

## **Étude des caractéristiques spectrales, spatiales et temporelles d'images satellites et du potentiel de la photo-interprétation pour cartographier les propriétés structurelles des écosystèmes arides du sud-ouest de Madagascar**

**Auteur :** Vander Linden, Arthur

**Promoteur(s) :** Bastin, Jean-François

**Faculté :** Gembloux Agro-Bio Tech (GxABT)

**Diplôme :** Master en bioingénieur : gestion des forêts et des espaces naturels, à finalité spécialisée

**Année académique :** 2021-2022

**URI/URL :** <http://hdl.handle.net/2268.2/15477>

---

### *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.*

---

**ÉTUDE DES CARACTÉRISTIQUES SPECTRALES,  
SPATIALES ET TEMPORELLES D'IMAGES  
SATELLITES ET DU POTENTIEL DE LA PHOTO-  
INTERPRÉTATION POUR CARTOGRAPHIER LES  
PROPRIÉTÉS STRUCTURELLES DES ÉCOSYSTÈMES  
ARIDES DU SUD-OUEST DE MADAGASCAR**

**ARTHUR VANDER LINDEN**

**TRAVAIL DE FIN D'ÉTUDES PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE  
MASTER BIOINGENIEUR EN GESTION DES FORÊTS ET DES ESPACES NATURELS**

**ANNÉE ACADÉMIQUE 2021-2022**

**PROMOTEUR(S): 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.

**ÉTUDE DES CARACTÉRISTIQUES SPECTRALES,  
SPATIALES ET TEMPORELLES D'IMAGES  
SATELLITES ET DU POTENTIEL DE LA PHOTO-  
INTERPRÉTATION POUR CARTOGRAPHIER LES  
PROPRIÉTÉS STRUCTURELLES DES ÉCOSYSTÈMES  
ARIDES DU SUD-OUEST DE MADAGASCAR**

**ARTHUR VANDER LINDEN**

**TRAVAIL DE FIN D'ÉTUDES PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE  
MASTER BIOINGENIEUR EN GESTION DES FORÊTS ET DES ESPACES NATURELS**

**ANNÉE ACADÉMIQUE 2021-2022**

**PROMOTEUR(S): JEAN-FRANÇOIS BASTIN**



## Remerciements

Mes remerciements s'adressent en premier lieu aux personnes qui m'ont accompagné dans la réalisation de ce mémoire de fin d'études et sans qui les pages qui suivent n'auraient jamais vu le jour. Mes propres réflexions ont toujours été plus fertiles grâce aux discussions communes et au partage d'idées, et dans ce sens j'ai eu énormément de chance de rencontrer des personnes stimulantes sur le plan intellectuel et humain tout au long de cette expérience. Jean-François, tout d'abord, je te remercie sincèrement pour la manière dont tu nous as encadrés, tu es un vrai leader qui transmet l'envie de se dépasser pour atteindre ses objectifs, et tu pousses toujours la réflexion un cran plus loin sans jamais freiner les initiatives personnelles (ce qui peut être problématique quand on pose trop de questions). J'aurais difficilement pu trouver un encadrement aussi complet que le tien, à l'équilibre parfait entre la stimulation intellectuelle et l'agence de voyages pour l'Afrique, encore merci pour ta confiance, ta disponibilité et pour m'avoir offert toutes ces opportunités. Ensuite, Guillaume, toi et moi on aura fait un beau bout de chemin ensemble et vécu une expérience inoubliable. Les trois mois à Madagascar et ensuite la réalisation de ce mémoire auraient eu un tout autre goût sans toi. C'était quand même exceptionnel ! Évidemment, arriver en terres inconnues avec un Vertex, une chemise « *à larges ventilations pour sentir la moindre brise d'air* », le chapeau de paille assorti et une boussole pour réaliser des inventaires au beau milieu des fourrés malgaches, ça peut poser des difficultés sans un minimum d'organisation et de facilitation. C'est pourquoi Felana, Olivia, toute l'équipe de l'ESSA forêt de l'Université d'Antananarivo, Gilberto et Hasina nos guides infaillobles, toutes les personnes du sud de Tuléar qui ont travaillé avec nous pour mener à bien nos inventaires (zaza kamo lambo), et tous ceux sans qui nous serions sûrement toujours coincés là-bas, je vous suis très reconnaissant de nous avoir accompagnés tout au long de l'aventure. J'imagine qu'un mémoire de fin d'études n'en serait pas un sans un contre-la-montre final, et on ne se bat jamais seul. Merci à l'équipe de nuit, et plus largement à tous celles et ceux présent.e.s au GP durant les heures de bureau, mais également dehors. Et surtout, pour finir, mais pas des moindres, merci à toi, Éva, mon cœur, de m'avoir supporté tout du long. Nos moments de détente (légèrement trop peu fréquents) durant cette période auront été précieux et tu m'auras à coup sûr permis d'éviter des carences nutritionnelles.

Ce mémoire de fin d'études est également l'aboutissement d'un projet entrepris il y a bien plus longtemps. Qui aurait cru que je serai là où j'en suis maintenant, avec autant d'expériences exceptionnelles au compteur ? Ce passage à Gembloux aura vraiment été fructueux. Merci celles et ceux qui ont cru en moi, et qui m'ont accompagné pendant toutes ces années, et plus particulièrement à mes parents, à Béa, mes sœurs, et à toi, Éva, qui te retrouve dans deux paragraphes, c'est dire !

## Organismes et subsides ayant permis la réalisation de ce travail de fin d'études

Unité Biodiversité et Paysages, Université de Liège Gembloux Agro-Bio Tech, Belgique

Mention Foresterie et Environnement de l'École Supérieure des Sciences Agronomiques, Université d'Antananarivo, Madagascar

Programme Erasmus



# Table des matières

<b>RÉSUMÉ</b>	<b>1</b>
<b>ABSTRACT</b>	<b>2</b>
<b>INTRODUCTION</b>	<b>3</b>
<b>MATÉRIEL ET MÉTHODE</b>	<b>7</b>
1. DESCRIPTION DE LA ZONE D'ÉTUDE	7
1.1. DESCRIPTION GÉNÉRALE	7
1.2. GEOMORPHOLOGIE, TOPOGRAPHIE, PEDOLOGIE ET OCCUPATION DU SOL	8
1.3. DESCRIPTION DE LA VEGETATION ET SAISONNALITE	9
2. DONNÉES DE RÉFÉRENCE	10
2.1. DONNEES DE TERRAIN : CONCEPTION DU PLAN D'ÉCHANTILLONNAGE ET RECOLTE DE DONNEES	11
2.2. DONNEES DE PHOTO-INTERPRETATION	15
2.3. DESCRIPTION DU JEU DE DONNEES DE REFERENCE	17
3. IDENTIFICATION DES GROUPES DE PRÉDICTEURS SPECTRAUX ET CRÉATION DES IMAGES COMPOSITES	18
3.1. DONNEES DE TELEDETECTION	18
3.2. VARIABLES DE SAISONNALITE	18
3.3. RESOLUTION SPATIALE	18
3.4. VARIABLES SPECTRALES	19
3.5. VARIABLES PHENOLOGIQUES	22
4. MODÉLISATION RANDOM FOREST ET ÉVALUATION	26
4.1. GÉNÉRALITÉS	26
4.2. VALIDATION CROISÉE SPATIALEMENT INDÉPENDANTE	26
4.3. MÉTRIQUES D'ÉVALUATION DE QUALITÉ ET D'INTERPRÉTATION DES MODÈLES	27
4.4. ANALYSES STATISTIQUES DES DIFFÉRENCES MOYENNES ENTRE MODÈLES	28
4.5. ÉVALUATION INDÉPENDANTE DES MODÈLES DE PHOTO-INTERPRÉTATION	28
4.6. ÉVALUATION DE L'IMPORTANCE ET SÉLECTION DES VARIABLES	28
5. MODÈLE INTÉGRATIF ET COMPARAISON AVEC LES CARTES GLOBALES	29
<b>RÉSULTAT</b>	<b>31</b>
1. VUE GLOBALE DES RÉSULTATS ET ÉVALUATION DU POTENTIEL DE LA PHOTO-INTERPRÉTATION	31
2. FACTEURS INFLUENÇANT LA CARTOGRAPHIE	35
2.1. SAISONNALITÉ	35
2.2. RESOLUTION SPATIALE	38
2.3. FACTEURS SPECTRAUX	39
2.4. PHENOLOGIE	42
3. MODÈLES INTÉGRATIFS	46
4. MISE EN PERSPECTIVE DES CARTES GLOBALES	47

<b>DISCUSSION</b>	<b>50</b>
<b>1. POTENTIEL DE LA PHOTO-INTERPRÉTATION</b>	<b>50</b>
<b>2. ÉTUDES DE L'INFLUENCE DES VARIABLES PRÉDICTIVES</b>	<b>51</b>
2.1. COUVERTURE DE LA CANOPEE	51
2.2. BIOMASSE	53
2.3. HAUTEUR DE CANOPEE	54
2.4. MODELES INTEGRATIFS	55
<b>3. REcul SUR LA MÉTHODOLOGIE ET SUR LES RÉSULTATS</b>	<b>56</b>
<b>4. TRANSFÉRABILITÉ ET IMPLICATIONS DE L'ÉTUDE DANS UN CONTEXTE GLOBAL ET LOCAL</b>	<b>57</b>
<b>CONCLUSION</b>	<b>59</b>
<b>CONTRIBUTION</b>	<b>59</b>
<b>DICTIONNAIRE DES ACRONYMES ET DES ABRÉVIATIONS</b>	<b>60</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>61</b>
<b>ANNEXES</b>	<b>67</b>
1. ÉTUDE DE SENSIBILITÉ RANDOM FOREST	67
2. RÉSULTATS DÉTAILLÉS	71
3. INFLUENCE DE LA SÉLECTION DE VARIABLES	76
4. TABLE DES INDICES DE VÉGÉTATION	77

## Résumé

Les fourrés xérophi les du sud-ouest de Madagascar sont des écosystèmes à très haute valeur conservatoire qui sont menacés par les activités humaines et les changements globaux. Ces écosystèmes souffrent d'un manque de représentativité à l'échelle régionale et globale, et cette situation complique la gestion participative et durable des espaces naturels, alors même qu'ils sont pourvoyeurs de nombreux services écosystémiques à la base de la subsistance de la population de cette partie enclavée de Madagascar. La modélisation des attributs structuraux par télédétection optique en zones arides est complexifiée par le manque de données représentatives et par la nécessité d'utiliser des prédicteurs adaptés en réponse aux spécificités structurelles et fonctionnelles de ces écosystèmes. L'acquisition d'informations par photo-interprétation en vue de calibrer des modèles de couverture de canopée est une technique qui permet d'augmenter considérablement le volume de données et qui offre des résultats prometteurs. Parallèlement, les paramètres à considérer en zones arides sont documentés par plusieurs études, mais la plupart ont été réalisées dans des écosystèmes aux caractéristiques structurelles distinctes des fourrés xérophi les. Cette étude aborde les caractéristiques spectrales, spatiales et temporelles déterminantes pour la caractérisation des propriétés structurelles des fourrés xérophi les et propose une méthodologie pour étendre les données récoltées par photo-interprétation à la biomasse aérienne et à la hauteur de canopée par extrapolation depuis les mesures d'ouverture de la canopée. La résolution spatiale et la saisonnalité sont les caractéristiques prépondérantes à considérer, tandis que les indices de végétation et les métriques phénologiques n'ont qu'un effet marginal. Parallèlement, la biomasse et la hauteur sont deux attributs structuraux fortement corrélés à l'ouverture de la canopée, qui peuvent être extrapolés aux données de photo-interprétation et engendrer des modèles avec des performances équivalentes aux approches basées sur des relevés de terrain. Cette étude souligne le rôle clé joué par la qualité des données de calibration dans la représentation finale des modèles et relève une des raisons expliquant la faible représentation des écosystèmes arides à l'échelle globale. La méthodologie développée pour cette étude est particulièrement pertinente pour le cas des fourrés xérophi les et pourrait constituer la base d'un programme de suivis affranchis d'une partie des contraintes liées à la récolte de données de terrain. En outre, les cartes produites offrent un outil qui, utilisé conjointement aux connaissances locales, permet l'identification de zones prioritaires de conservation et de réhabilitation écologique, venant ainsi en appui à la gestion du territoire.



# Abstract

The xerophit thickets of southwestern Madagascar are ecosystems with very high conservation value that are threatened by human activities and global changes. These ecosystems suffer from a lack of representativity at regional and global scales. This is a challenge for the participatory and sustainable management of natural areas, when these ecosystems provide many ecosystem services playing a large part in the livelihood of the population of this landlocked part of Madagascar. The use of optical remote sensing to model structural attributes in drylands is further complicated by the lack of representative data and the need to use predictors adapted to the structural and functional specificities of these ecosystems. The acquisition of information by photo-interpretation for calibrating canopy cover models is a technique that greatly increases the volume of data and shows promising results. The parameters to be considered in dryland areas are documented by several studies, but most of them have been carried out in ecosystems with structural characteristics distinct from xerophilous thickets. This study addresses the spectral, spatial and temporal characteristics that are decisive for characterising the structural properties of xerophytic thickets and proposes a methodology for extending data collected by photo-interpretation to above-ground biomass and canopy height by extrapolation from canopy opening measurements. Spatial resolution and seasonality are the predominant characteristics to be considered, while vegetation indices and phenological metrics only have a marginal effect. In addition, biomass and height are two structural attributes strongly correlated with canopy opening, which can be extrapolated to photo-interpretation data and generate models with performances equivalent to those obtained with approaches based on field survey data. This study highlights the key importance of high-quality calibration data for the final models representation and points out one of the reasons for the poor representation of dryland ecosystems on a global scale. The methodology developed is particularly relevant for xerophytic thickets and may form the basis of a monitoring programme unburdened by some of the constraints related to field data collection. The maps produced also provide a tool which, used together with community knowledge, helps to identify priority areas for conservation and ecological rehabilitation action plans, thus supporting land management.

## Introduction

Les zones arides sont définies par un indice d'aridité inférieur à 0,65 et s'étendent approximativement sur 41 % de la surface émergée de la planète (Food and Agriculture Organization of the United Nations, 2019). Elles sont habitées par plus de deux milliards de personnes, principalement dans des pays à revenu faible ou moyen inférieur, dont 90 % d'entre eux vivent sous le seuil de pauvreté (Niemeijer et al., 2007). Ces territoires abritent également de nombreux écosystèmes à très haute valeur conservatoires menacés par les activités humaines et les changements globaux (Crawhall et al., 2012). Parmi ceux-ci, les fourrés xérophiles du sud-ouest de Madagascar s'étendent sur 17 000 km<sup>2</sup> et sont considérés comme l'une des plus importantes écorégions de la planète (Olson et al., 2002). Leur état de conservation est évalué comme en danger critique, avec une perte de couverture forestière estimée à 45 % entre 1973 et 2013 (Brinkmann et al., 2014), alors même que les moyens de subsistance des populations de cette partie enclavée de Madagascar, extrêmement pauvres et marginalisées, dépendent entièrement de l'exploitation des ressources naturelles (Middleton et al., 2011; Hänke, 2016). Les principales causes de dégradation et de déforestation sont le changement d'affectation des terres pour la pratique de l'agriculture et la surexploitation des forêts pour la production de bois-énergie (Casse et al., 2004). Ces phénomènes engendrent des changements dans la structure et la composition des écosystèmes, notamment sous la forme de modifications de la densité de biomasse (Brinck et al., 2017), de la couverture de la canopée et de la structure tridimensionnelle de la végétation (Walters et al., 2017). Il s'en suit une dégradation des processus biologiques et des fonctions qui en découlent, dont nombreux services écosystémiques desquels bénéficient les habitants de la région (Vásquez-Grandón et al., 2018). Une connaissance précise et actualisée des indicateurs de l'état de l'écosystème est indispensable pour venir en appui à la gestion du territoire, par exemple lors de l'identification d'îlots de conservation ou de zones de restauration, et également en tant qu'aide à la décision pour organiser une cohabitation pérenne entre les humains et les autres espèces vivantes.

Parmi les approches qui permettent de décrire à grande échelle les écosystèmes et des paramètres indicateurs de leur état de conservation, la télédétection est une technique courante qui est particulièrement efficace (Thanh Noi et al., 2018). Historiquement, un des indicateurs communément utilisés est la couverture de la canopée, qui peut être définie comme la proportion de surface au sol couverte par la projection verticale de la canopée d'un arbre ou d'un buisson (Jongman et al., 2017). La biomasse, ainsi que la structure et la hauteur de canopée, sont également des indicateurs pertinents qui peuvent être dérivés de modèles empiriques construits avec des prédicteurs spectraux. Parmi ceux-ci, les approches non paramétriques et plus spécifiquement les algorithmes de *machine learning* se sont imposés grâce leur robustesse et à d'excellents résultats, potentiellement supérieurs aux autres techniques (Belgiu et al., 2016). Plusieurs cartes globales, qui renseignent notamment la couverture de canopée (Hansen et al., 2013; Sexton et al., 2013), la biomasse (Spawn et al., 2020) et la hauteur de canopée (Simard et al., 2011; Potapov et al., 2021; Lang et al., 2022), ont été produites en utilisant ces méthodes. Néanmoins, en plus d'importantes discordances entre les produits globaux, ces produits ont une qualité individuelle variable selon les régions et les écosystèmes considérés (Bos et al., 2019; Duncanson et al., 2019). Ces cartes globales sont en outre particulièrement peu précises, par exemple, pour la représentation du couvert forestier en zones arides (Bastin et al., 2017). Le déploiement en 2015 de la constellation de satellites multispectraux Sentinel 2 est une réelle opportunité pour améliorer la caractérisation et le monitoring à larges échelles des écosystèmes. En effet, ces satellites sont caractérisés par une résolution comprise entre 10 et 20 m pour les parties du spectres les plus utilisées, une durée de revisite de 5 jours, la présence de 13 bandes spectrales dont 4 d'entre-elles absente des satellites Landsat, et un accès public aux images. Cette situation offre de nouvelles

perspectives en télédétection optique, notamment dans le domaine de la caractérisation et du suivi des écosystèmes arides (Forkuor et al., 2018; E. D. Chaves et al., 2020; Eskandari et al., 2020; Shafeian et al., 2021).

Plusieurs hypothèses ont été posées pour tenter d'expliquer l'origine de la faible représentativité des écosystèmes arides à l'échelle mondiale. L'ambiguïté autour de la définition de la forêt est la principale cause de différences entre les cartes de couverture forestière (Sexton et al., 2016). En outre, les différences de propriétés spectrales entre formations forestières, mais également l'interférence de la signature spectrale d'autres écosystèmes tels que les prairies ou les zones humides, ont également été mises en avant (Bastin et al., 2017). Enfin, la localisation et la répartition des données utilisées pour construire les modèles sont suspectées d'être une source d'hétérogénéité dans les prédictions de modèles calibrés trop spécifiquement pour une région ou un groupe d'écosystèmes (E. D. Chaves et al., 2020). En effet, l'utilisation de données représentatives du domaine d'application des modèles lors de la phase de calibration permet d'améliorer la qualité globale des résultats. En outre, bien qu'il existe de nombreuses études plus localisées, le sud-ouest de Madagascar, et plus généralement les zones arides, souffrent d'un manque d'intérêt de la communauté scientifique, de données disponibles et de résultats fiables (Vieilledent et al., 2018; Smith et al., 2019; Schröder et al., 2021).

Cette situation se reflète dans la faible qualité de représentation des écosystèmes arides à l'échelle globale, principalement en raison d'une sous-représentation dans les réseaux de mesures de terrain de longs termes (Smith et al., 2019). Un accroissement considérable du volume de données disponibles pour les zones arides a été rendu possible par le développement de nouvelles techniques de récolte de données à distance. Par exemple, un inventaire de 210 000 parcelles de 0,5 ha, construit par photo-interprétation d'images à très haute résolution, a été utilisé pour évaluer l'étendue des forêts en zones arides (Bastin et al., 2017). Cette nouvelle estimation du couvert forestier en zone aride est 38 à 47 % plus importante que celles fournies, respectivement, par Hansen et al. (2013) et Sexton et al. (2013). Des algorithmes de détection automatique de couronne, basés sur le spectre visible, ont été appliqués en savanes sur des images à très haute résolution et ont également permis de récolter à distance des données pertinentes (Shafeian et al., 2021; Fremout, 2022; Gan et al., 2022). Les données acquises à distance sont généralement utilisées seules (Eskandari et al., 2020), ou en combinaison avec des données de terrain (Kattenborn et al., 2019; Ludwig et al., 2019), lors de la calibration de modèles de couverture forestière. De plus, les données de recouvrement acquises à distance ont été renseignées comme de meilleures qualités, car plus objectives et moins sujettes à l'effet opérateur (Morrison, 2016; Ludwig et al., 2019). Néanmoins, le recouvrement est le seul paramètre structurel que ces techniques permettent d'obtenir à distance, contrairement aux données de terrain qui renseignent également sur la biomasse, la hauteur et d'autres attributs écosystémiques. Il subsiste beaucoup d'incertitudes quant au potentiel de la photo-interprétation pour dériver d'autres paramètres. En effet, aucune étude n'a étudié les relations entre les données de terrain et de photo-interprétation sur de mêmes parcelles afin d'extrapoler des attributs mesurés sur le terrain depuis des données de recouvrement acquises par photo-interprétation. Cette technique pourrait considérablement augmenter la taille des jeux de données et participer à une amélioration des connaissances en zones arides.

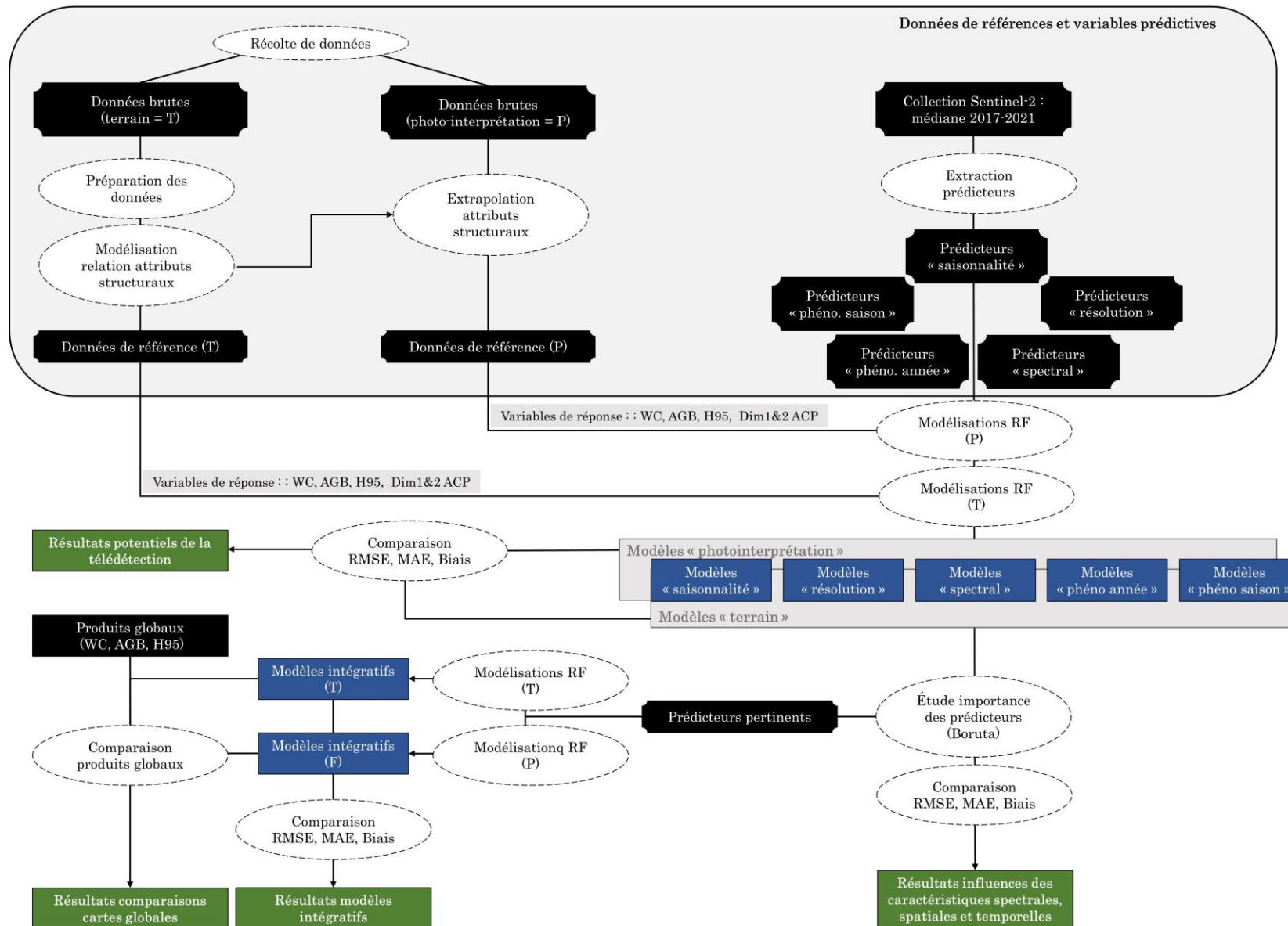
Les écosystèmes forestiers en zones arides se distinguent à plusieurs niveaux des forêts tropicales, boréales et tempérées, et leurs spécificités structurelles et fonctionnelles sont également responsables de leur faible représentation à l'échelle globale. Ils se démarquent notamment par des conditions écologiques et hydrologiques particulières, une différente organisation des strates de la végétation, un faible signal spectral influencé par la réflectance du sol et de plantes mortes ou sénescents, un découplage de l'activité

photosynthétique et de la signature spectrale, un agencement tridimensionnel hétérogène localement et régionalement, et l'imprévisibilité des saisons de végétation (Smith et al., 2019). Ces facteurs complexifient le développement de modèles globaux, ou l'extrapolation à partir de modèles non spécifiques. Plusieurs études se sont intéressées aux spécificités de la signature spectrale des zones arides. Il a été démontré, par exemple, que le rapport des réflectances entre le rouge et l'infrarouge, mis en évidence par l'indice de végétation par différence normalisée (NDVI), est fortement modifié sous influence du sol. Conjointement, la nature et l'hygrométrie du sol sous une canopée épars influence l'intensité du signal spectral. De manière générale, les indices de végétations montrent de moins bonnes performances en zone aride, tandis que les facteurs d'ajustement au sol ne permettent qu'une amélioration marginale des résultats (Smith et al., 2019). Néanmoins, un nombre croissant d'études réalisées en zone aride ont mis en avant de meilleurs résultats avec d'autres parties du spectre électromagnétique, principalement l'infrarouge à courte longueur d'onde — SWIR1 et SWIR2 (Forkuor et al., 2018) — , et de la première bande Sentinel-2 dans le proche infrarouge, le Red-Edge1 (Macintyre et al., 2020). Des indices de végétation calculés à partir de cette portion du spectre, lorsqu'ils sont utilisés, présentent de très bons résultats (E. D. Chaves et al., 2020).

Il a également été montré que la saisonnalité de l'activité végétale influence la réflectance dans le spectre visible et infrarouge, avec un contraste plus important pour un sol sans strate herbacée en saison sèche. L'identification de la saison adéquate est donc un facteur important à considérer lors de la caractérisation de ces écosystèmes (Higginbottom et al., 2018; Ludwig et al., 2019). Néanmoins, cette identification est complexifiée par l'imprévisibilité des pics d'activité végétale en réponse aux conditions météorologiques, et ce particulièrement dans le sud-ouest de Madagascar (Randriamahefasoa et al., 2017). Parallèlement à la saisonnalité, des métriques phénologiques calculées sur des séries temporelles denses (par exemple l'écart-type, la valeur maximum ou minimale) ont permis une amélioration significative de la qualité de représentation des savanes (Müller et al., 2015) et des forêts ouvertes (Karlson et al., 2015). L'utilisation de métriques calculées entre deux périodes sur l'ensemble des bandes spectrales, à partir d'une différence normalisée, a également engendré une amélioration de la représentation d'écosystèmes arides (Nagelkirk et al., 2020). Enfin, Shafeian et al. (2021) a mis en évidence l'influence de la résolution spatiale lors de la représentation d'écosystèmes arides caractérisés par une couverture épars d'arbres et de buissons.

La majeure partie des recherches sur les facteurs qui influencent la cartographie des zones arides ont été réalisées dans des écosystèmes parfois très différents des fourrés épineux du sud-ouest de Madagascar. Leurs caractéristiques structurelles — très hétérogènes selon les faciès considérés —, les distinguent notamment des savanes, des forêts denses sèches décidues, ou encore des zones de buissons éparses, au sein desquelles la plupart des études en zone aride ont été réalisées. Les fourrés épineux sont donc un cas d'étude particulièrement pertinent pour vérifier les hypothèses qui ont été exposées dans d'autres recherches. Dans ce contexte, l'objectif de cette étude est d'apporter une réponse aux questions suivantes :

- a. Quelle est l'influence des facteurs spectraux, de la résolution spatiale, de la saisonnalité et de la phénologie lors de la modélisation d'attributs forestiers à partir de prédicteurs optiques ?
- b. Est-ce que des paramètres structuraux peuvent être modélisés efficacement à partir de données de recouvrement obtenues par photo-interprétation afin d'être utilisés comme données de calibration de modèles ?
- c. Est-ce qu'un modèle calibré localement et qui inclut des facteurs spécifiques aux zones arides permet une meilleure représentation du recouvrement de la canopée, de la biomasse et de la hauteur de canopée que les cartes calibrées pour une échelle globale ?

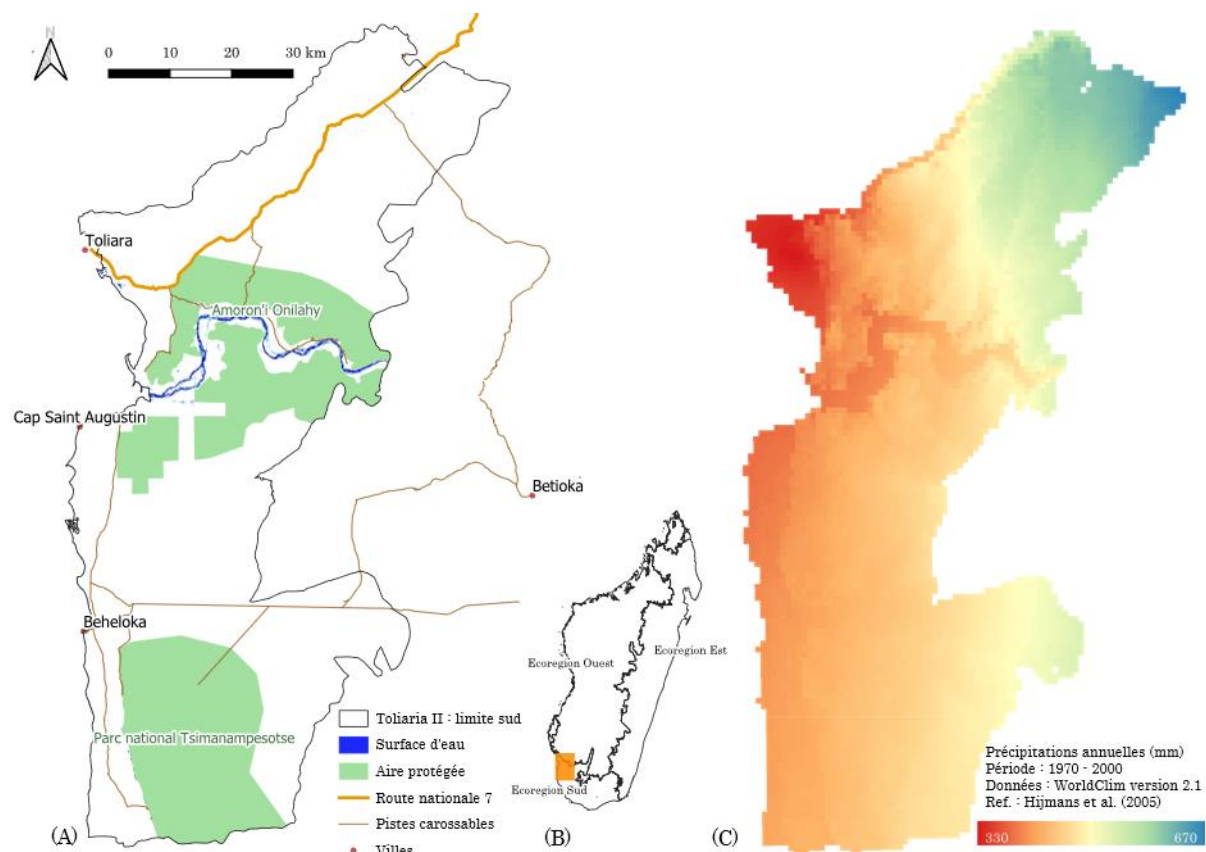


# Matériel et méthode

Une diagramme qui synthétise les principales étapes de la partie matériel et méthode de cette étude est présentée à la page précédente.

## 1. Description de la zone d'étude

### 1.1. Description générale

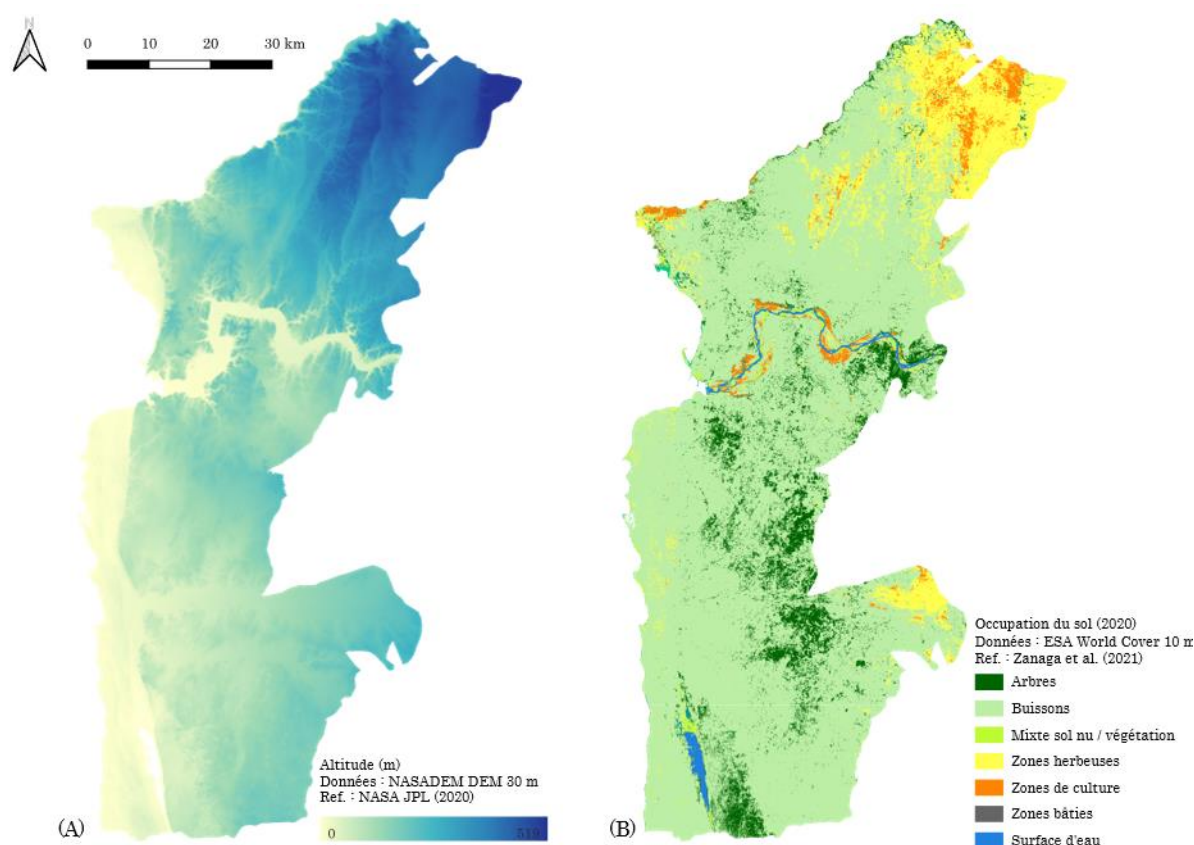


**Figure 1** : partie sud du district de Toliara II : (A) Contexte général, (B) Situation géographique à Madagascar dans l'écorégion Sud, (C) Précipitations annuelles moyennes pour la période 1970-2000 modélisées par WorldClim Bio v2.1 Hijmans et al. (2005)

La zone d'étude est située au sud-ouest de Madagascar et s'étend sur 4820 km<sup>2</sup> (figure 1.B). Elle est délimitée au nord par le fleuve Fiherenana et reprend les limites administratives du district de Toliara II. Le territoire est traversé par une unique route asphaltée, la route nationale 7, depuis Antananarivo vers Tuléar, ainsi que par quelques pistes sableuses carrossables. Une grande partie du territoire est constituée d'aires protégées, dont le parc national de Tsimanampetsotse fait partie (figure 1.A) Le climat de la zone d'étude est semi-aride et caractérisé par une très forte insolation. Les alizés, vents dominants de l'océan Indien en provenance de l'E-SE, arrivent déchargés de leur humidité à l'ouest de Madagascar et sont responsables d'un climat très stable à faible régime de nébulosité et de précipitation tout au long de l'année. Les précipitations sont principalement causées par des dépressions tropicales présentes dans le canal du Mozambique. L'irrégularité de ces dépressions, tant au niveau de leur localisation que de leur intensité, engendre une importante variabilité interannuelle des précipitations (Randriamahefasoa et al., 2017).

La zone d'étude est caractérisée localement par une importante variabilité spatiale des précipitations. Le principal gradient est situé dans l'axe E-O et trouve son origine dans l'élévation des reliefs centraux. Depuis la zone côtière, qui est la zone la plus aride avec environ 300 mm/an, les précipitations s'accroissent vers l'intérieur des terres pour atteindre jusqu'à 600 mm/an (figure 1.C). Le second gradient, qui s'étend dans l'axe N-S, est principalement lié à la latitude qui modifie la probabilité d'être influencé par des facteurs météorologiques (Hijmans et al., 2005). La résultante de ce gradient est située dans l'axe NE-SO. Les précipitations sont en moyenne concentrées durant l'été austral, entre les mois de décembre et de mars, tandis que le reste de l'année est remarquablement sec et ensoleillé (Randriamalala et al., 2016).

## 1.2. Géomorphologie, topographie, pédologie et occupation du sol



**Figure 2** : partie sud du district de Toliaria II : (A) Description topographique : modèle numérique de terrain NASADAM 30 m, NASA JPL (2020). Le domaine côtier se situe aux faibles élévations, représentées en jaune, tandis que le domaine calcaro-basaltique débute à partir des falaises et s'étend sur le reste de la zone d'étude. Les falaises sont présentes à la transition brutale d'altitude (B) Occupation du sol : ESA World Cover 10 m, Zanaga et al. (2021). La zone d'étude est caractérisée par l'omniprésence d'écosystèmes arbustifs et buissonnants et globalement peu bâtie.

La zone d'étude est scindée en deux domaines géomorphologiques. Le domaine côtier — ou « plaine côtière » — s'étend le long des côtes et étroitement entre les falaises du plateau calcaire. Il est composé d'éléments sédimentaires continentaux sableux. Ce domaine s'élève jusqu'à une vingtaine de mètres au-dessus du niveau de la mer (figure 2.A). Le domaine calcaro-basaltique — ou « plateau calcaire » — est essentiellement formé d'affleurements calcaires et basaltiques ponctués par des sols sableux de composition physico-chimique très diverse. Parmi ceux-ci, des sols superficiels à sables roux-beige sont majoritaires sur les plateaux et faibles pentes, tandis que des sols profonds et plus fertiles à sables roux constituent les vallons et bas de pente. Le plateau s'élève depuis les falaises occidentales (60-100 m) vers l'intérieur des terres pour atteindre environ 200 m d'altitude à l'ouest. Le N-E est caractérisé par une élévation jusqu'à



520 m de haut, vers un troisième domaine géomorphologique, le domaine de l'Isalo, qui est caractérisé par des grès acides (Sourdat, 1972).

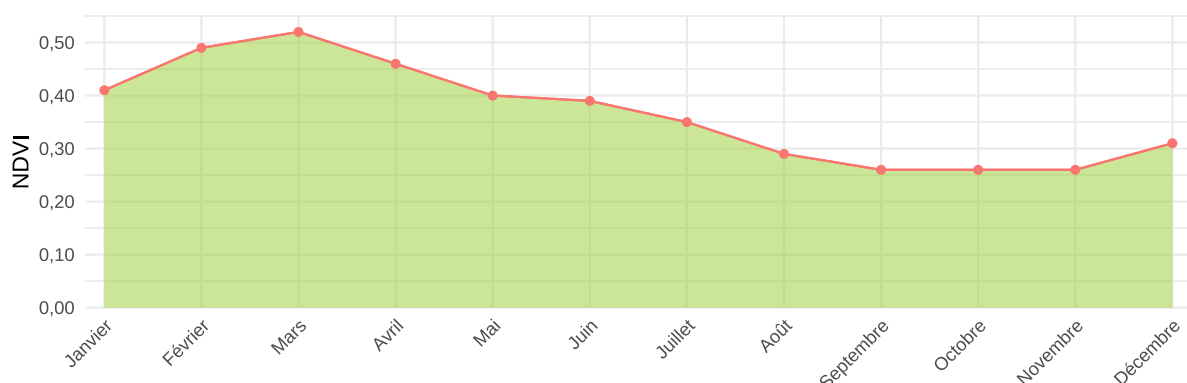
### 1.3. Description de la végétation et saisonnalité

Un riche cortège écosystémique s'est développé en réponse à ces nombreuses et subtiles variations des facteurs pédoclimatiques. L'agencement phytogéographique prend la forme d'une succession suivant les principaux gradients écologiques de la zone d'étude. La flore de la zone d'étude est caractérisée par des traits d'adaptation aux conditions xériques du milieu : des feuilles xéromorphes (épines, feuilles succulentes, etc.), des tiges succulentes (plantes cactiformes ou à latex), un port buissonnant, une faible hauteur. À l'échelle du paysage, les principaux écosystèmes rencontrés sont regroupés sous la typologie des fourrés xérophiles, dont la couverture de canopée varie avec le faciès de végétation (tableau 1). Ce terme fait référence à l'état climacique de la végétation ligneuse qui est une forme adaptée de la forêt dense sèche décidue aux conditions xériques du milieu (Lourenço et al., 1996).

**Tableau 1** : typologie et description des fourrés xérophiles du sud-ouest de Madagascar

Typologie	Espèces caractéristiques	Pédologie	Couverture de canopée	Hauteur de végétation
Bas fourré sur sol sableux	<i>Euphorbia stenoclada</i>	Sable blanc côtier	< 50 %	< 4 m
Bas fourré sur sol calcaire	-	Affleurement calcaire	variabilité locale	< 4 m
Haut fourré arbustif dense	<i>Didiera madagascariensis</i> <i>Adansonia fony</i> <i>Commiphora lamii</i>	Sable roux profond	> 95 %	< 6 m → 10 m
Haut fourré arbustif ouvert	<i>Didiera madagascariensis</i> <i>Commiphora monstruosa</i>	Sable roux-beige superficiel	60 – 70 %	< 6 m → 10 m

Les saisons d'activités de la végétation ont été identifiées sur base d'une série temporelle médiane pour la période 2017-2021 de l'indice de végétation par différence normalisée (NDVI, *normalized difference vegetation index*) calculé à partir des données Sentinel-2 pour la zone d'étude. La période de repos de végétation (saison de repos, NDVI médian = 0,26) s'étend sur les mois de septembre, octobre et novembre. La période de reprise de végétation (saison pré-végétation) débute ensuite et dure du mois de décembre à la fin du mois de février, où le pic de végétation (saison de végétation, NDVI médian = 0,52) a lieu en moyenne en mars. Il s'en suit alors une longue période de sénescence (saison post-végétation) entre les mois d'avril et d'août (figure 3).

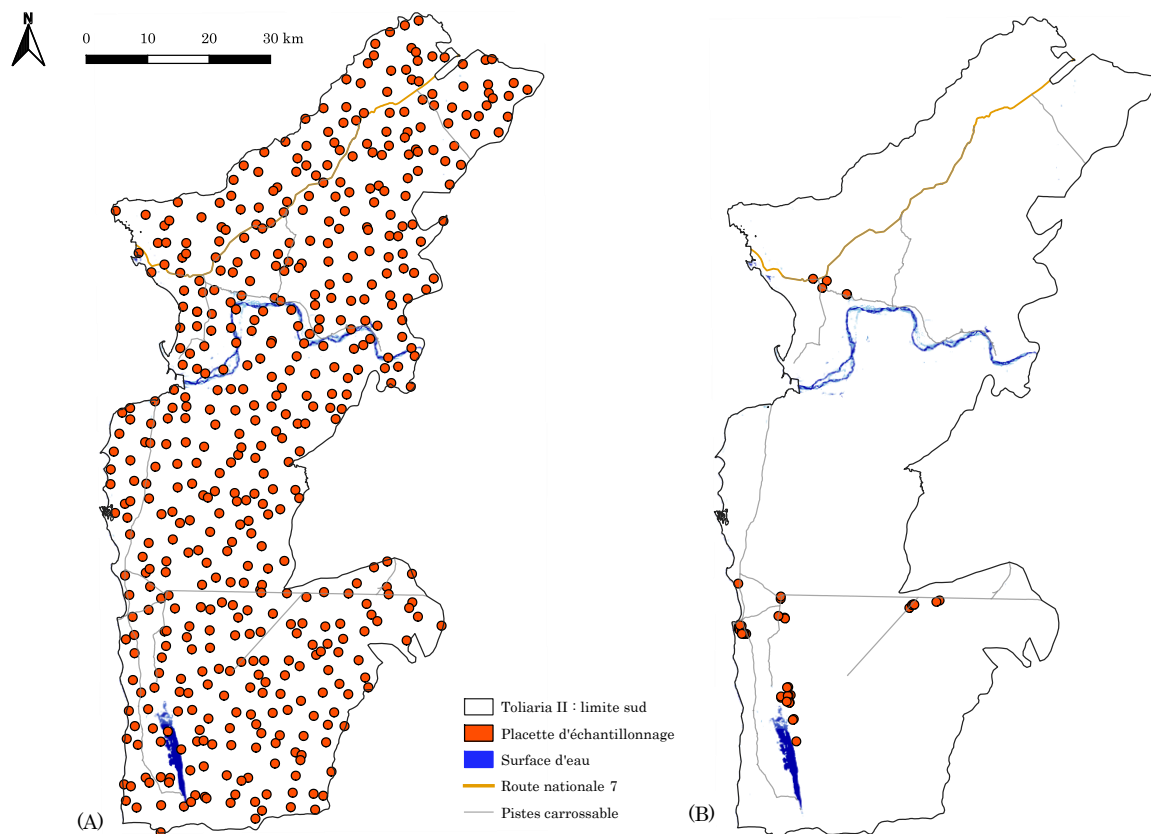


**Figure 3** : variations annuelles médianes du NDVI calculées entre 2017 et 2021 dans la zone d'étude à partir des données Sentinel-2. Chaque nœud de la courbe rouge correspond à une médiane mensuelle moyenne calculée sur la période 2017/2021. La saison de végétation a lieu pendant le mois de mars, et la saison de repos de végétation durant les mois de septembre, octobre et novembre. Les saisons pré-végétation et post-végétation sont situées à la transition entre la saison de végétation et la saison de repos.



## 2. Données de référence

Les données de référence sont constituées de deux jeux de données qui consistent en placettes d'échantillonnage réparties dans la zone d'étude. Le premier jeu de données comporte les informations récoltées sur le terrain au sein de 41 placettes d'échantillonnage de 900 m<sup>2</sup> sur lesquelles ont été réalisés un inventaire en plein de la végétation et une estimation du recouvrement de la végétation avec la méthode du *point-intercept sampling*, complété par une estimation par photo-interprétation (figure 4.B). Le second jeu de données est composé de 505 placettes d'échantillonnage de 2500 m<sup>2</sup> réparties dans la zone d'étude de manière stratifiée non alignée et sur lesquelles le recouvrement de la végétation a été mesuré par photo-interprétation au sein d'une grille systématique (figure 4.A). Une régression linéaire a été utilisée pour modéliser la relation qui lie les attributs structuraux mesurés sur le terrain (biomasse, hauteur de canopée et structure verticale de la végétation) et le recouvrement des parcelles de terrain estimé par photo-interprétation. Ces attributs ont ensuite été extrapolés aux données de photo-interprétation.



**Figure 4** : répartition géographique des données de référence. (A) placettes d'échantillonnage par photo-interprétation (505) dans un système stratifié non aligné, (B) placettes d'échantillonnage de terrain (41) positionnées après stratification selon la couverture en sol et les connaissances des populations locales quant à la localisation des écosystèmes étudiés

## 2.1. Données de terrain : conception du plan d'échantillonnage et récolte de données

### 2.1.a. Plan d'échantillonnage : stratification de la zone d'étude et itinéraire

La couverture du sol nu a été utilisée pour stratifier la zone d'étude et capter l'hétérogénéité à l'échelle des écosystèmes et du paysage. Cet indicateur a été utilisé conjointement avec les connaissances des populations locales concernant la localisation des écosystèmes présents dans la zone d'étude afin d'identifier des zones d'intérêts dont les conditions écologiques étaient homogènes. La couverture du sol nu présente la caractéristique de refléter le type d'écosystèmes et d'être un indicateur local de l'état de conservation des écosystèmes pour lesquels des valeurs de référence de couverture de canopée sont documentées (Lourenço, 1996). En outre, elle est mesurée sans équivoque avec des approches objectives sur le terrain et peut également être mesurée à distance par photo-interprétation sur des images à très haute résolution dont les résultats sont réputés fiables (Morrison, 2016). Cette approche permet donc une stratification objective de la zone d'étude basée sur des données récentes et des connaissances locales.

Les données de photo-interprétation ont été collectées sur la plateforme Collect Earth Online (CEO) (Saah et al., 2019). CEO est un système de visualisation et d'interprétation d'images satellites à très haute résolution (Bing Maps, Google satellite, Mapbox Satellite). Cet outil intègre des informations complémentaires qui peuvent être utilisées conjointement afin de faciliter l'interprétation des images. Un plan d'échantillonnage stratifié non aligné a été utilisé pour disposer aléatoirement 505 parcelles carrées de 2500 m<sup>2</sup> au sein d'une grille systématique dont les mailles ont une superficie de 9,5 km<sup>2</sup>. Cette stratégie combine les avantages de l'échantillonnage systématique tout en conservant une part d'aléatoire (Banko, 1998). Le recouvrement en sol nu a été estimé par interprétation de 49 points répartis systématiquement au sein de chaque unité d'échantillonnage. La surface totale échantillonnée représente 0,026 % de la zone d'étude (1,26 km<sup>2</sup>).

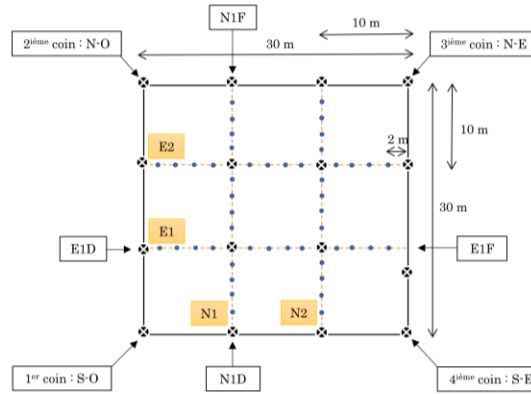
La plateforme Google Earth Engine (GEE) et l'algorithme Random Forest ont été utilisés pour modéliser la couverture du sol nu, dont les valeurs prédites ont été utilisées pour la stratification de l'échantillonnage de terrain. Les bandes brutes et les indices de végétation NDVI et EVI de quatre images composites construites sur une collection d'images Sentinel 2A de 2021 ont été utilisés comme prédicteurs spectraux. Les quatre images composites correspondent aux saisons de végétation (figure 3).

La zone d'étude a été stratifiée selon six niveaux de couverture en sol nu qui s'approchent de la distribution des percentiles de la distribution de recouvrement mesurée par photo-interprétation. Un itinéraire préliminaire a été établi après identification de zones d'inventaires potentielles, homogènes d'un point de vue structurel. L'itinéraire a été conçu pour respecter les contraintes de terrain et échantillonner tous les écosystèmes de la zone d'étude de manière équivalente en garantissant une bonne représentativité de la variabilité d'ouverture de la canopée au sein de ceux-ci.

### 2.1.b. Récolte de données *in situ*

Quarante-deux parcelles ont été inventoriées durant le mois d'avril 2022. Le protocole de récolte de données, inspiré de protocoles standardisés pour les zones semi-arides australiennes (TERN, *Terrestrial Ecosystem Research Network*), a été conçu pour localiser précisément les parcelles dans un système d'information géographique, déterminer la couverture de la végétation et la biomasse aérienne, et décrire la structure verticale et horizontale des écosystèmes.

### i. Mise en place des parcelles



**Figure 5** : schéma des placettes d'échantillonnage (900 m<sup>2</sup>) implantées durant la phase de récolte de données *in-situ*. Les lignes pointillées représentent les transects, sur lesquels a été positionné un point d'interception tous les deux mètres. Les points surmontés d'une croix blanche représentent les relevés GPS/GNSS matérialisés sur le terrain par un piquet surmonté d'un ruban coloré. Les placettes sont orientées dans l'axe nord-sud et subdivisées en 9 sous-placettes de 100 m<sup>2</sup> au sein desquelles l'inventaire en plein de la végétation a été réalisé.

L'homogénéité des conditions environnementales et de la structure de l'écosystème a été validée sur le terrain préalablement à la mise en place des parcelles. Les parcelles de 900 m<sup>2</sup> ont été subdivisées en sous-parcelles de 100 m<sup>2</sup>, orientées dans l'axe magnétique N-S. L'alignement des sous-parcelles a été réalisé avec une boussole SUUNTO, tandis que les distances ont été mesurées avec un dendromètre VERTEX III dont les ultrasons pénètrent la végétation. Chacun des seize coins, espacés de 10 m, ont été matérialisés à l'aide d'un piquet surmonté d'un ruban coloré et ont été géoréférencés avec un GPS portable multibande de type multi-GNSS (Garmin GPSMAP 65S) (figure 5).

### ii. Inventaire en plein

Un inventaire en plein de la végétation a été réalisé pour chaque individu dont la circonférence d'au moins une branche basale était supérieure ou égale à 5 cm. Les informations récoltées sont propres au type morphologique de la plante (figure 6), afin de permettre l'estimation de la biomasse selon les équations allométriques conçues par Randriamalala et al. (2022) pour les fourrés xérophiles du sud et sud-ouest de Madagascar. Outre l'identification de l'espèce et de la hauteur totale (H), les informations récoltées suivantes ont été récoltées :

- Type I : diamètre à 30 cm ( $D_{30}$ ), nombre de branches de plus d'un mètre ( $N_c$ ).
- Type II : diamètre à 30 cm de la tige la plus large ( $D_{30}^{MAX}$ ), nombre de branches de plus d'un mètre ( $N_c$ ).
- Type III : diamètre à 1,3 m (DBH).

Le nom vernaculaire local des espèces a été obtenu avec l'aide des populations locales et la correspondance avec les noms scientifiques a été réalisée en se référant à diverses sources, principalement d'autres études scientifiques. Néanmoins, toutes les espèces n'ont pas pu être identifiées.

### iii. Point-intercept sampling

La méthode du *point intercept sampling* a été utilisée pour estimer la couverture de la végétation en projection verticale au sol. Le long de quatre transects à l'intérieur de la parcelle, deux dans l'axe N-S et deux dans l'axe E-O, espacés de 10 m entre eux, un point d'interception a été positionné tous les 2 m, soit 14 points par transect. Les quatre points dédoublés situés à l'intersection des transects ne sont

comptabilisés qu'une fois, soit un total de 52 points par parcelle pour une densité de 577 points par hectare (figure 5). Le transect a été matérialisé à l'aide d'une corde de 10 m marquée tous les 2 m et tendue entre deux piquets. Chaque individu qui intercepte un point en projection verticale a été identifié, son type morphologique a été décrit et sa hauteur estimée à l'œil.

### 2.1.c. Préparation du jeu de données

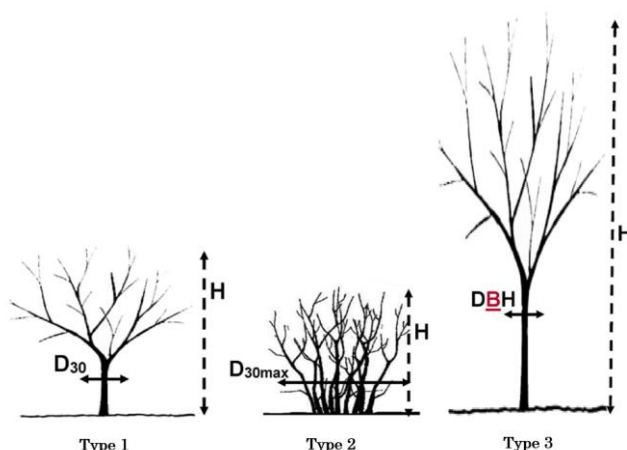
#### i. Localisation précise des parcelles dans un SIG

Une correction de la position des parcelles a été nécessaire afin de pallier les erreurs du GNSS et de l'opérateur. Le centroïde des parcelles a été calculé à partir de la longitude et latitude moyenne des seize points GPS relevés dans chaque parcelle. Un polygone carré de 2500 m<sup>2</sup> a ensuite été généré en tenant compte de la déclinaison magnétique dans la mesure de l'azimut. La déclinaison magnétique (21° 46 W) a été calculée avec le *World Magnetic Model*, en utilisant la longitude et latitude moyenne des parcelles (23°54'45'' S 43°43'45'' E) et la date moyenne de récolte de données (15 avril 2022).

#### ii. Détermination de la biomasse

Les équations allométriques utilisées pour l'estimation de la biomasse sont issues des travaux de Randrianamalala et al. (2022) dans lesquels trois équations sont établies pour chacun des types morphologiques présents dans ces écosystèmes (figure 6). Les équations allométriques de Vieilledent et al. (2012) établies pour les forêts sèches épineuses de Madagascar ont été utilisées pour les arbres de type 3 situés en dehors du domaine d'application des équations allométriques locales, principalement des baobabs (*Adansonia sp.*) (tableau 2).

Les valeurs de densité déterminées lors des travaux de Randrianamalala et al. (2022) ont été utilisées prioritairement. Lorsque les données n'étaient pas disponibles, les densités disponibles dans la base de données globale de densité du bois (*Global Wood Density database*) ont été utilisées (Flores et al., 2011). La densité au niveau de la famille a été utilisée lorsqu'aucune information n'a été trouvée au niveau du genre. Pour les espèces indéterminées, la densité moyenne des individus présents sur la même parcelle a été attribuée afin de ne pas perturber les moyennes locales.



**Figure 6** : description des types morphologiques de la végétation qui correspondent aux équations allométriques définies par Randrianamalala et al. (2022) pour les fourrés xérophiles du sud-ouest de Madagascar. Les paramètres structuraux utilisés pour déterminer la biomasse varient selon le type morphologique. Les plantes du type 1 ont une unique insertion basale et un fût de moins de 1,30 m de haut, le diamètre à 30 cm est relevé. Les plantes du type 2 ont de multiples insertions basales, le diamètre à 30 cm de la plus grosse branche est utilisé. Les plantes du type 3 ont une unique insertion basale et un fût de plus de 1,30 m, la mesure du diamètre est réalisée à 1,30 m (DBH). Figure présentée par Randrianamalala et al. (2022)

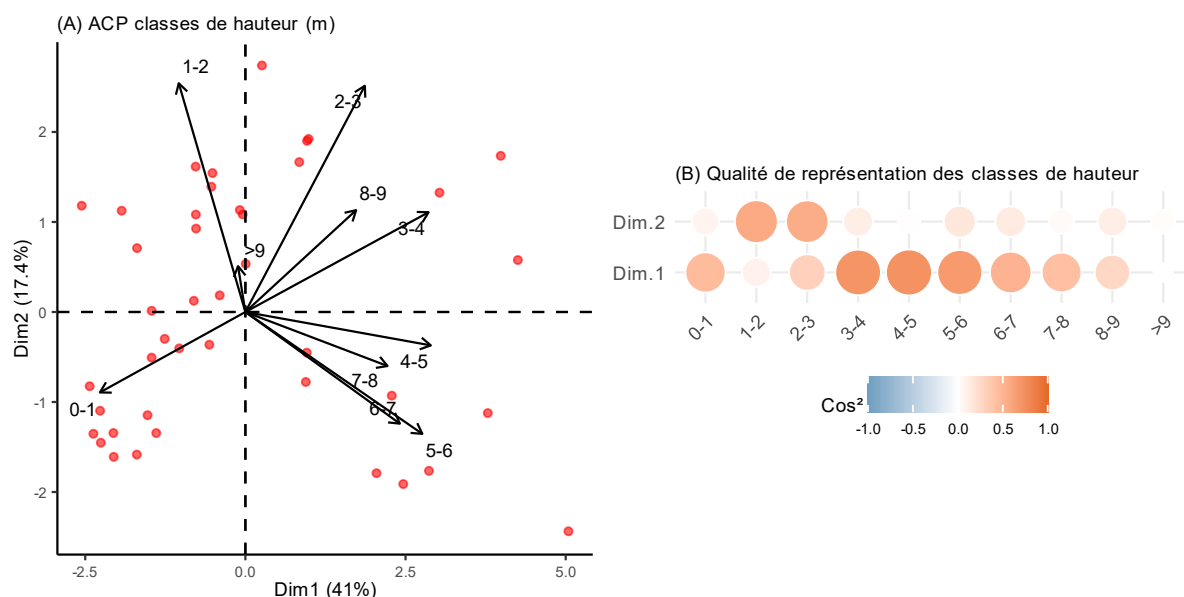
**Tableau 2** : équations allométriques utilisées pour l'estimation de la biomasse

Domaine d'application	Référence	Équation	R <sup>2</sup>	Biais
Type morphologique n°1 Fourrés xérophiles (Madagascar)	Randriamalala et al. (2022)	$\ln AGB = -2,629 + 2,103 \ln D_{30} + 1,043 \ln H$ $+ 1,464 \ln \rho - 1,537 \ln \rho \ln N_c$ $+ 0,712 \ln D_{30} \ln \rho \ln N_c$	0,82	34,0%
Type morphologique n°2 Fourrés xérophiles (Madagascar)	Randriamalala et al. (2022)	$\ln AGB = -2,629 + 2,103 \ln D_{30} + 1,043 \ln H$ $+ 1,464 \ln \rho - 1,537 \ln \rho \ln N_c$ $+ 0,712 \ln D_{30} \ln \rho \ln N_c$	0,84	14,0%
Type morphologique n°3 Fourrés xérophiles (Madagascar)	Randriamalala et al. (2022)	$\ln AGB = -2,629 + 2,103 \ln D_{30} + 1,043 \ln H$ $+ 1,464 \ln \rho - 1,537 \ln \rho \ln N_c$ $+ 0,712 \ln D_{30} \ln \rho \ln N_c$	0,83	19,3%
Forêts sèches épineuses (Madagascar)	Vieilledent et al. (2012)	$\ln AGB = -1,103 + 1,994 \ln DBH + 0,317 \ln H$ $+ 1,303 \ln \rho$	0,58	110,6%

$\ln$  = logarithme népérien, AGB = biomasse aérienne,  $\rho$  = infradensité au niveau du genre,  $D_{30}$  = diamètre à 0,30 m,  $D_{30 \max}$  = diamètre à 0,30 m de la tige la plus épaisse, DBH = diamètre à 1,30 m,  $N_c$  = nombre de branches de plus d'un mètre, H = hauteur

### iii. Hauteur de canopée et structure verticale de la végétation

L'indicateur utilisé pour représenter la hauteur de canopée est le 95<sup>e</sup> percentile de la hauteur relevée sur le terrain. La structure verticale de la végétation a été résumée en deux indicateurs qui résument la distribution des classes de hauteur : les coordonnées dans l'espace des deux premières dimensions d'une analyse en composantes principales appliquée à la distribution des classes de hauteurs (figure 7). Les individus de chaque parcelle ont été répartis par classe de hauteur de 1 m, avec la dernière classe ouverte (>9 m). Les coordonnées dans l'espace des deux premières composantes principales ont été attribuées à chaque placette. L'illustration de l'analyse et de la qualité de représentation des classes de hauteur sur les deux premières dimensions sont illustrées à la figure 7A et 7B. Le score d'une placette sur la première dimension, soit la première coordonnée dans l'espace de l'ACP, est principalement influencé par l'absence d'arbre de 1 à 3m et de plus de 9 m. La seconde coordonnée est influencée par les classes restantes. Les deux premières dimensions capturent, respectivement, 41% et 17,4% de la variance observée. Le positionnement de la plupart des parcelles est cohérent vis-à-vis des observations réalisées sur le terrain.



**Figure 7** : caractérisation de la structure verticale des écosystèmes à partir de (A) les coordonnées dans l'espace des deux premières dimensions d'une analyse en composantes principales (ACP) réalisée sur la distribution des classes de hauteur associée à chaque placette d'échantillonnage de terrain. Les placettes sont représentées par un point rouge. (B) Illustration de la qualité de représentation (corrélations,  $\cos^2$ ) des classes de hauteur dans les dimensions de l'ACP.

#### iv. Correction des données de couverture boisée

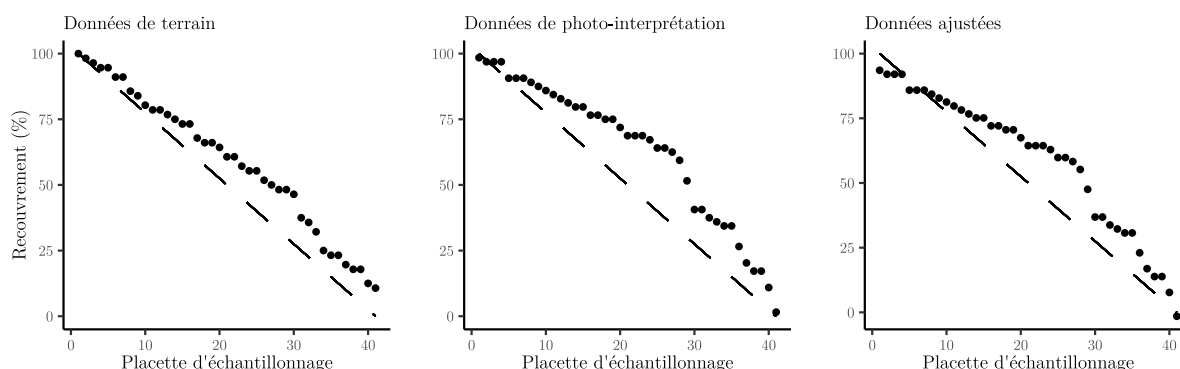
La couverture végétale des 42 parcelles échantillonnées sur le terrain a également été estimée par photo-interprétation, en suivant la méthodologie décrite précédemment. Il n'y a *a priori* aucune des deux méthodes d'estimation de la couverture végétale qui ne soit pas entachée d'une erreur. Par ailleurs, plusieurs études ont mis en évidence les divergences d'estimation du recouvrement selon les deux méthodologies, avec *a priori* des estimations plus fiables en utilisant des données d'images à très haute résolution (Kattenborn et al., 2019; Eskandari et al., 2020). La relation entre les données récoltées sur le terrain et celles acquises par photo-interprétation a donc été modélisée en tenant compte de cette double variabilité du jeu de données avec une régression linéaire modèle II de type *major axis regression* (MAR). Ce type de régression minimise la somme des carrés de la distance perpendiculaire entre chaque point et la ligne de régression, contrairement à une régression linéaire simple.

La couverture végétale estimée sur le terrain a été ajustée selon cette relation ( $r^2 = 0,81$ ), pour créer une nouvelle estimation de la couverture végétale ( $y$ ) comme fonction de celle mesurée par photo-interprétation ( $x$ ), telle que :

$$y = 0,030 + 0,981 x$$

#### v. Distribution des valeurs de recouvrement

La distribution des valeurs de recouvrement mesurées sur le terrain est proche d'une distribution équilibrée optimale, ce qui signifie que l'hétérogénéité à l'échelle du paysage a bien été capturée par la stratification sur base de la couverture en sol nu (figure 8).

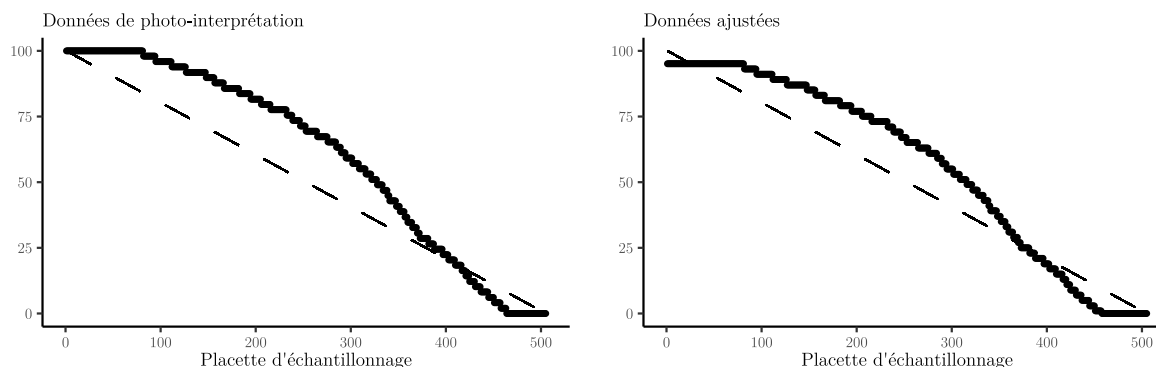


**Figure 8** : distribution du recouvrement des placettes d'échantillonnages de terrain. Les points représentent les placettes, classées par ordre décroissant de recouvrement. La ligne pointillée représente une distribution équilibrée.

## 2.2. Données de photo-interprétation

Le second jeu de données qui est utilisé dans le cadre de cette étude est constitué des données de photo-interprétation récoltées lors de la stratification de la zone d'étude. Les mesures de recouvrement ont été ajustées selon la relation modélisée entre le recouvrement ligneux mesuré sur le terrain et celui mesuré par photo-interprétation sur les parcelles de terrain. Les données de recouvrement initiales récoltées par photo-interprétation sont caractérisées par une surreprésentation des valeurs extrêmes – 0 et 100 % – et une faible représentation des valeurs comprises entre 30 et 60 % de couverture (figure 9). Le jeu de données de photo-interprétation final est construit après extrapolation des paramètres descriptifs supplémentaires (AGB, H95, structure verticale). La relation entre le recouvrement mesuré par photo-interprétation et les paramètres structuraux mesurés sur le terrain a été réalisée par régression linéaire.

Les deux jeux de données ont ensuite été utilisés conjointement afin de répondre aux objectifs de cette étude.



**Figure 9** : distribution du recouvrement des placettes d'échantillonnages de photo-interprétation. Les points représentent les placettes, classées par ordre décroissant de recouvrement. La ligne pointillée représente une distribution équilibrée.

## 2.2.a. Modélisation et extrapolation des paramètres additionnels

Le coefficient de corrélation entre la couverture végétale et les attributs structuraux mesurés sur le terrain a été utilisé en tant qu'indicateur du potentiel de modélisation. Une transformation logarithmique a été appliquée à la biomasse, qui évolue exponentiellement avec la couverture végétale et montre une saturation pour les valeurs de couverture boisée les plus élevées. Le potentiel de modélisation est important pour la biomasse et la hauteur, tandis qu'il est plus faible pour les composantes principales de distribution de hauteur (tableau 3).

**Tableau 3** : coefficients de corrélation calculés entre la couverture végétation et les autres attributs structuraux à partir des données récolées sur le terrain.

	Log AGB	H95	Dim 1 ACP H	Dim 2 ACP H
Couverture de la canopée	0,85	0,66	0,52	0,24

Log AGB = logarithme de la biomasse aérienne, H95 = 95<sup>e</sup> percentile de la distribution des hauteurs, Dim 1 ACP et Dim 2 ACP = dimensions 1 et 2 de l'analyse en composantes principales réalisée sur la distribution des classes de hauteurs

Les relations qui lient le recouvrement et les autres attributs structuraux ont été modélisées avec une régression linéaire simple. Trois des quatre paramètres à extrapoler montrent une relation très hautement significative ( $p$ -valeur  $< 0,001$ ) avec le recouvrement mesuré par photo-interprétation : le logarithme de la biomasse (adj.  $R^2 = 0,72$ ), la hauteur de canopée (adj.  $R^2 = 0,63$ ) et la première dimension de l'ACP (adj.  $R^2 = 0,25$ ). La relation avec la seconde composante principale (adj.  $R^2 = 0,03$ ) est moins évidente, avec une  $p$ -valeur de 0,136 (tableau 4).

**Tableau 4** : équations et paramètres de modélisation par régression linéaire des attributs écosystémiques mesurés sur le terrain depuis la couverture de canopée mesurée par photo-interprétation sur les parcelles de terrain

Équation régression linéaire	R <sup>2</sup> ajusté	p-valeur
$\text{Log AGB}_{\text{terrain}} = -0,512 + 2,342 * \text{WC}_{\text{photo-interprétation}}$	0,72	< 0,001
$\text{H95}_{\text{terrain}} = 1,174 + 4,650 * \text{WC}_{\text{photo-interprétation}}$	0,63	< 0,001
$\text{Dim1 H}_{\text{terrain}} = -2,389 + 4,057 * \text{WC}_{\text{photo-interprétation}}$	0,25	< 0,001
$\text{Dim2 H}_{\text{terrain}} = -0,716 + 1,203 * \text{WC}_{\text{photo-interprétation}}$	0,03	0,136

$\text{WC}_{\text{photo-interprétation}}$  = couverture de canopée mesurée par photo-interprétation sur les parcelles de terrain, Log AGB = logarithme de la biomasse aérienne, H95 = 95<sup>e</sup> percentile de la distribution des hauteurs, Dim 1 ACP et Dim 2 ACP = dimensions 1 et 2 de l'analyse en composantes principales réalisée sur la distribution des classes de hauteurs

### 2.3. Description du jeu de données de référence

La zone d'étude est caractérisée par une couverture boisée moyenne d'approximativement 60 % pour une biomasse aérienne moyenne de 18 t/ha et une hauteur de canopée moyenne de 4 m (tableau 5).

**Tableau 5** : statistiques descriptives des données de référence par variable de réponse

		Terrain (n = 41, 900m <sup>2</sup> )	Photo-interprétation (n = 505, 2500 m <sup>2</sup> )
Couverture boisée (%)	Moyenne ( $\pm \sigma$ )	59 ( $\pm 27$ )	57 ( $\pm 34$ )
	Médiane	64	65
Biomasse aérienne (t/ha)	Moyenne ( $\pm \sigma$ )	17,70 ( $\pm 18,40$ )	19,10 ( $\pm 19,20$ )
	Médiane	10,70	10,30
Hauteur de canopée (m)	Moyenne ( $\pm \sigma$ )	4,09 ( $\pm 1,90$ )	3,71 ( $\pm 1,79$ )
	Médiane	3,60	4,20
Dimension 1 ACP H	Moyenne ( $\pm \sigma$ )	0,01 ( $\pm 2,07$ )	0,15 ( $\pm 1,15$ )
	Médiane	- 0,56	0,25
Dimension 2 ACP H	Moyenne ( $\pm \sigma$ )	0,00 ( $\pm 1,35$ )	0,04 ( $\pm 0,34$ )
	Médiane	- 0,30	0,07

$\sigma$  = écart-type, n = nombre de placettes d'échantillonnage



### 3. Identification des groupes de prédicteurs spectraux et création des images composites

#### 3.1. Données de télédétection

Les données de télédétection utilisées sont issues de la série de satellites multispectraux Sentinel-2, dont la durée de revisite est de 5 jours. Les images de réflectance corrigées au niveau de la surface de la Terre (niveau 2A) ont été obtenues sur la plateforme Google Earth Engine (GEE). Les images avec plus de 20 % de nuages ont préalablement été filtrées et un masque construit sur base de la bande d'évaluation de la qualité *QA60* a été appliqué (algorithme *maskS2clouds* disponible sur GEE). Sauf indication contraire, la valeur médiane de réflectance entre 2017 et 2022 a été utilisée dans la construction des variables prédictives décrites ci-après. Les images composites ainsi obtenues sont représentatives de l'état moyen de l'activité végétale, fortement influencée par les variabilités interannuelles du climat. Les données spectrales utilisées sont situées dans le spectre visible, le proche infrarouge et l'infrarouge à courte longueur d'onde. Les bandes de résolution supérieure à 10 m ont été rééchantillonnées à 10 m directement dans GEE. Un tableau récapitulatif des modèles utilisés et des prédicteurs associés à ces modèles est présenté après chaque catégorie de paramètres étudiés. L'ensemble des prédicteurs et leurs définitions est présenté dans un tableau récapitulatif (tableaux 12).

#### 3.2. Variables de saisonnalité

L'étude de la saisonnalité est réalisée en comparant les performances des modèles construits à partir des prédicteurs spectraux de chacune des quatre périodes clés de l'activité végétale décrites ci-avant (repos, transition pré-végétation, végétation et transition post-végétation). Les prédicteurs sont les bandes spectrales brutes (R, G, B, NIR, SWIR1 et SWIR2) et trois indices de végétation : Le NDVI, sensible à la chlorophylle, le MSAVI2 qui comprend un facteur de correction qui minimise l'influence du sol, et l'EVI qui diminue le bruit atmosphérique et la saturation à haute biomasse. Les valeurs médianes de réflectance et d'indices de végétation, pour chacune des saisons, constituent les variables prédictives des modèles — soit 9 variables par modèle. Une cinquième image composite est constituée de l'ensemble des variables des quatre saisons, soit 36 variables (tableau 6)

**Tableau 6 :** Description des prédicteurs intégrés dans les modèles utilisés pour caractériser l'influence de la saisonnalité sur la représentation des attributs écosystémiques. Tous les modèles utilisent les mêmes prédicteurs à des saisons différentes, à l'exception du modèle *Annuel* qui intègre l'ensemble des prédicteurs des modèles saisonniers

Modèles	Saison de végétation	Période considérée (2017-2021)	Prédicteurs* Bandes Sentinel-2	Indices de végétation	Nb
Repos	Repos	01/09 – 30/11	B02, B03, B04, B8A, B11, B12	NDVI, EVI, MSAVI2	9
AvVeg	Pré-végétation	01/12 – 28/02			
Veg	Végétation	01/03 – 31/03			
ApVeg	Post-végétation	01/04 – 31/08			
Annuel	[↓]	[↓]	[↓]	[↓]	36

\* tableau détaillé des prédicteurs à la page 25 (tableau 12). [↓] = combinaison qui intègre tous les prédicteurs des lignes précédentes

#### 3.3. Résolution spatiale

L'impact de la résolution spatiale est évalué sur base d'une image composite construite par la médiane de la réflectance et des indices de végétation durant la période de repos de végétation. Ce choix repose sur

les résultats obtenus lors de l'étude des variables de saisonnalité pour le recouvrement de la végétation. La résolution initiale de 10 m a été rééchantillonnée jusqu'à 150 m, par tranche de 20 m, avec la réflectance moyenne des pixels agrégés. Les prédicteurs utilisés sont les bandes R, G, B, NIR, SWIR1, SWIR2, ainsi que les indices de végétation NDVI, MSAVI2 et EVI (tableau 7).

**Tableau 7** : Description des prédicteurs intégrés dans les modèles utilisés pour caractériser l'influence de la résolution spatiale sur la représentation des attributs écosystémiques. Tous les modèles utilisent les mêmes prédicteurs, pour une même date, à différentes résolutions

Modèles	Résolution	Période considérée (2017-2021)	Prédicteurs* Bandes Sentinel-2	Indices de végétation	Nb
Res.10	10 m	01/09– 30/11	B02, B03, B04, B8A, B11, B12	NDVI, MSAVI2, EVI	9
Res.30	30 m	01/09– 30/11			
Res.50	50 m	01/09– 30/11			
Res.70	70 m	01/09– 30/11			
Res.90	90 m	01/09– 30/11			
Res.110	110 m	01/09– 30/11			
Res.130	130 m	01/09– 30/11			
Res.150	150 m	01/09– 30/11			

\* tableau détaillé des prédicteurs à la page 25 (tableau 12)

### 3.4. Variables spectrales

L'étude des facteurs spectraux vise à caractériser l'influence de chaque partie du spectre électromagnétique prise isolément. Les bandes spectrales de Sentinel-2 ont été réparties en trois catégories selon leur nature : le visible, le proche infrarouge et l'infrarouge à courte longueur d'onde. Pour caractériser l'influence de ces catégories sur la représentation des attributs structuraux, quatre groupes de prédicteurs spectraux ont été mis en place. Parallèlement, l'influence des indices de végétation a été étudiée à partir de trois modalités. Ces groupes et modalités ont été croisés entre eux, de façon à obtenir 12 modèles qui ont pour objectif de caractériser l'influence des catégories spectrales et des indices de végétation et leurs intérêts dans le cadre de la modélisation des attributs structuraux (tableau 9). Trois analyses préliminaires ont été nécessaires afin d'identifier les prédicteurs spectraux, et plus particulièrement les indices de végétation, finalement intégrés dans ces modèles. Ces analyses ont été réalisées sur base des données de photo-interprétation, car elles offrent une meilleure représentativité du paysage que les données de terrain. Les données de réflectance utilisées consistent en la réflectance médiane en saison de repos de végétation sur la période 2017-2021.

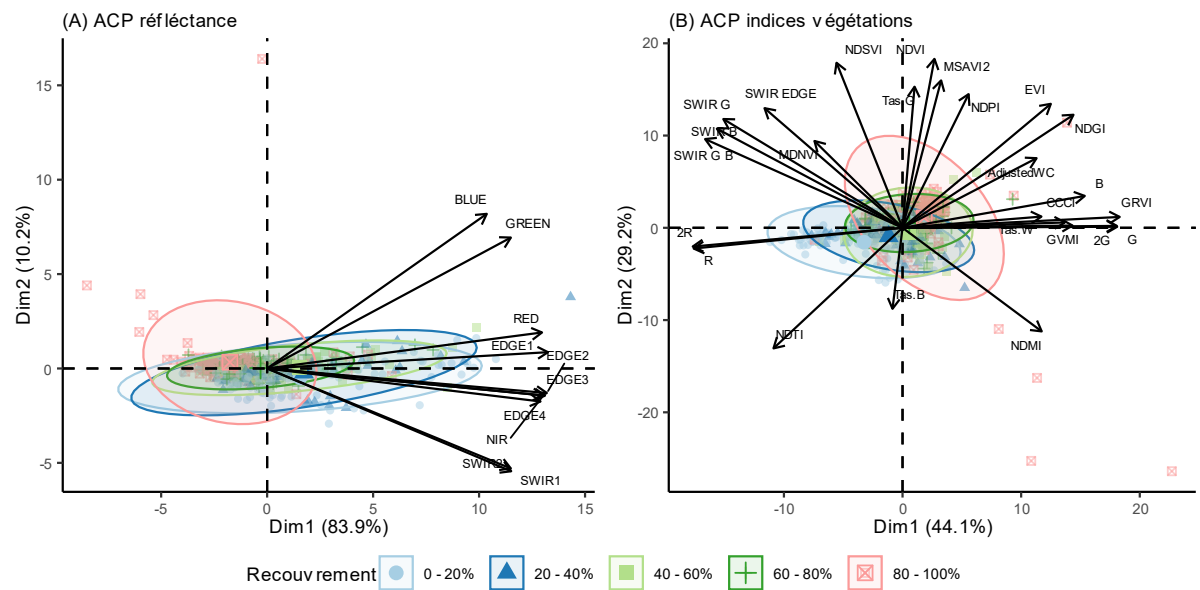
Dans un premier temps, une analyse en composantes principales (ACP) a été réalisée sur les bandes spectrales brutes de Sentinel-2 pour évaluer la séparabilité des placettes d'échantillonnage sur base de leurs signatures spectrales (figure 10.A). Dans un second temps, les valeurs de réflectance ont été analysées pour cinq classes de recouvrement (figure 11), et ensuite complétées par une analyse de la corrélation entre le degré d'ouverture de la canopée et la réflectance spectrale (tableau 8). Ces deux étapes ont permis d'identifier les parties du spectre électromagnétique qui peuvent permettre une meilleure discrimination des placettes sur base de l'ouverture de la canopée et un choix en conséquence d'une série d'indices de végétation. Enfin, une ACP a été réalisée sur la valeur des indices spectraux associés à chaque placette afin d'identifier, par groupe spectral, les indices de végétation les mieux représentés et les moins redondants sur les quatre premiers axes de variance (figure 10.B).

La première dimension de l'ACP réalisée sur les bandes spectrales brutes capture 83,9 % de la variance, avec une tendance de corrélation unidirectionnelle très marquée, principalement pour le rouge et le proche infrarouge. La seconde dimension (10,2 %) permet une discrimination sur base des bandes B et V d'une part, et des bandes SWIR d'autre part. Les parcelles sont globalement mal représentées dans l'espace de l'ACP et les classes de couverture ligneuse sont difficilement distinguables. Une légère démarcation est visible pour les parcelles au recouvrement le plus important.

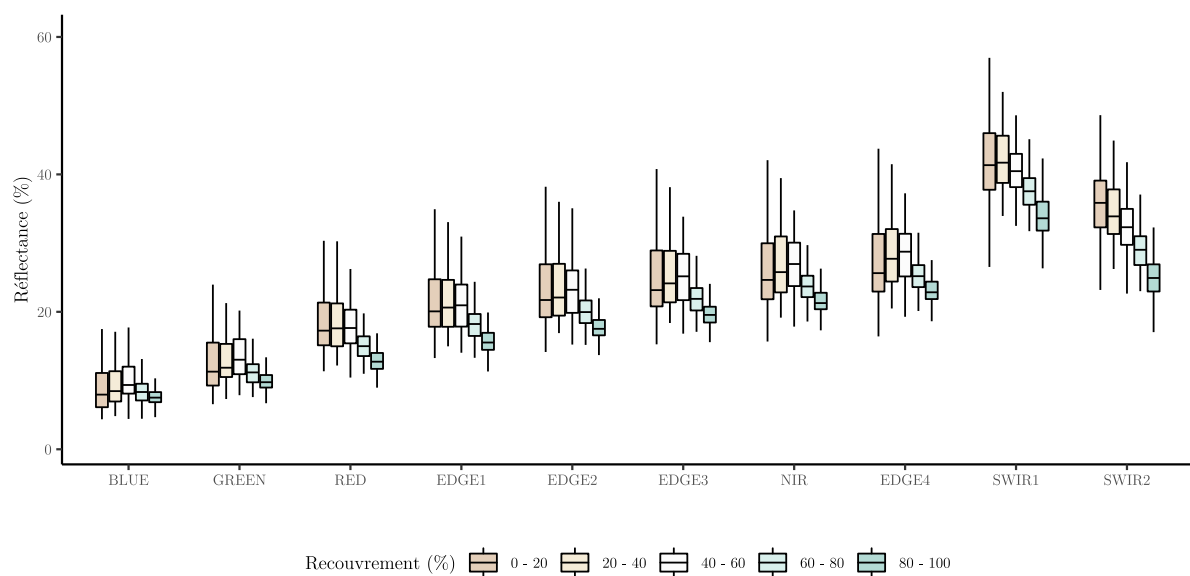
Sur base de l'analyse de la réflectance spectrale associée aux classes de recouvrement, les différences de signature spectrale les plus importantes en fonction du degré d'ouverture de la canopée ont été identifiées dans l'infrarouge à courte longueur d'onde (SWIR1 et SWIR2) et, dans une moindre mesure, dans le proche infrarouge (NIR & Red Edge 1). La variance entre les placettes d'une même classe de recouvrement est plus faible pour les classes les plus élevées (60-80 % et 80-100 %). Par ailleurs, ces classes de recouvrement se démarquent à partir de l'ensemble du spectre. La distinction des classes à faible recouvrement est plus subtile, avec une réflectance stable entre 0 et 60 % de recouvrement pour la plupart des bandes spectrales. Les bandes SWIR2, NIR et Red Edge 4 semblent les plus adaptées pour ces gammes de valeurs, ainsi que les bandes bleu et vert dans le spectre visible. L'analyse des corrélations de Spearman entre le degré d'ouverture de la canopée et la réflectance spectrale confirme les tendances préliminaires déjà observées. Une corrélation très importante est observée pour les bandes SWIR2, SWIR1, Red Edge 1 et Red Edge 1rouge. Les autres bandes sont moins corrélées à cet attribut.

Une série d'indices de végétation a été construite sur base de ces résultats préliminaires, avec pour objectif de caractériser l'influence des trois parties du spectre électromagnétique étudiées (annexe 4). Un groupe d'indices a donc été construit uniquement avec le spectre visible, un autre groupe avec seulement le proche infrarouge, un troisième groupe avec l'infrarouge à courte longueur d'onde et, enfin, un dernier groupe avec les trois catégories du spectre électromagnétique. Certains indices ont été élaborés spécifiquement pour cette étude lorsqu'aucune combinaison spectrale correspondant aux besoins de cette étude n'a été identifiée dans la littérature. C'est le cas, par exemple, d'un indice qui combine les bandes SWIR2 B et V, alors que ces bandes ont été identifiées comme les plus importantes pour distinguer les classes de recouvrement entre 0 et 60 %. La liste détaillée des indices de végétation — avec leur acronyme, formule — est présentée dans à l'annexe 4.

Enfin, la dernière étape préliminaire a consisté à sélectionner les indices les plus pertinents en fonction des objectifs de cette étude. Une seconde analyse en composantes principales a été réalisée uniquement sur base des valeurs d'indices de végétation associées à chaque parcelle. Les 4 premières dimensions capturent ensemble 91 % de la variance, respectivement 44 %, 30 %, 9 % et 8 %. Une meilleure représentation des parcelles dans l'espace de l'ACP, ainsi qu'une meilleure discrimination des classes de recouvrement, a été obtenue à la suite de l'utilisation des indices de végétation par rapport aux bandes spectrales brutes. L'indice de végétation le plus corrélé aux principaux axes de variance par groupe spectral étudié a été conservé pour la suite des analyses.



**Figure 10 :** représentation de l'espace des deux premières composantes principales de la variance spectrale au sein des placettes d'échantillonnage de photo-interprétation. Les classes de recouvrement associées aux placettes sont projetées sous la forme d'ellipses. (A) variance sur base des bandes brutes de Sentinel-2. (B) variance sur base d'une série d'indices de végétation.



**Figure 11 :** Distribution de la réflectance spectrale par classe de recouvrement (%) et par bande Sentinel-2

**Tableau 8** : analyse de la corrélation entre le degré d'ouverture de la canopée et la réflectance spectrale des bandes Sentinel-2

Bande spectrale	Coefficient de corrélation de Spearman	p-valeur
Bleu (B)	-0,18	< 0,001
Vert (G)	-0,31	< 0,001
Rouge (R)	-0,61	< 0,001
Red Edge 1	-0,59	< 0,001
Red Edge 2	-0,54	< 0,001
Red Edge 3	-0,51	< 0,001
Proche infrarouge (NIR)	-0,44	< 0,001
Red Edge 4	-0,40	< 0,001
Infrarouge à courte longueur d'onde (SWIR1)	-0,62	< 0,001
Infrarouge à courte longueur d'onde (SWIR2)	-0,77	< 0,001

**Tableau 9** : Description des prédicteurs intégrés pour chaque modèle qui vise à caractériser l'influence des prédicteurs spectraux sur la représentation des attributs écosystémiques. Le premier groupe de modèles (groupes de réflectance) vise à isoler l'influence de la partie du spectre électromagnétique. Le second groupe de modèles (catégorie de modèles) vise à caractériser l'influence des bandes brutes de Sentinel-2 utilisées seules, les indices de végétation utilisés seuls et la combinaison des bandes brutes et des indices de végétation.

	Groupes de réflectance	Visible = (*)	Visible + proche IR	Visible + IR à courte longueur d'onde	Combinaisons groupes de réflectance
Catégorie de modèles		<i>RGB</i>	<i>RGB_IR</i>	<i>RGB_SWIR</i>	<i>RGB_IR_SWIR</i>
Bandes brutes (S2)		- Bleu - Vert - Rouge	( <sup>*1</sup> ) + Red Edge 1 + NIR	( <sup>*1</sup> ) + SWIR1 + SWIR2	[→]
Indices de végétation		- GRVI - B	( <sup>*2</sup> ) + NDGI + NDVI + CCCI	( <sup>*2</sup> ) + SWIR G B + NDSVI	[→] + Tasseled Cap Br. + GVMi
Combinaison bandes brutes et indices de végétation		[↓]	[↓]	[↓]	[→] [↓]

(\*) = groupe du visible, (<sup>\*1</sup>) = bandes brutes du visible, (<sup>\*2</sup>) = indices de végétation du visible, [→] = combinaison qui intègre tous les prédicteurs des colonnes précédentes ; [↓] = combinaison qui intègre tous les prédicteurs des lignes précédentes

### 3.5. Variables phénologiques

#### 3.5.a. Métriques annuelles

L'approche par les métriques annuelles vise à caractériser les variations phénologiques observées à l'échelle d'un cycle de végétation complet. Les métriques ont été calculées sur les bandes spectrales et sur les indices de végétation : la valeur maximale, la valeur minimale, la somme, l'écart-type (E.T.), l'écart des valeurs maximales et minimales (amplitude), l'écart interquartile (Q.25-75) et l'écart interpercentile 5 et 95 (Q.5-95). La série temporelle utilisée est constituée d'images composites construites par périodes de 15 jours avec la médiane de la réflectance et des indices de végétation sélectionnés lors de l'étude préliminaire de la variabilité spectrale. En outre, une analyse préliminaire de la variabilité annuelle des prédicteurs a été réalisée pour éliminer ceux dont la variabilité est trop faible pour avoir un effet sur la représentation des attributs écosystémiques étudiés. Cette analyse préliminaire a été réalisée sur les parcelles issues de la photo-interprétation. Le coefficient de variation annuelle de chaque prédicteur a été calculé sur la série temporelle. Ce coefficient de variation annuelle a été mis en perspective avec les classes

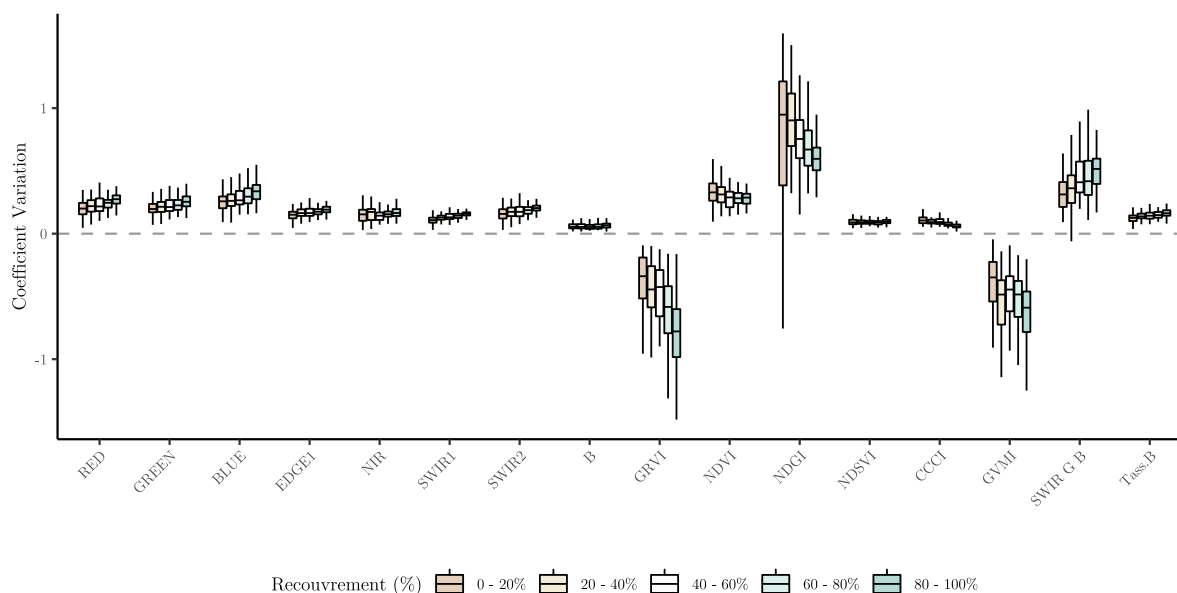
de couverture de la canopée (figure 12). Aucune bande spectrale brute ne montre de variation annuelle importante, tandis que quatre indices de végétation se démarquent clairement. Les métriques phénologiques annuelles ont été calculées sur les variations du GRVI, du GVMI, du NDGI et du SWIR G B.

Six modèles ont été utilisés pour cette partie (tableau 10). Un premier modèle est constitué des bandes brutes en saison de repos (BANDS). Quatre modèles incluent les bandes brutes en saison de repos et, respectivement, les métriques phénologiques calculées sur le GVMI, le GRVI, le NDGI et le SWIR G B. Le dernier modèle (FULL) utilise les bandes brutes et l'ensemble des métriques phénologiques.

**Tableau 10** : Description des prédicteurs intégrés dans les modèles utilisés pour caractériser l'influence des variations phénologiques annuelles sur la représentation des attributs écosystémiques. Les modèles intègrent les mêmes bandes brutes, ainsi que les mêmes métriques phénologiques mais calculées sur des cibles différentes. Le modèle intégré combine l'ensemble des métriques phénologiques.

Modèles	Cible des métriques phénologique	Période considérée (2017-2021)	Prédicteurs* Bandes Sentinel-2	Métriques phénologiques	Nb
BANDS	/				7
GVMI	GVMI	01/09– 30/11	B02, B03, B04, B05, B8A, B11, B12	Max, min, produit,	14
GRVI	GRVI			amplitude, Q.25-75,	14
NDGI	NDGI			Q.5-95 , E.T., somme	14
SWIR G B	SWIR G B				14
FULL	[↓]			[↓]	35

\* tableau détaillé des prédicteurs à la page 25 (tableau 12). [↓] = combinaison qui intègre tous les prédicteurs des lignes précédentes



**Figure 12** Distributions groupées par classe de recouvrement (%) et par prédicteur spectral du coefficient de variation calculé sur une série temporelle de 24 images d'une période 15 jours chacune.

### 3.5.b. Métriques saisonnières

L'approche par les différences normalisées entre saisons a pour objectif de mettre en évidence des variations phénologiques entre deux périodes spécifiques, qui seraient déterminantes pour caractériser les attributs écosystémiques étudiés. Les quatre saisons de végétation sont utilisées, ce qui donne six combinaisons possibles. La valeur absolue d'une différence normalisée est calculée pour chaque prédicteur spectral, telle que :

$$\Delta = \frac{|B_{saison\ n} - B_{saison\ m}|}{|B_{saison\ n} + B_{saison\ m}|}.$$

Les prédicteurs spectraux pour lesquels les métriques phénologiques saisonnières sont les bandes brutes et quatre indices de végétation. Les modèles sont comparés à un modèle qui n'intègre que les bandes brutes en saison de repos de la végétation (tableau 11).

**Tableau 11** : Description des prédicteurs intégrés dans les modèles utilisés pour caractériser l'influence des variations phénologiques saisonnières sur la représentation des attributs écosystémiques. Les modèles intègrent les bandes brutes en saison de repos et les métriques phénologiques (différences normalisées absolues) calculées entre deux saisons sur les bandes brutes et les indices de végétation.

Modèles	Saisons utilisées	Cible des métriques phénologique (X)	Prédicteurs* Bandes Sentinel-2	Métriques phénologiques	Nb
Repos	Repos				7
[ApVeg-Veg]	Post-végétation	Bandes brutes (B02, B03, B04, B05, B8A, B11, B12)  Indices de végétation (GRVI, GVM, NDGI, SWIR G B)	B02, B03, B04, B05, B8A, B11, B12 (saison de repos)	$\Delta = \frac{ X_{saison\ n} - X_{saison\ m} }{ X_{saison\ n} + X_{saison\ m} }$	18
[AvVeg-ApVeg]	Pré-végétation				
	Post-végétation				
[AvVeg-Veg]	Pré-végétation				
	Végétation				
[Repos-ApVeg]	Repos				
	Post-végétation				
[Repos-Veg]	Repos				
	Végétation				
[Repos-AvVeg]	Repos				
	Pré-végétation				

\* tableau détaillé des prédicteurs à la page 25 (tableau 12)

**Tableau 12** : résumé de l'ensemble des prédicteurs utilisés, leurs dénominations, acronymes et définitions

Type	Bandes, indices ou paramètres		Définition
Variables spectrales brutes	Bandes	B02	Bleu, 490 nm
		B03	Vert, 560 nm
		B04	Rouge, 665 nm
		B05	Red Edge, 704 nm
		B8A	Proche infrarouge (NIR), 865 nm
		B11	Infrarouge à courte longueur d'onde (SWIR1), 1610 nm
		B12	Infrarouge à courte longueur d'onde (SWIR2), 2190 nm
Indices de végétation	Visible	B	$B02 / (B02 + B03 + B04)$
		GRVI	$(B03 - B04) / (B03 + B04)$
	+ Proche infrarouge	NDVI	$(B8A - B04) / (B8A + B04)$
		CCCI	$[(B8A - B05) / (B8A + B05)] / [(B8A - B04) / (B8A + B04)]$
		NDGI	$[(0,62 * B03 + 0,38 * B8A) - B04] / [(0,62 * B03 + 0,38 * B8A) + B04]$
		EVI	$2,5 * [(B8A - B04) / (B8A + 6 * B04 - 7,5 * B02 + 1)]$
		MSAVI2	$[(2 * B8A + 1) - \sqrt{(2 * B8A + 1)^2 - 8 * (B8A - B04)}] / 2$
	+ Infrarouge à courte longueur d'onde	NDSVI	$(B11 - B04) / (B11 + B04)$
		GVMI	$(0,1 * B8A - 0,2 * B12) / (0,1 * B8A + 0,2 * B12)$
		SWIR G B (personnel)	$(B12 - B02 - B03) / (B12 + B02 + B03)$
	Spectre complet	Tasseled Cap Transformation Brithness	$- 0,3599 * B2 - 0,3533 * B3 - 0,4734 * B4 + 0,6633 * B8A + 0,0087 * B11 - 0,2856 * B12$
Métriques de phénologie saisonnière	$\Delta$ : Différences normalisées	B02, B03, B04, B05, B8A, B11, B12	Métriques calculées entre deux saisons (repos, végétation, transitions) sur la réflectance médiane sur la période 2017 – 2021.
		GRVI, GVMI, NDGI, SWIR G B	
Métriques de phénologie annuelle	Minimum	GRVI, GVMI, NDGI, SWIR G B	Métriques calculées avec une série temporelle de 24 images sur l'année composée de la réflectance médiane sur la période 2017 – 2021.
	Maximum		
	Produit		
	Amplitude		
	Écart interquartile		
	Écart interpercentile [95 – 5]		
	Écart type		
	Somme		



## 4. Modélisation Random Forest et évaluation

### 4.1. Généralités

L'algorithme Random Forest Regression (RFR) (Breiman, 2001), implémenté dans le package R *RandomForest* et recouvert par le package *Caret*, a été utilisé pour modéliser les attributs écosystémiques sur base des variables prédictives décrites ci-avant. Random Forest est un algorithme d'apprentissage automatique qui consiste en un ensemble d'arbres décisionnels. Chaque arbre est construit sur un sous-ensemble aléatoire des données d'apprentissage et, à chaque nœud de l'arbre, un sous-ensemble aléatoire des prédicteurs est sélectionné afin d'identifier, sur base de l'indice de Gini, le discriminateur qui minimise la variance résiduelle. La prédiction finale, dans le cas de la régression, est obtenue par la moyenne des prédictions fournies par l'ensemble des arbres décisionnels (Breiman, 2001).

L'algorithme Random Forest a principalement été choisi pour sa robustesse, notamment lors du traitement de relations non linéaires, et pour sa capacité à intégrer des variables explicatives nombreuses et corrélées — une situation classique en télédétection (Breiman, 2001; Chan et al., 2012). Une réduction de la dimensionnalité et la suppression de variables corrélées peuvent néanmoins améliorer la qualité de modélisation et son interprétabilité (Karlson et al., 2015). Cet algorithme est caractérisé par une forte sensibilité à la taille des données d'entraînement et à leur distribution spatiale, notamment en cas d'autocorrélation spatiale (Millard et al., 2015). Enfin, plusieurs études ont comparé Random Forest avec d'autres algorithmes utilisés dans le domaine de la télédétection, notamment des approches non paramétriques (Forkuor et al., 2018; Thanh Noi et al., 2018), et ont montré que l'algorithme Random Forest était le plus performant lors de l'estimation du recouvrement de la végétation en zone aride (Eskandari et al., 2020; Nagelkirk et al., 2020) ou de la biomasse forestière (Powell et al., 2010; Fassnacht et al., 2014).

Une analyse de sensibilité a été réalisée pour chacune des variables de réponse afin d'identifier les paramètres qui minimisent la complexité des modèles tout en conservant leur qualité et stabilité (Annexe 1). La taille des nœuds terminaux (*nodesize*) et le nombre de variables sélectionnées aléatoirement à chaque nœud (*mtry*) influencent peu la qualité des modèles réalisés sur une image composite Sentinel 2 en saison de repos de végétation, quelle que soit la variable de réponse. Les valeurs par défaut ont été conservées pour ces paramètres, à savoir minimum 5 échantillons par nœud terminal, et la racine carrée du nombre de prédicteurs sélectionnés aléatoirement à chaque nœud. Le nombre d'arbres décisionnels par modèle est fixé à 50, quantité qui permet le meilleur compromis entre stabilité des résultats et complexité des modèles. Les données ont été standardisées préalablement à la construction des modèles.

### 4.2. Validation croisée spatialement indépendante

L'évaluation des modèles a été réalisée par validation croisée (*cross-validation*). Cette évaluation a été répétée vingt fois par modèle avec une configuration spatiale unique. Les données ont été séparées en *folds* (groupes) spatialement indépendants afin d'éviter une surestimation de la qualité des résultats due à l'autocorrélation spatiale (Roberts et al., 2017). Le package *BlockCV*, implémenté sur R, a été utilisé pour générer les *folds* (Valavi et al., 2019). Un examen préliminaire de l'autocorrélation spatiale des prédicteurs a été réalisé afin de fournir une estimation quantitative de l'espacement minimum entre les parcelles. Cette distance a été estimée à 1600 m.

Les données de référence récoltées sur le terrain — 41 parcelles — sont regroupées dans l'espace et forment des grappes d'échantillonnage dispersées à travers le paysage. Elles ont été séparées en cinq *folds* spatialement indépendants générés avec une distance minimale de 2000 m. Les données de référence acquises par photo-interprétation — 505 parcelles — sont disposées systématiquement dans l'espace. Elles ont été séparées en dix *folds* avec une distance minimale de 10 000 m. L'étape de création des *folds* et d'évaluation des modèles a été répétée vingt fois pour permettre une évaluation plus robuste de la qualité des modèles. L'évaluation finale d'un modèle est la valeur moyenne des vingt évaluations croisées. Cela représente, par métrique d'évaluation, 200 estimations pour le jeu de données de photo-interprétation et 100 estimations pour le jeu de données de terrain. Les modèles finaux ont ensuite été construits sur l'ensemble du jeu d'entraînement.

### 4.3. Métriques d'évaluation de qualité et d'interprétation des modèles

L'évaluation de la qualité des modèles est réalisée sur base de l'écart quadratique moyen (RMSE, *root-mean-square error*). Ce paramètre mesure les différences entre les valeurs prédites par un modèle ( $P_i$ ) et la valeur réellement observée ( $O_i$ ), tel que :

$$RMSE = \sqrt{\frac{\sum_1^n (O_i - P_i)^2}{n}}$$

La RMSE dépend de l'échelle de données et est donc exprimée dans l'unité de la variable de réponse. Cette métrique peut être utilisée pour comparer la qualité de modèles construits avec différents prédicteurs pour une même variable de réponse. Au plus l'écart entre les valeurs prédites et observées est faible, au plus la RMSE tend vers zéro, valeur qui équivaut à une absence de variance résiduelle.

Une seconde métrique, l'erreur quadratique relative (RSE, *relative squared error*), est utilisée pour comparer la qualité de modèles construits pour des variables de réponse exprimées en unités différentes. Ce paramètre mesure les différences entre les valeurs prédites ( $P_i$ ) et observées ( $O_i$ ) d'un modèle par comparaison à un prédicteur simple, la moyenne des valeurs observées ( $\bar{O}$ ), tel que :

$$RSE = \sqrt{\frac{\sum_1^n (O_i - P_i)^2}{\sum_1^n (O_i - \bar{O})^2}}$$

La RSE est normalisée par l'erreur quadratique totale du prédicteur simple, la moyenne, et ne dépend donc pas de l'échelle de données. La RSE est comprise entre 0 et  $+\infty$ . Une valeur inférieure à 1 indique que le modèle apporte une meilleure estimation de la variable de réponse que la moyenne des valeurs observées. Une RSE nulle indique une absence de variance résiduelle. Une valeur supérieure à 1 indique que le modèle est un moins bon estimateur que la moyenne des valeurs observées.

Deux autres métriques sont utilisées pour la description et l'interprétation spécifique à un modèle : le biais moyen et l'erreur absolue moyenne (MAE, *mean absolute error*).

Le biais est la quantité moyenne par laquelle les valeurs prédites sont supérieures ou inférieures aux valeurs observées, tel que :

$$\text{Biais moyen} = \frac{\sum_1^n (O_i - P_i)}{n}$$

Un modèle non biaisé aura une valeur proche de zéro. Le signe du biais indique le sens vers lequel tend systématiquement l'erreur : un signe négatif pour une sous-estimation moyenne, un signe positif pour une

surestimation moyenne. La valeur du biais moyen indique la quantité moyenne de déviation systématique et est donc exprimée dans l'unité de la variable de réponse.

La MAE est représentée par la moyenne des différences absolues entre les valeurs prédites ( $P_i$ ) et observées ( $O_i$ ), telle que :

$$MAE = \frac{\sum_1^n |O_i - P_i|}{n}$$

La MAE permet une interprétation des résultats que la RMSE, car elle quantifie une erreur absolue du modèle et non une erreur quadratique. Cependant cette métrique n'a pas été utilisée pour l'évaluation des modèles, car la RMSE permet de favoriser les modèles qui minimisent les erreurs importantes lorsque la distribution des erreurs n'est pas gaussienne. La MAE est inférieure à la RMSE, sauf lorsque l'écart des erreurs est constant ou suit une distribution strictement gaussienne.

#### 4.4. Analyses statistiques des différences moyennes entre modèles

Une analyse de la variance (ANOVA) de la moyenne des RMSE associées à chaque modèle a été réalisée pour évaluer la significativité statistique des différences entre les modèles calibrés pour une même variable de réponse. La RMSE moyenne a été calculée sur les 20 répétitions de la validation croisée après sélection de variables. Les conditions d'application de l'ANOVA ont été testées au préalable. La normalité de la variable dépendante (RMSE) a été déterminée avec un diagramme Quantile-Quantile (QQplot), tandis que la variabilité au sein des groupes a été vérifiée avec un test de Levene. Un test des étendues de Tukey a été réalisé afin d'évaluer le niveau de différence deux à deux.

#### 4.5. Évaluation indépendante des modèles de photo-interprétation

Une évaluation indépendante a été réalisée pour tous les modèles construits sur base des données acquises par photo-interprétation. Cette évaluation compare les résultats obtenus par les modèles de photo-interprétation avec les valeurs observées sur les placettes de terrain.

#### 4.6. Évaluation de l'importance et sélection des variables

Chaque modèle a été construit deux fois : une première fois avec l'ensemble des variables prédictives et une deuxième fois avec un sous-ensemble de variables identifiées comme significativement influentes par l'algorithme Boruta, implémenté sur R par le package *Boruta* (Kursa et al., 2010). Cet algorithme permet, d'une part, de fournir une évaluation de l'importance des variables et, d'autre part, d'éliminer les variables les moins pertinentes — le bruit. La méthode d'évaluation de l'importance des variables de l'algorithme Boruta se distingue de celle calculée par défaut par l'algorithme Random Forest, qui est basée sur la diminution moyenne des impuretés (importance de Gini). Le test Boruta est construit par comparaisons répétées de l'importance des variables initiales avec le score obtenu par des *shadow features*. Une *shadow feature* est construite en mélangeant aléatoirement les valeurs d'une variable explicative, qui est ensuite ajoutée au jeu de données initiales. Il y a autant de *shadow features* qu'il y a de variables explicatives, de sorte que la taille du jeu de données est doublée. L'évaluation repose sur l'estimation du score  $Z$ , qui illustre la perte de précision causée par la permutation aléatoire des prédicteurs, normalisée par la perte moyenne et l'écart-type calculé entre les arbres décisionnels qui utilisent cette variable. Lors de la construction d'un modèle Random Forest, un point (*hit*) est attribué à une variable explicative lorsque son score  $Z$  est plus important que le meilleur score  $Z$  obtenu par une *shadow feature*. Cette situation signifie que la variable

apporte plus d'informations qu'une série de valeurs aléatoires. L'étape est répétée un grand nombre de fois afin de générer une distribution binomiale des *hits*. L'importance d'une variable est dès lors testée statistiquement avec une p-valeur seuil de 0,01. L'algorithme s'arrête lorsque toutes les variables ont été rejetées ou acceptées, ou lorsque le nombre d'itérations est terminé — les variables restantes sont alors identifiées comme intermédiaires. Dans le cadre de cette étude, les valeurs par défaut de l'algorithme ont été conservées et les variables intermédiaires ont été considérées — en suivant une logique parcimonieuse et non conservatrice — comme rejetées. Les variables significativement importantes sont utilisées pour construire les modèles finaux, tandis que le score d'importance obtenu est utilisé comme indicateur de leur utilité.

## 5. Modèle intégratif et comparaison avec les cartes globales

Un modèle qui intègre l'ensemble des prédicteurs pertinents identifiés durant cette étude a été réalisé pour chaque variable de réponse. Par catégorie de facteurs étudiés, les prédicteurs qui apportent une diminution de la variance résiduelle ont été identifiés sur base de la RMSE associée aux modèles et de l'importance des variables évaluée par l'algorithme Boruta. Les modèles finaux ont servi à produire cinq cartes, respectivement pour le recouvrement ligneux, la biomasse, la hauteur de canopée et les indicateurs de la structure verticale des écosystèmes. Ces modèles ont été comparés avec la série de modèles construits uniquement sur les bandes du visible et du proche infrarouge, ainsi qu'avec ceux qui utilisent les bandes brutes de Sentinel 2A.

Les cartes issues des prédictions du modèle intégratif calibré sur les données de photo-interprétation ont été comparées aux produits globaux de couverture de canopée de Hansen et al. (2013) et celui de Sexton et al. (2013), au produit de hauteur de canopée produit par Potapov et al. (2021), et celui de biomasse de Spawn et Gibbs (2020). Les produits globaux ont été comparés avec les données de photo-interprétation car les modèles calibrés sur celles-ci minimisent davantage la variance résiduelle que ne le font les modèles calibrés sur les données de terrain. En outre, les données de photo-interprétation sont plus nombreuses et mieux réparties sur la zone d'étude. Pour les produits globaux de hauteur et de biomasse, une comparaison additionnelle à la lumière des données de terrain a été réalisée pour évaluer la similarité des tendances observées. Les cartes finales ont été comparées aux cartes globales disponibles pour les variables de réponse. Le recouvrement modélisé dans cette étude a été comparé aux cartes de couverture forestière globale proposées par Hansen et al. (2013) et par Sexton et al. (2013). Les estimations de la hauteur de canopée ont été comparées avec la carte de hauteur de canopée mondiale publiée par Lang et al. (2022), tandis que les estimations de biomasse ont été comparées avec la carte de biomasse globale pour l'année 2010 publiée par Spawn et al. (2020). Les différences ont été mises en évidence en faisant la différence entre le recouvrement estimé dans cette étude et les produits globaux. Les cartes issues des prédictions du modèle intégratif et calibrées sur les données de photo-interprétations ont été comparées aux produits globaux de couverture de canopée produits par Hansen et al. (2013) et par Sexton et al. (2013), de hauteur produit par Potapov et al. (2021), et de biomasse produit par Spawn et Gibbs (2020). Les produits globaux ont été comparés avec les données de photo-interprétation car les modèles calibrés sur celles-ci minimisent plus la variance résiduelle que les modèles calibrés avec les données de terrain. En outre, les données de photo-interprétation sont plus nombreuses et mieux réparties sur la zone d'étude. Pour les produits globaux de hauteur et de biomasse, une comparaison additionnelle à la lumière des données de terrain a été réalisée pour évaluer la similarité des tendances observées.

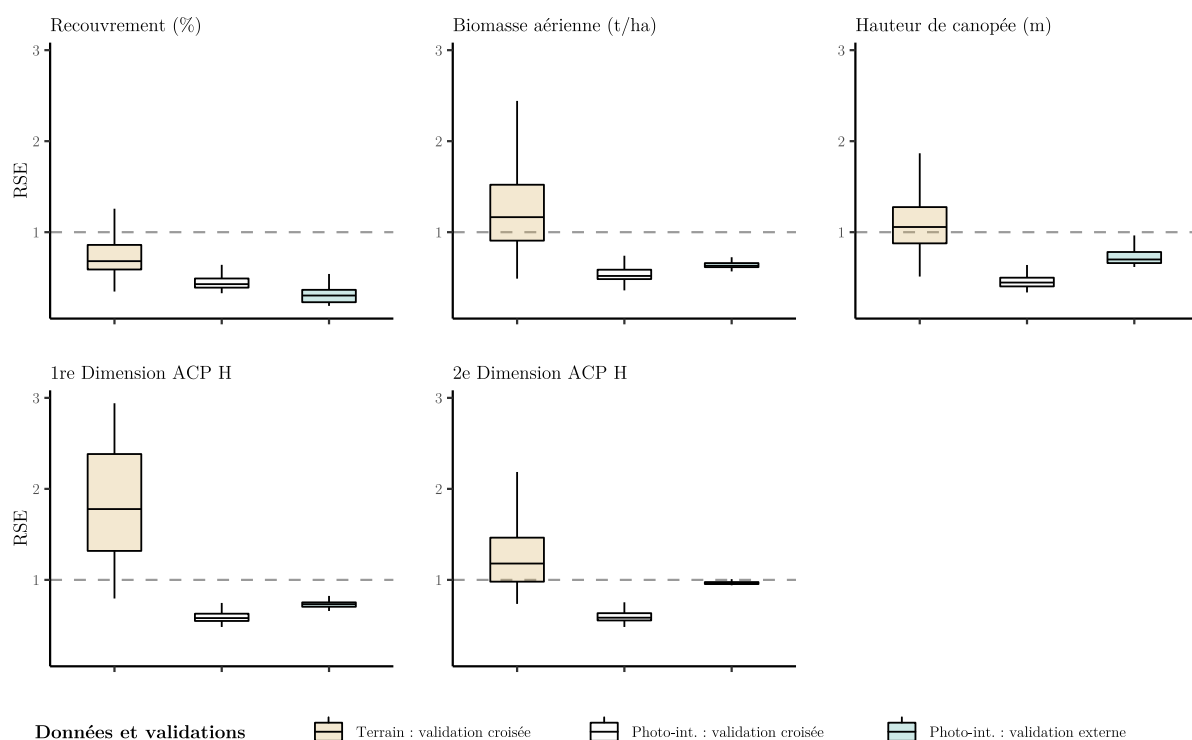
La comparaison du recouvrement a été effectuée en premier lieu sur la base d'une couverture de canopée supérieure à 10 % en tant compte de plusieurs niveaux de hauteurs. Ce seuil est utilisé notamment dans la définition de la forêt de la FAO. La comparaison a ensuite été réalisée en différences absolues de prédictions. Les produits globaux de Sexton et de Hansen renseignent la couverture de canopée des arbres de dimensions supérieures à 5 m. Le modèle intégratif de hauteur de canopée a été utilisé pour éliminer les zones de dimensions inférieures à 4 m et à 5 m afin de comparer le recouvrement pour un même seuil de hauteur. La hauteur de 4 m a été utilisée pour tenir compte de l'erreur absolue moyenne associée au modèle intégratif final développé dans le cadre de cette étude. La comparaison de la biomasse et de la hauteur a été effectuée uniquement sur base d'une différence absolue entre les produits globaux et les données de référence.

# Résultat

## 1. Vue globale des résultats et évaluation du potentiel de la photo-interprétation

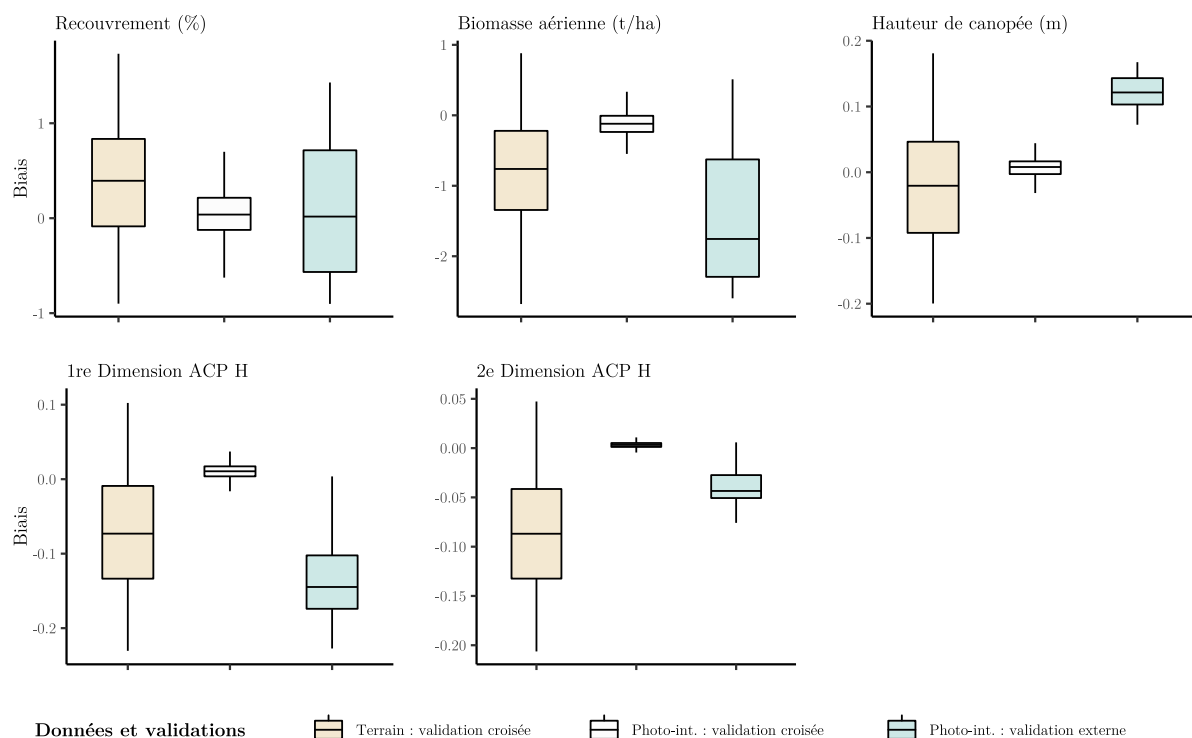
L'évaluation de la qualité de l'ensemble des modèles construits après sélection de variables avec l'algorithme Boruta est synthétisée dans les figures 13, 14 et 15, où sont représentés respectivement la RSE, la MAE et le biais. Les résultats sont groupés par type de validation et par jeu de données. La RSE est utilisée pour comparer les variables de réponse exprimées en unités différentes, tandis que la MAE et le biais permettent d'évaluer la qualité des modèles pour une réponse spécifique.

Les résultats obtenus lors de l'évaluation indépendante des modèles de photo-interprétation indiquent une RSE systématiquement plus faible que les modèles calibrés à partir des données de terrain et évalués en validation croisée. Les modèles qui utilisent les données de photo-interprétation ont tendance à être surévalués en validation croisée, mis à part pour l'estimation du recouvrement (figure 13). En outre, tous les modèles de photo-interprétation ont une RSE médiane inférieure à 1, tandis que ceux construits avec les données de terrain — à l'exception du recouvrement — ont une RSE médiane supérieure à 1 et sont donc, en moyenne, de moins bons estimateurs que la moyenne des données de terrain. De plus, les modèles de photo-interprétation ont une MAE plus faible que les modèles de terrain, à l'exception de la 2<sup>e</sup> dimension de l'ACP. Enfin, l'utilisation des données de photo-interprétation permet d'obtenir des modèles dont la variance est globalement plus faible.



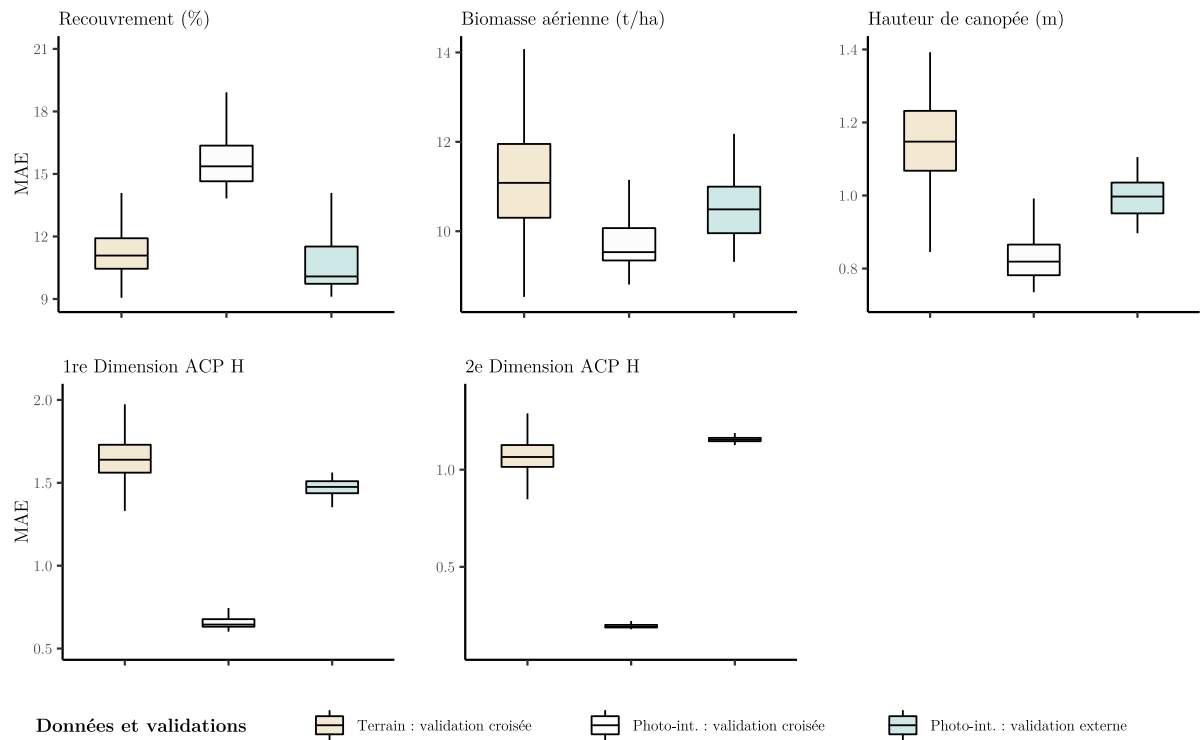
**Figure 13** : comparaison de la performance des modèles, sur base de la *relative squared error* (RSE, sans unité), par jeu de données (photo-interprétation ou terrain) et modalité d'évaluation (validation croisée ou externe), ainsi que par variable -réponse modélisée. Une RSE inférieure à 1 (ligne pointillée horizontale) indique que les modèles sont de meilleurs estimateurs que la moyenne des données de calibration. Un boxplot inclut l'évaluation de l'ensemble des modèles et groupes de prédicteurs après sélection de variables.

Les modèles qui prédisent le recouvrement sont peu biaisés (figure 14) et ont une MAE médiane de l'ordre de 10 à 11 %, pour une gamme de valeur interquartile inférieure à 2 % (figure 15). Les différences selon les données utilisées sont peu marquées. En revanche, le biais des modèles qui estiment les autres variables de réponse est plus important et diffère en fonction du jeu de données. Un biais négatif de l'ordre de 2 t/ha est observé lors de l'estimation de la biomasse avec les modèles de photo-interprétation, tandis que les modèles de terrain ont un biais négatif médian inférieur à 1 t/ha, pour une MAE médiane de l'ordre de 10 à 12 t/ha. Lors de l'estimation de la hauteur de la canopée, les modèles de terrain montrent également une absence de biais médian, tandis que les modèles de photo-interprétation surestiment en moyenne de 0,2 m, pour une MAE médiane comprise entre 0,9 et 1,1 m. Quant aux dimensions de l'ACP, les tendances sont inverses selon les données utilisées, bien que le biais soit généralement faible avec une MAE médiane de l'ordre de 1,5 pour la 1<sup>re</sup> dimension et de 1,0 pour la 2<sup>e</sup> dimension.



**Figure 14** : comparaison du biais de modélisation, par jeu de données (photo-interprétation ou terrain) et modalité d'évaluation (validation croisée ou externe), ainsi que par variable variables-réponses modélisées. Un boxplot inclut l'évaluation de l'ensemble des modèles et groupes de prédictors après sélection de variables.

Les modèles calibrés sur les données de photo-interprétation permettent une représentation avec une amplitude plus importante de valeurs prédites du recouvrement, de la biomasse et de la hauteur. L'effet inverse est observé pour les dimensions de l'ACP, avec une variance plus importante sur les modèles de terrain. Les modèles de photo-interprétation induisent une corrélation plus importante entre le recouvrement et les autres variables de réponse (tableau 13). Une analyse de corrélation sur un échantillon de 3500 points répartis aléatoirement sur les cartes prédites met en évidence un coefficient de corrélation supérieur à 88 % entre les données des modèles de photo-interprétation, tandis que ce coefficient est inférieur à 76 % pour les modèles calibrés sur les données de terrain. Cette importante propagation de la corrélation est également visible à l'échelle de chaque modèle, où le profil de tendance des variables de réponse est similaire (figure 16).

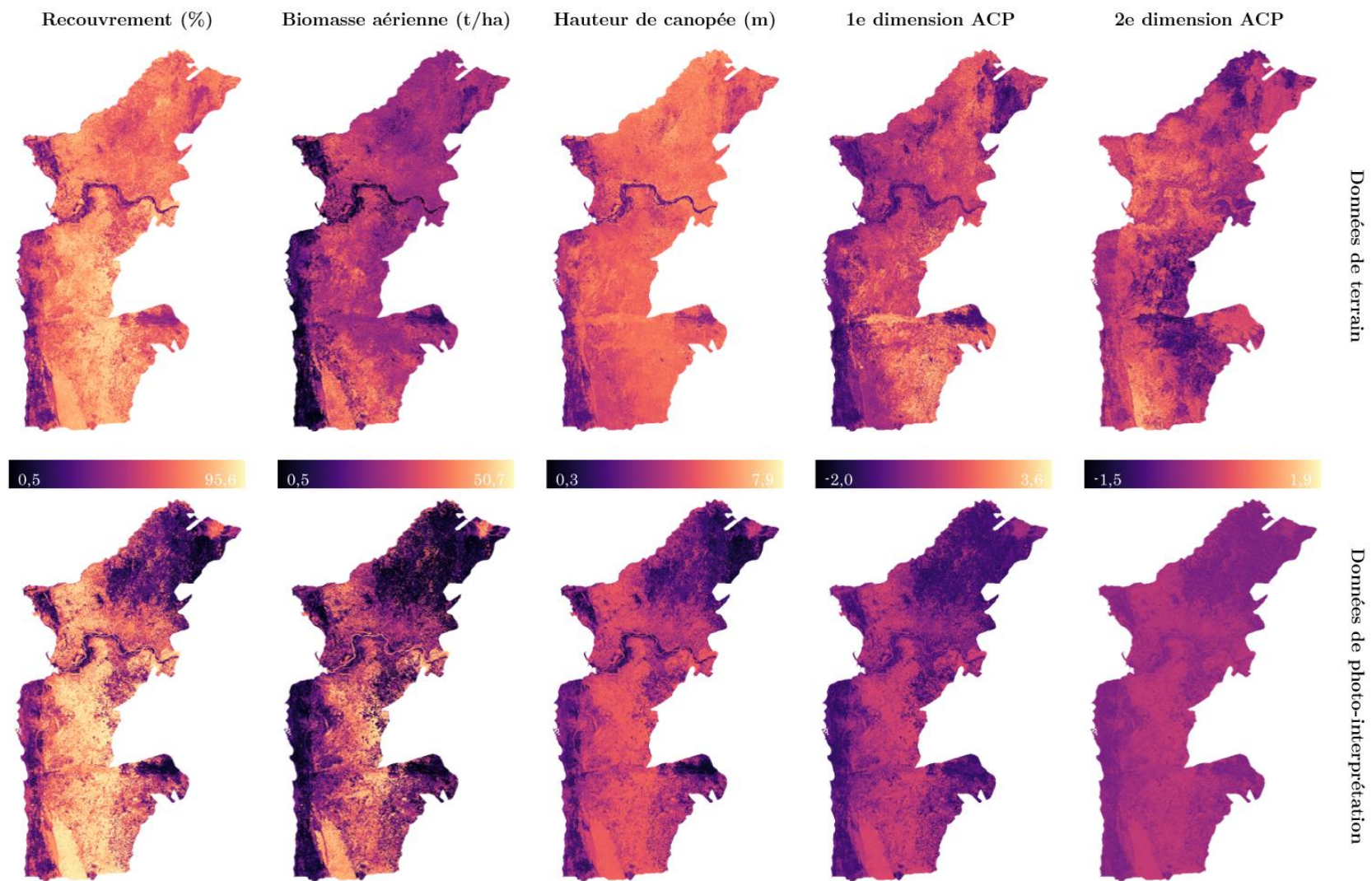


**Figure 15** : comparaison des erreurs absolues moyennes (MAE) associées à chaque modèle, par jeu de données (photo-interprétation ou terrain) et modalité d'évaluation (validation croisée ou externe), ainsi que par variable de réponse modélisée. Un boxplot inclut l'évaluation de l'ensemble des modèles et groupes de prédicteurs après sélection de variables.

**Tableau 13** : coefficients de corrélation calculés entre les valeurs de couverture de la canopée et des autres attributs écosystémiques prédites avec les modèles intégratifs calibrés avec les deux jeux de données pour les deux jeux de données (biomasse aérienne, hauteur de canopée, dimensions de l'analyse en composantes principales sur la distribution des classes de hauteur). Les valeurs ont été extraites par un échantillonnage de 3500 points répartis aléatoirement dans la zone d'étude.

Couverture de la canopée	Biomasse	H. canopée	Dim 1 ACP H	Dim 2 ACP H
Modèles terrain	0,76	0,69	0,63	-0,19
Modèles photo-interprétation	0,89	0,98	0,94	0,90





**Figure 16** : prédictions finales des modèles intégratifs calibrés sur les données de terrain (en haut) et sur les données de photo-interprétation (en bas) pour les cinq variables de réponse (de gauche à droite : recouvrement, biomasse aérienne, hauteur de canopée, 1<sup>re</sup> dimension et 2<sup>e</sup> dimension de l'analyse en composantes principales réalisée sur la distribution de la hauteur). L'échelle des données est identique entre modèles calibrés pour une même variable de réponse afin de mettre en évidence les différences de prédiction entre les deux jeux de données.

## 2. Facteurs influençant la cartographie

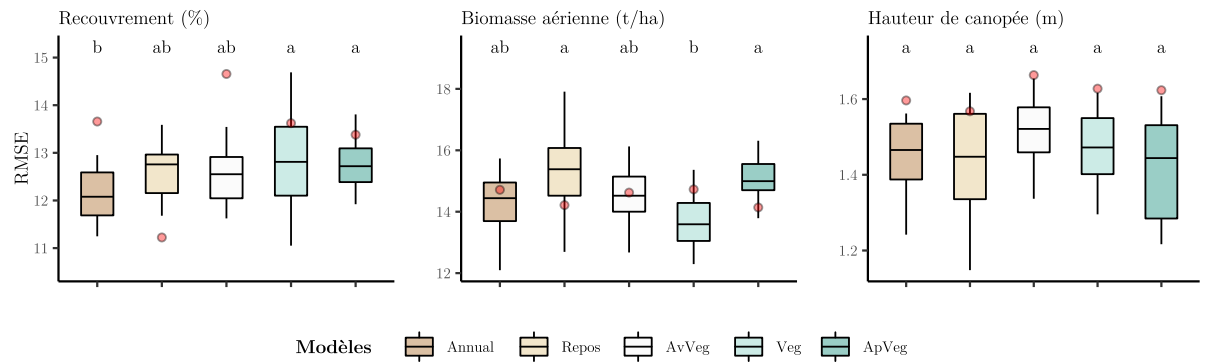
La caractérisation de l'influence des facteurs à considérer pour la représentation des attributs structuraux des écosystèmes est présentée sous la forme de boxplots pour la couverture de la canopée, la biomasse aérienne et la hauteur de canopée. Les modèles calibrés pour représenter la structure verticale de l'écosystème (dimensions ACP) ne sont pas présentés dans cette partie, car la propagation d'erreur est trop importante pour une étude fine de l'influence des facteurs de modélisation. Les résultats présentés proviennent des modèles calibrés avec les données de terrain, car ils permettent d'étudier l'influence des facteurs de modélisation sur l'estimation de la biomasse et de la hauteur de canopée avec une réponse moins corrélée au recouvrement de la végétation (tableau 14). Chaque boxplot intègre l'évaluation d'un modèle, construit à partir d'une combinaison de prédicteurs, dont l'évaluation croisée a été répétée 20 fois. Les différences entre les moyennes analysées par une ANOVA à un facteur sont présentées sous la forme de lettre au-dessus des boxplots. Le résultat de l'évaluation indépendante des modèles de photo-interprétation est également représenté sur les boxplots par un point rouge par modèle, afin d'illustrer les différences entre les jeux de données. Néanmoins, ces résultats ne seront pas utilisés directement pour évaluer l'influence des prédicteurs, mis à part pour la couverture de la canopée qui n'est pas un attribut extrapolé. En outre, l'analyse est complétée par une étude de l'importance des prédicteurs, évaluée avec l'algorithme Boruta, pour les groupes de modèles entre lesquels une différence significative a été mise en évidence. Les modèles calibrés pour représenter la structure verticale de l'écosystème (dimensions ACP) ne sont pas présentés dans cette partie car la propagation d'erreur est trop importante pour une étude fine de l'influence des facteurs de modélisation. Les résultats présentés proviennent des modèles calibrés avec les données de terrain, car ils permettent d'étudier l'influence des facteurs de modélisation sur l'estimation de la biomasse et de la hauteur de canopée avec une réponse moins corrélée au recouvrement de la végétation. Un tableau en annexe détaille l'ensemble des modèles pour chaque facteur et variable de réponse (Annexe).

### 2.1. Saisonnalité

La comparaison de la performance des modèles construits pour étudier l'influence de la saisonnalité est illustrée à la figure 17. Le modèle qui comprend l'ensemble des prédicteurs saisonniers se démarque en minimisant le plus la variance résiduelle lors de l'estimation du recouvrement ligneux à partir des données de terrain, mais aucune différence significative n'est identifiée pour la saison de repos de la végétation et pour la saison qui précède le pic de l'activité végétale. C'est néanmoins la saison de repos de la végétation qui se distingue fortement lors de la validation externe des modèles de photo-interprétation, avec une RMSE inférieure à 12 %. Pour l'estimation de la biomasse, la saison de végétation et, dans une moindre mesure, la période de transition pré-végétation obtiennent les meilleures performances, avec une RMSE médiane inférieure à 14 t/ha. Lors de l'estimation de la hauteur de canopée, toutes les saisons obtiennent des résultats équivalents, mais la saison de pré-végétation se distingue avec des modèles plus stables entre les répétitions. Les performances du meilleur modèle sont présentées dans le tableau 14.

Une analyse de l'importance des prédicteurs en fonction de la saison de végétation a été réalisée pour la couverture de la canopée et la biomasse aérienne (figure 18). Une faible variation globale de l'importance des prédicteurs est constatée entre les saisons de végétation. Pour la modélisation du recouvrement, l'importance des bandes brutes diminue au profit des indices de végétation durant la saison de végétation, mais la réflectance brute joue tout de même un rôle plus important que les indices de végétation.

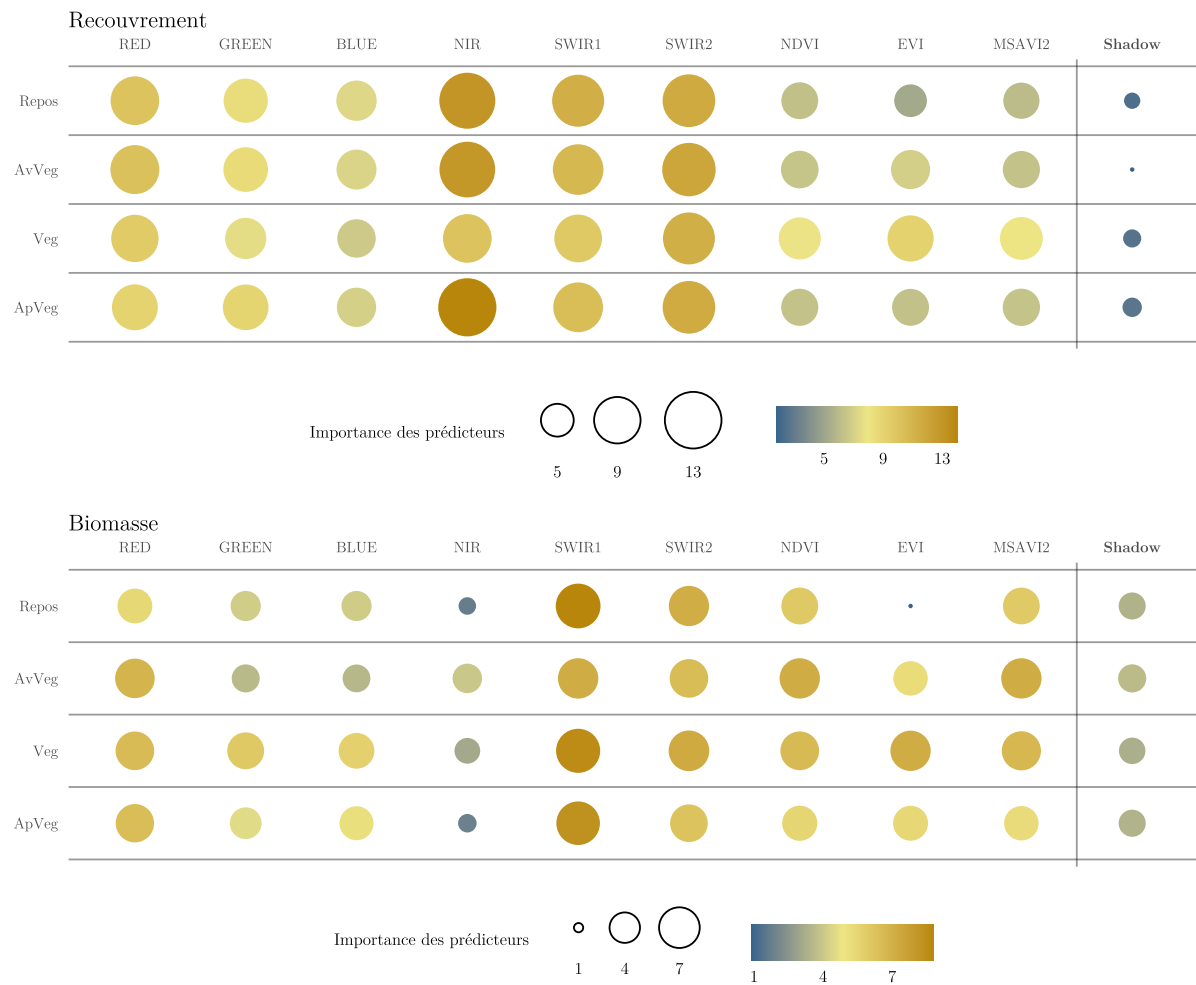
L'importance des variables lors de la modélisation de la biomasse varie plus fortement selon la saison considérée. La réflectance dans le proche infrarouge, le vert, le bleu, ainsi que les indices de végétation NDVI et EVI sont jugés plus importants durant la saison de végétation que durant les autres saisons. En outre, l'EVI et le proche infrarouge n'apportent pas d'information significative durant la saison de repos.



**Figure 17** : comparaison de l'écart quadratique moyen (RMSE) associé à chaque modèle qui vise à caractériser l'influence de la saisonnalité. Les modèles sont regroupés par variable de réponse. Un boxplot inclut les résultats de l'évaluation par validation croisée répétée 20 fois pour les modèles calibrés avec les données de terrain. Un point rouge qui représente l'évaluation externe des modèles calibrés avec les données de photo-interprétation est associé à chaque boxplot. Les lettres identiques représentent les modèles sans différence significative de performance identifiée par l'analyse de la variance (p-valeur = 0,05).

**Tableau 14** : Résultats détaillés du meilleur modèle de recouvrement pour l'étude de l'influence de la saisonnalité (exemple d'un boxplot). Les résultats de l'ensemble des modèles sont présentés en annexe.

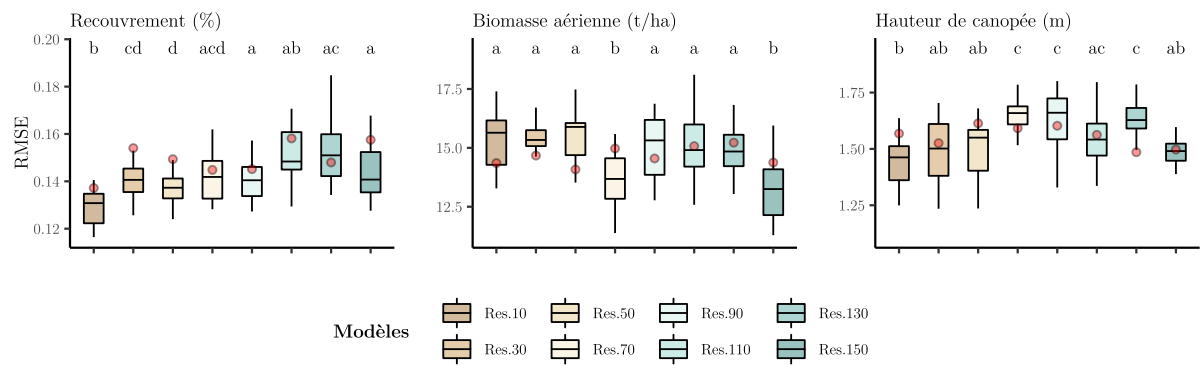
Variable de réponse : couverture de la canopée (%)						
Facteur étudié : saisonnalité						
Modèle : Repos (saison de repos)						
Prédicteurs : B02, B03, B04, B8A, B11, B12, NDVI, EVI, MSAVI2						
	Données de terrain → Validation croisée (20 répétitions)			Données de photo-interprétation → Validation externe		
	RMSE	MAE	Biais	RMSE	MAE	Biais
Moyenne	12,7	10,3	0,0	11,2	11,2	1,4
Médiane	12,7	10,3	0,0	-	-	-
Écart-type	0,6	0,7	0,7	-	-	-
Min	11,5	9,4	-0,9	-	-	-
Max	13,5	11,3	1,5	-	-	-



**Figure 18** : résultats de l'analyse de l'importance des prédicteurs par l'algorithme Boruta, regroupé par variable de réponse et par modèles utilisés pour caractériser l'influence de la saisonnalité. Les modèles sont placés en ligne et les prédicteurs en colonne. L'importance d'un prédicteur est à interpréter à la lumière du score maximum obtenu par les *shadow features* associées à ce modèle (Shadow). La taille et la couleur du cercle sont proportionnelles au score associé à un prédicteur. Les scores sont comparables entre les modèles d'une même variable de réponse, mais pas entre les variables de réponse.

2.2. Résolution spatiale

Les performances obtenues lors de la dégradation de la résolution, de 10 à 150 m par pas de 20 m, sont illustrés à la figure 19. Une perte de qualité, à l’exception de la modélisation de la biomasse, est constatée lorsque la résolution est dégradée. La résolution initiale de 10 m est la plus importante pour le recouvrement et pour la hauteur de canopée, bien que des résolutions plus grossières montrent des résultats comparables. Les résolutions de 70 m et de 150 m se démarquent des autres résolutions lors de l’estimation de la biomasse. Les performances du meilleur modèle sont présentées dans le tableau 15.



**Figure 19 :** comparaison de l’écart quadratique moyen (RMSE) associé à chaque modèle qui vise à caractériser l’influence de la résolution spatiale. Les modèles sont regroupés par variable de réponse. Un boxplot inclut les résultats de l’évaluation par validation croisée répétée 20 fois pour les modèles calibrés avec les données de terrain. Un point rouge qui représente l’évaluation externe des modèles calibrés avec les données de photo-interprétation est associé à chaque boxplot. Les lettres identiques représentent les modèles sans différence significative de performance identifiée par l’analyse de la variance (p-valeur = 0,05).

**Tableau 15 :** Résultats détaillés du meilleur modèle de recouvrement pour l’étude de l’influence de la résolution spatiale (exemple d’un boxplot). Les résultats de l’ensemble des modèles sont présentés en annexe.

Variable de réponse : couverture de la canopée (%)						
Facteur étudié : résolution spatiale						
Modèle : 10 m (résolution spatiale de 10 mètres)						
Prédicteurs : B02, B03, B04, B8A, B11, B12, NDVI, EVI, MSAVI2						
	Données de terrain → Validation croisée (20 répétitions)			Données de photo-interprétation → Validation externe		
	RMSE	MAE	Biais	RMSE	MAE	Biais
Moyenne	12,7	10,3	0,0	13,7	10,0	1,0
Médiane	13,1	10,4	-0,2	-	-	-
Écart-type	0,8	0,7	0,6	-	-	-
Min	11,6	9,2	-1,1	-	-	-
Max	14,1	11,7	1,5	-	-	-

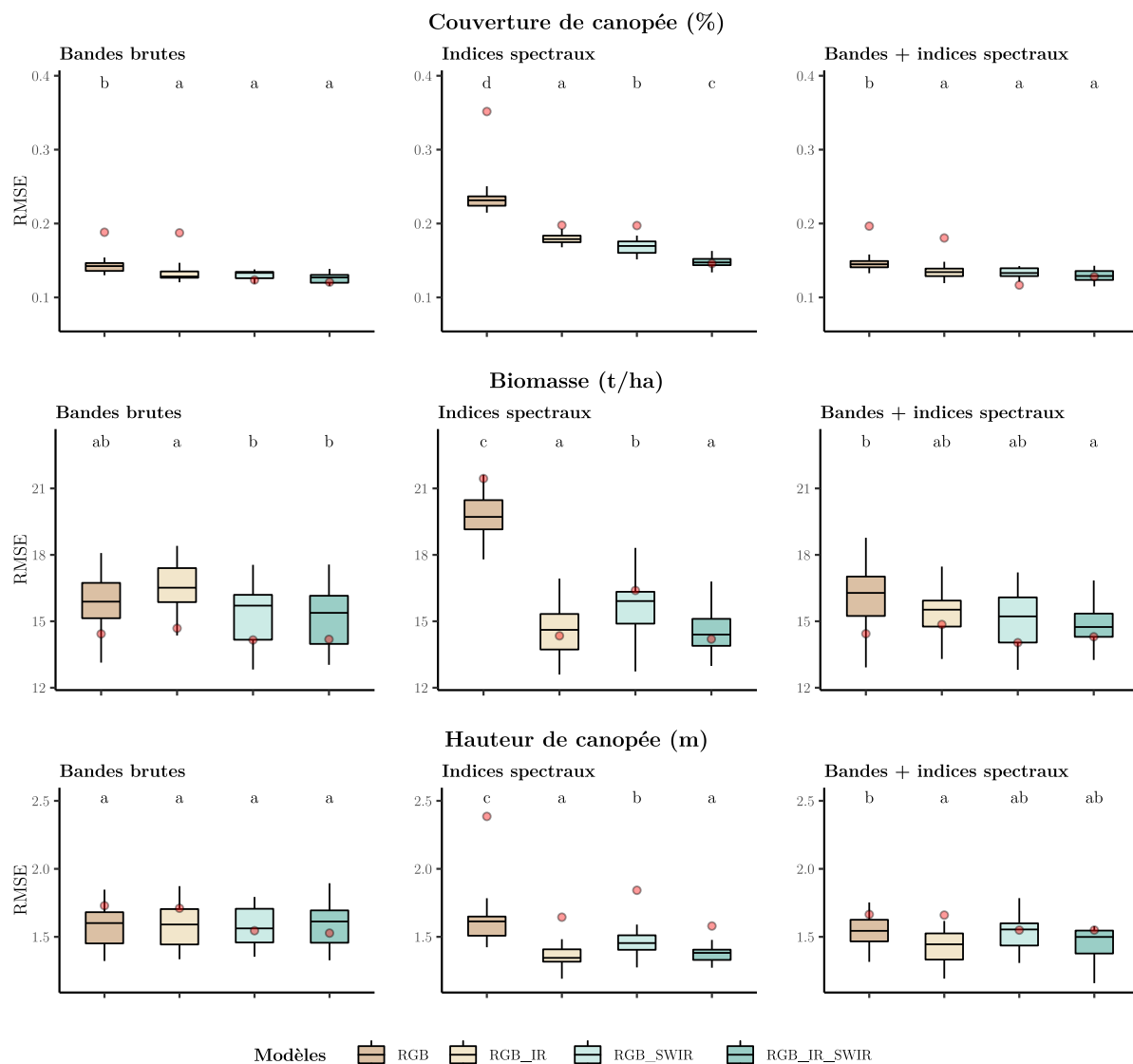
## 2.3. Facteurs spectraux

Les variables de réponse montrent un comportement différent face aux combinaisons spectrales étudiées, dont les résultats sont synthétisés à la figure 20 et les performances du meilleur modèle présentées dans le tableau 16. La couverture de la canopée est mieux représentée en combinant les infrarouges et le spectre visible qu'en utilisant le spectre visible seul. Les modèles de photo-interprétation sont plus spécifiquement influencés par l'infrarouge à courte longueur d'onde (SWIR). En outre, la combinaison des bandes brutes et des indices de végétation n'améliore pas la qualité des modèles, tandis que les indices seuls sont moins performants que les bandes brutes. Concernant la biomasse, la tendance est similaire, en moins prononcée. La meilleure représentation est également obtenue après l'ajout de l'infrarouge à courte longueur d'onde, tandis que l'ajout du proche infrarouge au spectre visible ne permet pas d'améliorer les modèles. A contrario, la partie du spectre utilisée ne semble pas influencer les modèles de hauteur de canopée lorsque les bandes brutes sont utilisées. La meilleure représentation est obtenue uniquement par l'utilisation du spectre visible. Néanmoins, de meilleurs résultats sont obtenus lorsque les indices de végétation qui utilisent les infrarouges sont combinés aux bandes brutes.

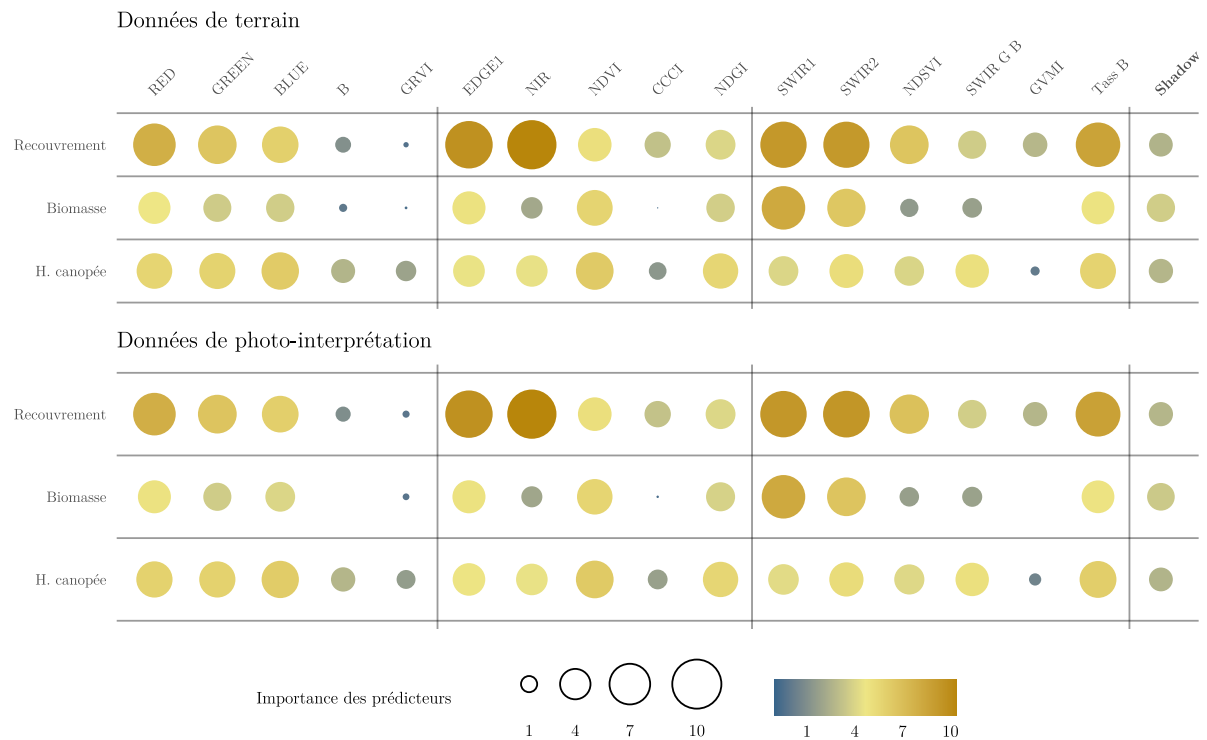
Ces tendances sont confirmées par l'analyse de l'importance des variables (figure 21). Les bandes spectrales brutes apportent généralement plus d'informations que les indices de végétation, quelle que soit la variable de réponse. L'ensemble du spectre apporte de l'information et les bandes sont fortement corrélées entre elles, avec un faible gain de qualité par ajout de bandes supplémentaires. Les infrarouges sont décrits comme particulièrement importants pour représenter le recouvrement. À l'inverse, le proche infrarouge est peu, voire pas, informatif lors de l'estimation de la biomasse en saison de repos, de même que les bandes bleues et vertes. L'estimation de la hauteur de canopée, pour laquelle la partie du spectre influence le moins la qualité de modélisation, montre une importance similaire entre les bandes et les indices, avec une légère prédominance du spectre visible. Parallèlement, un faible nombre d'indices de végétation apporte plus d'informations qu'une série aléatoire de données (*shadow features*). Néanmoins, certaines tendances peuvent être identifiées. L'indice *Tasseled Cap Brightness Transformation*, issu d'une combinaison linéaire de l'ensemble des bandes spectrales, figure systématiquement parmi les indices les plus informatifs. Il est surpassé par le NDVI pour l'estimation de la biomasse et de la hauteur de canopée. Le NDGI, issu d'une combinaison du rouge, du vert et du NIR, est relativement important pour l'ensemble des variables de réponse, tandis que le NDSVI, issu d'une combinaison du rouge et de l'infrarouge à courte longueur d'onde, est important lors de l'estimation du recouvrement. Les tendances d'importance des variables sont identiques selon le jeu de données utilisé.

**Tableau 16** : Résultats détaillés du meilleur modèle de recouvrement pour l'étude de l'influence des facteurs spectraux (exemple d'un boxplot). Les résultats de l'ensemble des modèles sont présentés en annexe.

Variable de réponse : couverture de la canopée (%)						
Facteur étudié : spectre électromagnétique						
Modèle : RGB + SWIR (bandes brutes)						
Prédicteurs : B02, B03, B04, B11, B12						
	Données de terrain → Validation croisée (20 répétitions)			Données de photo-interprétation → Validation externe		
	RMSE	MAE	Biais	RMSE	MAE	Biais
<b>Moyenne</b>	<b>13,2</b>	<b>10,6</b>	<b>0,0</b>	<b>12,0</b>	<b>9,2</b>	<b>-0,6</b>
Médiane	13,3	10,8	0,0	-	-	-
Écart-type	0,6	0,6	0,5	-	-	-
Min	12,1	9,5	-0,1	-	-	-
Max	14,0	11,6	1,0	-	-	-



**Figure 20 :** comparaison de l'écart quadratique moyen (RMSE) associé à chaque modèle qui vise à caractériser l'influence des prédicteurs spectraux. Les modèles sont regroupés par variable de réponse et par catégorie de prédicteurs spectraux (bandes brutes, indices de végétation, utilisation combinée des bandes brutes et indices de végétation). Un boxplot inclut les résultats de l'évaluation par validation croisée répétée 20 fois pour les modèles calibrés avec les données de terrain. Un point rouge qui représente l'évaluation externe des modèles calibrés avec les données de photo-interprétation est associé à chaque boxplot. Les lettres identiques représentent les modèles sans différence significative de performance identifiée par l'analyse de la variance (p-valeur = 0,05).



**Figure 21** : résultats de l'analyse de l'importance des prédictors par l'algorithme Boruta pour le modèle qui intègre l'ensemble des prédictors spectraux, regroupés par jeu de données de calibration, et par variable de réponse. Les modèles sont placés en ligne et les prédictors en colonne. Les prédictors sont regroupés par groupes de réflectance (visible, proche infrarouge, infrarouge à courte longueur d'onde). L'importance d'un prédictor est à interpréter à la lumière du score maximum obtenu par les *shadow features* associées à ce modèle (Shadow). La taille et la couleur du cercle sont proportionnelles au score associé à un prédictor. Les scores sont comparables entre les modèles d'une même variable de réponse, mais pas entre les variables de réponse.

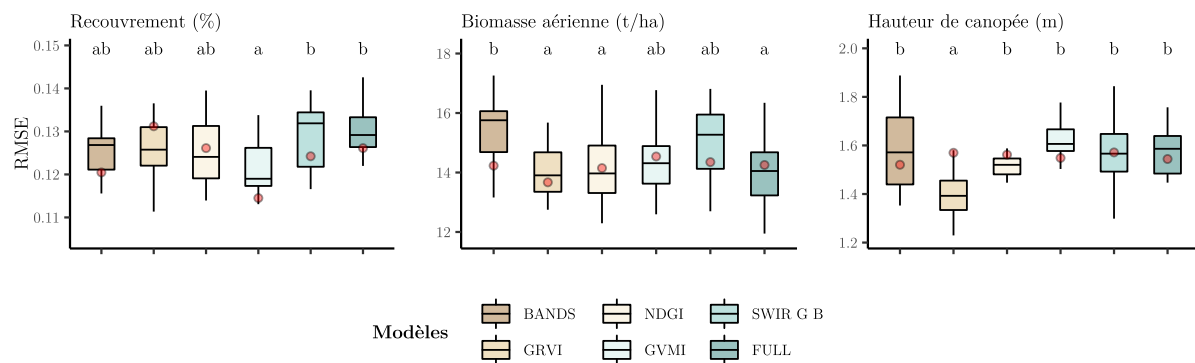


## 2.4. Phénologie

### 2.4.a. Métriques annuelles

L'étude de l'influence des composantes phénologiques annuelles est illustrée à la figure 22 et les performances du meilleur modèle présentées dans le tableau 17. La représentation du recouvrement avec les données de terrain n'est pas significativement améliorée par l'ajout de métriques phénologiques. Cependant, les variations annuelles du GRMI semblent améliorer la qualité des modèles construits à partir des données de photo-interprétation, bien que l'amplitude soit très faible. Lors de la modélisation de la biomasse, les variations annuelles du GRVI et du NDGI entraînent des modèles de meilleure qualité, tandis que seul le GRVI permet une amélioration de la modélisation de la hauteur de canopée.

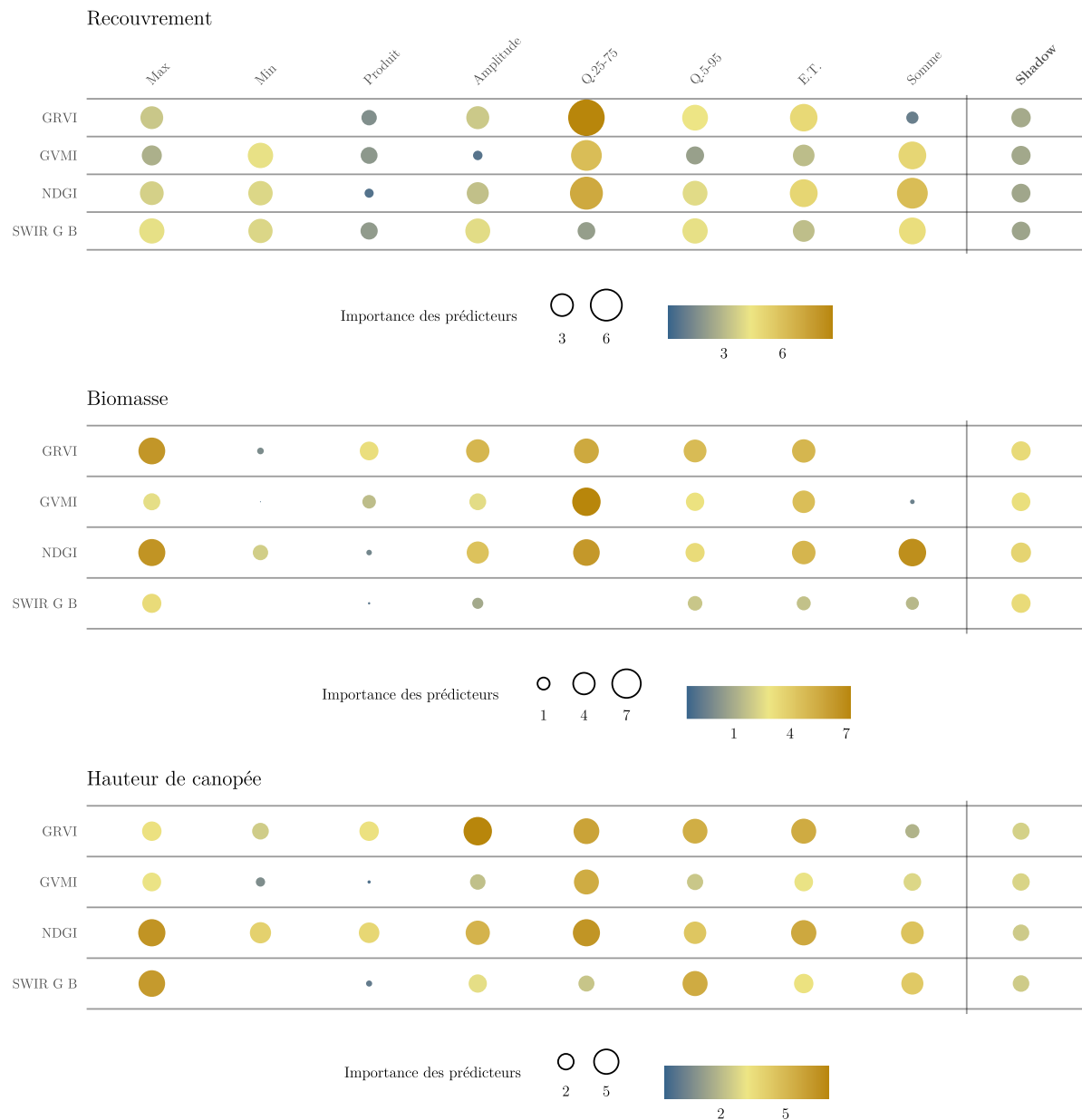
L'exploration de l'importance des variables met en évidence, lors de l'estimation de la biomasse, l'importance des valeurs maximales du GRVI et du NDGI, ainsi que l'écart interquartile de ces deux indices et du GVMI. Les autres métriques jouent un rôle plus marginal (figure 23). Pour la représentation de la hauteur de canopée, la plupart des métriques calculées sur le GRVI apportent de l'information, de même que les valeurs maximales du NDGI et du SWIR G B.



**Figure 22 :** comparaison de l'écart quadratique moyen (RMSE) associé à chaque modèle qui vise à caractériser l'influence de la phénologie annuelle. Les modèles sont regroupés par variable de réponse. Un boxplot inclut les résultats de l'évaluation par validation croisée répétée 20 fois pour les modèles calibrés avec les données de terrain. Un point rouge qui représente l'évaluation externe des modèles calibrés avec les données de photo-interprétation est associé à chaque boxplot. Les lettres identiques représentent les modèles sans différence significative de performance identifiée par l'analyse de la variance (p-valeur = 0,05).

**Tableau 17 :** Résultats détaillés du meilleur modèle de recouvrement pour l'étude de l'influence des variations phénologiques annuelles (exemple d'un boxplot). Les résultats de l'ensemble des modèles sont présentés en annexe.

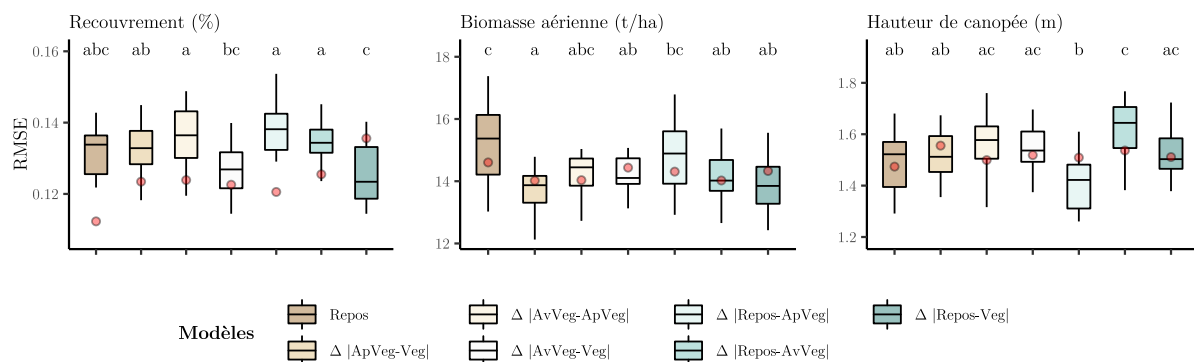
Variable de réponse : couverture de la canopée (%)						
Facteur étudié : variations phénologiques annuelles						
Modèle : GVMI (bandes brutes en saison de repos + métriques phénologiques calculées sur le GVMI)						
Prédicteurs : B02, B03, B04, B05, B8A, B11, B12, min, max, écart-type, amplitude, écart interquartiles, écart interpercentiles 5.95, somme, produit						
	Données de terrain → Validation croisée (20 répétitions)			Données de de photo-interprétation → Validation externe		
	RMSE	MAE	Biais	RMSE	MAE	Biais
<b>Moyenne</b>	<b>12,1</b>	<b>10,0</b>	<b>0,6</b>	<b>11,5</b>	<b>8,9</b>	<b>-1,5</b>
Médiane	11,9	9,8	0,1	-	-	-
Écart-type	0,7	0,7	0,5	-	-	-
Min	10,5	8,7	-0,3	-	-	-
Max	13,4	11,2	1,3	-	-	-



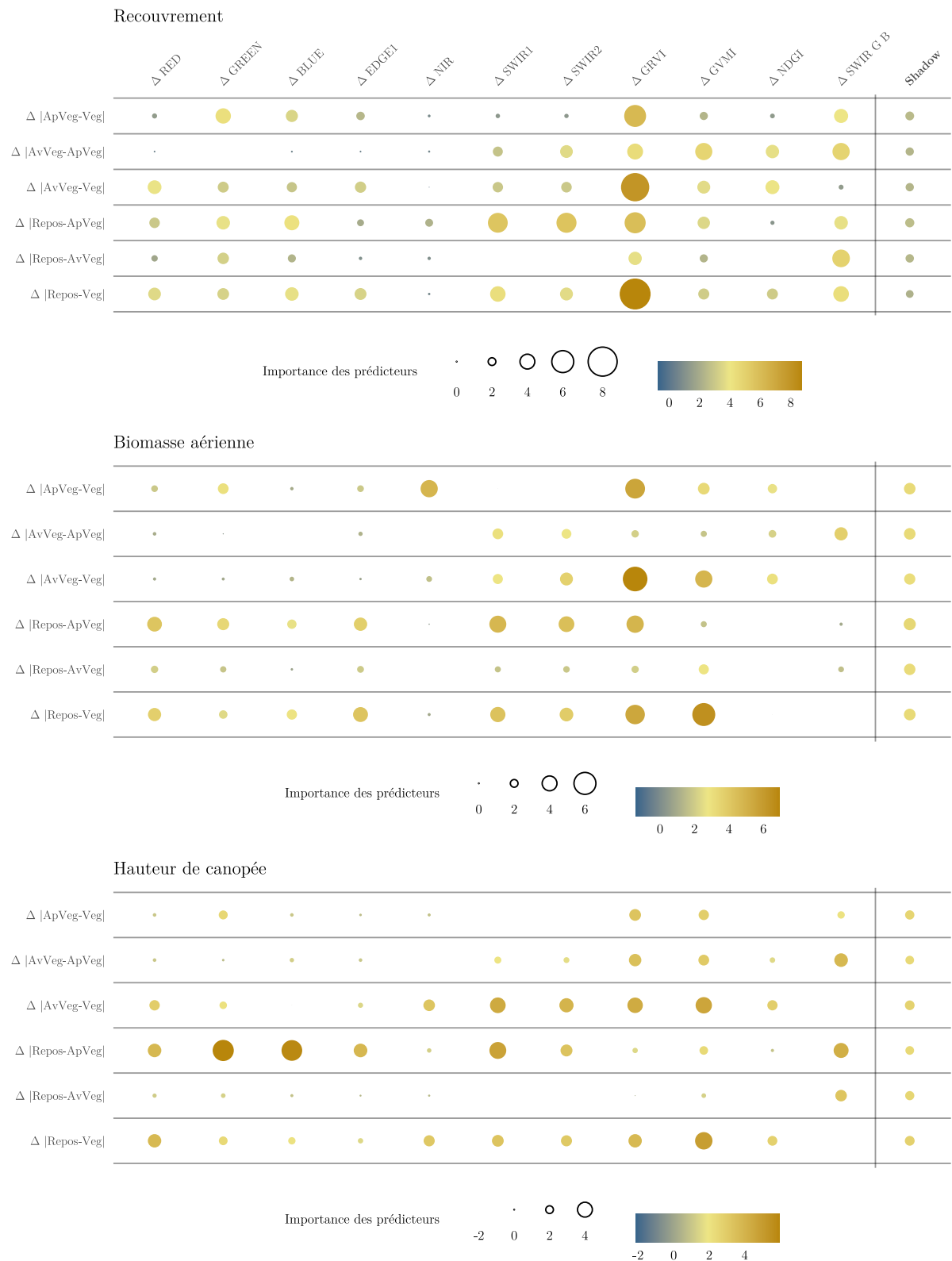
**Figure 23** : résultats de l'analyse de l'importance des prédicteurs par l'algorithme Boruta, regroupé par variable de réponse et par métriques phénologiques dont l'influence a été étudiées pour caractériser l'importance de la phénologie. Les modèles sont placés en ligne et les prédicteurs en colonne. L'importance d'un prédicteur est à interpréter à la lumière du score maximum obtenu par les *shadow features* associées à ce modèle (Shadow). La taille et la couleur du cercle sont proportionnelles au score associé à un prédicteur. Les scores sont comparables entre les modèles d'une même variable de réponse, mais pas entre les variables de réponse.

### 2.4.b. Métriques saisonnières

Les résultats obtenus lors de l'étude des métriques phénologiques calculées entre deux saisons sont présentés à la figure 25 et les performances du meilleur modèle présentées dans le tableau 18. Lors de la représentation de la couverture de la canopée, le modèle qui n'intègre aucune des métriques phénologiques est statistiquement de qualité équivalente à l'ensemble des modèles qui intègrent ces métriques. Le modèle construit sans métriques phénologiques et à partir des données de photo-interprétation est celui qui minimise le plus la variance résiduelle. Cependant, les variations annuelles du GRVI autour de la saison de végétation ont été identifiées comme influentes lors de l'analyse de l'importance des variables (figure 25). A contrario, les modèles de biomasse sont influencés positivement, en termes de variance résiduelle et de stabilité des modèles, par l'ajout des périodes phénologiques calculées à partir de la saison de végétation. Ce sont alors les variations du GRVI et, dans une moindre mesure, du NDVI et du GVMi qui permettent une meilleure représentation. La hauteur de canopée ne semble pas influencée par ces métriques, aucune ne permettant d'obtenir des résultats différents par rapport à l'utilisation des bandes brutes. L'utilisation de métriques phénologiques calculées entre la saison de repos de végétation et la transition post-végétation semble améliorer sensiblement les modèles, dans lesquels les variations dans le bleu et le vert semblent déterminantes.



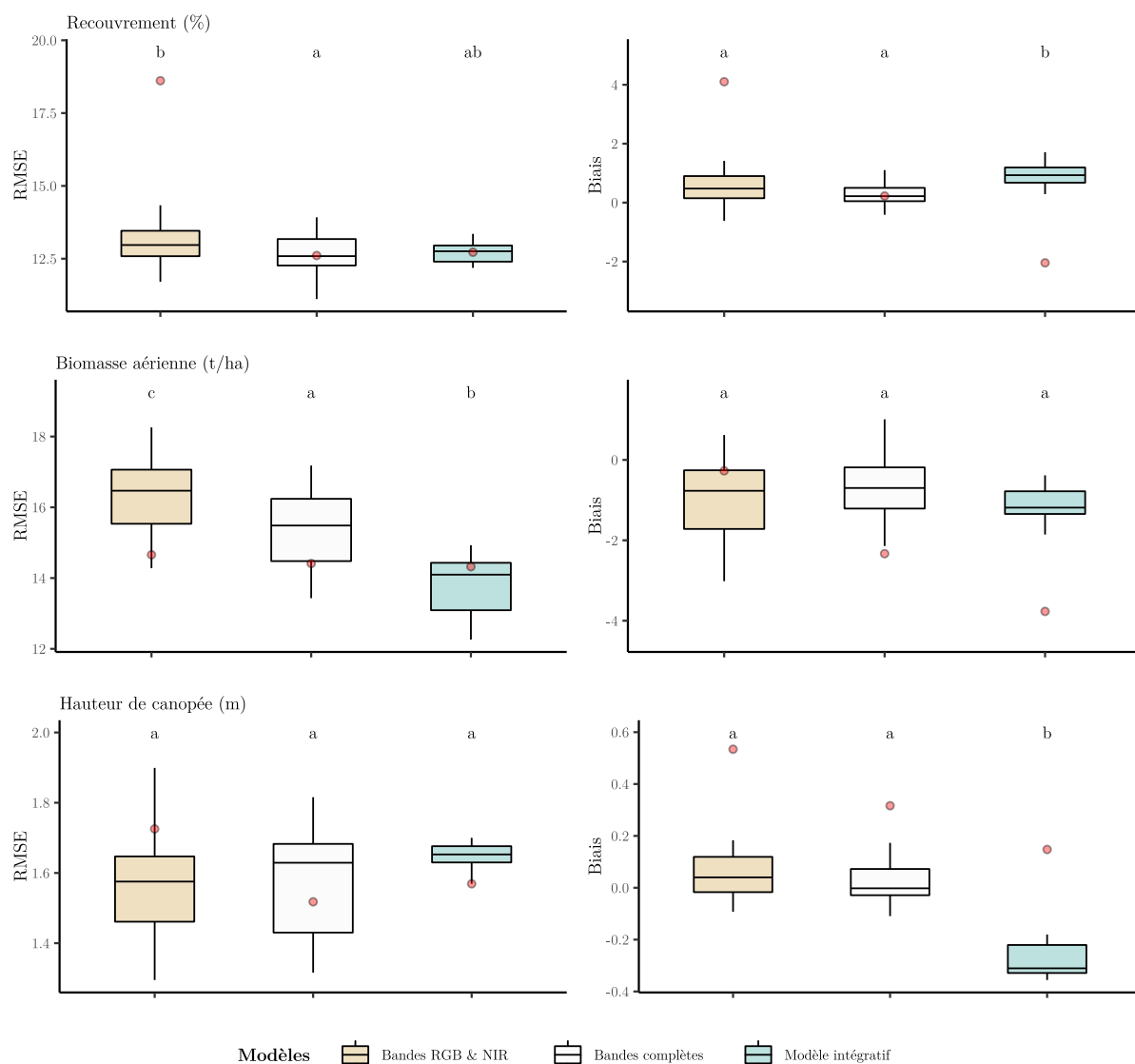
**Figure 24 :** comparaison de l'écart quadratique moyen (RMSE) associé à chaque modèle qui vise à caractériser l'influence de la phénologie saisonnière. Un modèle, dont le nom est construit tel que  $\Delta$ [saison 1 – saison 2], inclut les métriques phénologiques calculées entre deux saisons de végétation (différences normalisées pour chaque bande et indice de végétation). Les modèles sont regroupés par variable-réponse. Un boxplot inclut les résultats de l'évaluation par validation croisée répétée 20 fois pour les modèles calibrés avec les données de terrain. Un point rouge, qui représente l'évaluation externe des modèles calibrés avec les données de photo-interprétation, est associé à chaque boxplot. Les lettres représentent les groupes de modèles sans différence significative de performance identifiée par l'analyse de la variance (p-valeur = 0,05).



**Figure 25** : résultats de l'analyse de l'importance des prédicteurs évaluée par l'algorithme Boruta, regroupés par variable -réponse et par modèle. Un modèle, dont le nom est construit tel que  $\Delta$ [saison 1 – saison 2], inclut les métriques phénologiques calculées entre deux saisons de végétation (différences normalisées pour chaque bande et indice de végétation). Les modèles sont placés en ligne et les prédicteurs en colonne. Les prédicteurs représentés n'incluent pas les bandes brutes de Sentinel-2. L'importance d'un prédicteur est à interpréter à la lumière du score maximum obtenu par les *shadow features* associées à ce modèle (Shadow). La taille et la couleur du cercle sont proportionnelles au score associé à un prédicteur. Les scores sont comparables entre les modèles d'une même variable -réponse, mais pas entre les variables -réponses.

### 3. Modèles intégratifs

La performance du modèle intégratif de recouvrement de la canopée n'est pas significativement différente de celle des modèles calibrés sur les bandes brutes de Sentinel-2, que ce soit le modèle utilisant l'ensemble des bandes ou celui intégrant uniquement les bandes du spectre visible et proche infrarouge. En outre, ce modèle intégratif est plus biaisé que les autres ( $p$ -valeur  $< 0,05$ ). Parallèlement, la performance du modèle intégratif de biomasse aérienne est significativement supérieure à celle des autres modèles ( $p$ -valeur  $< 0,05$ ) et aucune différence de biais n'est observée. La performance du modèle intégratif de hauteur de canopée n'est pas significativement différente de celle des autres modèles et ce modèle est plus biaisé ( $p$ -valeur  $< 0,05$ ). Néanmoins il est plus stable que les modèles qui n'intègrent pas de prédicteurs additionnels.

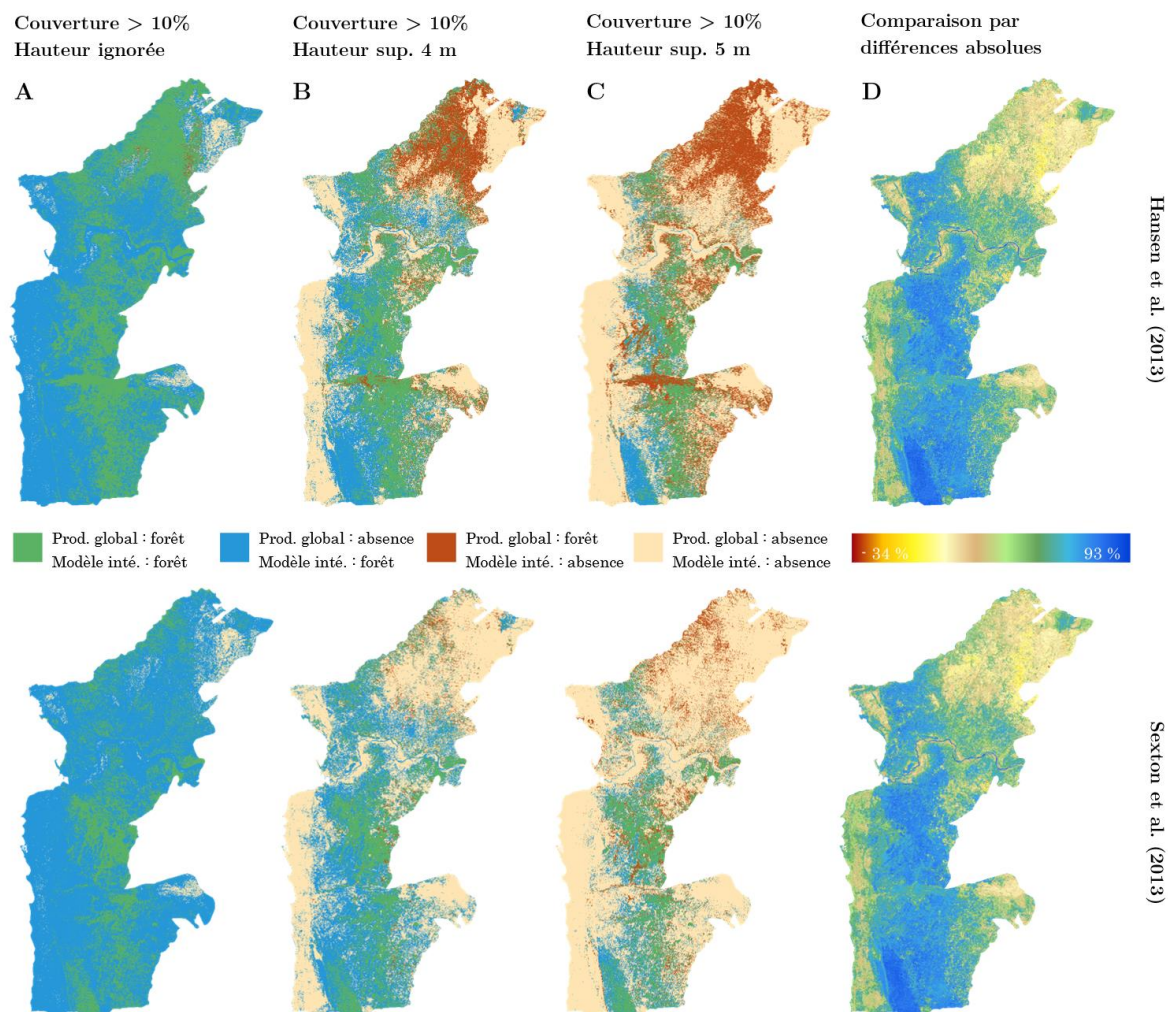


**Figure 26 :** comparaison du modèle intégratif avec deux modèles construits uniquement avec les bandes brutes de Sentinel-2. Le modèle *RGB & NIR* intègre les bandes du spectre visible et du proche infrarouge. Le modèle *Bandes complètes* utilise en plus les bandes de l'infrarouge à courte longueur d'onde. La comparaison est réalisée pour le recouvrement, la biomasse et la hauteur sur base de la RMSE et du biais moyen associés à chaque modèle. Un boxplot inclut l'évaluation répétée 20 fois pour chaque modèle. Un point rouge, qui représente l'évaluation externe des modèles calibrés avec les données de photo-interprétation, est associé à chaque boxplot. Une lettre différente est associée aux groupes de modèles présentant une différence significative de performance identifiée par l'analyse de la variance ( $p$ -valeur = 0,05).

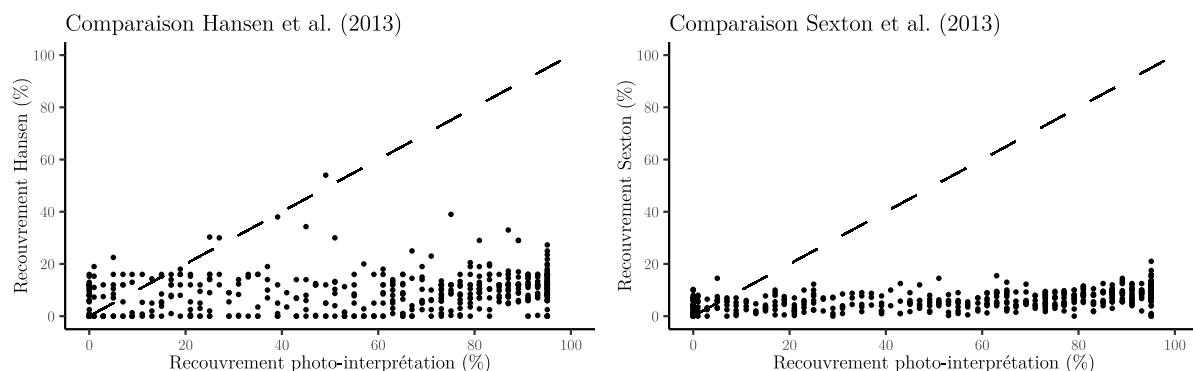
## 4. Mise en perspective des cartes globales

Les cartes de Hansen et al. (2013) et Sexton et al. (2013) sous-estiment fortement la couverture boisée, principalement dans les zones à haut taux de recouvrement. Une classification binaire construite avec les surfaces caractérisées par un recouvrement boisé supérieur à 10 % —base de la définition d’une forêt par la FAO— met en avant une sous-estimation généralisée de la surface forestière par les produits globaux. La carte de Hansen identifie des surfaces plus importantes que la carte de Sexton, notamment dans le nord et dans la bande sableuse latérale au centre de la zone d’étude (figure 27.A). Les zones éliminées après l’application d’un seuil de hauteur de 4 m au modèle intégratif sont situées principalement le long de la côte et dans la savane ouverte de l’est de la zone d’étude (figure 27.B). L’utilisation d’un seuil de hauteur de 5 m (figure 27.C) efface la plupart des différences entre le modèle intégratif et les produits globaux, bien qu’il subsiste des variations locales et une importante zone non identifiée par les produits globaux au sud de la zone d’étude, dans le parc national de Tsimanampetsotsa. La carte de Sexton est la seule qui capture une partie de la couverture végétale dans cette zone. La carte de Hansen identifie une grande étendue forestière qui est absente dans la carte de Sexton et dans le modèle intégratif. Après s’être affranchi de la classification binaire par seuil de 10 % de couverture boisée, les différences absolues montrent une grande similarité de prédictions entre les cartes de Hansen et de Sexton, ainsi qu’une importante sous-estimation générale de la couverture de canopée, principalement par Sexton (figure 27.D). La carte de Hansen sature sur toute la gamme de valeurs de recouvrement et ne détecte aucun taux de recouvrement supérieur à 25 %. La carte de Sexton est encore moins corrélée au recouvrement mesuré par photo-interprétation et sature à partir de 15 % de recouvrement (figure 28).

Par rapport aux valeurs extrapolées sur le jeu de données de photo-interprétation, le modèle d’estimation de la biomasse aérienne globale de Spawn et al. (2020) arrive à saturation à partir de 5 t/ha (figure 30) et ne permet pas une représentation adéquate de la biomasse dans la majorité de la zone d’étude (figure 29). Les zones ouvertes et à faible biomasse situées sur la bande côtière, dans la savane du centre-ouest et au nord de la zone d’étude sont celles où les différences sont les moins marquées. Le modèle d’estimation de la hauteur de canopée de Potapov et Gibbs (2022) sous-estime la hauteur sur l’ensemble de la zone d’étude par rapport aux valeurs extrapolées sur le jeu de données de photo-interprétation. Il y a de petites zones où la carte prédit des valeurs supérieures à 8-10 m et où le modèle intégratif ne prédit pas de hauteur de canopée (figure 29). Contrairement aux modèles globaux de hauteur de canopée et de biomasse, le modèle de Potapov et Gibbs ne montre pas de comportement de saturation, mais ne capte pas tous les écosystèmes de la zone d’étude (figure 30). Ce modèle affiche moins d’erreurs pour les écosystèmes les plus hauts que pour les écosystèmes de plus faibles dimensions. Il est par exemple incapable de représenter les écosystèmes côtiers, la majeure partie des fourrés sur sol calcaire et les écosystèmes du parc national de Tsimanampetsotsa. Les différences sont plus faibles en s’éloignant de la côte vers les écosystèmes aux dimensions les plus importantes. Les tendances observées sur les parcelles de photo-interprétation sont confirmées avec les données de terrain.



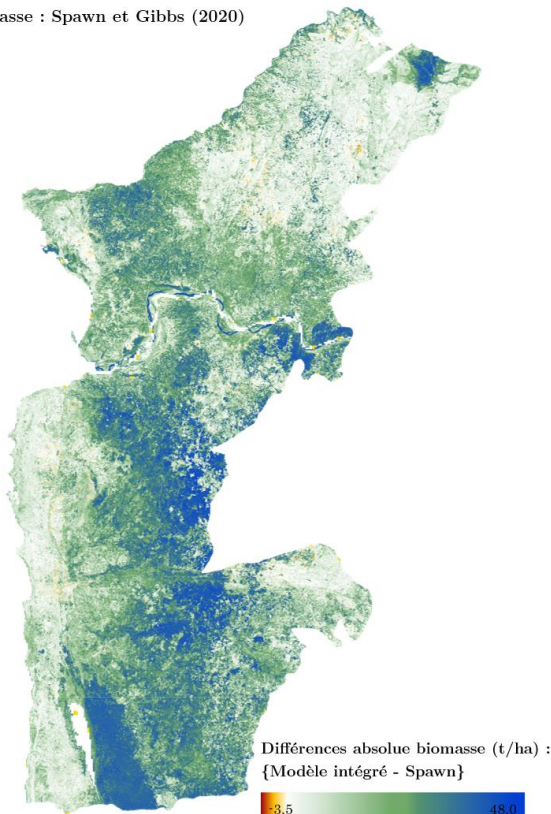
**Figure 27** : comparaison des prédictions de couverture de canopée par le modèle intégratif et par les produits globaux de Hansen Global Forest Change v1.9 (Hansen et al., 2013) et de Global Forest Cover Change (GFCC) Tree Cover 30 m 2015 (Sexton et al., 2013). La hauteur utilisée pour la définition de la forêt provient du modèle intégratif de hauteur construit dans le cadre de cette étude (RMSE = 1,6 m ; MAE = 1,0 m ; biais = 0,1 m). Le produit global de Hansen et al. (2013) est calibré pour des hauteurs à partir de 5 m. Le masque de forêt pour le produit de Hansen et al. (2013) inclut les pertes forestières estimées entre 2000 et 2019.



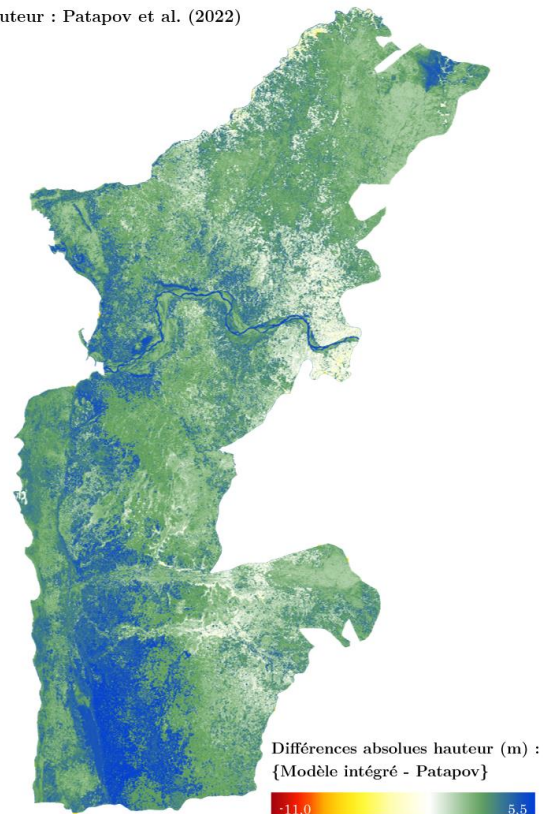
**Figure 28** : comparaison des valeurs de couverture de canopée mesurées par photo-interprétation et des valeurs prédites par Hansen et al. (2013) et par Sexton et al. (2013). Un point représente une placette d'échantillonnage de photo-interprétation, pour laquelle les valeurs moyennes de recouvrement prédites par les produits globaux ont été extraites. La ligne noire pointillée indique la correspondance parfaite entre les modèles.



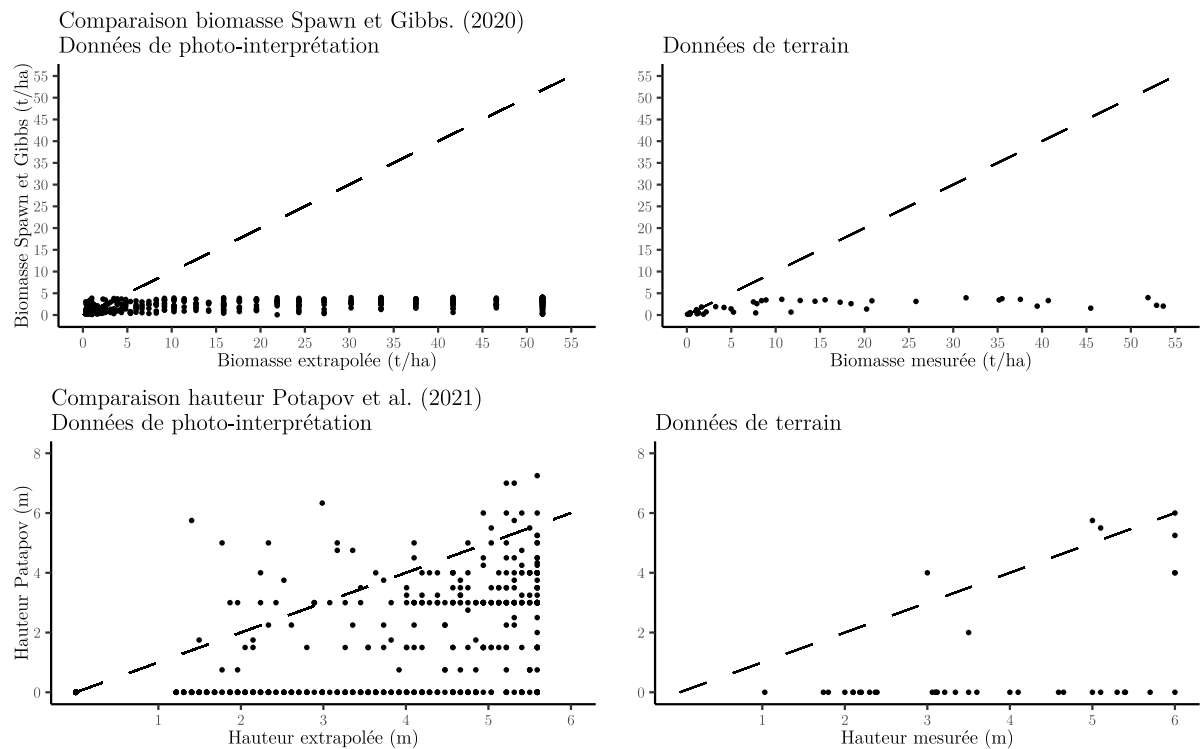
Biomasse : Spawn et Gibbs (2020)



Hauteur : Patapov et al. (2022)



**Figure 29** : comparaison des prédictions de biomasse aérienne (gauche) et de hauteur de canopée (droite) obtenues par le modèle intégratif calibré avec les données de photo-interprétation, et par les produits globaux de, respectivement, Spawn et Gibbs (2020) et de Potapov et al. (2022).



**Figure 30** : comparaison des valeurs de hauteur et de biomasse extrapolées au jeu de données de photo-interprétation et de terrain avec, respectivement, les valeurs prédites par Potapov et al. (2021) et par Spawn et Gibbs (2020). Un point représente une placette d'échantillonnage, pour laquelle les valeurs moyennes de hauteur et de biomasse prédites par les produits globaux ont été extraites. La ligne noire pointillée indique la correspondance parfaite entre les modèles.



# Discussion

## 1. Potentiel de la photo-interprétation

La première partie de cette étude a mis en évidence le potentiel de la photo-interprétation pour modéliser des attributs structuraux quantitatifs. En effet, les modèles de photo-interprétation minimisent davantage la variance résiduelle par rapport aux modèles de terrain, y compris pour les variables de réponse extrapolées que sont la biomasse et la hauteur de canopée. Bien que les variables de réponse extrapolées soient plus fortement corrélées à la couverture de la canopée, elles répondent indépendamment à l'influence des prédicteurs étudiés. Cependant, ces modèles sont légèrement plus biaisés pour les attributs extrapolés et la structure verticale des écosystèmes est moins bien représentée que par les modèles calibrés sur les données de terrain. L'extrapolation de la structure verticale de l'écosystème est plus complexe que pour la biomasse et la hauteur, car cet attribut est moins corrélé localement à l'ouverture de la canopée.

Les modèles construits avec le jeu de données de photo-interprétation ont été évalués avec les mesures réalisées sur les parcelles de terrain et ont été comparés aux résultats obtenus avec les modèles calibrés directement sur les données de terrain. Quelle que soit la variable de réponse considérée, les modèles qui utilisent les données de photo-interprétation minimisent en moyenne légèrement plus la variance résiduelle (MAE), sont de meilleurs estimateurs que la moyenne des données de validation (RSE). Néanmoins, ces modèles sont plus biaisés que les modèles de terrain, à l'exception des modèles relatifs au recouvrement ligneux. Ces résultats doivent être considérés à la lumière de deux points. Premièrement, les résultats sont influencés par la qualité de la modélisation des attributs forestiers extrapolés depuis la couverture de la canopée. Bien que ces modèles soient exempts de biais, ils n'ont pas été forcés à l'ordonnée à l'origine, ce qui implique que des valeurs biaisées ont été attribuées en l'absence de recouvrement, car les données d'absence de recouvrement sont mal représentées dans le jeu de données de terrain. Par exemple, une biomasse négative (-0,5 t/ha) et une hauteur de plus de 1 m sont associées aux écosystèmes complètement ouverts. Cette situation pourrait expliquer le biais plus important associé aux modèles de photo-interprétation et laisse une marge d'amélioration. Deuxièmement, les modèles évalués par validation croisée spatialement indépendante sont comparés à ceux évalués par validation externe. La validation croisée spatialement indépendante permet de minimiser le biais de surestimation de la qualité causé par l'autocorrélation spatiale. Néanmoins, dans le cadre de cette étude, cette technique a pu engendrer une sous-estimation de la qualité réelle des modèles calibrés sur les données de terrain. En effet, pour faire face aux contraintes logistiques de terrain, l'échantillonnage a été réalisé en grappes composées d'écosystèmes similaires, différents d'une grappe à l'autre. Lors de la création des groupes de parcelles pour la validation croisée, les parcelles ont été réunies en tenant compte, entre les groupes, d'une distance minimale équivalant à l'estimation de la distance d'autocorrélation. Ce choix induit que les écosystèmes majoritairement représentés dans une grappe sont surreprésentés au sein d'un groupe de validation croisée. À chaque répétition pendant une évaluation croisée, un faciès de végétation est donc sous-représenté lors de la calibration du modèle alors même que l'évaluation de la qualité est réalisée sur cet écosystème. L'ensemble des données, et donc des faciès de végétation, sont ensuite utilisés pour la calibration du modèle final. Ce dernier devrait donc être de meilleure qualité que celle estimée lors de la validation croisée. L'amplitude des différences est difficile à évaluer et cette situation implique que les modèles de photo-interprétation pourraient être de plus faible qualité que les modèles de terrain. Néanmoins, la comparaison visuelle entre les cartes modélisées à partir des données de terrain et celles modélisées à partir des données de photo-interprétation indique que les modèles de photo-interprétation permettent une meilleure représentation des

valeurs aux extrêmes, que ce soit au niveau du recouvrement, de la biomasse et de la hauteur de canopée. A contrario, les modèles de photo-interprétation ne permettent quasiment aucun contraste en ce qui concerne la structure verticale de la végétation (figure 16). Le potentiel de la photo-interprétation et de la modélisation d'attributs écosystémiques diffère fortement selon l'attribut considéré. La structure verticale des écosystèmes, quantifiée par les dimensions de l'ACP, est moins bien représentée que les autres attributs et donc source d'erreurs. En conséquence, cet attribut a été moins approfondi dans cette étude.

Sur les cartes finales, une importante corrélation est observée entre les attributs extrapolés depuis la couverture de la canopée et la couverture de canopée, situation qui découle de la nature des données de calibration de photo-interprétation. Il est donc difficile de distinguer la part d'indépendance des variables de réponse. Une piste d'interprétation se trouve dans le comportement des modèles en fonction des facteurs étudiés, ainsi que dans l'importance des prédicteurs lors de la construction de ces modèles. On constate que les variables de réponse sont influencées indépendamment du recouvrement, au moins en partie, par les prédicteurs étudiés. Néanmoins, la tendance de performance des modèles est identique entre les variables de réponse et le recouvrement de la canopée, ce qui n'est pas observé pour les modèles calibrés sur les données de terrain. Par conséquent, un crédit plus important a été accordé au jeu de données de terrain pour étudier la réponse des attributs aux prédicteurs spectraux.

## 2. Études de l'influence des variables prédictives

La seconde partie de cette étude s'est intéressée aux caractéristiques spectrales, spatiales et temporelles qui influencent la cartographie des propriétés structurelles des fourrés xérophiles, ainsi qu'aux spécificités et similarités des fourrés xérophiles par rapport aux écosystèmes arides majoritairement étudiés. La signature spectrale à l'échelle des pixels de Sentinel-2 (10 m) est une combinaison complexe entre l'influence du sol et des éléments de végétation, avec de fortes hétérogénéités locales et globales. On observe donc des tendances propres à chaque attribut structurel face à l'association de la saisonnalité et des facteurs spectraux utilisés, deux paramètres qui sont déterminants pour une représentation adéquate. Les nombreuses variations structurelles locales, dictées par des éléments de faibles dimensions, sont mieux représentées à une échelle plus fine. A contrario, l'utilisation de variables phénologiques et d'indices de végétation ne permet pas une meilleure représentation de ces écosystèmes très arides, dont la signature spectrale est faible et les amplitudes annuelles d'activité végétale difficiles à capter. Les paramètres considérés dans cette étude n'influencent que de manière marginale la modélisation des attributs écosystémiques et de très bons résultats sont obtenus avec un faible nombre de prédicteurs simples. L'utilisation de données qui reflètent l'hétérogénéité du paysage est une étape primordiale qui pourrait expliquer leur faible représentation à l'échelle globale.

### 2.1. Couverture de la canopée

Dans l'ensemble, il n'y a qu'une très faible influence des caractéristiques spectrales, spatiales et temporelles pour la caractérisation de la couverture de canopée. Des résultats légèrement meilleurs ont été observés lors de la modélisation du recouvrement de la canopée avec des prédicteurs spectraux en saison de repos de la végétation. D'autres études en écosystèmes arides soulignent la même tendance, mais avec de plus larges amplitudes entre la saison sèche et les autres saisons de végétation. La majorité de ces études ont été réalisées dans des écosystèmes où la strate herbacée induit une perturbation du signal spectral. La saison de repos, durant laquelle la strate herbacée est sénescence, permet alors un contraste plus important avec la végétation ligneuse sempervirente (Karlson et al., 2015; Forkuor et al., 2018; Higginbottom et al.,

2018; Nagelkirk et al., 2020; Gan et al., 2022). Les fourrés épineux sont composés d'une flore partiellement décidue et sont majoritairement exempts de strate herbacée, ce qui peut expliquer les faibles différences constatées entre les saisons de végétation. La très faible hygrométrie du sol en saison de repos, qui tend à minimiser les perturbations de la signature spectrale sous une canopée éparse, peut également expliquer ces meilleures performances durant la saison sèche (Smith et al., 2019).

L'échelle spatiale utilisée est un paramètre dont l'influence est assez marquée et qui engendre systématiquement une meilleure représentation à 10 m de résolution. Ces résultats divergent des études réalisées en savane ou en forêt ouverte, où l'utilisation d'une résolution plus grossière a deux effets principaux. Cela permet de minimiser l'influence des arbres caractérisés par une couronne plus importante que la taille des pixels, et minimise également l'impact de l'alternance du sol nu et de végétation sur des distances relativement importantes. Néanmoins, l'amélioration obtenue en moyenne se fait au prix d'une mauvaise représentation de plus fines variations locales et des individus de faibles dimensions (Higginbottom et al., 2018; Shafeian et al., 2021). Le cas des fourrés épineux est différent. Les arbres ont un diamètre de couronne inférieur à la taille des pixels et les variations de recouvrement sont réalisées sur de plus faibles distances (Lourenço, 1996). Des pixels de taille plus petite permettent donc de représenter correctement la majorité des éléments, même les baobabs dont la couronne est relativement petite comparée aux dimensions de l'arbre. C'est également le cas pour les bas fourrés sur dalle calcaire — où la présence de la végétation est directement dépendante des fissures dans le socle rocheux dont la densité est importante à l'échelle d'un pixel de 10 m. Parallèlement, les hauts fourrés ne nécessitent pas une réduction de la dimensionnalité, car ils sont homogènes à l'échelle locale et ont un taux de recouvrement important.

L'importance des bandes situées dans l'infrarouge et l'absence d'influence significative des indices de végétation ont été montrées par l'étude des facteurs spectraux. Cette caractéristique se retrouve également dans les autres écosystèmes arides, mais les fourrés xérophiles se distinguent sur certains aspects. L'infrarouge à courte longueur d'onde joue un rôle majeur dans la qualité de représentation, ce qui a déjà été mis en avant, en savane par Karlson et al. (2015), et en zones herbeuses ponctuées de buissons par Gan et al. (2022). Lors de l'étude préliminaire réalisée pour identifier les variations spectrales en fonction du niveau de recouvrement, les bandes SWIR, et plus particulièrement la bande SWIR2 (~2200 nm), étaient les prédicteurs les plus corrélés. En outre, l'infrarouge à courte longueur d'onde est particulièrement important en saison de repos et avant la saison de végétation, car il est corrélé à l'humidité du sol sous une canopée éparse, ainsi qu'au contenu en eau des feuilles. La réflectance à ~2200 nm montre des réponses particulièrement marquées pour les sols secs (Wang et al., 2008). Dès lors, ces résultats coïncident avec les précédentes études en zones arides. Néanmoins, une distinction peut être faite au niveau des indices de végétation. Alors que Zhang et al. (2013) et Ludwig et al. (2019) ont mis en avant l'importance du NDVI, un indice de végétation fortement corrélé à l'ouverture de la canopée en savane, celui-ci n'apporte que peu d'informations dans notre zone d'étude. Le NDSVI, qui remplace le NIR par le SWIR dans la différence normalisée, est jugé plus informatif, mais toujours d'importance marginale. En effet, les bandes brutes sont toutes jugées bien plus importantes que les indices de végétation. Ceci rejoint l'étude de Smith et al. (2019), qui ont montré une diminution de l'intérêt des indices avec l'ouverture de la canopée et l'aridité du sol (Smith et al., 2019).

Enfin, l'intégration de métriques phénologiques n'apporte que des améliorations marginales, avec une amplitude inférieure à 1 % pour l'écart quadratique moyen (RMSE). D'autres études ont montré des comportements différents, notamment des baisses significatives de la variance résiduelle en intégrant des

métriques calculées sur des séries temporelles et plus particulièrement en utilisant le NDVI (Karlson et al., 2015; Ludwig et al., 2019, 2019). Il est intéressant de constater qu'un indice de végétation pris isolément à un moment de l'année peut ne pas être informatif, mais que sa variation annuelle peut être corrélée aux degrés d'ouverture de la canopée et donc permettre une meilleure représentation de celle-ci. En outre, Gan et al. (2022) ont mis en évidence que l'importance des métriques phénologiques est variable selon l'intensité des périodes de sécheresse. Les variations annuelles et interannuelles de l'activité végétale sont très importantes dans les fourrés épineux, tandis que l'amplitude des variations phénologiques est faible. L'utilisation d'images composées de la réflectance médiane des cinq dernières années a pu masquer les périodes clés et gommer ces fines variations qui auraient potentiellement pu permettre une meilleure représentation. De futures études pourraient investiguer plus finement le rôle des métriques phénologiques afin de caractériser avec plus de fiabilité leur influence.

## 2.2. Biomasse

La saison de végétation permet d'obtenir les meilleurs résultats et les modèles les moins biaisés ( $< 0,1$  t/ha). La saison de pré-végétation montre des résultats similaires, bien que légèrement inférieurs. Dans cette zone d'étude, c'est durant la reprise de l'activité végétale que, sur les arbres décidus, a lieu l'apparition des feuilles, qui persistent durant le pic d'activité photosynthétique et de productivité. L'exploration du jeu de données a mis en évidence que cette flore décidue, majoritairement des arbres de type 3, sont responsables de la majeure partie de la biomasse stockée dans l'écosystème, tandis que la part des arbres de plus petites dimensions majoritairement sempervirents est marginale. La saison de végétation permettrait donc de mieux capter ces individus grâce à un contraste plus important entre les zones de faible et de forte productivité.

L'étude de l'influence du spectre est complexe et l'infrarouge joue un rôle ambigu lors de l'estimation de la biomasse. L'ajout des bandes NIR et Red Edge n'améliore pas significativement la performance des modèles basés sur le spectre visible. Néanmoins, lors de l'analyse de l'importance des variables, le NDVI et le NDGI, indices basés sur le proche infrarouge et le rouge, semblent tout de même apporter de l'information. Lorsque les indices sont utilisés seuls, ils permettent d'obtenir de meilleurs résultats qu'avec l'utilisation des bandes brutes. Par ailleurs, le SWIR joue un rôle important au niveau des variables, mais aucun indice basé sur celui-ci n'apporte d'information, à l'image de ce qui est constaté lors de l'étude du recouvrement de la canopée. Il est possible que le SWIR ne soit pas directement lié à l'estimation de la biomasse, mais intervienne en tant qu'indicateur du recouvrement, qui est lui-même fortement corrélé à la biomasse, et permette donc une meilleure estimation moyenne de la biomasse. Le spectre visible semble jouer un rôle important dans l'estimation de la biomasse et l'importance des indices identifiés pourrait être liée à l'utilisation du rouge dans le calcul de l'indice. Cette situation a également été observée lors de la caractérisation spectrale des plantes arides en Australie, où beaucoup d'écosystèmes similaires aux fourrés xérophiles sont présents. Le spectre visible est systématiquement identifié parmi les régions les plus importantes du spectre électromagnétique (Lewis, 2002). Les études réalisées au sujet de l'importance de l'infrarouge montrent des résultats divergents. Néanmoins, plusieurs études ont montré un découplage du NDVI et de la biomasse en région aride, notamment dans des écosystèmes arbustifs (Aranha et al., 2008). Il a également été montré que les indices de végétation basés sur le rouge et l'infrarouge sont moins précis à mesure que la proportion de végétation non photosynthétique augmente, ce qui est constaté dans notre zone d'étude durant la saison de repos (Berri et al., 2007).

Les faibles performances de l'optique pour modéliser la biomasse ont été soulignées dans plusieurs études, avec des modèles qui sont de plus faible qualité que ceux qui représentent la couverture de la canopée (Eisfelder et al., 2012; Karlson et al., 2015; Fremout, 2022). Les particularités structurelles des écosystèmes arides sont responsables d'une faible réflectance de la végétation et d'une forte interférence du signal du sol, caractéristique accentuée durant la période de repos de la végétation (Calvão et al., 2004). Il est intéressant de noter que dès 2006, les faibles performances de l'optique étaient déjà soulignées et que de meilleures estimations ont été obtenues en zone aride avec une modélisation de la biomasse à partir d'autres attributs structuraux, plutôt qu'à partir de prédicteurs optiques (Suganuma et al., 2006). Ceci explique en partie le fait que de meilleures estimations moyennes de la biomasse soient réalisées à partir d'un important jeu de données de calibration extrapolées depuis le recouvrement plutôt qu'à partir d'une faible quantité de données de terrain. En outre, les faibles performances de l'optique pour l'estimation de la biomasse expliquent les importantes erreurs d'estimation comparées aux valeurs moyennes de l'écosystème, avec une MAE de 9 à 12 t/ha pour une biomasse moyenne de 17,7 t/ha. D'autres études réalisées avec une méthodologie comparable dans des écosystèmes aux conditions écologiques similaires ont également obtenu de tels écarts. Par exemple, les études de Fremout et al. (2022), réalisées en forêt sèche, ont obtenu une MAE de 15,2 t/ha pour une biomasse moyenne de 36,2 t/ha.

Une des principales faiblesses de cette partie de l'étude est d'avoir pris la saison de repos de végétation pour étudier l'importance du spectre, alors même que la saison de végétation est plus propice à l'étude de la biomasse. C'est particulièrement visible dans l'étude de la phénologie, où les métriques calculées sur le GRVI et le NDGI — et plus particulièrement les valeurs maximales de ces indices — sont jugées particulièrement importantes. En outre, les modèles qui incluent des différences normalisées basées sur la saison de végétation permettent systématiquement d'obtenir de meilleurs résultats qu'avec les modèles construits sur la base de la saison repos seule. Ces résultats indiquent clairement que la saison de végétation est la plus propice à la caractérisation de la biomasse et que la phénologie pourrait potentiellement être moins informative lorsque la saison de végétation est utilisée directement. Parallèlement, les variations phénologiques importantes identifiées pour les indices qui utilisent le spectre visible uniquement (GRVI), ou qui sont combinés au proche infrarouge (NDGI), et jamais pour les indices qui utilisent l'infrarouge à courte longueur d'onde, renforcent l'hypothèse que cette partie du spectre est plutôt impliquée dans l'estimation du recouvrement que dans l'estimation de la biomasse. Ceci souligne également que le spectre visible joue un rôle prépondérant et que le proche infrarouge pourrait potentiellement jouer un rôle plus important en saison de végétation.

### 2.3. Hauteur de canopée

A l'image de la couverture de la canopée, la représentation de la hauteur de la canopée est très peu influencée par les paramètres étudiés dans cette étude. Les différences observées sont marginales, avec une différence de l'ordre de 0,3 m d'erreur absolue moyenne (MAE) entre les modèles calibrés avec les données de terrain. Cet attribut structurel n'est pas influencé significativement par la saisonnalité. Cependant, les modèles construits en saison de végétation sont les plus stables et les moins biaisés. La résolution spatiale est un facteur qui influence davantage la représentation de la hauteur de canopée et les caractériser en travaillant à de plus fines résolutions, jusqu'à 30 m, est déterminant. Cette caractéristique va à l'encontre de ce qui a été mis en évidence en savane, où une résolution plus grossière permettait de diminuer notamment l'influence de l'ombre (Hansen et al., 2016; Lang et al., 2019). L'effet d'ombre est limité dans le cas des fourrés épineux. Les zones les plus ouvertes sont également celles où les dimensions de la flore sont les moins importantes, ce qui limite les projections d'ombre au sol. De plus, la zone d'étude se situe

à hauteur du tropique du Capricorne, avec un soleil au zénith durant l'été austral limitant très fortement les perturbations du signal spectral causé par les effets d'ombre. Ce point pourrait également expliquer la meilleure stabilité des modèles construits durant la saison de végétation, et pourrait s'appliquer également à la représentation de la canopée et de la biomasse.

L'étude des facteurs spectraux met en évidence l'importance du spectre visible et, dans une moindre mesure, du spectre infrarouge. Lors de l'utilisation des bandes brutes, aucune différence significative n'est observée entre les modèles construits uniquement sur le spectre visible et ceux qui incluent les infrarouges. Les indices de végétation calculés sur le proche infrarouge permettent néanmoins d'améliorer la qualité et la stabilité des modèles. En outre, les prédicteurs identifiés comme les plus influents sont le bleu, le NDVI et, dans une moindre mesure, le vert, le rouge et le NDGI. D'autres études ont également observé un léger gain de qualité des résultats avec l'utilisation d'indices de végétation et de très bons résultats uniquement à partir du spectre visible (Chrysafis et al., 2017; Korhonen et al., 2017; Lang et al., 2019). L'hypothèse d'un lien entre la résolution spatiale et les facteurs spectraux a été mise en avant par Lang et al. (2019), qui soulignent que l'importance du spectre visible et du NIR pourrait être liée à la résolution spatiale plus fine dans cette partie du spectre électromagnétique. Les fourrés épineux se prêteraient également mieux à une résolution plus fine à cause de leur hétérogénéité structurelle à l'échelle d'un pixel de 10 m.

Les résultats et hypothèses concernant l'influence de la saisonnalité et de la méthodologie utilisée pour l'estimation de la biomasse peuvent également s'appliquer à la hauteur de canopée, avec l'obtention de modèles plus stables et moins biaisés durant la saison de végétation. Les métriques phénologiques incluent également la valeur maximale d'un indice de végétation et les variations saisonnières jugées importantes incluent la saison qui suit le pic d'activité végétale. Néanmoins, dans le cas de la hauteur de canopée, le spectre visible semble jouer un rôle plus important, notamment avec l'utilisation d'indices de végétation qui incluent le bleu et le vert. A contrario, GVMI qui est calculé sur le NIR et le SWIR n'a aucun impact.

## 2.4. Modèles intégratifs

Les améliorations marginales constatées tout au long de cette étude se reflètent dans la qualité des modèles qui intègrent l'ensemble des facteurs identifiés comme influents. En effet, à l'exception de la biomasse aérienne, les modèles intégratifs n'apportent aucune amélioration en termes d'erreur résiduelle et sont parfois plus biaisés. L'importance des bandes brutes de Sentinel 2 et la très faible influence des indices de végétation et des variables phénologiques ont été soulignées pour l'ensemble des variables de réponse. Ces résultats étaient globalement attendus au vu des analyses préliminaires réalisées pour guider les choix méthodologiques. Durant ces analyses préliminaires, les parcelles étaient mal représentées sur les principaux axes de variance, avec une faible séparabilité notamment au niveau de la couverture de la canopée. L'ajout d'indices de végétation ne permettait qu'une légère amélioration de la séparabilité entre les parcelles. En outre, les coefficients de variation sur les séries temporelles étaient assez faibles pour les bandes spectrales. Dans une logique de robustesse, d'adaptabilité, de commodité et de transférabilité, il pourrait être adéquat de limiter les modèles aux prédicteurs les plus pertinents, durant une période adéquate pour chaque variable de réponse et avec un jeu de données représentatives de l'hétérogénéité de la zone d'étude. Par exemple, Karlson et al. (2015) ont déterminé le nombre optimal de variables avec une approche par *backward selection* à partir de 31 prédicteurs et ont obtenu les meilleurs résultats avec des modèles construits avec cinq prédicteurs.

### 3. Recul sur la méthodologie et sur les résultats

Plusieurs études ont suggéré que les données de recouvrement dérivées d'images à très haute résolution peuvent être plus fiables et objectives que celles récoltées sur le terrain, car elles sont moins soumises à l'effet opérateur et permettent une meilleure représentativité de l'hétérogénéité du paysage (Morrison, 2016 ; Ludwig et al., 2019). Ce problème a été traité avec un système d'échantillonnage, le *point intercept sampling*, qui minimise les estimations subjectives. Nous avons ensuite ajusté les données avec un modèle de régression d'axe majeur, qui suppose une répartition des erreurs dans les deux jeux de données, sur base d'une photo-interprétation des parcelles échantillonnées sur le terrain. Les données de photo-interprétation ont également été ajustées avec le modèle de régression d'axe majeur, de manière à ne pas les considérer comme exemptes d'erreurs. Plusieurs éléments engendrent des difficultés lors de la photo-interprétation, notamment les petits éléments ligneux peu contrastés sur les images à très haute résolution, mais qui ont été observés sur le terrain. De plus, l'utilisation des images fournies par Google ou Bings peut entraîner des divergences d'estimations dues à des dates d'acquisition différentes selon les tuiles, dont la plupart datent de plusieurs années. Ces images pourraient ne pas être complètement représentatives d'un environnement très dynamique, ce qui est le cas dans notre zone d'étude, où se pratiquent l'agriculture sur brûlis ou le charbonnage. D'autre part, obtenir une représentativité de l'hétérogénéité du paysage lors de la récolte de données sur le terrain était un défi, compte tenu des contraintes logistiques et opérationnelles, du temps imparti et du budget disponible. La combinaison d'une stratification basée sur la couverture en sol nu, modélisée en préliminaire avec des données de photo-interprétation, et des connaissances locales sur la localisation des différents écosystèmes a permis de capter un maximum de la diversité paysagère. Tous les écosystèmes d'intérêt ont été échantillonnés, chacun d'entre eux à divers niveaux de couverture de la canopée et de dégradation. Néanmoins, cela a été réalisé au prix d'un faible nombre de répétitions et d'un important regroupement des parcelles dans l'espace, une situation qui a pu influencer les résultats et conclusions tirées durant cette étude.

Cette étude a mis en avant l'influence limitée de nombreux prédicteurs. Il semblerait qu'une meilleure représentation des attributs modélisés doive passer par des prédicteurs au-delà de l'optique, ou encore par l'utilisation d'un jeu de données de meilleure qualité. La qualité du jeu de données peut en effet être plus limitante que les techniques et prédicteurs utilisés pour des modèles empiriques (Hernández-Stefanoni et al., 2018 ; Réjou-Méchain et al., 2019 ; Nagelkirk et al., 2020). Plusieurs pistes sont envisageables pour améliorer la qualité des jeux de données de terrain et de photo-interprétation. L'une d'entre-elles est l'augmentation du nombre de placettes d'échantillonnage et une meilleure répartition de celles-ci au sein de la zone d'étude. Cela pourrait entraîner une stabilisation des modèles basés sur les données de terrain et permettre alors une interprétation plus fine de l'influence des facteurs étudiés. L'importante variabilité de la qualité de modélisation entre les répétitions indique que la taille du jeu de données de terrain est insuffisante. Ce n'est pas le cas des modèles calibrés avec les données de photo-interprétation, qui ont une variance bien plus faible.

Les modèles d'estimation de la biomasse — et, dans une moindre mesure, les modèles d'estimation de la hauteur de canopée — sont caractérisés par une importante variance résiduelle par rapport à l'amplitude des valeurs observées. Plusieurs éléments sont en cause pour expliquer cette situation. La hauteur de la végétation, paramètre utilisé dans la détermination de tous les attributs forestiers modélisés à l'exception du recouvrement, a été déterminée à vue d'œil, ce qui implique des estimations plus subjectives et soumises à un important effet opérateur. Il pourrait y avoir eu un effet de propagation d'erreur dès cette étape, ce qui pourrait expliquer les plus faibles résultats obtenus pour ces variables de réponse. En outre, les

estimations locales de la biomasse, mais également leur extrapolation à plus grande échelle, sont source d'importantes propagations d'erreur. Des biais apparaissent lors d'erreurs de mesure de la circonférence et à la suite d'identifications erronées de l'espèce, ainsi qu'en raison de l'utilisation des équations allométriques, elles-mêmes entachées d'erreurs et de biais, ou encore par manque de données concernant la densité du bois. Lors du passage à une échelle supérieure, les effets de divergences spatiales entre les mesures de terrain et les données de télédétection sont également une source d'erreur (Réjou-Méchain et al., 2019). Ce dernier point a été contrôlé grâce à l'utilisation d'un GPS-GNSS d'une précision moyenne inférieure à 2 m et à la récolte de 16 points par parcelle, qui ont été utilisés pour situer le centroïde des placettes dans le système d'information géographique. La précision de localisation et le contrôle de l'homogénéité au sein des parcelles et dans leur environnement proche devraient également avoir limité l'influence de la taille des parcelles, documentée comme une importante source d'erreur lors du passage à une échelle supérieure, surtout pour les plus petites placettes d'échantillonnage où les erreurs sont exacerbées (Hernández-Stefanoni et al., 2018).

Les résultats et conclusions tirées durant cette étude dépendent également des choix méthodologiques posés avant et durant la phase de calibration des modèles. En premier lieu, comme énoncé précédemment, l'utilisation d'informations spectrales agrégées sur les cinq dernières années a pu perturber les analyses de la phénologie. Par ailleurs, l'utilisation de la saison de repos pour caractériser l'influence de la phénologie sur la caractérisation de la hauteur et la biomasse aurait pu être répétée pour chaque saison. En parallèle, l'étude préliminaire des facteurs spectraux et phénologiques a été centrée trop spécifiquement sur la couverture de la canopée. D'autres prédicteurs pertinents auraient pu être identifiés avec une analyse préliminaire propre à chaque variable de réponse. Néanmoins, la discrimination des parcelles dans l'espace des analyses en composantes principales est indépendante des attributs structuraux mesurés sur les parcelles, car l'espace des composantes est construit uniquement sur la réflectance spectrale et les indices de végétation associée à chaque parcelle. Enfin, bien que les amplitudes moyennes des différences observées entre les modèles soient faibles, elles peuvent cacher de plus importantes variations locales. Cette situation a par exemple été constatée pour la couverture de la canopée estimée avec le modèle RGB, où un effet poivre et sel est présent mais a été gommé par agrégation à l'échelle des parcelles de 900 ou 2500 m<sup>2</sup> lors de l'évaluation. Une technique qui aurait pu permettre d'étudier l'influence de la taille des parcelles aurait été d'évaluer successivement les modèles à partir de sous-placettes de 100 m<sup>2</sup> jusqu'à la taille des placettes. Enfin, il est important de souligner que les résultats obtenus sont valables à l'échelle de la zone d'étude — et non par type d'écosystème — et qu'il pourrait être opportun d'approfondir l'analyse des résultats après stratification de la zone d'étude par faciès de végétation, bien que ceux-ci soient agencés en un complexe à fortes variations locales des écosystèmes et attributs structuraux.

#### **4. Transférabilité et implications de l'étude dans un contexte global et local**

Les fourrés xérophi les du sud-ouest de Madagascar ont des attributs structuraux et une réponse aux paramètres de modélisation qui les démarquent de la plupart des écosystèmes arides, notamment des savanes, et qui pourraient faciliter d'une certaine façon leur représentation à l'échelle globale en utilisant des données représentatives de leur hétérogénéité. Cette étude a mis en évidence des réponses marginales aux paramètres phénologiques, qui sont parmi les plus complexes à intégrer dans des modèles de télédétection optique. L'utilisation de la saison la plus adaptée à la variable de réponse semble être un paramètre suffisant pour une bonne représentation des fourrés xérophi les. En outre, l'influence des



paramètres spectraux est relativement faible alors que l'utilisation d'une résolution spatiale fine est importante. Les fourrés pourraient donc amplement bénéficier du potentiel de Sentinel-2 par rapport à d'autres satellites, et même du potentiel de la constellation Planet, au vu des faibles différences observées en s'affranchissant de l'infrarouge à courte longueur d'onde. De plus, l'absence de nébulosité qui caractérise la zone d'étude et la courte durée de revisite de Sentinel-2 et de Planet facilitent l'acquisition d'images de bonne qualité à la saison adéquate. Les paramètres qui semblent être les plus déterminants, eu égard aux limites de cette étude, sont la quantité et la qualité des données utilisées pour calibrer les modèles. La meilleure représentation du recouvrement obtenue avec le modèle calibré sur les données de photo-interprétation plutôt que sur les données de terrain illustre bien ce dernier point.

La comparaison des modèles intégratifs avec les produits globaux a mis en évidence une sous-représentation généralisée et de grande ampleur des fourrés xérophiles à l'échelle globale. Les conséquences sur l'estimation de la surface forestière sur base de la définition de la FAO ont été illustrées (figure 27). Ces cartes globales sont caractérisées par un point de saturation à partir de très faibles niveaux de recouvrement et de biomasse. La carte globale de hauteur de canopée de Potapov et al. (2022) s'est tout de même démarquée, avec une représentation plus réaliste des zones de dimensions plus importantes, mais peine à représenter les écosystèmes les plus bas. À l'échelle des fourrés xérophiles, la faible qualité des produits globaux pourrait trouver son origine non pas dans le défi que représente l'application des techniques de télédétection optique aux zones arides, mais plutôt dans un manque de représentativité de ces écosystèmes dans les jeux de données utilisés pour calibrer les modèles. L'acquisition à distance de données de recouvrement et l'extrapolation d'attributs structuraux depuis ces données représentent donc une opportunité d'amélioration des cartes globales. Avec une triple corrélation très prononcée entre le recouvrement, la hauteur et la biomasse, les fourrés xérophiles se prêtent particulièrement bien à l'extrapolation d'attributs structuraux à partir des mesures de recouvrement de la canopée. Cette piste pourrait également être investiguée dans d'autres types d'écosystèmes arides pour évaluer le potentiel de la démarche et entraîner potentiellement un accroissement considérable de la quantité de données disponibles et de la qualité des modèles qui en résulteraient. Cette méthodologie implique néanmoins l'utilisation de modèles robustes qui prédisent la relation entre le recouvrement et les autres attributs, et donc une phase de récolte de données de terrain pour établir les relations locales entre ces paramètres est nécessaire dans un premier temps.

Les hypothèses et conclusions de cette étude ont également des implications à l'échelle locale. Dans une zone où les relevés de terrain sont soumis à de nombreuses contraintes logistiques, administratives et sécuritaires, le suivi des écosystèmes pourrait être appuyé par des données acquises à distance et actualisées ponctuellement. Des images acquises par drone pourraient par exemple être utilisées pour construire localement des jeux de données, ce qui permettrait alors de se passer des images à très haute résolution fournies par des entreprises telles que Google et Bings, dont la mise à jour est imprévisible et qui peuvent devenir obsolètes dans un environnement dynamique. Cette étude a montré qu'au moins trois attributs structuraux pourraient localement être dérivés d'images acquises à distance et de modèles robustes. En outre, la relative aisance avec laquelle les fourrés xérophiles peuvent être représentés à l'échelle locale avec des prédictors spectraux simples renforce l'intérêt de cette opportunité. Les cartes produites offrent un outil qui, utilisé conjointement aux connaissances locales, permet l'identification de zones prioritaires de conservation et de réhabilitation écologique, venant ainsi en appui à la gestion participative des milieux naturels.

## Conclusion

Cette étude a montré que la résolution spatiale et la saisonnalité sont deux paramètres qui influencent la caractérisation des attributs écosystémiques, que le spectre infrarouge joue un rôle déterminant et que l'utilisation d'indices de végétation et de métriques phénologiques a un effet marginal. Cette étude a également mis en avant le potentiel de la photo-interprétation pour caractériser les attributs structuraux des fourrés xérophiles du sud-ouest de Madagascar. La couverture de la canopée, mesurée directement, ainsi que la biomasse aérienne et la hauteur de canopée, extrapolées par régression linéaire, ont été représentées avec une précision supérieure ou égale à celle obtenue avec des modèles calibrés sur données issues des relevés de terrain. Néanmoins, ces données sont fortement intercorrélées. L'importante sous-représentation de ces écosystèmes dans les cartes globales a également été mise en évidence. Cette situation pourrait être principalement imputable à l'absence d'une utilisation de données représentatives du domaine d'application des modèles globaux, et moins à l'absence d'une prise en considération de prédicteurs spécifiques aux zones arides.

## Contribution

Toutes les données de calibration des modèles qui ont été utilisées dans le cadre de cette étude ont été récoltées en collaboration avec Guillaume Renard, étudiant à l'Uliège GxABT master STE, et Felana Nantenaina, doctorante à l'Université d'Antananarivo ESSA forêt, entre les mois de février et mai 2022. Les codes informatiques (R, Google Earth Engine, etc.), les analyses, les cartes et les graphiques ont été conçus personnellement à l'issue d'une réflexion collaborative ou individuelle, ainsi qu'avec l'aide des membres de la communauté internationale partageant leurs expériences et connaissances sur internet. Les données satellitaires et géographiques, ainsi que les algorithmes et programmes informatiques, ont été produits par des tiers.

## Dictionnaire des acronymes et des abréviations

$\Delta   \mathbf{x} - \mathbf{y}  $	Différence normalisée absolue entre $\mathbf{x}$ et $\mathbf{y}$	<b>MAR</b>	Régression d'axe majeur, <i>Major axis regression</i>
$\Delta$	Différence normalisée	<b>mtry</b>	Paramètre RandomForest
$\rho$	Densité	<b>Nc</b>	Nombre d'arbres
$\sigma$	Écart-type	<b>NIR</b>	Proche infrarouge, <i>near infrared</i> , bande S2
<b>ACP</b>	Analyse en composantes principales	<b>Nodesize</b>	Paramètre RandomForest
<b>AGB</b>	Biomasse aérienne, <i>above ground biomass</i>	<b>Q.25-75</b>	Écart interquartile
<b>ANOVA</b>	Analyse de la variance	<b>Q.5-95</b>	Écart interpercentile 5-95
<b>BLUE, B</b>	Bleu, bande spectrale S2	<b>QQplot</b>	Graphique Quantile-Quantile
<b>CEO</b>	Collect Earth Online	<b>RED, R</b>	Rouge, bande spectral Sentinel-2
<b>D<sub>30</sub></b>	Diamètre à 30 cm	<b>RED-EDGE</b>	Bande spectre Sentinel-2
<b>D<sub>30</sub><sup>MAX</sup></b>	Diamètre à 30 cm de la branche la plus épaisse	<b>Res.</b>	Résolution
<b>DBH</b>	Diamètre à hauteur de poitrine	<b>RFR</b>	Random Forest Regression
<b>Dim.</b>	Dimension d'une ACP	<b>RMSE</b>	Écart quadratique moyen, <i>Root mean squared error</i>
<b>E.T.</b>	Écart-type	<b>RSE</b>	Erreur quadratique relative, <i>Relative Squared Error</i>
<b>GEE</b>	Google Earth Engine	<b>S2</b>	Sentinel-2
<b>GNSS</b>	Système de positionnement par satellites global navigation satellite systems	<b>Shadow</b>	<i>Shadow feature</i> , Boruta
<b>GPS</b>	Global Positioning System	<b>SIG</b>	Système d'information géographique
<b>GREEN, G</b>	Vert, bande spectrale S2	<b>SWIR</b>	Infrarouge à courte longueur d'onde, <i>Short-Wave Infrared</i> , bande S2
<b>H95</b>	95 <sup>e</sup> percentile de la distribution de hauteur	<b>Tass.B</b>	Tasseled Cap Brithness transformation
<b>MAE</b>	= Hauteur de canopée Erreur absolue moyenne, <i>Mean absolute error</i>		

## Références bibliographiques

- Aranha J.T., Viana H.F. & Rodrigues R., 2008. VEGETATION CLASSIFICATION AND QUANTIFICATION BY SATELLITE IMAGE PROCESSING. A CASE STUDY IN NORTH PORTUGAL 9.
- 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 43.
- Bastin J.-F., Berrahmouni N., Grainger A., Maniatis D., Mollicone D., Moore R., Patriarca C., Picard N., Sparrow B., Abraham E.M., Aloui K., Atesoglu A., Attore F., Bassüllü Ç., Bey A., Garzuglia M., García-Montero L.G., Groot N., Guerin G., Laestadius L., Lowe A.J., Mamane B., Marchi G., Patterson P., Rezende M., Ricci S., Salcedo I., Diaz A.S.-P., Stolle F., Surappaeva V. & Castro R., 2017. The extent of forest in dryland biomes. *Science* **356**(6338), DOI:10.1126/science.aam6527.
- Beeri O., Phillips R., Hendrickson J., Frank A.B. & Kronberg S., 2007. Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sens. Environ.* **110**(2), 216–225, DOI:10.1016/j.rse.2007.02.027.
- Belgiu M. & Drăguț L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31, DOI:10.1016/j.isprsjprs.2016.01.011.
- Bos A.B., De Sy V., Duchelle A.E., Herold M., Martius C. & Tsendbazar N.-E., 2019. Global data and tools for local forest cover loss and REDD+ performance assessment: Accuracy, uncertainty, complementarity and impact. *Int. J. Appl. Earth Obs. Geoinformation* **80**, 295–311, DOI:10.1016/j.jag.2019.04.004.
- Breiman L., 2001. Random Forests. *Mach. Learn.* **45**(1), 5–32, DOI:10.1023/A:1010933404324.
- Brinck K., Fischer R., Groeneveld J., Lehmann S., Dantas De Paula M., Pütz S., Sexton J.O., Song D. & Huth A., 2017. High resolution analysis of tropical forest fragmentation and its impact on the global carbon cycle. *Nat. Commun.* **8**(1), 14855, DOI:10.1038/ncomms14855.
- Brinkmann K., Noromiarilanto F., Ratovonamana R.Y. & Buerkert A., 2014. Deforestation processes in south-western Madagascar over the past 40 years: what can we learn from settlement characteristics? *Agric. Ecosyst. Environ.* **195**, 231–243, DOI:10.1016/j.agee.2014.06.008.
- Calvão T. & Palmeirim J.M., 2004. Mapping Mediterranean scrub with satellite imagery: biomass estimation and spectral behaviour. *Int. J. Remote Sens.* **25**(16), 3113–3126, DOI:10.1080/01431160310001654978.
- Casse T., Milhøj A., Ranaivoson S. & Romuald Randriamanarivo J., 2004. Causes of deforestation in southwestern Madagascar: what do we know? *For. Policy Econ.* **6**(1), 33–48, DOI:10.1016/S1389-9341(02)00084-9.
- Chan J.C.-W., Beckers P., Spanhove T. & Borre J.V., 2012. An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery. *Int. J. Appl. Earth Obs. Geoinformation* **18**, 13–22, DOI:10.1016/j.jag.2012.01.002.
- Chrysafis I., Mallinis G., Siachalou S. & Patias P., 2017. Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem. *Remote Sens. Lett.* **8**(6), 508–517, DOI:10.1080/2150704X.2017.1295479.

- Crawhall N., Davies J., Dudley N., Masumi G., Hanwood W., MacKinnon K., Poulsen L., Schulte-Herbrüggen B. & Smith J., 2012. Conserving dryland biodiversity, Nairobi, Kenya, IUCN, UNCCD, UNEP-WCMC.
- Duncanson L., Armston J., Disney M., Avitabile V., Barbier N., Calders K., Carter S., Chave J., Herold M., Crowther T.W., Falkowski M., Kellner J.R., Labrière N., Lucas R., MacBean N., McRoberts R.E., Meyer V., Næsset E., Nickeson J.E., Paul K.I., Phillips O.L., Réjou-Méchain M., Román M., Roxburgh S., Saatchi S., Schepaschenko D., Scipal K., Siqueira P.R., Whitehurst A. & Williams M., 2019. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. *Surv. Geophys.* **40**(4), 979–999, DOI:10.1007/s10712-019-09538-8.
- E. D. Chaves M., C. A. Picoli M. & D. Sanches I., 2020. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sens.* **12**(18), 3062, DOI:10.3390/rs12183062.
- Eisfelder C., Kuenzer C. & Dech S., 2012. Derivation of biomass information for semi-arid areas using remote-sensing data. *Int. J. Remote Sens.* **33**(9), 2937–2984, DOI:10.1080/01431161.2011.620034.
- Eskandari S., Reza Jaafari M., Oliva P., Ghorbanzadeh O. & Blaschke T., 2020. Mapping Land Cover and Tree Canopy Cover in Zagros Forests of Iran: Application of Sentinel-2, Google Earth, and Field Data. *Remote Sens.* **12**(12), 1912, DOI:10.3390/rs12121912.
- Fassnacht F.E., Hartig F., Latifi H., Berger C., Hernández J., Corvalán P. & Koch B., 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sens. Environ.* **154**, 102–114, DOI:10.1016/j.rse.2014.07.028.
- Flores O. & Coomes D.A., 2011. Estimating the wood density of species for carbon stock assessments. *Methods Ecol. Evol.* **2**(2), 214–220, DOI:10.1111/j.2041-210X.2010.00068.x.
- Food and Agriculture Organization of the United Nations, 2019. *Trees, forests and land use in drylands: the first global assessment: full report*.
- Forkuor G., Dimobe K., Serme I. & Tondoh J.E., 2018. Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience Remote Sens.* **55**(3), 331–354, DOI:10.1080/15481603.2017.1370169.
- Fremout T., 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 Sens. Environ.* **14**.
- Gan L., Cao X., Chen X., He Q., Cui X. & Zhao C., 2022. Mapping Shrub Coverage in Xilin Gol Grassland with Multi-Temporal Sentinel-2 Imagery. *Remote Sens.* **14**(14), 3266, DOI:10.3390/rs14143266.
- Hänke H., 2016. Livelihoods on the edge: farming household income, food security and resilience in southwestern Madagascar.
- Hansen M.C., Potapov P.V., Goetz S.J., Turubanova S., Tyukavina A., Krylov A., Kommareddy A. & Egorov A., 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. *Remote Sens. Environ., Landsat 8 Science Results* **185**, 221–232, DOI:10.1016/j.rse.2016.02.023.
- Hansen M.C., Potapov P.V., Moore R., Hancher M., Turubanova S.A., Tyukavina A., Thau D., Stehman S.V., Goetz S.J., Loveland T.R., Kommareddy A., Egorov A., Chini L., Justice C.O. & Townshend J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**(6160), 850–853, DOI:10.1126/science.1244693.

- Hernández-Stefanoni J.L., Reyes-Palomeque G., Castillo-Santiago M.Á., George-Chacón S.P., Huechacona-Ruiz A.H., Tun-Dzul F., Rondon-Rivera D. & Dupuy J.M., 2018. Effects of Sample Plot Size and GPS Location Errors on Aboveground Biomass Estimates from LiDAR in Tropical Dry Forests. *Remote Sens.* **10**(10), 1586, DOI:10.3390/rs10101586.
- Higginbottom T.P., Symeonakis E., Meyer H. & van der Linden S., 2018. Mapping fractional woody cover in semi-arid savannahs using multi-seasonal composites from Landsat data. *ISPRS J. Photogramm. Remote Sens.* **139**, 88–102, DOI:10.1016/j.isprsjprs.2018.02.010.
- Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**(15), 1965–1978, DOI:10.1002/joc.1276.
- Jongman R.H.G., Skidmore A.K., Múcher C.A., Bunce R.G.H. & Metzger M.J., 2017. Global terrestrial ecosystem observations: why, where, what and how? *GEO Handb. Biodivers. Obs. Netw.* 19–38, DOI:10.1007/978-3-319-27288-7\_2.
- Karlson M., Ostwald M., Reese H., Sanou J., Tankoano B. & Mattsson E., 2015. Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sens.* **7**(8), 10017–10041, DOI:10.3390/rs70810017.
- Kattenborn T., Lopatin J., Förster M., Braun A.C. & Fassnacht F.E., 2019. UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote Sens. Environ.* **227**, 61–73, DOI:10.1016/j.rse.2019.03.025.
- Korhonen L., Hadi, Packalen P. & Rautiainen M., 2017. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sens. Environ.* **195**, 259–274, DOI:10.1016/j.rse.2017.03.021.
- Kursa M.B. & Rudnicki W.R., 2010. Feature Selection with the **Boruta** Package. *J. Stat. Softw.* **36**(11), DOI:10.18637/jss.v036.i11.
- Lang N., Jetz W., Schindler K. & Wegner J.D., 2022. A high-resolution canopy height model of the Earth.
- Lang N., Schindler K. & Wegner J.D., 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sens. Environ.* **233**, 111347, DOI:10.1016/j.rse.2019.111347.
- Lewis M., 2002. Spectral characterization of Australian arid zone plants **28**(2), 13.
- Lourenço W.R. (Ed.), 1996. *Biogéographie de Madagascar: actes du Colloque International Biogéographie de Madagascar; Société de Biogéographie; Muséum; ORSTOM; Paris (France), du 26 au 28 sept. 1995 = Biogeography of Madagascar, Colloques et séminaires / ORSTOM, Paris: ORSTOM Ed, 588.*
- Lourenço W.R. & Société de biogéographie (Eds.), 1996. *Biogéographie de Madagascar: actes du Colloque international biogéographie de Madagascar, Société de biogéographie, Muséum, ORSTOM, Paris, France, du 26 au 28 septembre 1995, Collection Colloques et séminaires, Paris: ORSTOM, 588.*
- Ludwig M., Morgenthal T., Detsch F., Higginbottom T.P., Lezama Valdes M., Nauß T. & Meyer H., 2019. Machine learning and multi-sensor based modelling of woody vegetation in the Molopo Area, South Africa. *Remote Sens. Environ.* **222**, 195–203, DOI:10.1016/j.rse.2018.12.019.
- Macintyre P., van Niekerk A. & Mucina L., 2020. Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification. *Int. J. Appl. Earth Obs. Geoinformation* **85**, 101980, DOI:10.1016/j.jag.2019.101980.

- Middleton N., Stringer L., Goudie A. & Thomas D., 2011. The forgotten Billion - MDG Achievement in the Drylands, United Nations Development Programme.
- Millard K. & Richardson M., 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sens.* **7**(7), 8489–8515, DOI:10.3390/rs70708489.
- Morrison L.W., 2016. Observer error in vegetation surveys: a review. *J. Plant Ecol.* **9**(4), 367–379, DOI:10.1093/jpe/rtv077.
- Müller H., Rufin P., Griffiths P., Barros Siqueira A.J. & Hostert P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sens. Environ.* **156**, 490–499, DOI:10.1016/j.rse.2014.10.014.
- Nagelkirk R.L. & Dahlin K.M., 2020. Woody Cover Fractions in African Savannas From Landsat and High-Resolution Imagery. *Remote Sens.* **12**(5), 813, DOI:10.3390/rs12050813.
- NASA JPL, 2020. NASADEM Merged DEM Global 1 arc second V001.
- Niemeijer D., Puigdefabregas J. & White R., 2007. Dryland Systems. In: *Millennium Ecosystem Assessment*. 625–656.
- Olson D.M. & Dinerstein E., 2002. The Global 200: Priority Ecoregions for Global Conservation. *Ann. Mo. Bot. Gard.* **89**(2), 199, DOI:10.2307/3298564.
- Potapov P., Li X., Hernandez-Serna A., Tyukavina A., Hansen M.C., Kommareddy A., Pickens A., Turubanova S., Tang H., Silva C.E., Armston J., Dubayah R., Blair J.B. & Hofton M., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* **253**, 112165, DOI:10.1016/j.rse.2020.112165.
- Powell S.L., Cohen W.B., Healey S.P., Kennedy R.E., Moisen G.G., Pierce K.B. & Ohmann J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sens. Environ.* **114**(5), 1053–1068, DOI:10.1016/j.rse.2009.12.018.
- Randriamahefasoa T.S.M. & Reason C.J.C., 2017. Interannual variability of rainfall characteristics over southwestern Madagascar. *Theor. Appl. Climatol.* **128**(1–2), 421–437, DOI:10.1007/s00704-015-1719-0.
- Randriamalala J.R., Radosy H.O., Razanaka S., Randriambanona H. & Hervé D., 2016. Effects of goat grazing and woody charcoal production on xerophytic thickets of southwestern Madagascar. *J. Arid Environ.* **128**, 65–72, DOI:10.1016/j.jaridenv.2016.01.002.
- Randriamalala J.R., Radosy H.O., Ramanakoto M., Razafindrahanta H., Ravoninjatovo J.-M., Haingomanantsoa R.S. & Ramanananantoandro T., 2022. Allometric models to predict the individual aboveground biomass of shrubs of Malagasy xerophytic thickets. *Journal of Arid Environments* 202, 104751, DOI:10.1016/j.jaridenv.2022.104751.
- Réjou-Méchain M., Barbier N., Coutron P., Ploton P., Vincent G., Herold M., Mermoz S., Saatchi S., Chave J., de Boissieu F., Féret J.-B., Takoudjou S.M. & Pélissier R., 2019. Upscaling Forest Biomass from Field to Satellite Measurements: Sources of Errors and Ways to Reduce Them. *Surv. Geophys.* **40**(4), 881–911, DOI:10.1007/s10712-019-09532-0.
- Roberts D.R., Bahn V., Ciuti S., Boyce M.S., Elith J., Guillerá-Arroita G., Hauenstein S., Lahoz-Monfort J.J., Schröder B., Thuiller W., Warton D.I., Wintle B.A., Hartig F. & Dormann C.F., 2017. Cross-

- validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**(8), 913–929, DOI:10.1111/ecog.02881.
- Saah D., Johnson G., Ashmall B., Tondapu G., Tenneson K., Patterson M., Poortinga A., Markert K., Quyen N.H., San Aung K., Schlichting L., Matin M., Uddin K., Aryal R.R., Dilger J., Lee Ellenburg W., Flores-Anderson A.I., Wiell D., Lindquist E., Goldstein J., Clinton N. & Chishtie F., 2019. Collect Earth: An online tool for systematic reference data collection in land cover and use applications. *Environ. Model. Softw.* **118**, 166–171, DOI:10.1016/j.envsoft.2019.05.004.
- Schröder J.M., Ávila Rodríguez L.P. & Günter S., 2021. Research trends: Tropical dry forests: The neglected research agenda? *For. Policy Econ.* **122**, 102333, DOI:10.1016/j.forpol.2020.102333.
- Sexton J.O., Noojipady P., Song X.-P., Feng M., Song D.-X., Kim D.-H., Anand A., Huang C., Channan S., Pimm S.L. & Townshend J.R., 2016. Conservation policy and the measurement of forests. *Nat. Clim. Change* **6**(2), 192–196, DOI:10.1038/nclimate2816.
- Sexton J.O., Song X.-P., Feng M., Noojipady P., Anand A., Huang C., Kim D.-H., Collins K.M., Channan S., DiMiceli C. & Townshend J.R., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *Int. J. Digit. Earth* **6**(5), 427–448, DOI:10.1080/17538947.2013.786146.
- Shafeian E., Fassnacht F.E. & Latifi H., 2021. Mapping fractional woody cover in an extensive semi-arid woodland area at different spatial grains with Sentinel-2 and very high-resolution data. *Int. J. Appl. Earth Obs. Geoinformation* **105**, 102621, DOI:10.1016/j.jag.2021.102621.
- Simard M., Pinto N., Fisher J.B. & Baccini A., 2011. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res. Biogeosciences* **116**(G4), DOI:10.1029/2011JG001708.
- Smith W.K., Dannenberg M.P., Yan D., Herrmann S., Barnes M.L., Barron-Gafford G.A., Biederman J.A., Ferrenberg S., Fox A.M., Hudson A., Knowles J.F., MacBean N., Moore D.J.P., Nagler P.L., Reed S.C., Rutherford W.A., Scott R.L., Wang X. & Yang J., 2019. Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities. *Remote Sens. Environ.* **233**, 111401, DOI:10.1016/j.rse.2019.111401.
- Sourdat M., 1972. Le Sud-Ouest de Madagascar : morphogenèse et pédogenèse.
- Spawn S.A. & Gibbs H.K., 2020. Global Aboveground and Belowground Biomass Carbon Density Maps for the Year 2010. *ORNL DAAC* DOI:10.3334/ORNLDAAC/1763.
- Suganuma H., Abe Y., Taniguchi M., Tanouchi H., Utsugi H., Kojima T. & Yamada K., 2006. Stand biomass estimation method by canopy coverage for application to remote sensing in an arid area of Western Australia. *For. Ecol. Manag.* **222**(1–3), 75–87, DOI:10.1016/j.foreco.2005.10.014.
- Thanh Noi P. & Kappas M., 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **18**(1), 18, DOI:10.3390/s18010018.
- Valavi R., Elith J., Lahoz-Monfort J.J. & Guillera-Arroita G., 2019. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**(2), 225–232, DOI:10.1111/2041-210X.13107.
- Vásquez-Grandón A., Donoso P.J. & Gerding V., 2018. Forest Degradation: When Is a Forest Degraded? *Forests* **9**(11), 726, DOI:10.3390/f9110726.



- Vieilledent G., Vaudry R., Andriamanohisoa S.F.D., Rakotonarivo O.S., Randrianasolo H.Z., Razafindrabe H.N., Rakotoarivony C.B., Ebeling J. & Rasamoelina M., 2012. A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. *Ecological Applications* 22(2), 572–583, DOI:10.1890/11-0039.1.
- Vieilledent G., Grinand C., Rakotomalala F.A., Ranaivosoa R., Rakotoarijaona J.-R., Allnutt T.F. & Achard F., 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. *Biol. Conserv.* **222**, 189–197, DOI:10.1016/j.biocon.2018.04.008.
- Walters M. & Scholes R.J. (Eds.), 2017. *The GEO Handbook on Biodiversity Observation Networks*, Cham: Springer International Publishing.
- Wang L., Qu J.J., Hao X. & Zhu Q., 2008. Sensitivity studies of the moisture effects on MODIS SWIR reflectance and vegetation water indices. *Int. J. Remote Sens.* **29**(24), 7065–7075, DOI:10.1080/01431160802226034.