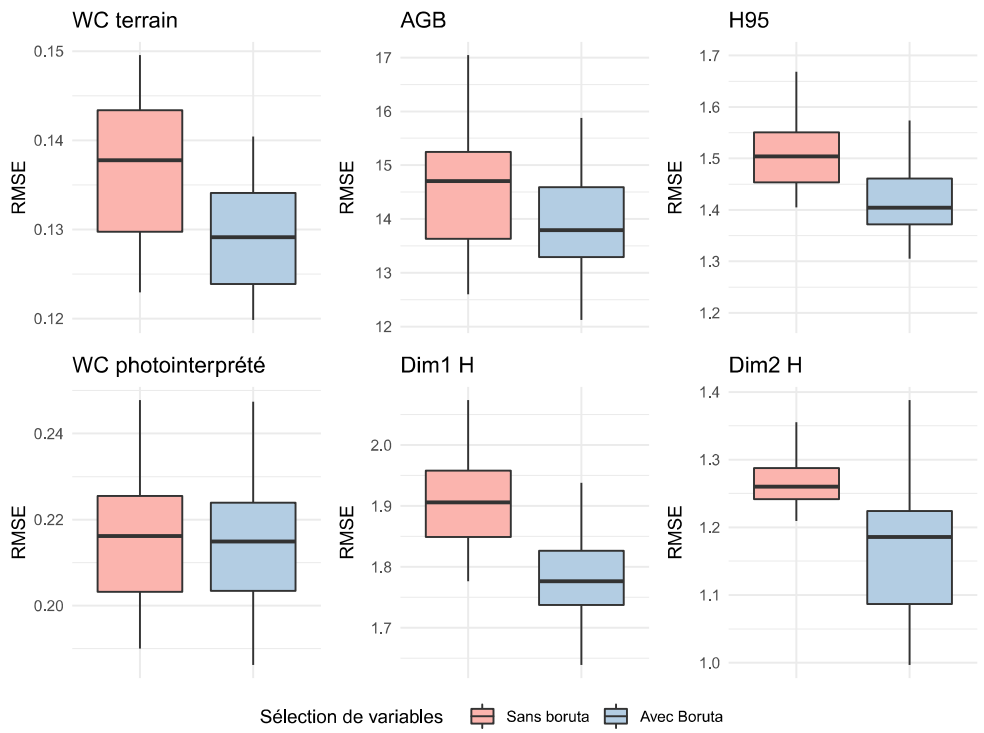


Figure 12 – Comparaison des RMSE des modèles avec ou sans sélection de variables Boruta pour les différentes variables réponses. L'écart interquartile correspond aux variations de RMSE selon les capteurs et saisonnalités employés.

3.2 Validation jeu de photointerprétation

Afin d'évaluer la pertinence des modèles construits à l'aide de la photointerprétation, notamment des variables réponses extrapolées, les modèles de photointerprétation ont été appliqués aux parcelles de terrain. La Figure 13 présente les résultats. Il est observable que pour l'ensemble des variables réponses, la validation des modèles de photointerprétation sur les 41 plots de terrain donne des performances similaires ou légèrement supérieures aux modèles de terrain en CV. Les jeux de données extrapolés aux 505 placettes peuvent ainsi être utilisés pour les différentes variables par la suite.

La comparaison entre les modèles de terrain en CV et de photointerprétation en CV indique des RMSE nettement plus faibles pour les modèles de photointerprétation pour les variables réponses H95, Dim1H et Dim2H. Concernant le WC, le RMSE du jeu de photointerprétation est environ deux fois supérieur aux RMSE des jeux de terrain en CV et de photointerprétation validés sur les parcelles de terrain. Néanmoins, la valeur du RSE du WC de photointerprétation à la Figure 14 est plus faible que celui du WC de terrain.

Il est visible à la Figure 14 que les valeurs de RSE des modèles de photointerprétation validés sur les parcelles de terrain sont toutes inférieures aux parcelles de terrain validées en CV avec un écart interquartile plus réduit que les parcelles de terrain validées en CV.

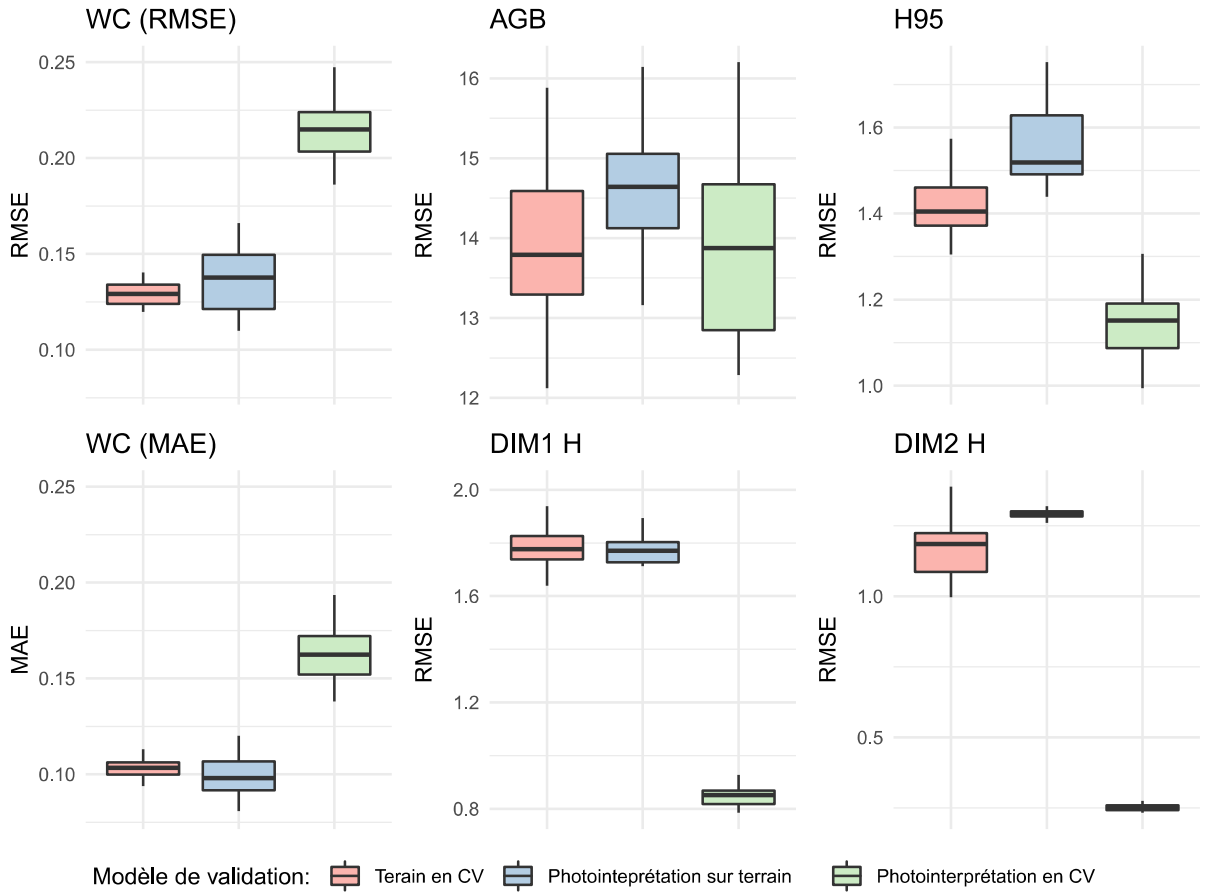


FIGURE 13 – Comparaison des RMSE des modèles de terrain validés en CV, des modèles de photointerprétation validés en CV et des modèles de photointerprétation validés sur les données de terrain en fonction des variables réponse. L'écart interquartile correspond aux variations de RMSE selon les capteurs et saisonnalités employés.

3.3 Influence variable réponse

Les RSE des variables réponses sont comparées à la Figure 14. La comparaison est réalisée sur les jeux de données de terrain validés en CV, des modèles de photointerprétation validés en CV et des modèles de photointerprétation validés sur les parcelles de terrain. La ligne pointillée à l'ordonnée égale à un est la valeur à laquelle la somme des carrés des écarts (MSE) des modèles est plus grande que celle d'un modèle utilisant la moyenne comme valeur prédite.

Il est observable que, pour le jeu de données de terrain, la variable la mieux expliquée est la couverture ligneuse, suivie de la hauteur de canopée. Ensuite, l'AGB et la Dim2H possèdent des médianes presque égales, mais l'écart interquartile de la Dim2H est plus important, traduisant plus de variabilité des performances en fonction de la saisonnalité du capteur employé. Enfin, la performance la plus faible est la Dim1H avec une RSE supérieure à deux.

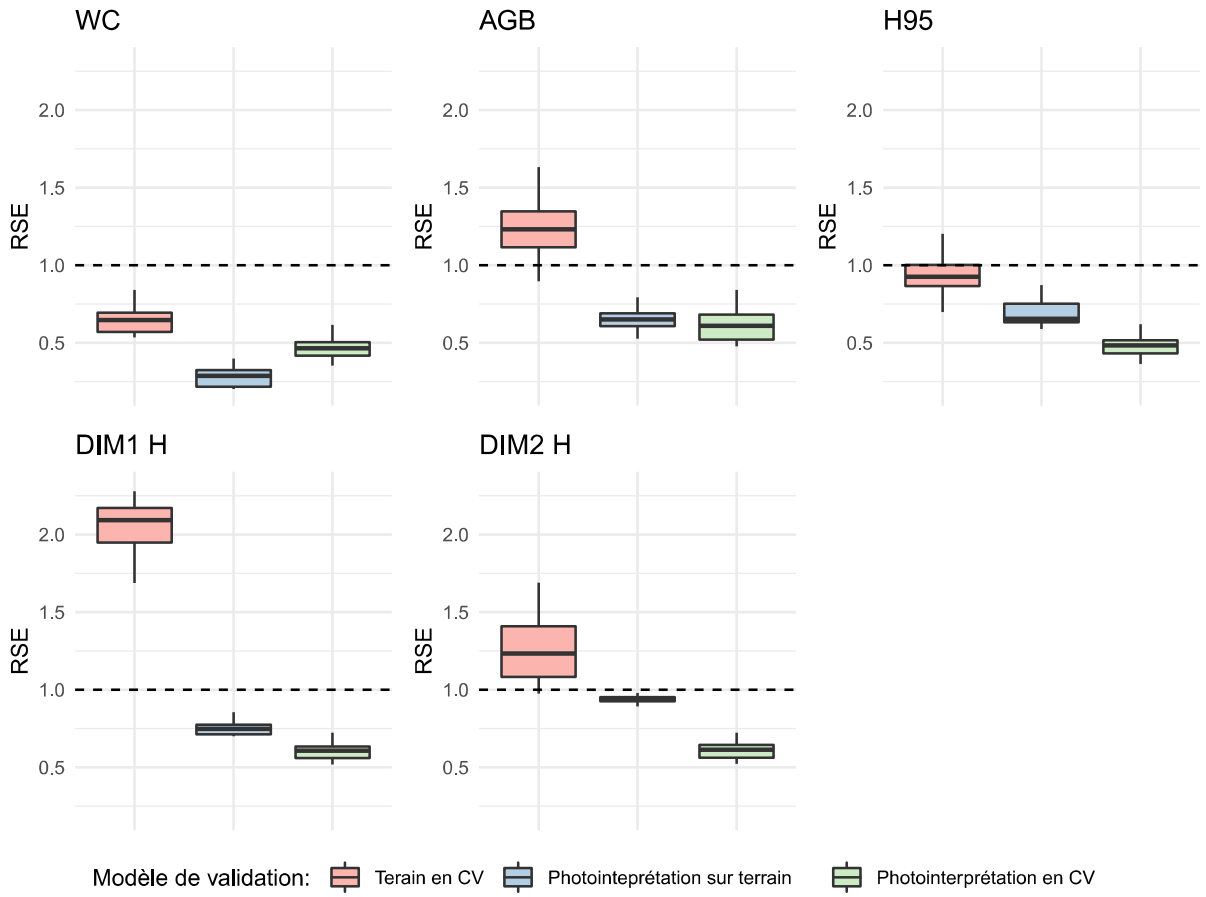


FIGURE 14 – Comparaison des RSE des modèles de terrain validés en CV, des modèles de photointerprétation validés en CV et des modèles de photointerprétation validés sur les plots de terrain en fonction des variables réponse. L'écart interquartile correspond aux variations de RMSE selon les capteurs et saisonnalités employés.

3.4 Influence prétraitements

Pour le WC pour les jeux de données de terrain et de photointerprétation, la Figure 15 affiche une RMSE médiane légèrement supérieure pour les données prétraitées pour le jeu de terrain et inversement pour le jeu de photointerprétation validé sur les données de terrain. L'inspection de la Figure 16 indique que les prétraitements diminuent la précision de trois saisonnalités sur six pour le jeu de terrain tandis qu'ils diminuent la précision de quatre saisonnalités sur sept, pour le jeu de données de photointerprétation. Ainsi, des tendances contraires ressortent des deux jeux de données et il est difficile de tirer des conclusions sur l'effet bénéfique ou délétère des prétraitements sur les données Sentinel-1, d'autant plus que les écarts de RMSE sont faibles, l'écart maximal étant de 1,2 % pour les modèles de terrain, l'effet des prétraitements reste limité. Les prétraitements sur les métriques AGB et Dim1H et Dim2H ne présentent pas non plus d'effets marqués. A l'inverse, pour H95, les prétraitements semblent bénéfiques, avec une RMSE plus faible de 9,6 cm pour les données prétraitées.

3.5 Comparaison satellites

Etant donné qu'il n'y a pas de conclusion sur les effets des prétraitements et que la fusion des données Sentinel-1 et Sentinel-2 a été effectuée avec les données prétraitées, les données Sentinel-1 prétraitées sont utilisées pour la comparaison des satellites.

Concernant le WC du jeu de terrain, les données Sentinel-2 possèdent une RMSE médiane plus faible que les données Sentinel-1, tandis que pour le jeu de photointerprétation, il n'y a pas de médiane de RMSE qui se démarque, néanmoins, il est observable que pour la photointerprétation, l'écart interquartile est supérieur pour les données Sentinel-1, traduisant plus de variabilités entre les saisonnalités, ce qui est observable à la Figure 16. Il est également observable que la combinaison des données S-1 et S-2 augmente les performances des modèles de terrain et de photointerprétation. Il est à noter que les différences de performances observées sont minimes, les médianes des RMSE étant toutes comprises entre 12 et 15 %. Pour l'AGB, les valeurs médianes ne permettent pas de distinguer un capteur par rapport aux autres. La comparaison des RMSE des modèles par saisonnalité indique des valeurs de RMSE Sentinel-1 inférieures à Sentinel-2 pour les saisons dry et drying (cf. Figure 16) et supérieures pour les saisonnalités grow, growing et multi. Bien que les médianes soient de valeurs similaires, le modèle monocapteur avec la RMSE la plus faible est obtenu avec Sentinel-2 et le meilleur modèle est obtenu avec la combinaison des deux capteurs. Concernant H95, Sentinel-1 possède une RMSE médiane plus faible que Sentinel-2 et que les capteurs fusionnés. Pour les dimensions 1 et 2, les deux satellites ont des RMSE similaires, et leur fusion semble légèrement améliorer les performances.

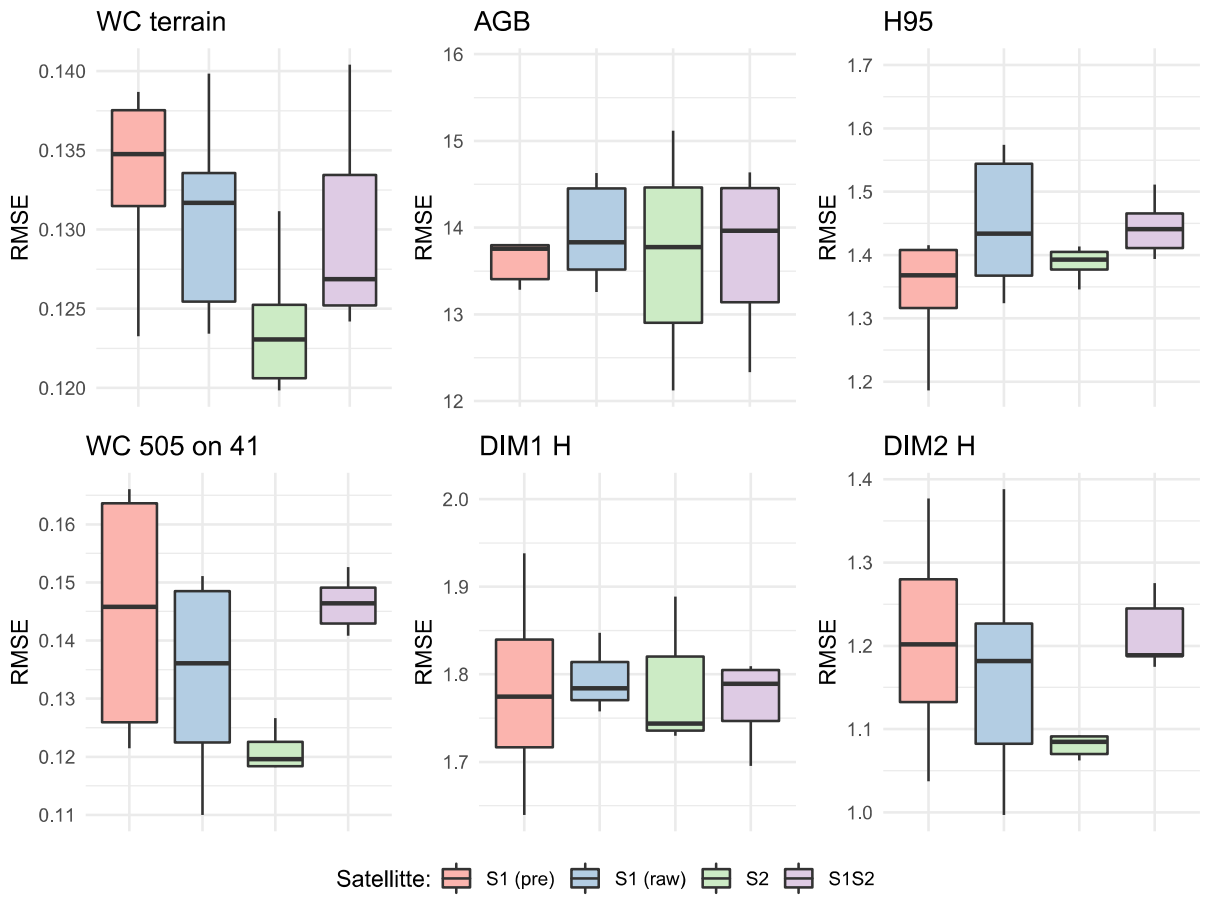


FIGURE 15 – Valeurs de RMSE des modèles de terrain et de photointerprétation pour le WC en fonction de la saison, du prétraitement. L'écart interquartile est dû aux différences des RMSE entre les saisonnalités employées. Les échelles de RMSE sont différentes pour les deux graphiques.

3.6 Influence saisonnalité

Comme illustré à la Figure 16, concernant le WC, le jeu de terrain ne permet de distinguer de tendances, le jeu de photointerprétation validé sur les parcelles de terrain, quant à lui, met en avant les saisons grow et multi pour Sentinel-1 et la fusion s1-s2, tandis qu'il montre que les saisons drying et growing sont plus défavorables pour Sentinel-1. Pour Sentinel-2, les saisonnalités dry et drying paraissent les plus adaptés. Concernant l'AGB, les saisonnalités multi et grow semblent les plus favorables pour Sentinel-2, et les saisonnalités multi et dry paraissent les plus adaptées à Sentinel-1. La saison growing, quant à elle, parait comme la plus défavorable pour la cartographie de l'AGB pour les deux Sentinel, et la saisonnalité year parait plus défavorable pour Sentinel-2. Pour ce qui est de l'H95, la saisonnalité multi apparaît comme la plus favorable pour les deux Satellites. Pour Sentinel-1, la modalité monosaisonnale la plus adaptée parait être la saison dry, tandis que pour Sentinel-2, grow et year présentent les RMSE les plus faibles. Pour les différentes variables réponses, les modèles multisaisonnaux diminuent généralement les RMSE et la fusion radar-optique semble diminuer l'influence de la saisonnalité.

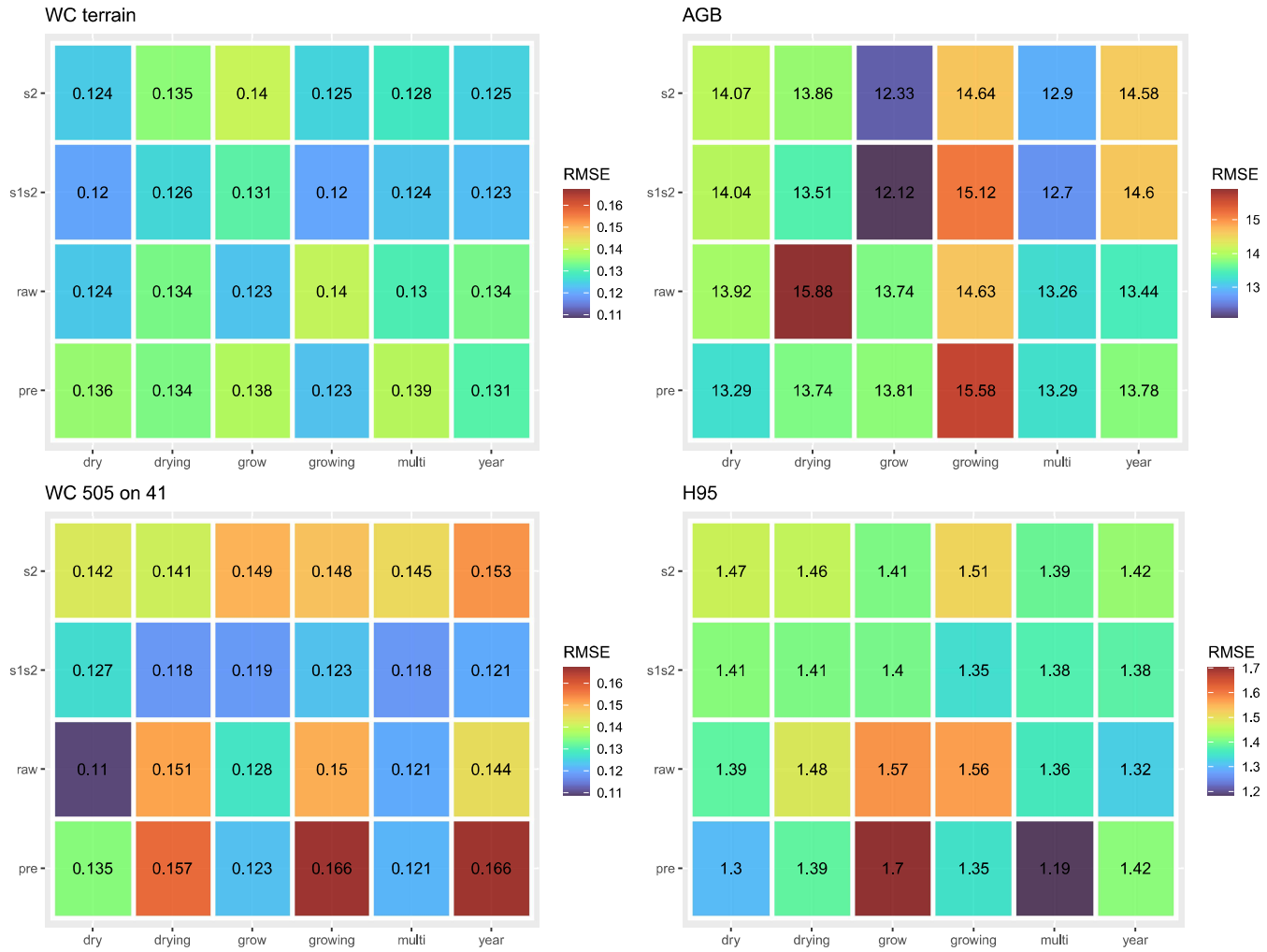


FIGURE 16 – Valeurs de RMSE en fonction de la saison et du capteur des modèles de terrain pour les variables réponses WC, AGB et H95 validées en CV, et du modèle de photointerpétration WC validé sur les données de terrain.

3.7 Importance des variables

3.7.1 Couverture ligneuse

Les variables sélectionnées par Boruta pour les modèles multisaisonnaux de WC de terrain sont présentées à la Figure 17. Pour les données Sentinel-1, un total de 29 bandes ont été sélectionnées. Il est observable que parmi les indices, huit produits VV+VH ont été sélectionnés alors qu'aucun quotient VV-VH n'a été sélectionné. Parmi les métriques temporelles, la moyenne annuelle a été sélectionnée pour chaque polarisation et chaque orbite. Parmi les métriques de texture, seule la corrélation a été sélectionnée une fois. Concernant les bandes, il n'y a pas de tendance qui se marque au niveau des orbites, les orbites ascendantes et descendantes sont sélectionnées chacune 8 fois. Pour les polarisations, les deux polarisations sont également sélectionnées huit fois chacune, mais parmi les bandes ayant obtenu les scores les plus élevés, les polarisations VH sont plus fréquentes. Toutes les saisonnalités hors métriques temporelles ont été sélectionnées hormis les bandes annuelles (year) et les deux seules bandes ayant un score Z supérieur à 6 appartiennent à la saison grow.

Pour les données Sentinel-2, les bandes possédant un score Z supérieur à 4 sont toutes des bandes SWIR, NIR, et REDEDGE, ces bandes sont sélectionnées 25 fois sur 37 bandes au total. Parmi les sept

bandes possédant un score Z supérieur à 5, cinq appartiennent à la saison dry. Les indices sélectionnés sont NDVI et MSAVI2 pour les saisons dry et drying. Les métriques texturales sélectionnées sont le "sum average" également pour les saisons dry et drying.

Pour les données Sentinel-1 et Sentinel-2 combinées, les 18 bandes avec les scores Z les plus élevés parmi les 41 bandes sélectionnées sont toutes des bandes optiques SWIR, NIR et REDEDGE. Pour Sentinel-2, une métrique de texture, le NDVI sum average, est sélectionnée et l'indice MSAVI2 est sélectionnée pour. Huit prédicteurs Sentinel-1 sont sélectionnés, parmi lesquels : un indice, le produit VV+VH, qui a le score le plus élevé des prédicteurs S-1, deux métriques temporelles, les moyennes annuelles des orbites ascendantes et descendantes de la polarisation VH et cinq bandes.

L'importance et la sélection des variables pour le WC photointerprété (cf. Annexe D) présentent des similarités et quelques différences avec les modèles de terrain. Parmi les 41 variables sélectionnées par Boruta pour Sentinel-1. Parmi les similarités, la saison grow obtient également les scores les plus élevés, le produit VV+VH est également sélectionné 8 fois et la moyenne annuelle a été sélectionnée pour chaque polarisation et chaque orbite. A contrario, le quotient VV-VH est sélectionné sept fois alors qu'il n'avait pas été sélectionné pour le jeu de terrain. Autre différence, la métrique texturale "dissimilarité" est sélectionnée sept fois tandis que pour le jeu de terrain la seule métrique texturale sélectionnée était la corrélation sélectionnée une fois. Concernant Sentinel-2, les bandes SWIR, NIR ET REDEDGE sont sélectionnées plusieurs fois comme pour le jeu de terrain, mais un poids plus important des indices de végétation et des métriques de texture est observé. La fusion Sentinel-1 et Sentinel-2, quant à elle, ne présente plus uniquement des bandes optiques avec les scores les plus élevés, le score Z le plus élevé étant à ici une bande VH.

3.7.2 Biomasse aérienne

Les variables sélectionnées pour l'AGB sont présentées à la Figure 18. Pour Sentinel-1, six bandes brutes sont sélectionnées, parmi celles-ci, cinq sont polarisées VH. Le produit VV+VH est également sélectionné pour la saison grow en orbite ascendante et possède le troisième score Z le plus élevé. La moyenne annuelle pour les orbites ascendantes et descendantes de la polarisation VH est sélectionnée. La métrique texturale est calculée pour la saison drying en polarisation VV et on orbite descendante. Au total, hormis la métrique de texture et le produit VVVH, touts les prédicteurs sélectionnés sont en polarisation VV.

Pour Sentinel-2, parmi les six bandes possédant les scores Z les plus élevés, cinq sont des bandes SWIR. Les trois indices de végétation sont sélectionnés. Pour les métriques de texture, le sum average est sélectionné pour les saisons dry et drying.

La fusion des bandes Sentinel-1 et 2 sélectionne moins de variables que pour les satellites utilisés séparément. Cinq bandes SWIR, une bande RED et une bande VH en orbite ascendante sont sélectionnées.

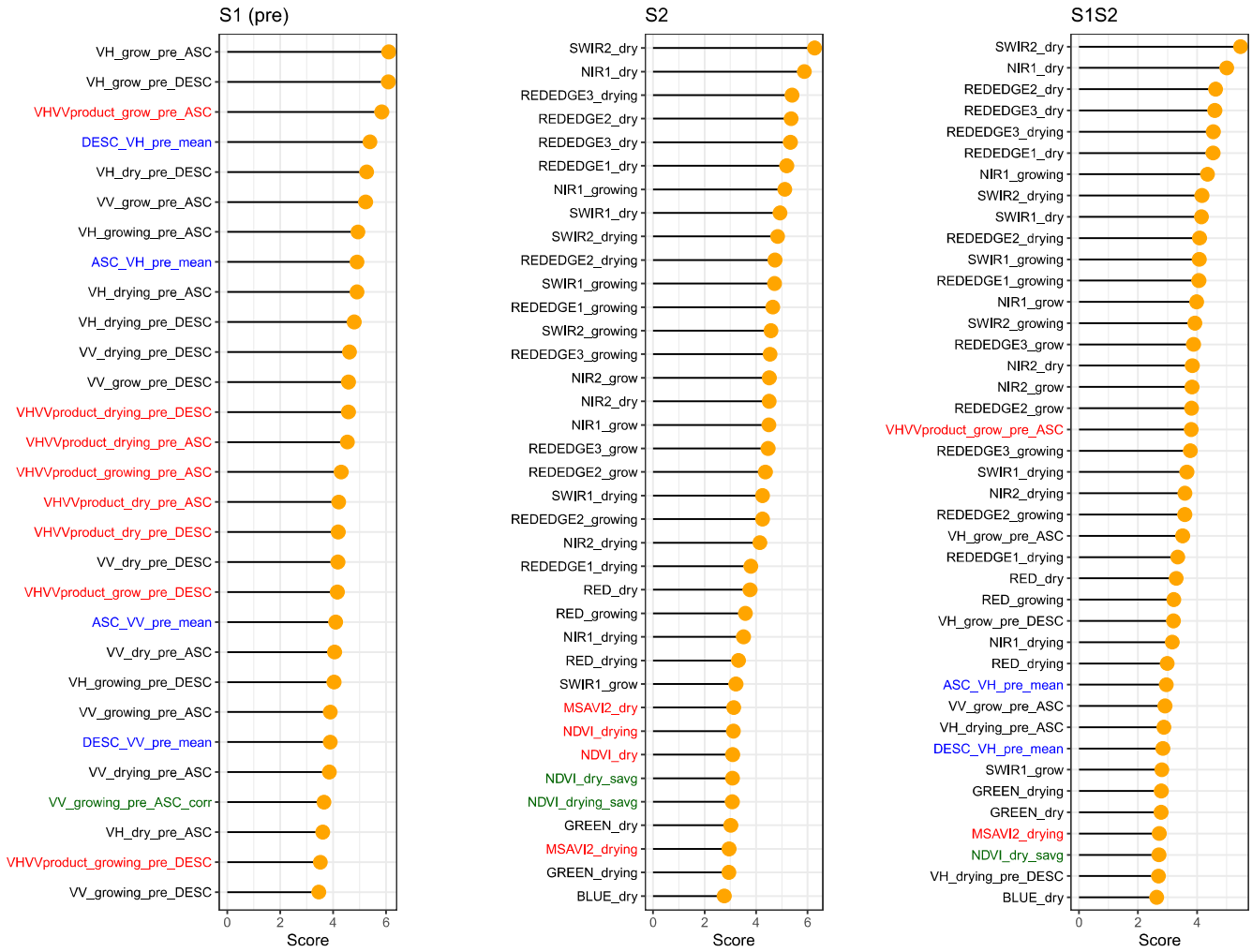


FIGURE 17 – Comparaison des scores Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour le WC. Les prédicteurs en noir sont des bandes, en rouge des indices dérivés, en bleu des métriques temporelles et en vert des indicateurs de texture.

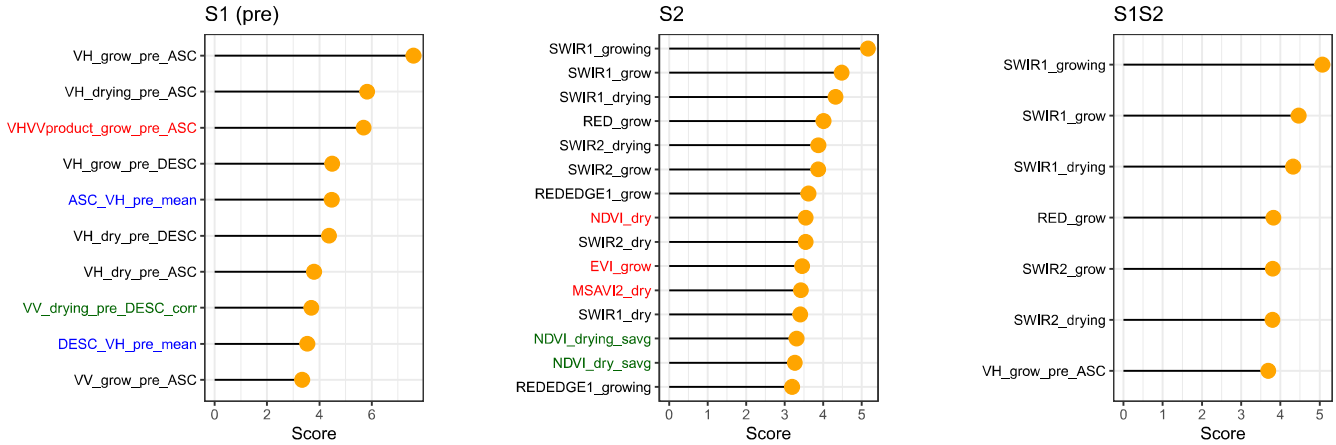


FIGURE 18 – Comparaison des scores Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour l'AGB. Les prédicteurs en noir sont des bandes, en rouge des indices dérivés, en bleu des métriques temporelles et en vert des indicateurs de texture.

3.7.3 Hauteur de canopée

Les variables sélectionnées sont présentées à la Figure 19. Parmi les variables sélectionnées pour Sentinel-1, il est remarquable que le quotient VH-VV a obtenu les deux scores le plus élevés et a été sélectionné trois fois. Le produit VH+VV a quant à lui été sélectionné une fois. Un nombre supérieur de bandes VH ont été sélectionnées comparé aux bandes VV ont . Les bandes VH ont été sélectionnées au nombre supérieur de fois aux bandes VV et ont obtenu un score moyen plus élevé. La métrique de texture shade a été sélectionnée une fois.

Quant à Sentinel-2, les bandes BLUE ont obtenu les scores les plus élevés. Les bandes SWIR, GREEN, RED et dans une moindre mesure NIR paraissant également importante. L'indice EVI a obtenu le score le plus élevé parmi les indices, suivis de MSAVI2 et NDVI. La métrique de texture sum average a été sélectionnée deux fois.

Pour la combinaison des capteurs, seules des variables prédictives Sentinel-2 ont été sélectionnées, principalement de la saison drying. Les variables ayant obtenu les scores Z les plus élevés sont BLUE, SWIR et EVI.

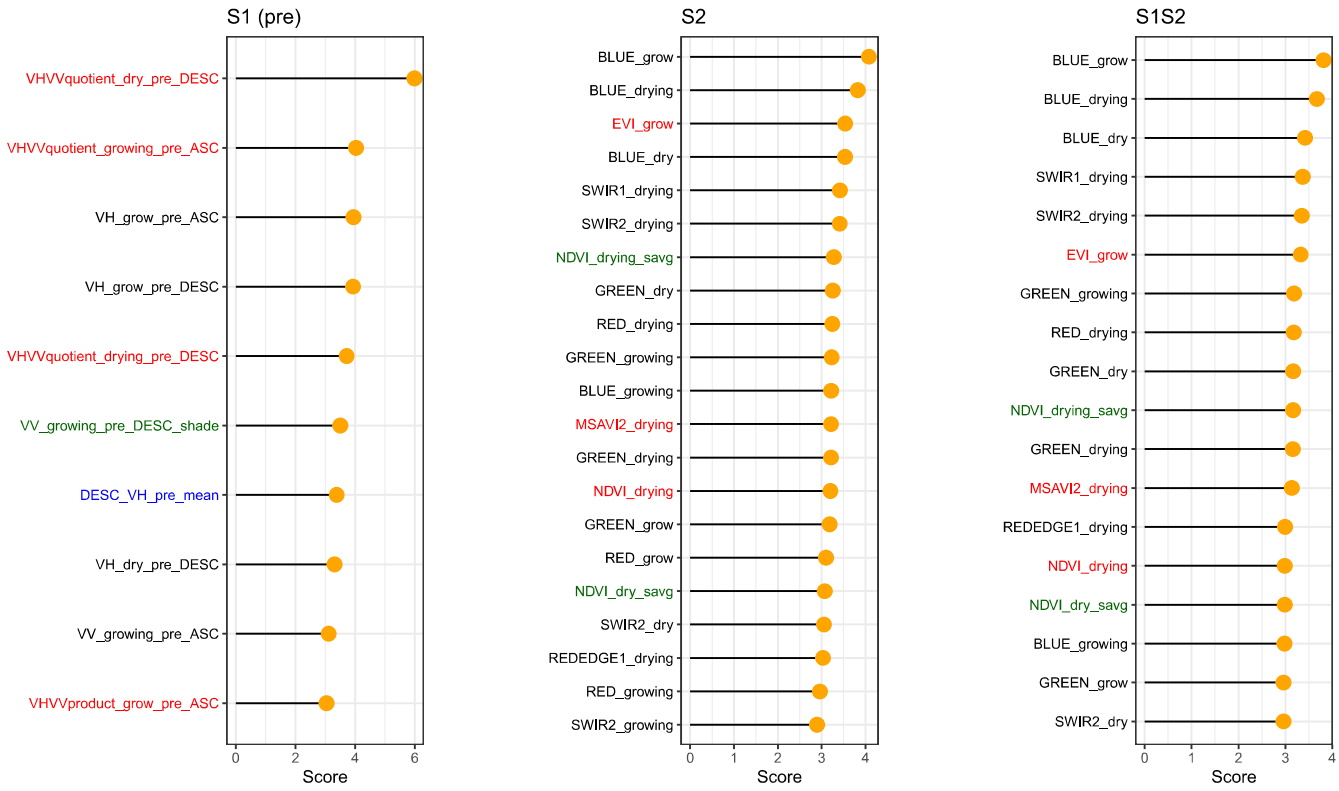


FIGURE 19 – Comparaison des scores Z des variables sélectionnées par Boruta pour les modèles multisaisonnaux Sentinel-1 (prétraité), Sentinel-2 et la fusion des deux pour l'H95. Les prédicteurs en noir sont des bandes, en rouge des indices dérivés, en bleu des métriques temporelles et en vert des indicateurs de texture.

3.8 Cartes prédites

La Figure 20 présente les cartes finales de WC, AGB et H95. La carte de WC est réalisée avec les données de photointerprétation et les cartes d'AGB et H95 sont réalisées avec les données de terrain. Les modèles multisaisons et multicapteurs ont été utilisés pour les cartes de WC et d'AGB, tandis que le modèle Sentinel-1 multisaison a été utilisé pour la carte d'H95.

		WC [-]	AGB [T.ha^{-1}]	H95 [m]
	Moyenne	0,568	17,668	4,120
S-1 (pre)	Médiane	0,602	16,502	4,283
	Ecart-type	0,217	12,236	1,252
	Moyenne	0,571	20,676	4,693
S-2	Médiane	0,605	21,857	4,852
	Ecart-type	0,261	10,506	0,939
	Moyenne	0,577	20,474	4,770
S-1 & S-2	Médiane	0,623	21,410	4,915
	Ecart-type	0,266	11,440	0,959

Table 9 – Moyennes, médianes et écart-types des cartes prédites avec les modèles multisaisons

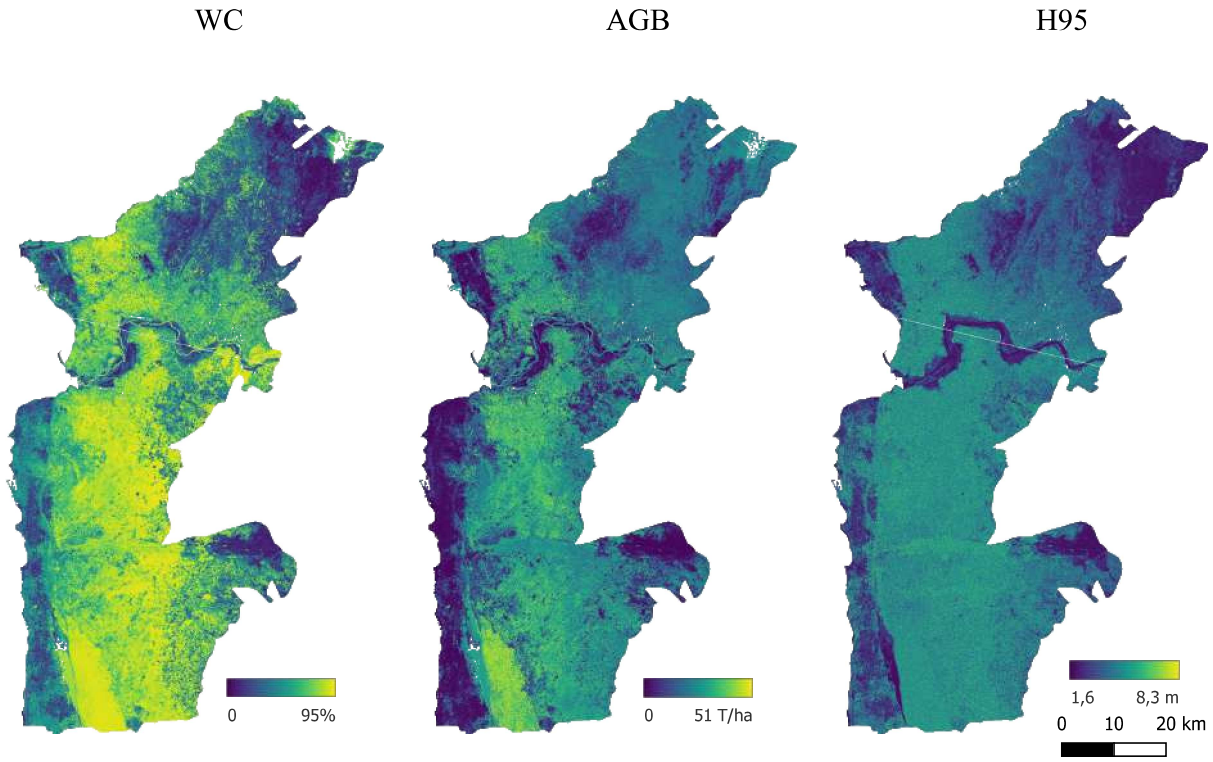


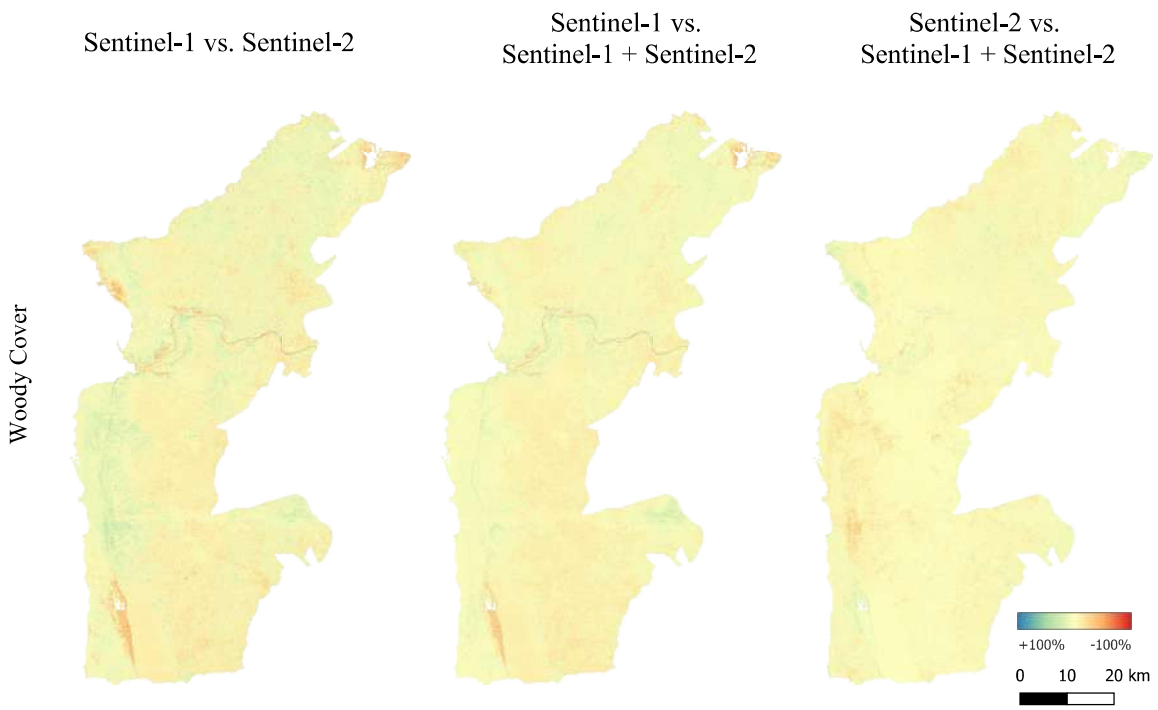
Figure 20 – Cartes finales de WC, AGB et H95

3.9 Répartition spatiale des différences entre satellites

La Figure 20 illustre les différences de WC entre les capteurs. Il est visible que Sentinel-1 estime plus de WC dans les zones côtières plus ouvertes que Sentinel-2 et Sentinel-2 estime plus de WC dans les zones plus denses. Les différences d'estimation du WC semblent ainsi plus marquées pour Sentinel-2. Les comparaisons entre Sentinel-1 et s1+s2 illustrent des tendances similaires avec une estimation de WC supérieure dans les zones plus denses pour s1+s2.

Pour l'AGB, il est visible que Sentinel-1 estime davantage d'AGB à la jonction des zones côtières et du plateau ainsi qu'au bord du fleuve Onilahy. Sentinel-2 quant à lui estime davantage d'AGB dans les zones plus denses ainsi qu'au nord/est de la zone d'étude. Les tendances entre Sentinel-1 et les capteurs fusionnés sont similaires. Les différences entre Sentinel-2 et s1s2 sont quant à elles moins marquées.

Pour H95, la comparaison des deux satellites indique une estimation supérieure de Sentinel-2 dans les zones ouvertes à l'Est correspondant à des zones de savanes anthropiques ou de cultures. Il est observable que les différences entre Sentinel-2 et la fusion des capteurs sont minimes, ce qui est sans doute dû à la sélection de variables qui n'a sélectionné que des variables Sentinel-2.



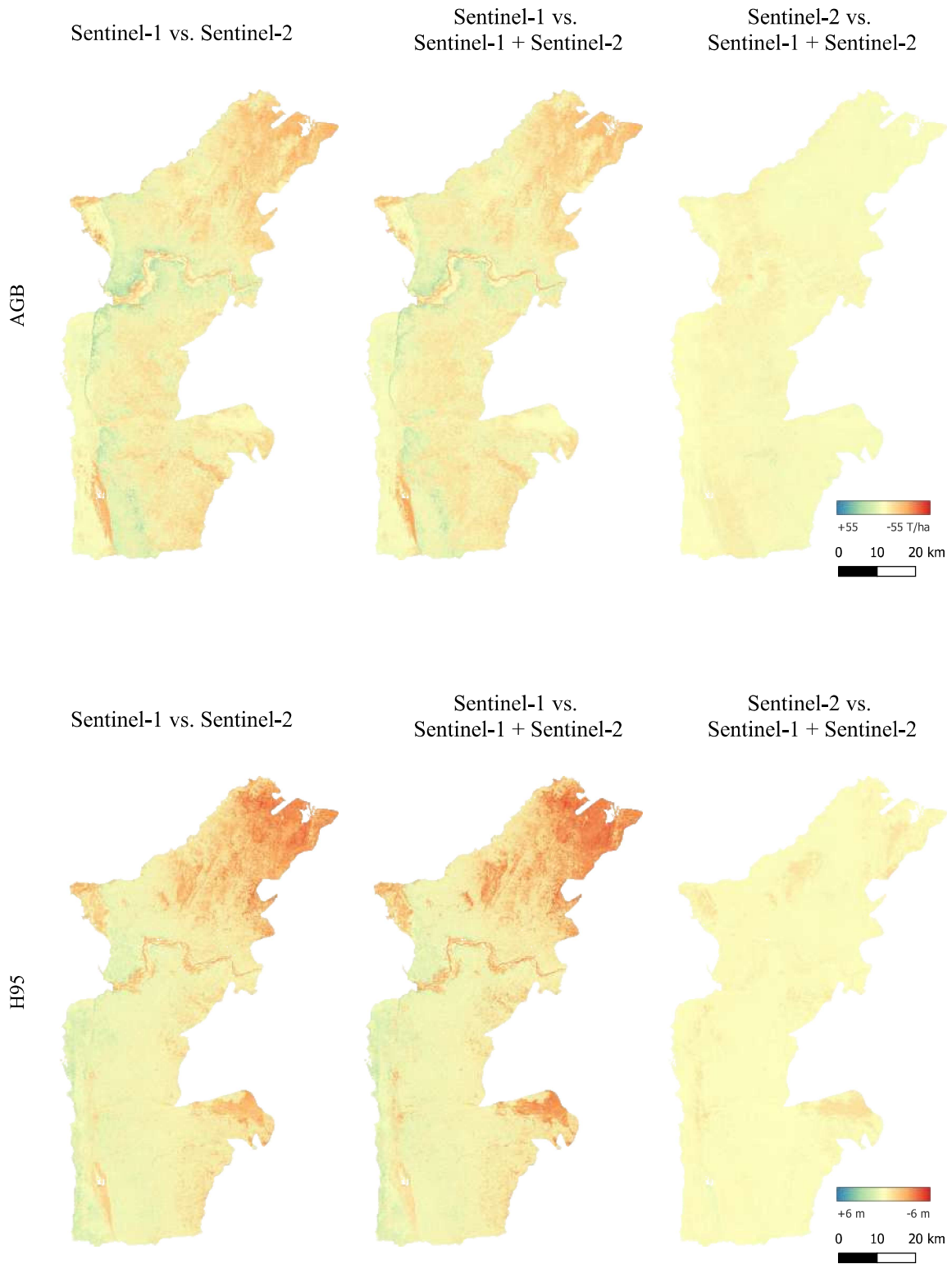


FIGURE 20 – Différences en fonction des trois capteurs pour les variables réponses WC (jeu de photointerprétation), AGB, et H95 avec les modèles multisaisons. Les différences positives indiquent des estimations plus élevées pour le jeu de données avec le capteur portant le premier nom (par exemple, Sentinel-1 dans la carte gauche), tandis que les différences négatives indiquent des valeurs plus élevées dans le produit portant le second nom (par exemple, Sentinel-2 dans ce cas).

4 Discussion

Avant d'entamer la discussion, les résultats principaux sont très brièvement résumés. Concernant, les variables réponses, les valeurs de RSE dans l'ordre croissant sont : WC, H95, AGB, Dim2H et Dim1H. La validation des modèles de photointerprétation sur les données de terrain présentent des RMSE similaires ou légèrement supérieures. Les prétraitements supplémentaires n'ont pas présenté de tendances marquées sur les RMSE. Les meilleurs modèles monocapteurs sont obtenus à l'aide de S-2 pour AGB, et à l'aide de S-1 pour WC et H95. Les effets de la saisonnalité varient en fonction des capteurs et de variables réponses employés. La sélection de variables augmente la précision des modèles. Les bandes et indices obtiennent en moyenne les scores les plus importants pour Sentinel-1. La fusion des capteurs sélectionne davantage de variables Sentinel-2 pour les jeux de données de terrain. La répartition spatiale des différences entre les satellites indique davantage de WC estimé dans les zones ouvertes par S-1 et davantage de WC estimé dans les zones plus denses par S-2. Ce dernier estime également des valeurs élevées de H95 dans les zones ouvertes à l'est et au nord-est de la zone d'étude.

4.1 Potentiel de photointerprétation

La validation des modèles de photointerprétation sur les données de terrain (validation externe) présente des valeurs de RMSE égales ou légèrement supérieures et des RSE plus faibles que les données de terrain en cross-validation (cf. Figures 13 et 14). Les résultats de cette validation externe indiquent donc la possibilité d'extrapoler des valeurs de WC aux autres variables réponses avec des performances similaires.

Pour le WC, des RMSE et MAE bien plus importantes sont observées pour les modèles de photointerprétation en CV que pour les modèles de terrain en CV et les modèles de photointerprétation validés sur les données de terrain, tout en présentant des valeurs de RSE plus faibles. Une partie de ces différences est possiblement due aux différences de distribution des valeurs de WC de terrain et de photointerprétation. Ce dernier jeu de données possède une médiane supérieure et plus de données aux extrêmes : 49 valeurs minimales (0) et 81 valeurs maximales (0,951). Ainsi, bien que les RMSE de photointerprétation en CV soient supérieurs, la comparaison des RSE réaffirme le potentiel de l'évaluation du WC par photointerprétation.

Les résultats indiquent ainsi la pertinence de tirer avantage des corrélations existantes entre les variables réponses (cf. Figure 8) afin de pouvoir décrire divers paramètres structuraux tout en augmentant la taille et en améliorant la distribution spatiale du jeu de données. Néanmoins, les conclusions tirées ici n'ont pu être corroborées par la littérature.

La comparaison des RSE illustre un effet potentiel lié à la taille et à la répartition des jeux de données. En effet, il est remarquable que la RSE de validation croisée des modèles de photointerprétation est systématiquement inférieure à la RSE des modèles de terrain en CV. Une explication possible est que la répartition spatiale non équilibrée du jeu de terrain aboutit à des validations de jeux d'entrainement parfois très différents. Le nombre supérieur d'observations et la répartition plus équilibrée du jeu de photointerprétation permettent un nombre supérieur de folds et une répartition plus équilibrée de ceux-ci. L'Annexe E illustre une répartition des folds pour les jeux de terrain et de photointerprétation.

4.2 Comparaisons variables réponses

La comparaison des RSE des données de terrain (cf. Figure 14) classe les performances des modèles de manière décroissante dans l'ordre suivant : WC, H95, AGB, Dim2H et Dim1H. L'écart interquartile du WC de terrain est le plus faible. La relation entre le WC et les valeurs de réflectance et de rétrodiffusion

des capteurs optiques et radar semble ainsi moins impactée par les saisonnalités et capteurs employés avec des différences de RSE sont moins marquées que pour les autres variables réponses.

La précision des modèles d'AGB est considérablement plus faible que celle des modèles de WC. Bien qu'une part de cette différence peut être expliquée par la relation moins directe entre les valeurs de réflectance et rétrodiffusion des capteurs et l'AGB, cette dernière étant une variable composite, la différence peut être aussi due aux erreurs liées à l'estimation de l'AGB. En effet, d'une part, la densité de bois n'a pu être trouvée pour tous les individus. D'autre part, l'équation de Vieilledent et al. (2012) pour les forêts sèches épineuses de Madagascar a été utilisée en dehors de son domaine de validité (5 à 48 cm) principalement pour les baobabs (*Adansonia sp.*). L'utilisation d'une équation allométrique spécifique à ce genre telle que celle de Malimbwi et al. (2018) construite sur les baobabs de Tanzanie pourrait peut-être améliorer la précision de l'estimation de biomasse. D'autres challenges pour la caractérisation de l'AGB, plus courants, tels que la présence de pixels mixtes (contenant du sol et de la végétation), ou plus spécifiques à la zone d'étude, tels que la grande variabilité de compositions d'espèces associées aux gradients de la zone et à l'ampleur de la dégradation influencent également la précision. En effet, les parcelles sont caractérisées par grande variabilité qui se traduit par une gamme d'AGB allant 0,35 $T.ha^{-1}$ à 66,29 $T.ha^{-1}$.

Concernant l'H95, une RMSE médiane des modèles de 1,4 m est observée pour une hauteur médiane de 3,6m et une hauteur moyenne de 4,1m. Les performances sont similaires à Kacic et al. (2021) qui a obtenu une RMSE médiane de 1,6m pour une hauteur de canopée moyenne de 5,3m dans le Chaco paraguayen. Dim2H possède des performances légèrement inférieures à la biomasse, mais ne représente que 17,4 % de la variabilité des classes de hauteur définies. Quant à Dim2H , elle possède un RSE supérieur à deux et représente 41 % de la variabilité des classes de hauteur définies.

Les performances de ces deux métriques de la structure verticale des parcelles semblent indiquer que ces dernières sont moins adaptées à la zone d'étude et ne sont donc pas discutées dans les sous-sections suivantes. A l'inverse, les performances des variables WC, H95 et dans une moindre mesure AGB, semblent indiquer ces dernières comme plus appropriées pour la description et l'évaluation de la végétation de la zone.

4.3 Facteurs d'influence des performances de Sentinel-1

Bien que la sous-section précédente a mis en exergue le potentiel de la photo-interprétation, et qu'un nombre d'observations supérieur est avantageux pour la modélisation, la discussion des résultats est réalisée sur les jeux de données de terrain. Pour cause, les variables réponses extrapolées, qui sont à 100% corrélées entre elles, ne le sont en réalité pas. Il parait dès lors imprudent de tirer des conclusions sur celles-ci.

Contrairement aux résultats escomptés, l'implémentation des prétraitements (Mullissa et al., 2021) n'a pas eu d'effet marqué sur la précision des modèles. Cela peut être dû d'une part à une configuration non optimale des prétraitements. En effet, le choix du filtre speckle et des paramètres associés a reposé, comme recommandé par Mullissa et al. (2021), sur une évaluation visuelle de la réduction du speckle tout en conservant des détails, et bien que la configuration choisie a paru optimale par l'opérateur, cette dernière a pu être biaisée. Une seconde explication est possible. Les parcelles sur lesquelles repose l'évaluation des modèles ne sont que peu concernées par les prétraitements supplémentaires autres que le filtre speckle, à savoir la normalisation radiométrique de terrain et la correction de bordure de bruit additionnelle. Or, les données ici utilisées ont été agrégées par des médianes temporelles, réduisant ainsi le speckle (Glaude & Orban, 2022). Les différences de speckle entre les jeux de données prétraités et non prétraités seraient alors minimes. Bien que les RMSE semblent peu impactés, la comparaison visuelle

d'images monodates, avec et sans prétraitements, suggère l'importance de ces derniers. Les effets de la topographie sont atténués et les zones de distorsion, pour lesquelles les prédictions sont peu fiables, sont masquées, empêchant des interprétations erronées de ces dernières. L'Annexe F présente une visualisation des l'effets des prétraitements. De plus, bien que pour ce travail l'agrégation temporelle réduit l'utilité du filtre speckle, celui-ci peut se montrer important pour un suivi de la zone, à l'instar de Reiche et al. (2021) qui utilise des données Sentinel-1 avec les prétraitements supplémentaires Mullissa et al. (2021) pour repérer des perturbations dans le bassin du Congo.

L'importance des variables a montré que pour le WC et l'AGB, les bandes possèdent les scores les plus élevés. À l'inverse, pour H95, les quotients VH-VV possèdent les scores les plus élevés et les indicateurs dérivés constituent la majorité des indicateurs sélectionnés. Bien que les bandes ont obtenu les scores Z les plus élevés pour le WC et l'AGB, quasiment la moitié des prédicteurs sélectionnés étaient des indicateurs dérivés. Le produit VH+VV s'est montré particulièrement adapté pour le WC (jeu de terrain et photointerprétation) tandis que le quotient VH-VV s'est montré adapté pour le WC de photointerprétation et l'H95. Concernant l'AGB, les résultats obtenus sont en contradictions avec Forkuor et al. (2020) qui a trouvé que les indicateurs dérivés étaient le plus important pour cartographier l'AGB dans des zones arides d'Afrique de l'Ouest. Dans le cas du quotient VH-VV, Laurin et al. (2018) a trouvé cet indicateur comme étant le plus important pour cartographier l'AGB. Bien que pour notre zone, le quotient se soit montré surtout important pour H95 et WC de photointerprétation et non pour AGB, cela peut s'expliquer par le fait que ces deux métriques sont corrélées à l'AGB et possèdent des performances supérieures. Parmi les autres indicateurs dérivés, la texture s'est révélée surtout importante pour le WC de photointerprétation et dans une moindre mesure pour les autres variables réponses. Les résultats obtenus sont en accord moyen avec Mishra et al. (2019a) qui soulignait l'importance "vitale" des textures pour les données SAR utilisées pour de la classification. Pour les métriques temporelles, la seule qui a été sélectionnée pour les différentes variables réponse est la moyenne annuelle. Les autres métriques temporelles explicitant les variations intra-annuelles apparaissent donc peu importantes pour les écosystèmes. Les bandes, les indices VH+VV et VV-VH et dans une moindre mesure la texture semblent ainsi importants pour la cartographie de métriques de structure pour notre zone d'étude.

Pour les différentes variables réponses, la polarisation VH s'est montrée systématiquement plus adaptée que la polarisation VV. Ces résultats sont corroborés par ceux de Forkuor et al. (2020), pour qui les bandes VH étaient également plus importantes que les bandes VV pour la cartographie de l'AGB, ainsi que par ceux de Heckel et al. (2020) qui a obtenu les mêmes tendances pour la classification de zones de forêts/non-forêts en Afrique du Sud. La polarisation croisée VH des données SAR en bande C est plus sensible à la rétrodiffusion volumique, et donc plus apte à déterminer la structure de la végétation (Heckel et al., 2020).

L'impact de la saisonnalité varie en fonction des variables réponses. Pour le WC de photointerprétation [5], la saison grow semble la plus adaptée. L'hypothèse est que lors de cette saison la couronne de la végétation et les feuilles qui la composent influencent davantage la réponse de la bande C, le rendant plus sensible au WC. Concernant les modèles d'AGB et H95, la saison dry semble être la modalité monosaisonnale la plus adaptée. Ces résultats sont appuyés par les résultats de Forkuor et al. (2020) et Laurin et al. (2018) qui ont également trouvé de meilleurs résultats pour la cartographie de l'AGB en saison sèche. D'une part, cela peut s'expliquer par le fait que l'humidité du sol influence la réponse de la bande C et peut réduire la sensibilité des données SAR à la biomasse. D'autre part, lors de la saison de végétation dry, les branches de la végétation décidue sont davantage exposées, ce qui permet une meilleure pénétration dans la canopée des ondes. Par ailleurs, la sensibilité de Sentinel-1 à l'humidité

5. La présence de contradictions dans les valeurs de RMSE en le WC de terrain et photointerprété a conduit au choix du WC de photointerprétation en validation externe pour discuter de la saisonnalité à cause de son RSE inférieur.

du sol peut être la cause des performances globalement plus faibles de la saison growing pour le WC photointerprété, l'AGB et l'H95. En effet, la croissance de la végétation coïncide avec la saison des pluies. Lors de cette période, le foliage de la végétation ligneuse n'est pas encore totalement développée et l'influence de l'humidité du sol serait plus proportionnellement importante que pour la saison grow. Il faut toutefois garder à l'esprit que les différences inter-saisonnales sont faibles et que les précisions des modèles, surtout AGB et H95, restent limités, nous mettant en garde contre la surintérprétation des résultats. Les modèles multisaisonnaux améliorent la précision du WC de photointerprétation, de H95 et et égalent la précision du meilleur modèle mono-saisonnal d'AGB. L'utilisation de modèles multisaisonnaux apparaît dès lors pertinente pour notre zone d'étude.

Une limite identifiée quant à l'évaluation de l'impact de la saisonnalité est le choix de la saison de végétation maximale, qui a été choisie en sélectionnant les deux moyennes mensuelles de NDVI les plus élevées à savoir février et mars. Néanmoins, la Figure 9 indique que la valeur de NDVI moyen d'avril n'est que légèrement inférieure à celle de mars et aurait pu être incluse dans la saison de végétation maximale. Des différences plus marquées entre les saisons grow et drying auraient potentiellement pu être mises davantage en avant.

4.4 Sentinel-1 versus Sentinel-2

Concernant le WC photointerprété et H95, Sentinel-1 a fourni des modèles plus performants que Sentinel-2. Plus spécifiquement pour le WC, l'inspection de la carte des différences entre S-1 et S-2 (cf. Figure 20) semble indiquer une surestimation de S-2 par rapport à S-1 dans les zones les plus denses. L'inspection de la régression linéaire entre les valeurs observées et prédites (cf. Annexe G) indique que le modèle Sentinel-1 multisaisons a tendance à surestimer le WC dans les zones plus ouvertes et que le modèle Sentinel-2 multisaisons a tendance à sous-estimer le WC dans les zones plus ouvertes et à le surestimer dans les zones plus denses. Les plus faibles performances de Sentinel-2 pour le WC photointerprété peuvent être expliquées d'une part, par l'influence de la strate basse, notamment du couvert herbacé qui, bien que moins important que dans des savanes, est présent dans certaines zones dégradées. Ceci pourrait être une explication pour le RMSE plus élevé de Sentinel-2 en saison grow. Une seconde explication possible est l'influence du sol. En effet, la Figure 21, qui présente une jonction entre un sol ocre et un sol roux avec une couverture ligneuse qui semble similaire sur les deux sols, montre des différences d'estimation entre Sentinel-1 et Sentinel-2 (modèles multisaisonnaux). Ce dernier estime davantage de couverture dans la zone de sol roux tandis que Sentinel-1 en estime davantage dans le sol ocre à côté.

Pour H95, les plus faibles performances de Sentinel-2 peuvent être en partie expliquées par les mêmes raisons expliquant les plus faibles performances pour WC. Un élément remarquable est la surestimation de Sentinel-2 dans les zones ouvertes à l'Est et au Nord-Est de la zone qui apparaît à la Figure 20. Ces surestimations spécifiques à ces zones peuvent être également liées au fait qu'aucun inventaire de terrain n'ait été réalisé dans ces zones. Cette lacune pourrait être comblée à l'aide des données GEDI à l'instar de Potapov et al. (2021) et Kacic et al. (2021) qui les ont utilisées comme données de références pour modéliser H95.

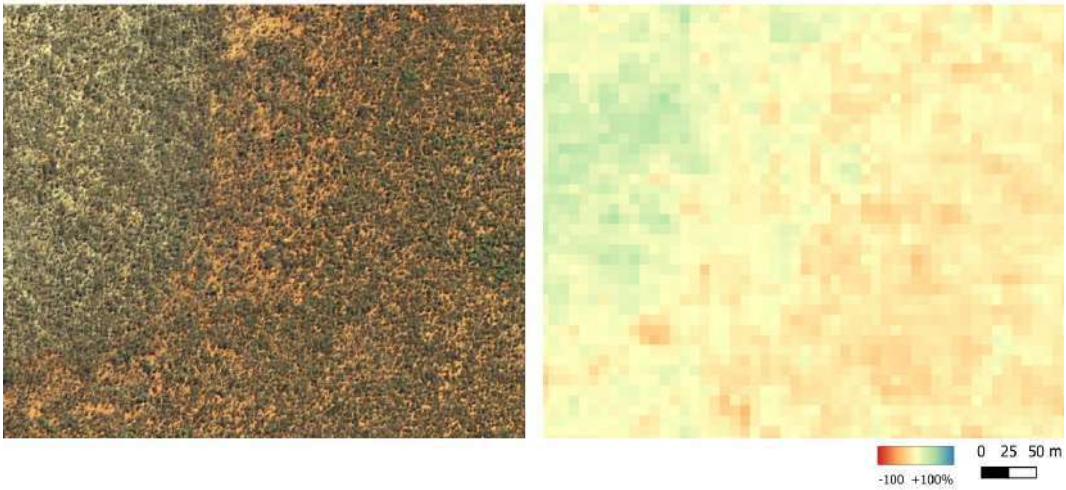


FIGURE 21 – Influence du sol sur la différence entre les valeurs de WC Sentinel-1 et Sentinel-2 (modèles multisaisonnaux). Les valeurs positives indiquent des valeurs plus élevées de Sentinel-1, tandis que des valeurs négatives indiquent des valeurs de Sentinel-2 plus élevées.

A Contrario d'H95 et WC, Sentinel-2 a fourni de meilleurs modèles pour l'AGB. Ce résultat agrée avec d'autres études qui ont également comparé les données optiques et radar pour la cartographie de l'AGB (Forkuor et al., 2020; Fremout et al., 2022). Une raison possible est la limitation de la bande C pour cartographier l'AGB. La bande C a une capacité de pénétration limitée dans la végétation et le signal est sujet à la saturation. Une solution possible à cette limitation, serait l'inclusion de données de longueurs d'onde plus longues, telles que les données à bande L ou bande P, qui sont caractérisées par une pénétration de la végétation plus importante et pourraient améliorer la cartographie de l'AGB, notamment de la zone de plus forte biomasse. Laurin et al. (2018) a obtenu de meilleurs résultats en combinant les données Sentinel-1 et ALOS2 (bande L) qu'avec Sentinel-1 seul pour cartographier l'AGB en Italie centrale.

La combinaison des données Sentinel-1 (prétraitées) et Sentinel-2 a presque systématiquement amélioré la précision des modèles de WC et d'AGB. D'autres études ayant étudié la fusion radar-optique pour diverses métriques structurales de la végétation ont obtenu des résultats similaires (Laurin et al., 2018; Forkuor et al., 2020; Fremout et al., 2022; Baumann et al., 2018). Une autre tendance observée intéressante est que la fusion des données Sentinel-1 et Sentinel-2 diminue l'influence de la saisonnalité sur les résultats. Ces conclusions peuvent être expliquées par les différences des techniques d'imagerie et les informations qu'elles apportent. Les données SAR sont plus influencées par la structure de la végétation tandis que les données optiques sont plus influencées par la densité de la canopée et les caractéristiques liées au foliage. Ces informations complémentaires bénéficient à la précision lorsqu'elles sont combinées. Toutefois, il est remarquable que pour H95, la fusion des capteurs n'a augmenté les performances que pour deux saisonnalités et l'a diminué pour trois. Ces résultats sont vraisemblablement dus au fait que la sélection de variables donne plus d'importance aux variables S-2 même si les modèles monocapteurs S-2 ont présenté de plus faibles précisions.

En réalité, lors de la fusion des modèles de terrain multisaisons S-1 et S-2, Boruta sélectionne majoritairement des variables prédictives Sentinel-2. Pour l'AGB, une seule variable S-1 est sélectionnée lors de la fusion des capteurs, cela peut néanmoins s'expliquer en partie par le fait que le modèle S-2 multi possède un RMSE plus faible que S-1 multi et que leur fusion obtient une RMSE encore plus faible. Pour le WC de terrain, la fusion augmente également les performances des modèles multisaisonnaux. A contrario, pour H95, le modèle S-1 multi obtient des performances plus élevées que S-2 multi, mais uniquement des variables S-2 sont sélectionnées par Boruta lors de la fusion des deux capteurs.

Rudnicki et al. (2015) a testé la sensibilité de Boruta [6]. Il en a conclu que la sensibilité diminue lorsque la classification devient plus compliquée. Selon l'auteur, il est possible d'augmenter la sensibilité sans baisser la spécificité en augmentant le nombre d'arbres dans l'algorithme RF qui délivre l'importance des variables. Cette solution a été testé mais ne s'est pas montrée concluante. La combinaison des variables sélectionnées individuellement pour les capteurs dans un RF a également été testé mais ne s'est pas montrée concluante (RMSE : 1,29m), l'importance des variables mettant en avant les variables Sentinel-2 (cf. Annexe H). L'importance des variables calculées en utilisant la diminution moyenne de la précision (Mean decrease accuracy, abrégé MDA) présente ainsi des limites. Une piste d'amélioration est l'utilisation d'une autre métrique d'évaluation de l'importance pouvant contourner certaines limitations de la MDA (par exemple Bénard et al. (2021)). En outre, il est remarquable que la sélection de variables Boruta sur les jeux de données de terrain (cf. Figure 17) et de photointerprétation (cf. Figure 25) pour WC accorde plus d'importance pour les prédicteurs Sentinel-1 dans le modèle de photointerprétation que celui de terrain, illustrant ainsi une effet de la taille du jeu de données sur la sélection de variables.

4.5 Comparaison produits existants

Bien que certains modèles présentent des limites, leur comparaison avec des produits à des échelles plus globales montre des performances supérieures de nos modèles pour WC, AGB et H95.

Métrique	Modèle	RMSE	R^2
AGB	Bouvet et al. (2018)	35,58 $T.ha^{-1}$	0,13
	AGB s1s2 multi	12,7 $T.ha^{-1}$	0,76
H95	Potapov et al. (2021)	3,91 m	0,083
	H95 pre multi	1,19 m	0,68
WC	Vieilledent et al. (2018)	0,345	0,44
	WC s1s2 multi	0,118	0,84

Table 10 – Comparaison des RMSE et R^2 des meilleurs modèles multisaisonnaux avec des produits plus globaux

4.6 Perspectives suplémentaires

En plus des perspectives déjà évoquées dans la discussion, à savoir l'extrapolation des variables métriques corrélées, l'amélioration de l'estimation de l'AGB in situ, l'intégration de données radar à plus grandes longueurs d'ondes afin de contourner les limitations de la bande C, l'intégration de données de références GEDI et l'amélioration de la mesure de l'importance des variables, une autre perspective pourrait être l'utilisation de techniques d'inventaires innovantes.

En effet, la taille et la répartition du jeu de données de terrain a constitué une limite de ce travail, (p. ex. pour H95, importance des variables) et bien que l'extrapolation des données de terrain aux placettes de photointerprétation semble intéressante pour contourner cette limite, d'autres approches permettant plus de précision méritent d'être investiguées. En particulier, la littérature ainsi que l'expérience du terrain semble indiquer l'utilisation d'aéronefs sans équipage (UAV) pour générer des données de référence comme prometteuse (Emilien et al., 2021; Navarro et al., 2019; Kattenborn et al., 2019). Ils pourraient permettre la réalisation d'inventaires moins laborieux, moins destructifs et ainsi augmenter la capacité d'échantillonnage.

6. La sensibilité est définie comme le nombre d'attributs réellement pertinents reconnus comme tel par l'algorithme divisé par nombre total d'attributs réellement pertinents.

5 Conclusion

Cette étude a cartographié le WC, l'AGB et l'H95 et deux métriques de répartition des hauteurs construites sur une ACP (Dim1H et Dim2H) dans le du Sud-Ouest de Madagascar en utilisant des données multitemporelles Sentinel-1, Sentinel-2, des indices, des métriques de texture et des métriques temporelles, sur base de données de terrain et de photointerprétation dans une régression Random Forest, tout en prenant en compte l'autocorrélation spatiale lors de la validation croisée. Les types de végétations considérés sont principalement fourrés xérophiles, ainsi que des forêts sèches et des formations de transition. Différentes combinaisons de capteurs et de saisons ont été testées afin d'atteindre les objectifs de l'étude. La comparaison des performances des variables réponses ont mis en avant WC, AGB et H95 pour décrire la zone d'étude. L'analyse des modèles a montré le potentiel des données Sentinel-1 pour caractériser les métriques structurales de la zone d'étude et a même obtenu de meilleurs résultats que Sentinel-2 pour WC et H95. Les bandes Sentinel-1 et les indices dérivés se sont avérés être d'importants prédicteurs tandis que les métriques de texture et de temporalité ont semblé moins primordiales. En outre, l'apport de la fusion des saisons est mis avant, ayant augmenté ou égalé la précision des modèles monosaison. La fusion des capteurs a, dans la plupart des cas, augmenté la précision des modèles à l'exception notable d'H95, lié à une possible limite d'évaluation de l'importance des variables. Un second effet bénéfique de la fusion des capteurs est une influence moindre des effets de la saisonnalité, les caractéristiques des capteurs se montrant complémentaires.

Bien que cette étude constitue une première tentative de caractérisation locale de la zone, les données produites permettent de décrire à haute résolution spatiale les écosystèmes du Sud-Ouest de Madagascar avec une précision supérieure à d'autres produits existants. Ces nouveaux jeux de données, ainsi que leurs futures améliorations, pourront soutenir des projets de conservation et de gestion durable de ces écosystèmes.

6 Contribution de l'étudiant

L'étudiant a contribué aux différentes étapes de cette étude : photointerprétation et stratification de la zone d'étude à l'aide de cartes préliminaires, élaboration de protocoles de récolte de données de terrain pour répondre aux objectifs, récolte des données de terrain, traitement et analyse des données, rédaction et communication des résultats.

7 Bibliographie

Abera T.A. et al., 2022. Land cover map for multifunctional landscapes of taita taveta county, kenya, based on sentinel-1 radar, sentinel-2 optical, and topoclimatic data. *Data*, 7(3) :36.

Anchang J.Y. et al., 2020. Toward Operational Mapping of Woody Canopy Cover in Tropical Savannas Using Google Earth Engine. *Frontiers in Environmental Science*, 8.

Archibald S. & Scholes R., 2007. Leaf green-up in a semi-arid african savanna-separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, 18(4) :583–594.

Argamosa R.J.L. et al., 2018. Modelling above ground biomass of mangrove forest using sentinel-1 imagery. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4(3).

Baetens L., Desjardins C. & Hagolle O., 2019. Validation of copernicus sentinel-2 cloud masks obtained from maja, sen2cor, and fmask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sensing*, 11(4) :433.

Banko G., 1998. A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory.

Bastin J.F. et al., 2017. The extent of forest in dryland biomes. *Science*, 356(6338) :635–638.

Baumann M. et al., 2018. Mapping continuous fields of tree and shrub cover across the gran chaco using landsat 8 and sentinel-1 data. *Remote sensing of environment*, 216 :201–211.

Bénard C., Da Veiga S. & Scornet E., 2021. Mda for random forests : inconsistency, and a practical solution via the sobol-mda. *arXiv preprint arXiv :2102.13347*.

Biau G. & Scornet E., 2016. A random forest guided tour. *Test*, 25(2) :197–227.

Bouvet A. et al., 2018. An above-ground biomass map of african savannahs and woodlands at 25 m resolution derived from alos palsar. *Remote sensing of environment*, 206 :156–173.

Bruggisser M. et al., 2021. Potential of sentinel-1 c-band time series to derive structural parameters of temperate deciduous forests. *Remote Sensing*, 13(4) :798.

Carrasco L., O'Neil A.W., Morton R.D. & Rowland C.S., 2019. Evaluating combinations of temporally aggregated sentinel-1, sentinel-2 and landsat 8 for land cover mapping with google earth engine. *Remote Sensing*, 11(3). doi :10.3390/rs11030288.

Carrière S.D. et al., 2018. Sustainable groundwater resources exploration and management in a complex geological setting as part of a humanitarian project (mahafaly plateau, madagascar). *Environmental Earth Sciences*, 77(21) :1–16.

Cartus O. et al., 2021. Sentinel-1 coherence for mapping above-ground biomass in semiarid forest areas. *IEEE Geoscience and Remote Sensing Letters*, 19 :1–5.

Conners R.W., Trivedi M.M. & Harlow C.A., 1984. Segmentation of a high-resolution urban scene using texture operators. *Computer vision, graphics, and image processing*, 25(3) :273–310.

Emilien A.V., Thomas C. & Thomas H., 2021. Uav & satellite synergies for optical remote sensing applications : A literature review. *Science of remote sensing*, 3 :100019.

Faye G. et al., 2020. Complementarity of Sentinel-1 and Sentinel-2 Data for Mapping Agricultural Areas in Senegal. *Advances in Remote Sensing*, 9(3) :101–115. doi :10.4236/ars.2020.93006. Number : 3 Publisher : Scientific Research Publishing.

Ferry L., L'Hote Y. & Wesselink A., 1998. Les précipitations dans le sud-ouest de madagascar. *IAHS PUBLICATION* :89–96.

Fick S.E. & Hijmans R.J., 2017. Worldclim 2 : new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12) :4302–4315.

Forkuor G. et al., 2020. Above-ground biomass mapping in west african dryland forest using sentinel-1 and 2 datasets-a case study. *Remote Sensing of Environment*, 236 :111496.

Francken N. & Minten B., 2005. Security and perceptions on justice in rural madagascar. *Dynamics in Social Service Delivery and the Rural Economy of Madagascar : Descriptive Results of the 2004 Commune Survey. Ilo Policy Report.*

Frantz D. et al., 2021. National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment*, 252 :112128. doi :10.1016/j.rse.2020.112128.

Fremout T. et al., 2022. Site-specific scaling of remote sensing-based estimates of woody cover and aboveground biomass for mapping long-term tropical dry forest degradation status. *Remote Sensing of Environment*, 276 :113040.

Gan L. et al., 2022. Mapping shrub coverage in xilin gol grassland with multi-temporal sentinel-2 imagery. *Remote Sensing*, 14(14) :3266.

Ghasemi N., Sahebi M.R. & Mohammadzadeh A., 2011. A review on biomass estimation methods using synthetic aperture radar data. *International Journal of Geomatics and Geosciences*, 1(4) :776.

Glaude Q. & Orban A., 2022. The dark side of the remote sensing : current sar remote sensing missions and applications. *Bulletin de la Société Géographique de Liège.*

Gorelick N. et al., 2017. Google earth engine : Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202 :18–27.

Hanisch S., 2015. Improving cropping systems of semi-arid south-western Madagascar under multiple ecological and socio-economic constraints. Ph.D. thesis.

Hänke H. & Barkmann J., 2017. Insurance function of livestock, farmers coping capacity with crop failure in southwestern madagascar. *World Development*, 96 :264–275.

Hansen M. et al., 2003. Global percent tree cover at a spatial resolution of 500 meters : First results of the modis vegetation continuous fields algorithm. *Earth Interactions*, 7(10) :1–15.

Haralick R.M., Shanmugam K. & Dinstein I.H., 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) :610–621.

Hasan M.J., Kim J., Kim C.H. & Kim J.M., 2020. Health state classification of a spherical tank using a hybrid bag of features and k-nearest neighbor. *Applied Sciences*, 10(7) :2525.

Heckel K. et al., 2020. Predicting forest cover in distinct ecosystems : The potential of multi-source sentinel-1 and-2 data fusion. *Remote Sensing*, 12(2) :302.

Ienco D., Interdonato R., Gaetano R. & Ho Tong Minh D., 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158 :11–22. doi :10.1016/j.isprsjprs.2019.09.016.

Kacic P., Hirner A. & Da Ponte E., 2021. Fusing Sentinel-1 and -2 to Model GEDI-Derived Vegetation Structure Characteristics in GEE for the Paraguayan Chaco. *Remote Sensing*, 13(24) :5105. doi : 10.3390/rs13245105. Number : 24 Publisher : Multidisciplinary Digital Publishing Institute.

Kattenborn T. et al., 2019. Uav data as alternative to field sampling to map woody invasive species based on combined sentinel-1 and sentinel-2 data. *Remote sensing of environment*, 227 :61–73.

King J.R. & Jackson D.A., 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics : The official journal of the International Environmetrics Society*, 10(1) :67–77.

Kohavi R. & John G.H., 1997. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2) :273–324.

Kuhn M., 2015. Caret : classification and regression training. *Astrophysics Source Code Library* :ascl–1505.

Kursa M.B. & Rudnicki W.R., 2010. Feature selection with the boruta package. *Journal of statistical software*, 36 :1–13.

Laurin G.V. et al., 2018. Above-ground biomass prediction by sentinel-1 multitemporal data in central italy with integration of alos2 and sentinel-2 data. *Journal of Applied Remote Sensing*, 12(1) :016008.

Le Toan T., Beaudoin A., Riom J. & Guyon D., 1992. Relating forest biomass to sar data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2) :403–411.

Lee J.S. et al., 1994. Speckle filtering of synthetic aperture radar images : A review. *Remote sensing reviews*, 8(4) :313–340.

Lee J.S. et al., 2008. Improved sigma filter for speckle filtering of sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1) :202–213.

Legendre P., 1993. Spatial autocorrelation : trouble or new paradigm ? *Ecology*, 74(6) :1659–1673.

Legendre P. & Oksanen M.J., 2018. Package 'lmodel2'. *See https ://CRAN. R-project. org/package= lmodel2.*

Louis J. et al., 2016. Sentinel-2 sen2cor : L2a processor for users. In Proceedings living planet symposium 2016. Spacebooks Online, 1–8.

Main R. et al., 2016. Hyper-temporal c-band sar for baseline woody structural assessments in deciduous savannas. *Remote Sensing*, 8(8) :661.

Malimbwi R., Eid T. & Chamshama S., 2018. Allometric tree biomass and volume models in tanzania. *Morogoro : Department of Forest Mensuration and Management, Sokoine University of Agriculture.*

Mathieu R. et al., 2013. Toward structural assessment of semi-arid african savannahs and woodlands : The potential of multitemporal polarimetric radarsat-2 fine beam images. *Remote Sensing of Environment*, 138 :215–231.

McNicol I.M., Ryan C.M. & Mitchard E.T., 2018. Carbon losses from deforestation and widespread degradation offset by extensive growth in african woodlands. *Nature communications*, 9(1) :1–11.

Mishra V.N. et al., 2019a. Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data. *Earth Science Informatics*, 12(1) :71–86. doi :10.1007/s12145-018-0369-z.

Mitchard E.T. et al., 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 Sensing of Environment*, 115(11) :2861–2873.

Mullissa A. et al., 2021. Sentinel-1 sar backscatter analysis ready data preparation in google earth engine. *Remote Sensing*, 13(10) :1954.

Nagelkirk R.L. & Dahlin K.M., 2020. Woody cover fractions in african savannas from landsat and high-resolution imagery. *Remote Sensing*, 12(5) :813.

NASA, 2020. Nasadem merged dem global 1 arc second v001. *NASA EOSDIS Land Processes DAAC.*

Navarro J.A. et al., 2019. Integration of uav, sentinel-1, and sentinel-2 data for mangrove plantation aboveground biomass monitoring in senegal. *Remote Sensing*, 11(1) :77.

Nematchoua M.K., Ricciardi P., Orosa J.A. & Buratti C., 2018. A detailed study of climate change and some vulnerabilities in indian ocean : A case of madagascar island. *Sustainable cities and society*, 41 :886–898.

Neudert R., Goetter J.F., Andriamparany J.N. & Rakotoarisoa M., 2015. Income diversification, wealth, education and well-being in rural south-western madagascar : Results from the mahafaly region. *Development Southern Africa*, 32(6) :758–784.

Nkeumoe Numbisi F., Vancoillie F. & De Wulf R., 2018. Multi-date sentinel1 sar image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape. 42 :339–346.

Potapov P. et al., 2021. Mapping global forest canopy height through integration of gedi and landsat data. *Remote Sensing of Environment*, 253 :112165.

Quegan S. & Yu J.J., 2001. Filtering of multichannel sar images. *IEEE Transactions on geoscience and remote sensing*, 39(11) :2373–2379.

Randriamalala J.R. & Hervé D., 2022. Goat breeding : A possible sustainable way to manage xerophytic thickets in southwestern madagascar. *Land*, 11(3) :405.

Randriamalala J.R. et al., 2016. Effects of goat grazing and woody charcoal production on xerophytic thickets of southwestern madagascar. *Journal of Arid Environments*, 128 :65–72.

Randriamalala J.R. et al., 2022. Allometric models to predict the individual aboveground biomass of shrubs of malagasy xerophytic thickets. *Journal of Arid Environments*, 202 :104751.

Ratovonamana R.Y., Rajeriarison C., Roger E. & Ganzhorn J.U., 2011. Phenology of different vegetation types in tsimanampetsotsa national park, southwestern madagascar. *Malagasy Nature*, 5 :14–38.

RColorBrewer S. & Liaw M.A., 2018. Package 'randomforest'. *University of California, Berkeley : Berkeley, CA, USA.*

Reiche J. et al., 2021. Forest disturbance alerts for the congo basin using sentinel-1. *Environmental Research Letters*, 16(2) :024005.

Rudnicki W.R., Wrzesień M. & Paja W., 2015. All relevant feature selection methods and applications. In Feature Selection for Data and Pattern Recognition, Springer. 11–28.

Ryan C.M. et al., 2012. Quantifying small-scale deforestation and forest degradation in african woodlands using radar imagery. *Global Change Biology*, 18(1) :243–257.

Sanchez A.H. et al., 2020. Comparison of cloud cover detection algorithms on sentinel–2 images of the amazon tropical forest. *Remote Sensing*, 12(8) :1284.

Seddon N. et al., 2000. Conservation issues and priorities in the mikea forest of south-west madagascar. *Oryx*, 34(4) :287–304.

Sexton J.O. et al., 2013. Global, 30-m resolution continuous fields of tree cover : Landsat-based rescaling of modis vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6(5) :427–448.

Skakun S. et al., 2022. Cloud mask intercomparison exercise (cmix) : An evaluation of cloud masking algorithms for landsat 8 and sentinel-2. *Remote Sensing of Environment*, 274 :112990.

Smith W.K. et al., 2019. Remote sensing of dryland ecosystem structure and function : Progress, challenges, and opportunities. *Remote Sensing of Environment*, 233 :111401.

Tarrio K. et al., 2020. Comparison of cloud detection algorithms for sentinel-2 imagery. *Science of Remote Sensing*, 2 :100010.

Valavi R., Elith J., Lahoz-Monfort J.J. & Guillera-Arroita G., 2018. blockcv : An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *bioRxiv* :357798.

Van Tricht K., Gobin A., Gilliams S. & Piccard I., 2018. Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping : A case study for belgium. *Remote Sensing*, 10(10) :1642.

Vieilledent G. et al., 2012. A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. *Ecological Applications*, 22(2) :572–583.

Vieilledent G. et al., 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in madagascar. *Biological Conservation*, 222 :189–197.

Vollrath A., Mullissa A. & Reiche J., 2020. Angular-based radiometric slope correction for sentinel-1 on google earth engine. *Remote Sensing*, 12(11) :1867.

White A. et al., 2012. Ausplots rangelands.

Xue J. & Su B., 2017. Significant remote sensing vegetation indices : A review of developments and applications. *Journal of sensors*, 2017.

8 Webographie

ESA, . Users guides - sentinel-2 msi level-2a). https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a. Consulté le 28-06-2022.

ESA, 2022. Copernicus sentinel-1b anomaly (6th update). https://sentinels.copernicus.eu/web/sentinel/-/copernicus-sentinel-1b-anomaly-6th-update/1.1. Consulté le 20-06-2022.

FAO, . FAO/UNESCO Soil Map of the World. https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/.

FAO, 2022. Open Foris Homepage. https://openforis.org/. Consulté le 20-03-2022.

Google, 2021. Sentinel-1 algorithms). https://developers.google.com/earth-engine/guides/sentinel1. Consulté le 20-06-2022.

Zupanc, 2017. Improving cloud detection with machine learning). https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13.