

Cartographie et dynamiques du couvert ligneux des fourrés xérophiles du sud-ouest de Madagascar : apports méthodologiques à partir du District de Toliara II



Felana Nantenaina RAMALASON

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLoux AGRO-BIO TECH

**Cartographie et dynamiques du couvert
ligneux des fourrés xérophiles du sud-ouest de
Madagascar : apports méthodologiques à
partir du District de Toliara II**

Felana Nantenaina RAMALASON

Dissertation originale présentée en vue de l'obtention du grade de doctorat en
sciences agronomiques et ingénierie biologique

Promoteurs : Jean-François Bastin (Université de Liège – Gembloux Agro-Bio Tech)
Olivia Lovanirina Rakotondraso (Université d'Antananarivo)

Année civile : 2026

© Felana Nantenaina RAMALASON, Juillet 2026

Citation : Felana Nantenaina RAMALASON, 2026. Cartographie et dynamiques du couvert ligneux des fourrés xérophiles du sud-ouest de Madagascar : apports méthodologiques à partir du District de Toliara II. Université de Liège – Gembloux Agro-Bio Tech, Belgique

Photo de couverture

Photographies aériennes par drone des fourrés xérophiles du District de Toliara II, sud-ouest de Madagascar (2023). De gauche à droite : (1) vue nadirale des fourrés xérophiles sur dalle calcaire ; (2) vue nadirale des fourrés xérophiles sur sable roux ; (3) fourrés xérophiles en bordure du lac Tsimanapetsotsa (Parc National de Tsimanapetsotsa), avec quelques individus résiduels d'*Adansonia sp.* © Ramalason Felana Nantenaina, 2023.

**Le SEIGNEUR m'a dit : « Je vais t'enseigner et te montrer
la route à suivre. Je vais te donner un conseil en gardant les
yeux fixés sur toi**

Psaumes 32:8 PDV2017 (Parole de Vie 2017)

**The LORD says, “I will guide you along the best pathway
for your life. I will advise you and watch over you.**

Psalms 32:8 NLT (New Living Translation)

**“Hanome saina anao sy hampianatra anao izay lalan-
kalehanao Aho; hitsinjo anao ny masoko ka hanolo-tsaina
anao Aho.”**

Salamo 32:8 MG1865

Abstract

The arid ecosystems of the Toliara II District, in southwestern Madagascar, dominated by xerophytic thickets, exhibit an exceptional level of plant endemism (95%), but are subject to increasing human pressures and remain difficult to map using satellite imagery because of their atypical optical properties. This thesis proposes an integrated methodological framework to map and monitor their woody cover over a 36-year period, combining very high-resolution photointerpretation, multi-source satellite data (Sentinel-1, Sentinel-2, AlphaEarth Foundations), and Landsat time series processed in Google Earth Engine.

Local studies rely predominantly on binary forest/non-forest classifications that fail to capture the gradual degradation of xeric formations. The evaluation of four global mapping products — GFCC, GFW, CGLS-LC100 and Dynamic World — against 41 independent field surveys reveals substantial inter-product discrepancies in tree cover estimation (coefficient of variation: 78.94%). Dynamic World provides the best performance ($R^2 = 0.70$), yet remains inferior to the locally calibrated models developed in this thesis ($R^2 = 0.75$ – 0.88).

Comparing four satellite configurations and two sampling strategies shows that the spatial distribution of plots — not sensor sophistication — determines model accuracy. Accordingly, 505 photo-interpreted plots yield markedly better results than 41 field surveys ($R^2 = 0.88$ vs. 0.46 – 0.66). The minimum threshold is 15.5 dispersed plots per 1,000 km², with the optimal range falling between 20.7 and 41.4. Dispersed sampling achieves equivalent performance to clustered sampling with half the number of plots.

Over 36 years, the net cover loss is moderate (-3.2% , or $-15,185$ ha), yet conceals a profound structural transformation: intact thickets (cover $> 80\%$) declined from 62.8% to 29.0% of the landscape, while degraded thickets (60 – 80%) expanded from 18.5% to 39.5% . Four successive phases characterize this trajectory: slight expansion (1989–2001), accelerated degradation (2001–2013), attenuated decline (2013–2019) and partial recovery (2019–2025); these phase-level rates, while indicative of broad trends, are not individually significant at the 95% confidence level (Theil–Sen regression), with the structural shift from intact to degraded thickets standing as the most robust result of this analysis. Cover loss accelerates beyond 8.9 cumulative fire events over the 25-year record — equivalent to one fire every three years —, beyond which the resilience capacity of the vegetation is progressively exhausted; management recommendations aim to bring this frequency down to one fire every five years or less to keep plots within the partial-resilience zone. Cover loss accelerates beyond 8.9 cumulative fire events and decreases steadily with distance from settlements. In pastoral areas where fire frequency remains below this threshold, cover is maintained or slightly increases, suggesting that extensive small-ruminant grazing does not in itself drive directional cover loss.

This thesis establishes the first empirically quantified framework for calibration sampling optimization in tropical arid zones, demonstrates the primacy of spatial

representativeness over instrumental precision, and confirms that a continuous cover variable is essential to reveal the structural degradation that binary approaches conceal. Future directions include validating the protocol in other arid ecosystems, integrating Transformer vision models for individual crown detection, and coupling cover maps with allometric equations to estimate carbon stocks over 36 years, paving the way for integration of these formations into REDD+ mechanisms.

Keywords: woody cover, xerophytic thickets, multi-source remote sensing, calibration sampling, Landsat time series, Madagascar

Résumé

Les écosystèmes arides du district de Toliara II, dans le sud-ouest de Madagascar, dominés par les fourrés xérophiles, présentent un taux d'endémisme végétal exceptionnel (95 %), mais sont soumis à des pressions humaines croissantes et restent difficiles à cartographier par satellite en raison de leurs propriétés optiques atypiques. Cette thèse propose un cadre méthodologique intégré pour cartographier et suivre leur couvert ligneux sur 36 ans, combinant photo-interprétation à très haute résolution, données satellitaires multi-sources (Sentinel-1, Sentinel-2, AlphaEarth Foundations) et séries temporelles Landsat traitées sous Google Earth Engine.

Les études locales reposent majoritairement sur des classifications binaires forêt/non-forêt incapables de restituer la dégradation graduelle de ces formations. L'évaluation de quatre produits cartographiques globaux — GFCC, GFW, CGLS-LC100 et Dynamic World — confrontés à 41 relevés terrain révèle des écarts considérables entre produits (coefficient de variation : 78,94 %). Dynamic World offre les meilleures performances ($R^2 = 0,70$), mais reste inférieur aux modèles calibrés localement développés dans cette thèse ($R^2 = 0,75-0,88$).

La comparaison de quatre configurations satellitaires et de deux stratégies d'échantillonnage montre que c'est la répartition spatiale des placettes — et non la sophistication des capteurs — qui détermine la précision des modèles. Ainsi, 505 placettes photo-interprétées produisent des résultats nettement meilleurs que 41 relevés terrain ($R^2 = 0,88$ contre $0,46-0,66$). Le seuil minimal est de 15,5 placettes dispersées par 1 000 km², le seuil optimal entre 20,7 et 41,4. Un échantillonnage dispersé atteint des performances équivalentes à un échantillonnage groupé avec deux fois moins de placettes.

Sur 36 ans, la perte nette de couvert est modérée (-3,2 %, soit -15 185 ha). Elle masque cependant une transformation profonde : les fourrés intacts (couvert > 80 %) ont reculé de 62,8 % à 29,0 % du paysage, au profit des fourrés dégradés (60-80 %), passés de 18,5 % à 39,5 %. Quatre phases se succèdent : expansion légère (1989-2001), dégradation accélérée (2001-2013), déclin atténué (2013-2019) et récupération partielle (2019-2025) ; ces taux, bien qu'indicatifs des grandes tendances, ne sont pas individuellement significatifs au seuil de 95 % (test de Theil-Sen), la transformation structurelle (fourrés intacts vers dégradés) constituant le résultat le plus robuste de cette analyse. La perte s'accélère au-delà de 8,9 feux cumulés sur 25 ans — soit environ un feu tous les trois ans —, seuil au-delà duquel la capacité de résilience de la végétation s'épuise progressivement ; la gestion proposée vise à ramener cette fréquence à un feu tous les cinq ans ou moins pour maintenir les parcelles dans la zone de résilience partielle. La perte décroît par ailleurs régulièrement avec la distance aux habitations. Dans les zones pastorales où la fréquence des feux reste sous le seuil critique, le couvert se maintient, suggérant que le pâturage extensif n'induit pas en lui-même de perte directionnelle.

Cette thèse établit le premier cadre empirique quantifié pour l'optimisation de l'échantillonnage en zone aride tropicale, démontre la primauté de la représentativité

spatiale sur la précision instrumentale, et confirme qu'une variable continue de couvert est indispensable pour révéler la dégradation structurelle que les approches binaires occultent. Les perspectives portent sur la validation du protocole dans d'autres écosystèmes arides, l'intégration de modèles Transformer pour la détection des couronnes individuelles, et le couplage des cartes de couvert avec des équations allométriques pour estimer les stocks de carbone sur 36 ans, ouvrant la voie aux mécanismes REDD+.

Mots-clés : couvert ligneux, fourrés xérophiiles, télédétection multi-sources, échantillonnage de calibration, séries temporelles Landsat, Madagascar

Remerciements

Au terme de ce travail, je rends grâce à Dieu Tout-Puissant pour la force, la santé et la persévérance qu'Il m'a accordées tout au long de la réalisation de cette thèse.

Cette thèse est le fruit d'une collaboration entre Madagascar et la Belgique, rendue possible grâce au soutien de nombreuses personnes et institutions à qui j'exprime ma profonde gratitude.

Sur le plan institutionnel et financier, je remercie le programme PACODEL pour la bourse d'impulsion qui a permis d'initier ce travail doctoral et de réaliser un premier séjour à l'Université de Liège — Gembloux Agro-Bio Tech. Je remercie en particulier les responsables administratifs du PACODEL pour l'organisation rigoureuse des voyages et des mobilités, qui ont rendu ces séjours possibles dans les meilleures conditions. Je remercie l'Académie de Recherche et de l'Enseignement Supérieur (ARES) pour les bourses de mobilité accordées sur quatre années successives, ainsi que pour le financement du projet de recherche et de développement AGRIFO — Appui à la Gestion Durable et à la Restauration des Fourrés Xérophiles du Sud-Ouest de Madagascar — qui a soutenu l'ensemble de la recherche et les séjours en Belgique. Ces financements m'ont également offert l'opportunité de découvrir de nouveaux pays et d'enrichir mon parcours scientifique au-delà des frontières de Madagascar.

Je tiens à exprimer ma sincère reconnaissance à mon promoteur nord, le Professeur Jean-François Bastin, pour avoir supervisé ce travail avec rigueur et engagement tout au long de ces années. Sa disponibilité malgré de nombreuses responsabilités, son expertise technique, son sens critique et sa capacité à ouvrir des perspectives scientifiques ambitieuses ont été déterminants dans l'aboutissement de cette thèse et dans la publication des travaux dans des revues internationales de référence. Je remercie également mon promoteur sud, le docteur Olivia Lovanirina Rakotondrasoa, pour sa confiance, son soutien constant, ses conseils avisés et le partage de son expérience tout au long de ce parcours.

Mes remerciements vont aux membres du comité de thèse — le Professeur Jan Bogaert, le docteur Harifidy Rakoto Ratsimba, le docteur Ghislain Vieilledent et le Professeur Jeroen Meersmans — pour le suivi attentif de cette recherche, leurs recommandations constructives et leur disponibilité tout au long des différentes étapes de ce travail. Je tiens à exprimer une gratitude particulière au Professeur Jan Bogaert pour ses précieuses orientations, ses conseils avisés, ses encouragements constants et son soutien indéfectible tout au long de la réalisation de cette thèse.

Je tiens également à remercier les membres du jury qui ne faisaient pas partie du comité de thèse, le Professeur Jean-Louis Doucet, en sa qualité de président du jury, ainsi que le Professeur Zo Rabemananjara. Je leur suis reconnaissante d'avoir accepté d'évaluer ce travail et de contribuer dans l'amélioration du manuscrit. Le temps consacré à la lecture de ce manuscrit, leurs remarques et leurs questions contribuent à enrichir la qualité scientifique de cette recherche.

Je remercie le Professeur Josoa Randriamalala pour le partage de son expertise et de sa connaissance approfondie de la zone d'étude, qui ont contribué à ancrer ce travail dans les réalités écologiques et sociales du sud-ouest de Madagascar.

Cette thèse doit beaucoup à la contribution de plusieurs étudiants de Gembloux Agro-Bio Tech. Arthur Vander Linden et Guillaume Renard, dans le cadre de leurs travaux de fin d'études de master, ont apporté un appui précieux lors de la collecte des données terrain en 2022 et de la photo-interprétation sur la plateforme Collect Earth Online. Elise Cambier et Emma Daffe ont contribué aux campagnes d'acquisition d'images par drone. Leur implication et leur sérieux ont directement enrichi la qualité des données mobilisées dans cette thèse.

Je remercie chaleureusement l'équipe PIRATES — Planetary Innovative Research Across Terrestrial Ecological Systems — ainsi que l'ensemble de l'unité Biodiversité, Écosystèmes et Paysages de Gembloux Agro-Bio Tech pour les échanges scientifiques et le partage de compétences techniques. Je remercie plus particulièrement l'équipe administrative PATO de cette unité, et en particulier Emilie et Cynthia, pour leur disponibilité, leur efficacité et leur aide précieuse dans la gestion des aspects administratifs tout au long de ces années.

À Madagascar, je remercie l'équipe sud du projet AGRIFO pour son appui logistique et scientifique sur le terrain, ainsi que le laboratoire LLANDDEV pour sa contribution à la mise en œuvre des activités de recherche. Ma gratitude va également aux guides locaux qui ont facilité les campagnes de terrain dans des conditions parfois difficiles, ainsi qu'à l'association ANALASOA pour son accompagnement lors des missions dans le Parc National de Tsimanapetsotsa.

Je remercie les doctorants du projet AGRIFO — Aina, Mbolatiana et Sismondi — pour les échanges scientifiques et la solidarité qui ont marqué ce parcours commun, ainsi que les doctorants du réseau sud — Nadège, Roselande, Léa, Lisette, Timothée, Médard, John, Héritier, Henri, Nathan, Dieudonné, Salomon, Carlo, — et l'ensemble des étudiants de la maison nord-sud, pour la convivialité et le soutien mutuel qui ont accompagné ces années de formation.

Enfin, mes remerciements les plus profonds vont à ma famille, mes parents et mes sœurs, Rinah et Evah. Je remercie tout particulièrement mon mari, Avotra, pour son soutien indéfectible, ses encouragements constants, ses prières et sa patience tout au long de ces années de travail. Sa présence a été un appui essentiel dans les moments de doute comme dans les moments de satisfaction.

À toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de cette thèse et dont le nom n'est pas explicitement mentionné ici, j'adresse mes sincères reconnaissances.

Table des matières

Abstract	i
Résumé	iii
Remerciements	v
Table des matières	vii
Liste des figures.....	xiv
Liste des tableaux	xx
Liste des acronymes et abréviations	xxi
Glossaire	xxiii
Chapitre 1: Introduction générale	1
1.1. Les écosystèmes arides dans la surveillance mondiale de la végétation....	2
1.2. Limites des méthodes cartographiques existantes pour les zones arides ..	4
1.3. Problématique : trois lacunes cumulatives	6
1.4. Questions de recherche, hypothèses et objectifs	7
1.5. Structure de la thèse	10
Chapitre 2 : Cadre théorique et revue de littérature.....	12
2.1. Cartographie des végétations arides : défis et enjeux globaux.....	13
2.1.1 Biais des produits globaux dans les milieux ouverts	13
2.1.2 Définitions du couvert forestier et hétérogénéité inter-produits	14
2.1.3 Implications pour les zones arides tropicales	15
2.2. L'étendue mondiale des forêts en zones sèches.....	16
2.2.1 Une sous-estimation systématique à l'échelle planétaire.....	16
2.2.2 Méthodologie de référence : photo-interprétation	17
2.2.3 Distribution géographique mondiale des forêts sèches.....	17
2.2.4 Implications cartographiques : pourquoi les produits globaux échouent en zones arides.....	19
2.3. Méthodes de télédétection appliquées aux zones arides	20
2.3.1 Classification supervisée et Random Forest	20
2.3.2 Séries temporelles Landsat et Google Earth Engine.....	21
2.3.3 Photo-interprétation à très haute résolution (Collect Earth Online)	22
2.3.4 Indices spectraux et harmonisation inter-capteurs.....	22
2.4. Recension des travaux antérieurs sur les méthodes utilisées dans la cartographie	23
2.4.1 Référence bibliographique.....	23

2.4.2	Abstract.....	23
2.4.3	Résumé.....	24
2.4.4	Introduction.....	24
2.4.5	Matériels et méthodes	25
2.4.6	Résultats.....	25
2.4.7	Discussion.....	29
2.4.8	Conclusion	29
2.5.	Optimisation du plan d'échantillonnage de calibration.....	30
2.5.1	Densité d'échantillonnage.....	30
2.5.2	Configuration spatiale : systématique versus stratifiée-aléatoire.....	30
2.5.3	Interaction entre précision et coût logistique	31
2.6.	Dynamiques des écosystèmes xérophiles : concepts clés	32
2.6.1	États de dégradation et résilience écologique	32
2.6.2	Régimes de feu en zone aride	33
2.6.3	Pression anthropique et conversion d'usage des terres	34
2.7.	Cas du sud-ouest de Madagascar : lacunes identifiées.....	34
2.7.1	Carences en données de cartographie historique précise	35
2.7.2	Absence d'analyse systématique des dynamiques en longue série.....	35
2.7.3	Besoin d'optimisation des protocoles de calibration	35
2.8.	Synthèse du Chapitre 2.....	36
Chapitre 3: Description de la zone d'étude.....		37
3.1.	Localisation géographique et délimitation de la zone d'étude	38
3.2.	Cadre bioclimatique	39
3.2.1	Caractérisation de l'aridité.....	39
3.2.2	Saisonnalité des précipitations.....	40
3.3.	Relief et hydrographie.....	42
3.3.1	Relief.....	42
3.3.2	Hydrographie	43
3.4.	Géologie et pédologie	43
3.4.1	Substrats géologiques	43
3.4.2	Types de sols et relation végétation–sol	45

3.5.	Végétation	47
3.5.1	Gradient structural est–ouest et types de formations	47
3.5.2	Composition floristique et endémisme remarquable	47
3.5.3	Valeur de conservation de la faune	48
3.6.	Contexte socio-économique et démographique	49
3.7.	Pressions anthropiques sur les écosystèmes xérophiles	49
3.7.1	Agriculture sur brûlis (hatsake)	49
3.7.2	Production de charbon de bois	50
3.7.3	Feux de pâturage et pratiques pastorales	50
3.7.4	Gouvernance forestière et crises politiques	51
3.8.	Historique des études cartographiques antérieures dans la zone	52
Chapitre 4 : Méthodologie générale		54
4.1.	Vue d’ensemble de la démarche	55
4.2.	Données de référence	57
4.2.1	Inventaires terrain	57
4.2.2	Photo-interprétation	58
4.3.	Données prédictives	59
4.3.1	Sentinel-1/2 et AlphaEarth Foundations	59
4.3.2	Série temporelle Landsat multi-capteurs 1989-2025	60
4.3.3	Variables spectrales et temporelles	61
4.4.	Modélisation Random Forest	62
4.4.1	Paramètres d’entraînement	62
4.4.2	Importance des variables	63
4.6.	Analyse des dynamiques temporelles	64
4.6.1	Classification des états de dégradation du couvert	64
4.6.2	Tendances 1989-2025	65
4.6.3	Identification des points de changement	65
4.7.	Analyse des facteurs explicatifs	65
4.7.1	Régimes de feux — MODIS MCD64A1	65
4.7.2	Gradient de distance aux habitations — Google Open Buildings	66
4.8.	Articulation entre les articles	66

4.9.	Synthèse du chapitre 4.....	68
Chapitre 5: Potential of global vegetation maps in capturing xerophytic vegetation cover: insights from Madagascar’s arid ecosystems.....		
5.1.	Référence bibliographique	70
5.2.	Abstract	70
5.3.	Introduction.....	70
5.4.	Materials and methods	72
5.4.1	Study area	72
5.4.2	Data collection	72
5.4.3	Analytical framework	77
5.4.4	Methodological limitations related to seasonal and temporal inconsistencies.....	78
5.5.	Results.....	79
5.5.1	Heterogeneity of existing remote sensing products	79
5.5.2	Comparison of products estimates with field data.....	82
5.6.	Discussion	88
5.6.1	Inter-product heterogeneity causes and implications for arid ecosystems	88
5.6.2	Systematic biases in vegetation cover estimation across global products	88
5.6.3	Ecological factors affecting model performance: insights from vegetation typology	89
5.6.4	Resolution, seasonality and deforestation: interpreting product limitations	89
5.7.	Conclusion.....	90
Chapitre 6: The importance of the sampling design in mapping woody cover in arid ecosystems.....		
6.1.	Référence bibliographique	92
6.2.	Abstract	92
6.3.	Introduction.....	92
6.4.	Materials and methods	95
6.4.1	Study area	95
6.4.2	Satellite data acquisition and preprocessing	96
6.4.3	Calibration data sources.....	97

6.4.4	Comparative modeling and validation.....	98
6.4.5	Spatial and statistical analyses.....	99
6.4.6	Sensitivity analysis to sampling density	99
6.4.7	Evaluation of metrics and critical threshold identification	100
6.5.	Results.....	102
6.5.1	Spatial representativeness outweighs measurement precision.....	102
6.5.2	Calibration density determines spatial pattern fidelity	103
6.5.3	Sampling strategy and size control model convergence	105
6.5.4	Operational thresholds depend on spatial distribution.....	107
6.6.	Discussion	108
6.6.1	Superiority of photointerpretation for model calibration.....	108
6.6.2	Critical thresholds and spatial distribution of samples	109
6.6.3	Multi-sensor complementarity: gains conditional on sample adequacy	110
6.6.4	Limitations and methodological perspectives.....	111
6.6.5	Implications for sustainable management of arid ecosystems.....	112
6.6.6	Transferability and operational recommendations	112
6.7.	Conclusion	113
Chapitre 7: Spatio-temporal analysis of structural degradation of xerophytic thickets masked by limited net area loss: a 36-year Landsat assessment (1989–2025).....		115
7.1.	Référence bibliographique	116
7.2.	Abstract.....	116
7.3.	Introduction.....	116
7.4.	Materials and Methods.....	118
7.4.1	Study area	118
7.4.2	Satellite data acquisition and preprocessing	119
7.4.3	Calibration and spatial cross-validation data.....	120
7.4.4	Field validation data	120
7.4.5	Random Forest woody cover modelling.....	121
7.4.6	Woody cover change analysis	121
7.4.7	Classification of ecological degradation states.....	121
7.4.8	Anthropogenic drivers of woody cover dynamics.....	123

7.5.	Results	124
7.5.1	Woody cover modelling and mapping	124
7.5.2	Temporal phases of woody cover loss and recovery	125
7.5.3	Long-term woody cover dynamics	126
7.5.4	Impacts of fire regimes on woody cover.....	129
7.5.5	Distance gradient from built-up areas.....	133
7.6.	Discussion	135
7.6.1	Nonlinear degradation trajectories over 36 years	135
7.6.2	Fire as driver of structural degradation	136
7.6.3	Human proximity gradient and resource-use zonation	137
7.6.4	Methodological advances and transferability	138
7.6.5	Methodological limitations and perspectives.....	139
7.6.6	Implications for management and conservation	140
7.7.	Conclusion.....	140
Chapitre 8: Discussion générale		142
8.1.	D'un diagnostic local à une approche intégrée	143
8.2.	État de la zone d'étude et enjeux pour les acteurs concernés.....	144
8.2.1	Zonation spatiale des priorités à partir des cartes de dynamiques	144
8.2.2	La maîtrise des feux comme levier d'intervention prioritaire	147
8.2.3	Un dispositif de monitoring actualisable à faible coût.....	147
8.2.4	Valorisation des résultats par et pour les communautés locales	148
8.3.	Contribution aux Objectifs de Développement Durable	150
8.4.	Apports méthodologiques transversaux	150
8.4.1	La calibration locale comme réponse aux limites des produits globaux.	150
8.4.2	La distribution spatiale de l'échantillon prime sur la précision de mesure	153
8.4.3	La variable continue de couvert comme outil de détection précoce de la dégradation	154
8.5.	Dynamiques du couvert ligneux sur 36 ans : résultats et interprétations	155

8.5.1	Le glissement vers des états dégradés : une dynamique invisible aux classifications binaires.....	155
8.5.2	Les feux et la proximité aux habitations comme moteurs principaux de la dégradation	157
8.5.3	Une récupération partielle depuis 2019	158
8.6.	Transférabilité du cadre méthodologique.....	159
8.6.1	Conditions de transférabilité et seuils opérationnels	159
8.6.2	Incertitudes résiduelles et précautions d'interprétation.....	160
8.7.	Limites méthodologiques transversales	161
8.8.	Aspects innovants de la thèse.....	163
8.9.	Perspectives.....	163
Chapitre 9: Conclusion générale		166
Références		170
Annexes		193

Liste des figures

Figure 1-1: Discordance entre signal satellitaire et réalité terrain dans les fourrés xérophiles du District de Toliara II. (a) L'image Google Earth (résolution ~0,5 m, 2023) présente une végétation très fragmentée, quasi-indiscernable du sol nu, en raison des signatures spectrales atypiques des espèces végétales présentes dans les fourrés xérophiles. (b) La photographie par drone prise sur le même site révèle un couvert ligneux dense et continu. 5

Figure 1-2: Architecture générale de la thèse en neuf chapitres. Chapitre 1 (gris) : introduction générale et questions de recherche. Chapitres 2-4 (vert) : cadre théorique (incluant publication 1 en section 2.4), contexte de la zone d'étude (Toliara II) et méthodologie générale. Chapitres 5-7 (jaune et rose) : trois articles empiriques — Chapitre 5 (Article 2, *Journal of Arid Environments* 2025), Chapitre 6 (Article 3, *GIScience & Remote Sensing* 2026), Chapitre 7 (Article 4, en révision, *Discover Forests*). Chapitre 8 (bleu) : discussion générale et perspectives. Chapitre 9 (gris) : conclusion..... 11

Figure 2-1 : Distribution mondiale des forêts (vert) et des formations non forestières à couvert arboré $\geq 10\%$ (jaune) dans les zones arides (Bastin et al., 2017), avec localisation de de la zone d'étude. (A) Carte mondiale, avec localisation de Madagascar (encadré rouge). (B) Couvert forestier (2014) au sein des zones arides de Madagascar (UNCCD/CBD) et zoom sur la zone d'étude du District de Toliara II, montrant le même couvert forestier..... 19

Figure 2-2: Illustration des deux stratégies spatiales d'échantillonnage de calibration comparées dans l'étude. 32

Figure 3-1: Localisation géographique du District de Toliara II et de la zone d'étude. À gauche : zone d'étude effective (en jaune) au sein du District de Toliara II, avec le réseau routier et les limites communales. À droite : situation du district dans la Région Atsimo Andrefana et positionnement régional. En médaillon : localisation dans Madagascar..... 39

Figure 3-2 : Diagramme ombrothermique du District de Toliara II (normales climatologiques 1991–2020) 41

Figure 3-3: Caractérisation climatique du District de Toliara II : (A) précipitation moyenne annuelle ; (B) température moyenne annuelle ; (C) classes d'aridité (UNEP/UNCCD). Sources : CHIRPS v2.0, ERA5-Land, Trabucco & Zomer (2019). 42

Figure 3-4 : Modèle numérique de terrain (MNT) du District de Toliara II. Le relief est globalement bas et plat sur la majeure partie du district (0-150 m), à l'exception d'un massif plus élevé (jusqu'à plus de 400 m) au nord-est, en bordure du district. Source : Copernicus DEM GLO-30 (ESA), via Google Earth Engine..... 43

- Figure 3-5 : Groupes de sols (FAO-UNESCO) et substrats observés sur le terrain — District de Toliara II. Le district est dominé par les Lithosols (63 %), avec des Regosols calcaires, Arenosols et Fluvisols en complément. Les 67 points superposés montrent les substrats relevés sur le terrain par drone (2024-2025) : dalle calcaire, sable blanc, sable roux et sable roux avec roches apparentes. Source : Carte mondiale des sols FAO-UNESCO (1974), via GeoServer resiliencemada.gov.mg ; relevés de terrain par drone, 2024-202544
- Figure 3-6: Diversité des formations végétales du District de Toliara II selon les substrats pédologiques. 1. Formation arbustive à arborée sur sol ferrugineux en saison humide, couvert relativement fermé (est). 2. Même formation en saison sèche, illustrant la défeuillaison marquée des ligneux. 3. Fourré xérophile à *Alluaudia procera* sur sable roux en saison humide. 4 et 5. Fourrés xérophiles très ouverts sur dalle calcaire affleurante, à arbustes bas et contournés. 6. Végétation basse et éparse à *Euphorbia stenoclada* sur sable blanc côtier46
- Figure 3-7: Transect schématisé est-ouest illustrant le gradient structural et floristique de la végétation du District de Toliara II en fonction des substrats pédologiques.....48
- Figure 3-8: Illustrations des pressions anthropiques sur les fourrés xérophiles du District de Toliara II.51
- Figure 4-1 : Cadre méthodologique général de la thèse. Les quatre objectifs spécifiques (OS1–OS4) sont représentés par des cadres colorés organisés selon une progression logique : recension des méthodes (OS1), évaluation des produits globaux (OS2), optimisation de l'échantillonnage (OS3), et analyse des dynamiques long terme (OS4). La rangée inférieure présente les éléments méthodologiques communs aux quatre articles : données de référence, configurations prédictives, modélisation Random Forest et analyse des dynamiques.56
- Figure 4-2 : Comparaison des deux sources de données de référence : inventaire terrain (41 placettes, validation) et photo-interprétation CEO (505 placettes, calibration).....57
- Figure 4-3 : Construction de la série temporelle Landsat harmonisée (1989-2025) à partir de 1 527 scènes multi-capteurs, produisant 35 cartes annuelles de couvert ligneux.....61
- Figure 5-1: Example of mixed woody vegetation structure showing tall trees (>5 m) and understory shrubs (<5 m) in a dry forest formation of southwestern Madagascar.....73
- Figure 5-2: Map showing the location of the study area in the Toliara II District, southwestern Madagascar. The red line indicates the regional boundary, and the black line outlines the district (FTM, 2015). The study area is highlighted in yellow. Black dots represent 505 photo-interpreted plots (2500 m² each) collected using Collect Earth Online (CEO) in 2021, which were used to generate a reference map for

stratified plot selection. Green stars indicate the 41 inventory plots selected along a wood cover gradient and surveyed during the 2022 field campaign (April–May)... 76

Figure 5-3: Flowchart of the methodological framework. Left: Assessment of inter-product heterogeneity, including FAO-based classification, chi-square tests for forest area comparison, ANOVA/ Kruskal-Wallis test, and relative standard deviation maps. Right: Comparison with field data. A vegetation cover map from 505 photo-interpreted plots (RF model) guided the selection of 41 inventory plots (PIS method). Accuracy was assessed via regression and mean-relative error. Plots were grouped by floristic and structural typologies (PCA), and error differences were tested using ANOVA or Wilcoxon tests. ANOVA = Analysis of Variance; RF = Random Forest; PIS = Point Intercept Sampling; MAEr = Mean Absolute Error relative to observed mean. 78

Figure 5-4: Vegetation cover estimates and variability in the Toliara II District (Madagascar) from four global products: GFCC (2015), GFW (2019), CGLS-LC100 (2019), and DW (2019). Panel A: Maps of tree, shrub, and total woody cover (%), along with their coefficients of variation (CV). Green shades represent vegetation density; red, yellow–blue gradients show spatial variability (CV, %), with red indicating high variation. Panel B: CV distributions for each cover type. Tree cover shows the highest variability (mean CV = 78.94%), followed by shrub cover (63.08%) and woody cover (20.75%). CV was calculated as (standard deviation / mean) × 100 across products. 80

Figure 5-5: Heterogeneity in forest area estimates across 4 global vegetation products GFCC (2015), GFW (2019), DW (2019), and CGLS-LC100 (2019) in Toliara II District, Madagascar. Panel (A) shows the spatial distribution of forested areas as defined by FAO (2020), while panel (B) highlights discrepancies in total forest area among the datasets. In a context of active deforestation, older datasets such as GFCC (2015) would be expected to report substantially higher forest cover than more recent products. However, this is not observed, raising concerns about the internal consistency of these globally referenced datasets. 81

Figure 5-6 : Linear regression plots comparing tree cover, shrub cover, and wood cover estimates from global remote sensing products to 2022 field data across the Toliara II District, Madagascar. Each panel shows the product-specific regression line (in color) and the 1:1 reference line (in black), representing perfect agreement with field data. The position and slope of the regression lines highlight trends of over- or underestimation for each vegetation type. Statistical indicators (R^2 , p-value, MAEr) summarize the strength and accuracy of the relationship between remote sensing estimates and field observations. 83

Figure 5-7 : : Principal Component Analysis (PCA) biplots showing clustering of 42 vegetation plots in the Toliara II District, Madagascar. A. PCA biplot based on Hellinger-transformed species composition, with clusters defined by floristic composition. Cluster 1 (purple) represents plots dominated by dry forest species, and Cluster 2 (green) includes plots characteristic of succulent thickets. B. PCA biplot

based on structural attributes (mean height, diameter, and stem count), with clusters defined by vegetation structure. Cluster 1 (red) includes multi-stemmed, open formations; Cluster 2 (blue) corresponds to denser, woody formations. Arrows indicate the main variables influencing axis separation85

Figure 5-8 : Illustrations of the diversity of xerophytic ecosystems captured by drone in August 2023. These images showcase different types of xerophytic vegetation in the study area, highlighting the variation in species and ecosystem structure within the region.....86

Figure 5-9: Relative errors (%) in tree, shrub, and wood cover across vegetation typologies in the Toliara II District, Madagascar. (A) Errors by structural types (developed vs. multi-stem). (B) Errors by floristic composition (diversified vs. succulent). Boxplots compare deviations between field data and estimate from four global products (GFCC 2015, GFW 2019, DW 2019, CGLS-LC100 2019), with relative errors limited to $\pm 300\%$. The results highlight that some global datasets better capture specific vegetation types, emphasizing the importance of selecting context-appropriate products for ecological analyses.....87

Figure 6-1: Photointerpretation of woody cover using very high-resolution orthophotographs. Examples of (A) dense woody cover ($47/49=0.96$) and (B) sparse woody cover ($16/49=0.33$). Each 50×50 m plot (red square) contains 49 observation points in a 7×7 grid. Red circles indicate woody vegetation presence; yellow circles indicate absence. A total of 505 plots were photo-interpreted from 2021–2022 orthophotographs using Collect Earth Online (Bey et al., 2016), following environmental stratification across topographic, climatic, vegetation, and edaphic gradients.98

Figure 6-2: Methodological framework. The workflow integrates multi-source data collection (field plots, photointerpretation, and four satellite configurations: Sentinel-1(COPERNICUS/S1_GRD), Sentinel-2 (COPERNICUS/S2), Sentinel-1+2, and AlphaEarth), two sampling strategies (dispersed vs. clustered with $n=10-505$), and two calibration sources (photo-calibrated with variable n and field-calibrated with 41 plots). Models are validated using spatial cross-validation where the 41 field plots are partitioned into 5 folds (k-means) and evaluated based on performance metrics (RMSE, R^2 , Error%), prediction maps with uncertainty, and optimal thresholds and sample sizes.....101

Figure 6-3: Performance comparison between field and photointerpretation calibration for mapping woody cover with Sentinel-1+2 (A, B) and AlphaEarth (C, D). (A, C) Observed vs. predicted scatter plots demonstrate photointerpretation superiority (505 plots: RMSE = 0.11, $R^2 = 0.88$ for Sentinel-1+2; RMSE = 0.10, $R^2 = 0.88$ for AlphaEarth) over field calibration alone (41 plots: RMSE = 0.17, $R^2 = 0.66$ for Sentinel-1+2; RMSE = 0.21, $R^2 = 0.46$ for AlphaEarth). (B, D) Relative error distribution by spatial validation blocks reveals reduced and stable variability for photointerpretation calibration versus field approaches for both satellite data sources.102

Figure 6-4 : Spatial predictions of woody cover, frequency distributions, and prediction uncertainty. (A–D) Predicted woody cover maps, (E–H) frequency distributions comparing observed (grey) and predicted values (teal: field-based calibration; pink: photo interpretation-based calibration), and (I–L) prediction uncertainty maps for field-based calibration (A, C, E, G, I, K) and photo interpretation-based calibration (B, D, F, H, J, L) using Sentinel-1+2 (A, B, E, F, I, J) and AlphaEarth (C, D, G, H, K, L). Color scales indicate woody cover percentage (0–100%) and prediction uncertainty (0–40%). North arrow shown in panels A and I, scale bar shown in panels D and L apply to all map panels. Geographic coordinate system: WGS 84 (EPSG:4326)..... 104

Figure 6-5: Performance metrics across photo interpretation sample sizes and sampling strategies. RMSE (top), R^2 (middle), and relative error (bottom) for dispersed (left) and clustered (right) sampling using Sentinel-1, Sentinel-2, Sentinel-1 + Sentinel-2, and AlphaEarth. Dashed lines with triangles: field-based calibration; solid lines with circles: photo interpretation-based calibration. Vertical line indicates baseline field calibration (41 plots). Error bars show standard deviations over 100 iterations. 106

Figure 7-1: Methodological workflow for woody cover mapping and degradation analysis. The approach combines photointerpretation (n=505) and field validation (n=41) to calibrate a Random Forest model subsequently applied to the full Landsat time series (1989-2025). Temporal dynamics of woody cover are analyzed in relation to fire regimes and proximity to human built up areas. 124

Figure 7-2 : Spatial distribution of woody cover classes at six selected time steps and net woody cover change (1989–2025). Classification maps illustrate the progressive expansion of degraded classes (beige to orange tones) and the contraction of intact woody thickets (dark green) over the 36-year period, most pronounced in the northern and eastern parts of the study area. The change map (Δ woody cover, 1989–2025) quantifies pixel-level variation in woody cover percentage; red tones indicate loss and green tones indicate gain 128

Figure 7-3 : Temporal dynamics of woody cover class proportions (1989–2025). (A) Six-year interval trends in the relative contribution of five woody cover classes to total landscape area. (B) Heatmap of annual proportions by class; cell values indicate percentage of total area. Both panels show a marked decline of intact thickets between 2001 and 2013, followed by partial stabilization. The 2003 data gap reflects insufficient image availability. Values for 1992, 2000, 2009–2010, and post-2022 may be affected by phenological bias due to image compositing outside the standard January–June window..... 129

Figure 7-4: Spatial patterns of fire frequency and woody cover change (2000–2025). (A) Cumulative fire frequency derived from the MODIS MCD64A1 product (Bastin et al., 2025; Giglio et al., 2018); built-up areas (Google Open Buildings v3) are shown in blue. (B) Net woody cover change (2000–2025); red tones indicate loss and green tones indicate gain. The spatial co-occurrence of high fire frequency zones (northern

and central-eastern sectors) and areas of severe woody cover loss confirms the spatial association between fire regime and vegetation degradation.131

Figure 7-5: Relationship between fire regime and woody cover change (2000–2025). (A) Hexagonal density plot of cumulative fire frequency versus Δ woody cover; the polynomial regression curve (red) identifies a threshold at 8.9 fire events, corresponding to the inflection point beyond which degradation accelerates nonlinearly. (B) Density distribution of Δ woody cover for burned (median = -18.0%) and unburned (median = -0.1%) pixels, illustrating the magnitude of fire-driven loss. (C) Woody cover recovery trajectory as a function of time since last fire (years); the regression line shows no significant recovery trend over the observation period...133

Figure 7-6: Woody cover distribution by distance from built-up areas (1989–2025). A. Mean woody cover (%) averaged over all years within each 1-km concentric buffer zone; error bars indicate \pm standard error across years. The fitted curve (LOESS regression, shaded band = 95 % confidence interval) shows a rapid increase from 57 % at 1 km to approximately 67 % beyond 6–7 km. Annual rates of woody cover change (FAO method) range from $-0.166\% \text{ yr}^{-1}$ at 1 km to $-0.090\% \text{ yr}^{-1}$ at 10 km (Table 3). B. Proportion of woody cover classes per management zone averaged over 1989–2025. No intact thickets ($> 80\%$) were detected within 10 km of built-up areas. Built-up area locations derived from Google Open Buildings v3 (Sirko et al., 2021).134

Figure 8-1 : Zonation spatiale des priorités d'intervention dérivée des cartes de dynamiques du couvert ligneux (1989-2025) et du gradient de distance aux habitations. Le centre regroupe les clusters d'habitations et le réseau routier ; le gradient du rouge (fronts de dégradation active) au jaune (zones de récupération potentielle) puis au vert (noyaux résilients) traduit la diminution de la pression de dégradation avec l'éloignement aux habitations.146

Liste des tableaux

Tableau 1-1 : Articulation des questions, hypothèses, objectifs et chapitres	10
Tableau 2-1: Synthèse des méthodes de télédétection des écosystèmes des zones arides au niveau international et local	27
Tableau 3-1 : Statistiques de l'indice d'aridité (IA) par classe climatique au sein du District de Toliara II, calculées à partir du raster à résolution de 1 km (EPSG:32738 ; 4 747 pixels valides). L'IA est défini comme le rapport précipitations annuelles / évapotranspiration potentielle annuelle (P/ETP). La classification suit la nomenclature UNCCD (Trabucco and Zomer, 2019 ; UNCCD, 2024).	40
Tableau 4-1 : Variables spectrales et temporelles utilisées dans les modèles Random Forest (série Landsat 1989-2025).	62
Tableau 4-2: Matrice d'articulation des quatre articles : hypothèses, données et méthodes.....	67
Tableau 5-1: Comparison of GFW, GFCC, CGLS-LC100, and DW products based on key parameters relevant to global vegetation mapping. The table summarizes definitions, data sources, resolution, seasonality, availability, and methods as described by the original developers. No models were trained by the authors; this is a descriptive comparison of existing datasets	74
Tableau 6-1: Performance metrics across photo interpretation sample sizes and sampling strategies. RMSE (top), R ² (middle), and relative error (bottom) for dispersed (left) and clustered (right) sampling using Sentinel-1, Sentinel-2, Sentinel-1 + Sentinel-2, and AlphaEarth. Dashed lines with triangles: field-based calibration; solid lines with circles: photo interpretation-based calibration. Vertical line indicates baseline field calibration (41 plots). Error bars show standard deviations over 100 iterations.	107
Tableau 7-1: Woody cover classification system along an ecological degradation gradient. Continuous woody cover maps were reclassified into five discrete classes representing distinct ecological states, validated against field measurements and very high-resolution satellite image photointerpretation across 41 plots	121
Tableau 7-2: Annual rates of woody cover change by phase (1989–2025). A ₁ and A ₂ denote woody surface area (ha) at the beginning and end of each phase, respectively. Woody surface area corresponds to pixels with woody cover exceeding 10%. Annual rates were calculated following Puyravaud (2003): $r = (1/\Delta t) \times \ln(A_2/A_1) \times 100$. Years 2002–2003 were excluded due to insufficient Landsat scene availability (Scan Line Corrector failure).....	126
Tableau 7-3: Net woody cover change by distance class from built-up areas (1989–2025). Annual deforestation rate calculated using the FAO method (Puyravaud, 2003). Intensity zones are defined as follows: intensive (≤ 2 km), moderate (3–6 km), peripheral (>6 km).	135

Liste des acronymes et abréviations

AGRIFO	Appui à la Gestion Durable et à la RestauratIon des FOurrés xérophiles du sud-ouest de Madagascar
AlphaEarth	AlphaEarth Foundations (modèle géospatial d'embedding)
ANOVA	Analysis of Variance / Analyse de la Variance
ARES	Académie de recherche et d'enseignement supérieur
CBD	Convention sur la Biodiversité Biologique
CEO	Collect Earth Online
CGLS-LC100	Copernicus Global Land Service Land Cover 100m
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CNN	Convolutional Neural Network / Réseau de neurones convolutif
CV	Coefficient of Variation / Coefficient de Variation
DEM	Digital Elevation Model / Modèle Numérique de Terrain
DW	Dynamic World
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
FRA	Forest Resources Assessment
GEE	Google Earth Engine
GELOSE	Gestion Locale Sécurisée
GFCC	Global Forest Cover Change
GFW	Global Forest Watch
GLO-30	Copernicus DEM GLO-30 (résolution 30m)
GRD	Ground Range Detected (Sentinel-1)
IA	Indice d'Aridité
IEFN	Inventaire Écologique Forestier National
INSTAT	Institut National de la Statistique (Madagascar)
IUCN	International Union for Conservation of Nature
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LaSRC	Land Surface Reflectance Code
LUCC	Land Use and Land Cover Change / Changement d'Utilisation et de Couverture des Terres
MAE	Mean Absolute Error
MAEr	Mean Absolute Error relative to observed mean
MEFT	Ministère de l'Environnement et des Forêts et du Tourisme
MODIS	Moderate Resolution Imaging Spectroradiometer

MRV	Measurement, Reporting and Verification / Mesure, Notification et Vérification
MSE	Mean Squared Error
NBR	Normalized Burn Ratio
NDBI	Normalized Difference Built-up Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
ODD	Objectifs de Développement Durable
OLI	Operational Land Imager (Landsat 8/9)
ONE	Office National pour l'Environnement (Madagascar)
PCA	Principal Component Analysis / Analyse en Composantes Principales
PI	Point-Intercept (méthode)
PIR	Proche Infrarouge
QA	Quality Assessment (masques de qualité Landsat)
R ²	Coefficient de détermination
REDD+	Reducing Emissions from Deforestation and forest Degradation
RF	Random Forest
RMSE	Root Mean Square Error / Racine de l'Erreur Quadratique Moyenne
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperture Radar / Radar à Synthèse d'Ouverture
SAVI	Soil Adjusted Vegetation Index
SVM	Support Vector Machine
SWIR	Short Wave Infrared / Infrarouge à Ondes Courtes
TC	Tree Cover
TERN	Terrestrial Ecosystem Research Network (Australie)
TM	Thematic Mapper (Landsat 5)
UNCCD	United Nations Convention to Combat Desertification
UNEP	United Nations Environment Programme / Programme des Nations Unies pour l'Environnement
UICN	Union Internationale pour la Conservation de la Nature
USGS	United States Geological Survey
VOI	Vondron'Olona Ifotony (communautés locales)
WC	Wood Cover
YSF	Years Since Fire / Années depuis le dernier feu

Glossaire

Afin d'éviter toute ambiguïté terminologique, les principaux concepts mobilisés dans cette thèse sont présentés et définis ci-après.

Fourré xérophile (*xerophytic thicket*) : Formation végétale arbustive du sud-ouest malgache adaptée à la sécheresse. Dominée par les familles endémiques Didiereaceae et Euphorbiaceae, qui possèdent des feuilles réduites et épaisses ainsi que des capacités de stockage d'eau pour survivre à l'aridité (Cornet et al., 1976; Hervé et al., 2022). C'est l'écosystème d'étude central de cette thèse.

Couvert ligneux (*woody cover*) : Proportion de surface couverte par la canopée (arbres et arbustes confondus), exprimée en pourcentage continu (0-100 %). Cette approche continue, plutôt que les classifications binaires forêt/non-forêt, permet de capturer les changements graduels du couvert particulièrement importants en zones arides (Bastin et al., 2017). C'est la variable cartographiée dans cette thèse.

Dense / Continu : Qualificatifs relatifs, et non des seuils absolus, utilisés par référence aux classes de couvert définies dans cette thèse (fourré intact > 80 %, dégradé 60–80 %, etc.) plutôt qu'à une valeur fixe. La densité ou la continuité d'une formation s'apprécie à l'échelle du pixel Landsat (30 m), l'échelle spatiale d'analyse retenue dans cette thèse ; un même couvert peut ainsi apparaître "dense" à cette résolution tout en présentant une hétérogénéité visible à une échelle plus fine (placette terrain par exemple).

Arbre / Arbuste : Arbres = ligneux de hauteur > 5 m ; Arbustes = ligneux de hauteur < 5 m. Ce seuil de 5 m reprend la convention retenue par la FAO dans sa définition opérationnelle de la forêt (FAO, 2020). Dans le sud-ouest de Madagascar, les Didiereaceae et Euphorbiaceae constituent un continuum entre ces catégories.

Couvert ligneux succulent : Compartiment du couvert ligneux défini ci-dessus, regroupant les espèces aux tiges succulentes gorgées d'eau et dépourvues de feuillage pendant une grande partie de l'année (notamment Didiereaceae et Euphorbiaceae), dont la signature de réflectance est intermédiaire entre la végétation arborée dense et le sol nu, distincte de celle du bois ligneux *sensu stricto*. Cette distinction n'exclut pas ces espèces du couvert ligneux mesuré dans cette thèse ; elle signale une hétérogénéité spectrale interne au compartiment dont il faut tenir compte dans l'interprétation des cartes.

Placette : Unité spatiale délimitée sur laquelle on mesure directement le couvert, la hauteur et la composition floristique, selon les principes standards d'inventaire forestier par placettes (Picard et al., 2012). Les placettes constituent la base de calibration empirique de tous les modèles de cartographie de cette thèse.

Aridité / Indice d'aridité (P/PET) : Rapport entre précipitations moyennes sur 30 ans et évapotranspiration potentielle, base de la classification UNCCD/CBD des zones arides (ARID, SE-AR, DSHUM) utilisée dans cette thèse (Middleton, N. J. and Thomas, D. S. G., 1997).

Télé-détection : Acquisition d'information sur un objet ou une surface sans contact direct, par capteurs aéroportés ou satellitaires (Lillesand et al., 2015).

Photo-interprétation : Interprétation visuelle manuelle d'images aériennes ou satellites à très haute résolution spatiale, utilisée ici comme méthode de calibration/validation complémentaire aux placettes de terrain, notamment pour les périodes antérieures à l'archive Landsat continue (Bastin et al., 2017).

Classification supervisée (Random Forest) : Algorithme d'apprentissage automatique par agrégation d'arbres de décision (forêt d'arbres de décision), utilisé pour la classification et la régression du couvert ligneux (Breiman, 2001).

Calibration : Ajustement des paramètres d'un modèle statistique sur un sous-ensemble de données connu (ici, les placettes de terrain) pour qu'il reproduise au mieux les observations (Klemeš, 1986).

Validation : Test indépendant de la capacité du modèle déjà calibré à prédire correctement sur des données qu'il n'a pas « vues » pendant l'ajustement — la séparation calibration/validation garantit qu'un modèle ne se contente pas de reproduire les données sur lesquelles il a été réglé (Klemeš, 1986).

Validation croisée (cross-validation) : Division répétée du jeu de données en sous-ensembles d'entraînement et de test (k-fold), permettant d'estimer la robustesse d'un modèle indépendamment d'un découpage calibration/validation unique, particulièrement utile face à un nombre limité de placettes (Stone, 1974).

Métriques de validation (R^2 , RMSE, MAE, erreur relative, biais) : Ensemble d'indicateurs statistiques standards d'évaluation de la performance d'un modèle par comparaison aux données de terrain — R^2 (variance expliquée), RMSE (racine de l'erreur quadratique moyenne), MAE (erreur absolue moyenne), erreur relative (erreur normalisée par rapport à la valeur observée, exprimée en %, permettant de comparer la performance entre classes de couvert) et biais (sur- ou sous-estimation systématique) (D. N. Moriasi et al., 2007).

Dégradation : Réduction du couvert ligneux, quantifiée par la diminution du pourcentage de canopée (arbres + arbustes ≥ 1 m) entre deux périodes. Mesurée en taux annuel (%/an) ou en changement de classe écologique inférieure (Chapitre 7).

Résilience écologique : Capacité d'un écosystème à absorber une perturbation et à retrouver sa structure et son fonctionnement initiaux, ou à demeurer dans le même bassin d'attraction (Holling, 1973a).

Tipping points (seuils de basculement) : Seuils critiques au-delà desquels un écosystème bascule vers un nouvel état stable, souvent irréversible à court/moyen terme (Scheffer et al., 2001). Pour les fourrés xérophiiles : à quel niveau de couverture l'écosystème perd-il sa résilience ?

Approche continue vs binaire : L'approche continue (couvert 0-100 %) détecte simultanément dégradation graduelle et tipping points, à l'inverse des cartographies binaires forêt/non-forêt (Hansen et al., 2013) qui se révèlent invisibles aux changements graduels et tendent à sous-estimer la végétation ligneuse clairsemée des biomes arides (Bastin et al., 2017). Cette thèse privilégie l'approche continue pour capturer les dynamiques non-linéaires des écosystèmes arides.

Chapitre 1

Introduction générale

1.1. Les écosystèmes arides dans la surveillance mondiale de la végétation

Les écosystèmes arides et semi-arides couvrent environ 40 à 45 % de la surface terrestre émergée et assurent la subsistance de plus de deux milliards de personnes à travers le monde (Reynolds et al., 2007; UNCCD, 2024). Ces biomes fournissent des services écosystémiques essentiels : production de biomasse ligneuse, maintien des cycles hydrologiques et conservation d'une biodiversité endémique remarquable (D'Odorico et al., 2013; Maestre et al., 2016). On estime qu'environ 50 % du cheptel mondial dépend des pâturages des zones arides pour son alimentation (FAO, 2020b). Malgré cette importance fonctionnelle, les zones arides demeurent systématiquement sous-représentées dans les inventaires forestiers mondiaux et les dispositifs de surveillance de la végétation.

Ces déficits de surveillance ne sont pas sans conséquences. La dégradation des terres arides — diminution persistante de la productivité et des fonctions écologiques sous l'effet conjoint des variations climatiques et des activités humaines — affecte aujourd'hui environ 24 % des terres sèches mondiales (UNCCD, 2024). En Afrique, où les zones arides s'étendent sur plus d'un milliard d'hectares et abritent quelque 400 millions de personnes directement dépendantes des ressources ligneuses, cette dégradation se traduit par des pertes de productivité documentées et une réduction des stocks de carbone que les bilans forestiers mondiaux peinent à quantifier, faute de données cartographiques adaptées (FAO, 2020b; UNEP, 2021). L'absence d'outils de surveillance fiables pour ces écosystèmes conduit ainsi à sous-estimer l'ampleur de leur dégradation dans les rapports nationaux de déforestation et les mécanismes internationaux de financement tels que REDD+.

Avant l'essor des technologies de télédétection satellitaire, la cartographie de la végétation reposait essentiellement sur des approches de terrain et de photo-interprétation. Entre les années 1960 et 1990, les méthodes classiques de cartographie se focalisaient principalement sur (i) les relevés floristiques et phytosociologiques de terrain (Cornet et al., 1976; Koechlin et al., 1974) (ii) l'interprétation visuelle de photographies aériennes et de cartes géomorphologiques (Faramalala, 1988), et (iii) les inventaires forestiers manuels. Bien que ces approches soient scientifiquement solides, elles restaient limitées par leur coût élevé, leur lourdeur logistique et une couverture spatiale souvent incomplète. L'arrivée de la télédétection satellitaire dans les années 1980–1990, notamment avec les images Landsat TM (30 m de résolution), a permis d'élargir considérablement l'échelle des cartographies. Cependant, une limite majeure persistait : les écosystèmes arides, en raison de leurs signatures spectrales particulières, demeuraient difficilement discriminables par les algorithmes de classification développés principalement pour des milieux plus humides.

Cette marginalisation est documentée de manière rigoureuse par Bastin et al., (2017), qui ont démontré, à partir de la photo-interprétation de 213 000 placettes réparties dans les biomes secs, que les forêts en zones arides couvrent 1 079 millions d'hectares — soit environ 40 % de plus que les estimations antérieures (FAO, 2010; Hansen et al., 2013). Ce résultat, publié dans *Science*, révèle que la végétation

ligneuse des milieux arides est structurellement invisible aux méthodes conventionnelles. Les mécanismes précis de cet échec — seuils de définition du couvert, confusion spectrale sol-végétation, résolution spatiale insuffisante, phénologie caducifoliée — constituent le cœur de la problématique adressée par cette thèse et sont développés en détail au Chapitre 2 (§2.2). À l'échelle de Madagascar, cette sous-estimation se concentre particulièrement dans les fourrés xérophiles du Domaine du Sud, dont le District de Toliara II constitue le cas d'application central de cette thèse.

Plus récemment, Bastin et al., (2025) ont cartographié les alternatives naturelles de couvert végétal à l'échelle mondiale, confirmant que les biomes secs tropicaux sont parmi les plus concernés par les écarts entre couverture observée et potentiel naturel. Lewin et al., (2024) ont par ailleurs montré que les zones arides comptent parmi les milieux les plus menacés sur le plan de la biodiversité vertébrée, ce qui renforce l'urgence d'améliorer les outils de détection et de suivi de ces végétations. Ces travaux soulignent collectivement que les déficits de connaissance sur les végétations arides ne sont pas anecdotiques : ils ont des conséquences directes sur les politiques de conservation et les bilans carbone mondiaux.

Dans ce contexte, Madagascar représente un cas d'étude particulièrement révélateur. L'île figure parmi les pays à plus forte déforestation tropicale (Vieilledent et al., 2018) et héberge une biodiversité qualifiée d'« extraordinary » par Antonelli et al., (2022) dans *Science*, avec notamment 95 % d'endémisme spécifique dans les végétations xérophiles du sud-ouest. Ces formations — fourrés à *Didiereaceae* et *Euphorbiaceae*, forêts sèches à *Commiphora* — subissent une dégradation accélérée sous l'effet de pressions anthropiques conjuguées. Leurs dynamiques à long terme restent pourtant insuffisamment documentées, et les outils cartographiques disponibles s'avèrent structurellement inadaptés à leurs particularités spectrales et physiologiques, comme l'illustre ce travail de thèse.

C'est dans ce contexte que s'inscrit cette thèse, réalisée dans le cadre du projet AGRIFO (Appui à la Gestion Durable et à la Restauration des Fourrés xérophiles du sud-ouest de Madagascar), financé par l'Académie de Recherche et d'Enseignement Supérieur de Belgique (ARES). Le projet vise à renforcer les capacités de recherche de l'Université d'Antananarivo et d'accompagner les populations cibles du Sud-Ouest de Madagascar vers une gestion durable des fourrés xérophiles. AGRIFO repose sur une approche multidisciplinaire structurée autour de quatre thèses doctorales complémentaires portant sur la télédétection, l'agroécologie, l'élevage caprin et la socio-économie. Cette organisation vise à appréhender les enjeux écologiques et socio-économiques des fourrés xérophiles de manière intégrée.

La présente thèse constitue le volet télédétection du projet. Elle a pour objectif d'améliorer la caractérisation de la végétation ligneuse des fourrés xérophiles à travers l'évaluation de produits de télédétection existants, le développement de cartes à haute résolution de la couverture végétale et l'analyse de leur dynamique spatio-temporelle. Par ses objectifs, cette recherche contribue principalement à l'Objectif de Développement Durable 15 (Vie terrestre) en fournissant des informations spatiales fiables pour le suivi, la conservation et la restauration des écosystèmes terrestres. Elle

contribue également à l'ODD 13 (Mesures relatives à la lutte contre les changements climatiques) en améliorant la quantification des composantes de la végétation impliquées dans le stockage du carbone et le suivi des changements environnementaux. Enfin, les connaissances produites peuvent soutenir indirectement l'ODD 2 (Faim zéro) en favorisant une gestion durable des ressources végétales utilisées par les populations locales.

Au-delà de ses contributions scientifiques, cette thèse vise une valorisation opérationnelle directe : fournir aux décideurs politiques, gestionnaires d'aires protégées et communautés locales une base cartographique fiable et des méthodologies pour la gestion durable des fourrés xérophiiles.

1.2. Limites des méthodes cartographiques existantes pour les zones arides

1.2.1 Le défi spectral des végétations xérophiiles

La cartographie du couvert végétal par télédétection repose sur trois composantes interdépendantes : les données satellitaires utilisées comme prédicteurs spectraux, les données de calibration mobilisées pour entraîner et valider les modèles, et les algorithmes de traitement. Des progrès considérables ont été réalisés sur chacune d'elles — déploiement des archives Landsat harmonisées via Google Earth Engine (Gorelick et al., 2017), généralisation du random forest (Belgiu and Drăguț, 2016), accès à l'imagerie à très haute résolution. Ces avancées se sont cependant déployées principalement dans les contextes des forêts tropicales humides et des zones tempérées, laissant les défis propres aux milieux arides tropicaux largement non résolus.

Le premier de ces défis est spectral. La végétation xérophiile présente des signatures spectrales atypiques qui déjouent les algorithmes entraînés sur des végétations conventionnelles. Les espèces des familles Euphorbiaceae (*Euphorbia stenoclada*) et Didiereaceae (*Alluaudia procera*, *Didiera madagascariensis*) — dominantes dans le sud-ouest de Madagascar — sont aphyllées ou micro-phyllées : leurs troncs et tiges succulentes reflètent le rayonnement solaire d'une manière proche de celle du sol nu. En saison sèche, lorsque la végétation est en dormance foliaire, le signal spectral peut être pratiquement identique à celui d'un sol dénudé, conduisant à des sous-estimations massives du couvert réel (Abel et al., 2023; De Marzo et al., 2021). La Figure 1.1 illustre ce paradoxe : une même formation végétale paraît quasi-absente sur une image Google Earth à très haute résolution, alors qu'une photographie prise par drone révèle un couvert ligneux dense et continu. C'est précisément ce décalage qui explique pourquoi Lasry et al., (2004) avaient prédit, à partir d'images SPOT, une déforestation totale des forêts sèches de Mikea vers 2026 — une prédiction infirmée par les inventaires réalisés dans le cadre de cette thèse.

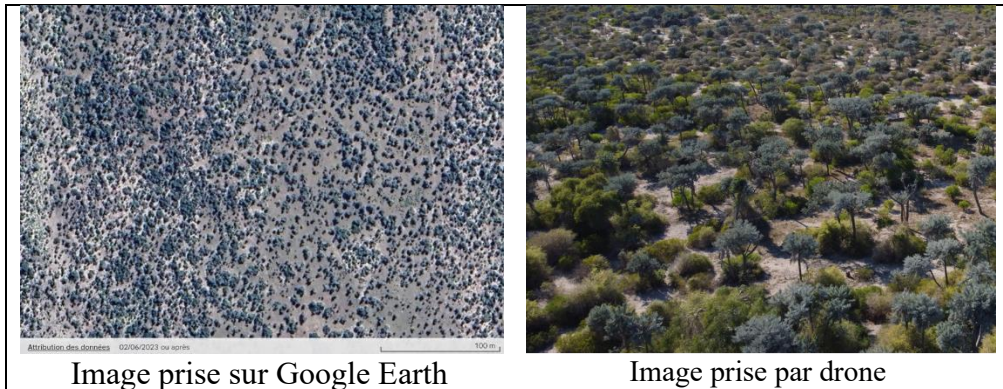


Figure 1-1: Discordance entre signal satellitaire et réalité terrain dans les fourrés xérophiles du District de Toliara II. (a) L'image Google Earth (résolution $\sim 0,5$ m, 2023) présente une végétation très fragmentée, quasi-indiscernable du sol nu, en raison des signatures spectrales atypiques des espèces végétales présentes dans les fourrés xérophiles. (b) La photographie par drone prise sur le même site révèle un couvert ligneux dense et continu.

1.2.2 La question des produits globaux en milieu aride

Le second défi tient à la multiplicité des produits globaux disponibles et à l'incertitude sur leur performance respective dans les écosystèmes xérophiles. Quatre produits de référence — GFCC, GFW, CGLS-LC100 et Dynamic World — mobilisent des définitions, des résolutions et des algorithmes profondément différents. Leurs divergences dans les milieux ouverts, et les mécanismes techniques qui les expliquent, sont analysés en détail au Chapitre 3 (§3.1). Leur performance n'avait pas été évaluée empiriquement dans le contexte du sud-ouest de Madagascar avant les travaux conduits dans cette thèse — c'est l'objet du Chapitre 6.

1.2.3 L'échantillonnage de calibration comme facteur limitant

Le troisième défi concerne l'échantillonnage de calibration des modèles de cartographie. La précision d'un modèle par apprentissage automatique dépend fondamentalement de la qualité et de la représentativité des données d'entraînement. En milieu aride, les inventaires forestiers de terrain sont coûteux, logistiquement contraints et spatialement biaisés vers les zones accessibles (McRoberts et al., 2018). Faute de consensus empirique sur la densité optimale de placettes, les études publiées rapportent des densités d'échantillonnage variant de 35 à 1 000 placettes par 1 000 km² et des tailles de placettes de 0,5 à 1 hectare, sans qu'aucun protocole optimal ne soit établi (Bastin et al., 2017; Goldblatt et al., 2017; Van Passel et al., 2020). La photo-interprétation via des outils comme Collect Earth Online (Bey et al., 2016) constitue une alternative prometteuse, permettant de produire des échantillons plus nombreux et spatialement représentatifs (Pengra et al., 2007). Mais la question de savoir si ces échantillons photo-interprétés sont équivalents — ou supérieurs — aux inventaires de terrain pour calibrer des modèles de couvert ligneux en milieu aride

n'avait pas été tranchée empiriquement avant ce travail. C'est précisément ce que cette thèse vise à démontrer au Chapitre 6.

À l'échelle de Madagascar, les études cartographiques sur les écosystèmes du sud-ouest cumulent ces trois limites. Brinkmann et al., (2014) ont utilisé des classifications binaires forêt/non-forêt sur images Landsat couvrant les communes de Beheloka, Betioky-Sud et des plaines côtières du plateau Mahafaly sur 40 ans (1973–2013), approche qui ne capture ni la diversité structurale des formations ni la dégradation graduelle du couvert. Randriamalala et al., (2015) ont cartographié les dynamiques spatiales des fourrés xérophiles de Betioky-Sud et Soalara-Sud à partir d'images Landsat 7 multi-dates (1990, 2000, 2010), en modélisant les transitions post-forestières par matrices markoviennes ; cette approche, bien que ancrée dans les fourrés xérophiles du District de Toliara II, reste limitée à des classifications catégorielles sur trois dates. Ces limitations locales reproduisent à l'échelle malgache le biais global documenté par Bastin et al., (2017) : les formations ligneuses ouvertes y sont systématiquement sous-représentées, privant les gestionnaires et les politiques de conservation d'une base de données fiable pour orienter leurs interventions.

1.2.4 Nouvelles approches et limites persistantes

Parallèlement à l'évolution des produits globaux, les conditions techniques de la cartographie des écosystèmes arides ont été profondément transformées par l'émergence de deux avancées complémentaires. D'une part, la plateforme Google Earth Engine (GEE ; Gorelick et al., 2017) a rendu accessibles des pétaoctets d'archives satellitaires précalculées, permettant de traiter des séries temporelles de plusieurs décennies à l'échelle mondiale sans recourir à des infrastructures de calcul lourdes. Cette démocratisation du calcul distribué a directement rendu possible des analyses comme celles de Bastin et al. (2017) ou les travaux de reconstruction temporelle présentés au Chapitre 8 de cette thèse.

D'autre part, l'essor des méthodes d'apprentissage automatique et, plus récemment, de l'apprentissage profond (deep learning) a ouvert de nouvelles possibilités pour l'estimation du couvert végétal. Des initiatives comme AlphaEarth — mobilisée dans l'Article 3, chapitre 6 de cette thèse — illustrent l'ambition de ces approches : combiner des images à très haute résolution et des architectures neuronales pour produire des estimations de couvert à des échelles et des précisions inédites. Malgré ces avancées, des incertitudes structurelles persistent dans les milieux arides. Les modèles entraînés sur des données majoritairement issues de forêts tropicales humides ou tempérées restent peu performants dans les formations ligneuses ouvertes, où la confusion spectrale sol-végétation et la phénologie caducifoliée limitent la transférabilité des algorithmes. C'est précisément ce constat qui motive les hypothèses et les protocoles développés dans cette thèse.

1.3. Problématique : trois lacunes cumulatives

La synthèse précédente fait apparaître trois lacunes qui se renforcent mutuellement et structurent la problématique de cette thèse.

La **première lacune est documentaire**. La littérature cartographique sur les végétations ligneuses du sud-ouest de Madagascar reste dominée par des méthodes

classiques à résolution spatiale et temporelle limitée, centrées sur des classifications binaires forêt/non-forêt qui occultent la diversité structurale des formations xérophiles et leur dégradation graduelle. Cette lacune documentaire a des conséquences opérationnelles directes : elle empêche de choisir les méthodes les plus adaptées aux réalités spectrales et physionomiques de ces végétations.

La **deuxième lacune est méthodologique**. Elle se situe à deux niveaux. D'une part, la performance des produits globaux dans les écosystèmes xérophiles malgaches n'est pas connue : on ignore dans quelle mesure ces produits divergent entre eux et dans quelle mesure cette divergence est expliquée par les caractéristiques floristiques et structurales des formations. D'autre part, les conditions d'un échantillonnage de calibration efficace pour cartographier le couvert ligneux en milieu aride — en termes de source, de densité et de configuration spatiale — n'ont pas été établies empiriquement dans ce contexte.

La **troisième lacune est dynamique**. La trajectoire du couvert ligneux de Toliara II sur plusieurs décennies n'a jamais été analysée à partir d'une variable continue traitant le couvert comme un gradient quantitatif. Cette approche est pourtant indispensable pour détecter la dégradation graduelle — réduction progressive du taux de recouvrement sans changement de catégorie — qui constitue probablement la forme dominante de changement végétal dans ces écosystèmes sous pression (Brandt et al., 2018; Watson and Venter, 2019). Les effets de seuil (tipping points) dans la réponse aux perturbations anthropiques et aux régimes de feux, documentés dans d'autres systèmes arides (Bond and Keeley, 2005; Pausas and Ribeiro, 2017), n'ont jamais été caractérisés pour ces végétations.

Ces trois lacunes sont cumulatives : l'absence de bilan méthodologique local empêche de choisir les approches les plus pertinentes ; l'absence d'évaluation des produits globaux dans ce contexte limite la fiabilité des données mobilisables ; et l'inadéquation des protocoles de calibration réduit la précision des modèles de dynamique végétale.

1.4. Questions de recherche, hypothèses et objectifs

En réponse à ces lacunes, la question centrale de cette thèse est : Comment se présente l'état et la dynamique de la couverture forestière du sud-ouest de Madagascar si on utilisait une méthode de photo-interprétation et des images à haute résolution ? Cette question se décline en deux sous-questions de recherche majeures, chacune structurant deux sous-hypothèses et leurs objectifs correspondants.

SQ1 — Quels sont les biais méthodologiques des approches cartographiques existantes, et comment les produits globaux restituent-ils le couvert xérophile du sud-ouest de Madagascar ? Cette première question adresse le diagnostic méthodologique. Elle examine les limites des approches classiques utilisées localement et évalue la performance des outils disponibles globalement.

H1 — Approches existantes insuffisantes et performance hétérogène

La littérature cartographique locale privilégie des méthodes classiques à résolution spatiale/temporelle limitée qui négligent les végétations ouvertes. Les produits

globaux, bien qu'accessibles, présentent une performance inégale dans ce contexte régional.

H1.1 — Biais des approches classiques. La cartographie en zones arides pâtit de biais méthodologiques profonds. Les méthodes locales sont dominées par des approches sur images basse résolution avec classifications binaires forêt/non-forêt, qui produisent des biais systématiques en milieux xérophiles. Ces classifications ignorent la réalité biophysique des formations arides : structures ouvertes, dégradation graduelle, et couverture végétale continue plutôt que discrète. Comme l'a démontré Bastin et al. (2017) dans leur recension mondiale couvrant 213 000 placettes, les approches binaires classiques ont systématiquement sous-estimé l'étendue des forêts sèches à l'échelle planétaire (→ +40 % d'apports nouveaux). Cette lacune documentaire se répercute opérationnellement : elle empêche de choisir les méthodes les plus adaptées aux réalités spectrales des végétations xérophiles.

H1.2 — Hétérogénéité des produits globaux. Bien que des produits de couvert ligneux existent à l'échelle mondiale Global Forest Cover Change (GFCC, Sexton et al., 2013), Global Forest Watch (GFW, Hansen et al., 2013), Copernicus Global Land Service (CGLS-LC100, Buchhorn et al., 2020) et Dynamic World (DW, Brown et al., 2022), leur performance en écosystèmes xérophiles reste inconnue. Ces produits reposent sur des définitions divergentes du couvert forestier, des seuils de classification différents, et des données satellitaires d'âges variables, ce qui génère une hétérogénéité inter-produits substantielle. En zones arides spécifiquement, les formations sont atypiques du point de vue physiognomique (structure ouverte, couverture faible, mosaïques paysagères complexes), ce qui rend les modèles globaux particulièrement vulnérables aux erreurs de commission et d'omission.

SQ2 — Cette deuxième question aborde les conditions opérationnelles d'une cartographie robuste et la reconstruction des dynamiques écologiques longues. Elle combine aspects méthodologiques et empiriques.

H2 — Déterminants opérationnels et dynamiques non-linéaires

La précision des modèles dépend fondamentalement du plan d'échantillonnage plutôt que des capteurs utilisés. Le couvert ligneux de Toliara II a suivi un déclin non-linéaire révélant des seuils critiques au-delà desquels l'écosystème perd sa résilience.

H2.1 — Plan d'échantillonnage déterminant. La densité, la configuration spatiale et la source des données de référence constituent les facteurs clés de précision des modèles de cartographie du couvert ligneux, indépendamment de la sophistication des capteurs. Contrairement à l'intuition qui privilégie l'investissement dans de meilleures données satellitaires, la littérature croissante sur l'échantillonnage en télédétection (Stehman & Foody, 2009; Olofsson et al., 2014) démontre que la représentativité spatiale des placettes — et non la résolution radiométrique des capteurs — détermine la fiabilité des modèles. Cette hypothèse repose sur l'observation que même les satellites de très haute résolution ne peuvent corriger une configuration d'échantillon biaisée ou clairsemée.

H2.2 — Déclin non-linéaire et tipping points. Le couvert ligneux de Toliara II a suivi un déclin non-linéaire (1989–2025), avec des réponses non-proportionnelles aux régimes de feux et à la pression anthropique, révélant des seuils écosystémiques

critiques. Les écosystèmes arides sont connus pour exhiber des dynamiques non-linéaires caractérisées par des points de basculement (tipping points) au-delà desquels les changements deviennent irréversibles (Holling, 1973; Scheffer et al., 2001; Bond & Keeley, 2005). Une approche en variable continue — plutôt que binaire — est indispensable pour détecter la dégradation graduelle (réduction progressive du taux de recouvrement sans changement de catégorie), qui constitue probablement la forme dominante de changement végétal sous pression anthropique (Brandt et al., 2018; Watson & Venter, 2019). Les effets de seuil dans la réponse aux perturbations anthropiques et aux régimes de feux, documentés dans d'autres systèmes arides (Pausas & Ribeiro, 2017), n'ont jamais été caractérisés pour les fourrés xérophiiles malgaches.

Ainsi, la thèse a donc pour objectif d'analyser l'état et la dynamique du couvert ligneux du sud-ouest de Madagascar en identifiant les biais méthodologiques, optimisant l'échantillonnage de calibration, et reconstruisant 36 ans de changements, pour proposer une cartographie fiable et des méthodes transférables aux zones arides tropicales.

La thèse poursuit quatre objectifs articulés selon la même progression logique que les hypothèses :

O1.1 — Synthétiser et identifier les approches méthodologiques utilisées pour cartographier les écosystèmes forestiers des zones arides à l'échelle internationale et dans le contexte spécifique du sud-ouest de Madagascar, afin d'identifier les lacunes et les biais (Section 2.4).

O1.2 — Évaluer la performance des principaux produits globaux de couvert végétal — Global Forest Cover Change (GFCC, Sexton et al., 2013), Global Forest Watch (GFW, Hansen et al., 2013), Copernicus Global Land Service (CGLS-LC100, Buchhorn et al., 2020) et Dynamic World (DW, Brown et al., 2022) — dans les végétations xérophiiles du District de Toliara II, en les confrontant à des données de terrain indépendantes et en identifiant les facteurs floristiques et structuraux qui expliquent les désaccords (Chapitre 5).

O2.1 — Développer et valider un cadre quantitatif pour l'optimisation du plan d'échantillonnage de calibration des modèles de couvert ligneux en milieu aride, en comparant source des données de référence, configuration spatiale et densité d'échantillonnage, et en dérivant des seuils opérationnels transférables (Chapitre 6).

O2.2 — Reconstruire et analyser les dynamiques du couvert ligneux de Toliara II sur 36 ans (1989–2025) à partir d'une série temporelle Landsat harmonisée, en caractérisant les phases de changement, les transitions entre états de dégradation, et en quantifiant les relations entre dynamiques du couvert, régimes de feux et gradient de pression anthropique (Chapitre 7).

Le Tableau 1.1 ci-dessous synthétise le lien entre les questions de recherche (SQ), les hypothèses (H avec leurs sous-hypothèses SH.i), les objectifs spécifiques (O) et les chapitres correspondants.

Tableau 1-1 : Articulation des questions, hypothèses, objectifs et chapitres

Question	Hypothèse	Sous-hypothèses	Objectifs	Chapitre
SQ1 : Biais méthodologiques existants et performance produits globaux	H1 : Approches insuffisantes et performance hétérogène	H1.1 : Biais des approches classiques H1.2 : Hétérogénéité des produits globaux	O1.1: Recension méthodes cartographiques O1.2: Évaluation produits globaux	Ch.2 (§2.4, Article 1) Ch.5 (Article 2)
SQ2 : Conditions opérationnelles cartographie robuste et dynamiques longues	H2 : Déterminants opérationnels et dynamiques non-linéaires	H2.1 : Plan d'échantillonnage déterminant H2.2 : Déclin non-linéaire et tipping points	O2.1: Optimisation plan d'échantillonnage O2.2: Reconstruction dynamiques 36 ans	Ch.6 (Article 3) Ch.7 (Article 4)

1.5. Structure de la thèse

Cette thèse suit une progression logique en neuf chapitres : après l'introduction générale (Chapitre 1) qui pose la question centrale et les deux questions de recherche avec leurs hypothèses et objectifs, les Chapitres 2 à 4 fournissent le cadre théorique, contextuel et méthodologique. Le Chapitre 2 présente l'état de l'art en télédétection des zones arides et inclut en section 2.4 l'Article 1 (recension des méthodes cartographiques, publié dans *Les Cahiers de l'ED GRND 2025*, adressant O1.1). Le Chapitre 3 caractérise le District de Toliara II comme zone d'étude, tandis que le Chapitre 4 expose la méthodologie générale applicable à tous les articles empiriques. Les trois articles empiriques (Chapitres 5–7) structurent ensuite la réponse aux hypothèses : le Chapitre 5 (Article 2, *Journal of Arid Environments 2025*) évalue les produits globaux de couvert végétal (O1.2), le Chapitre 6 (Article 3, *GIScience & Remote Sensing 2026*) optimise le plan d'échantillonnage de calibration (O2.1), et le Chapitre 7 (Article 4, en cours de révision chez *Discover Forests*) reconstruit les dynamiques du couvert sur 36 ans et caractérise les seuils écosystémiques critiques (O2.2). Le Chapitre 8 constitue la discussion générale et perspectives, structurée selon quatre axes : d'abord l'analyse de l'état du système étudié (District de Toliara II) et les conséquences de la dégradation observée, ensuite l'importance de ces résultats pour les acteurs locaux et régionaux (décideurs, gestionnaires, communautés), puis les aspects innovants et contributions scientifiques des travaux, en transition logique vers les perspectives. Le Chapitre 9, enfin, constitue une conclusion synthétique des réponses apportées aux questions de recherche et à la transférabilité des méthodes.

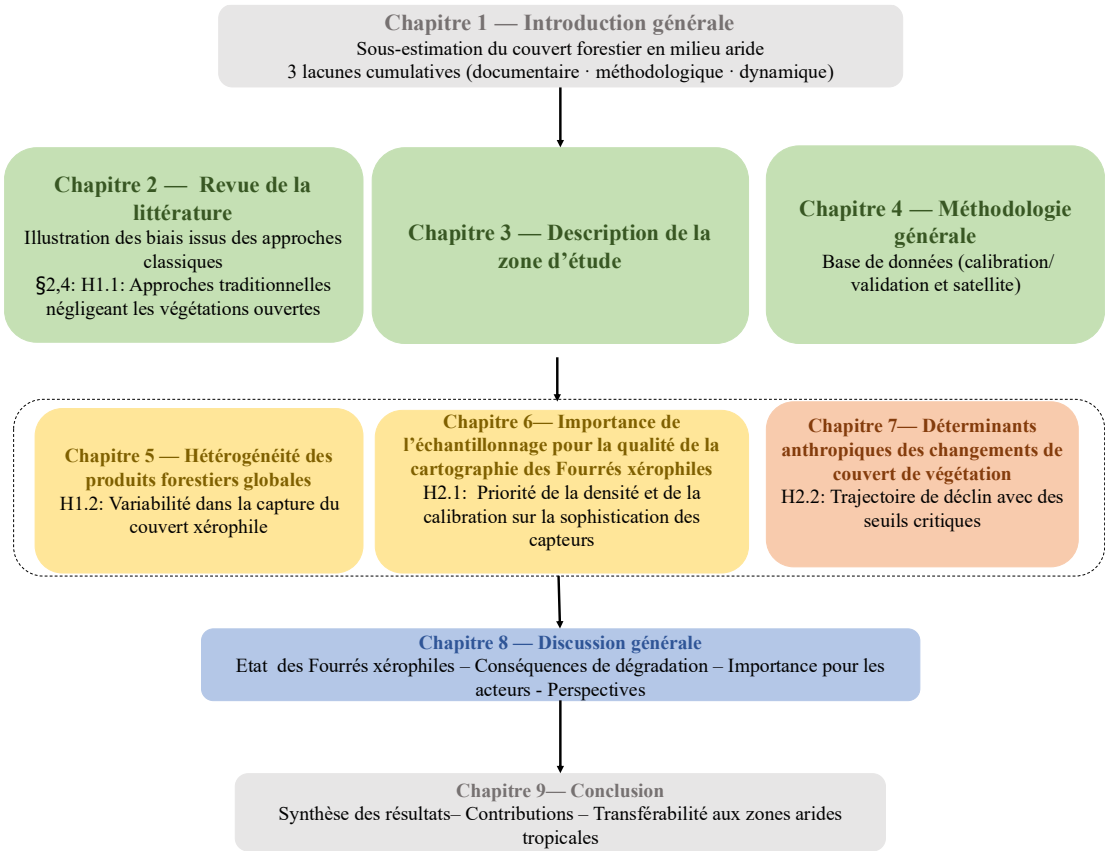


Figure 1-2: Architecture générale de la thèse en neuf chapitres. Chapitre 1 (gris) : introduction générale et questions de recherche. Chapitres 2-4 (vert) : cadre théorique (incluant publication 1 en section 2.4), contexte de la zone d'étude (Toliara II) et méthodologie générale. Chapitres 5-7 (jaune et rose) : trois articles empiriques — Chapitre 5 (Article 2, Journal of Arid Environments 2025), Chapitre 6 (Article 3, GIScience & Remote Sensing 2026), Chapitre 7 (Article 4, en révision, Discover Forests). Chapitre 8 (bleu) : discussion générale et perspectives. Chapitre 9 (gris) : conclusion

Chapitre 2

Cadre théorique et revue de littérature

Ce chapitre pose les bases théoriques et conceptuelles qui structurent les quatre études dans le cadre de cette thèse. Il est organisé en six sections. La section 2.1 analyse les mécanismes de dégradation des écosystèmes xérophiles dans une perspective théorique et comparative, en distinguant les processus irréversibles (déforestation par hatsake) des processus de dégradation sélective (carbonisation) et des perturbations récurrentes (feux pastoraux). La description factuelle des pressions anthropiques dans le contexte spécifique du District de Toliara II a été présentée au Chapitre 2 (§2.7) ; le présent chapitre en développe les fondements conceptuels pour préparer l'interprétation des dynamiques quantifiées au Chapitre 8. La section 2.2 développe les apports de Bastin et al. (2017), travail fondateur qui a documenté à l'échelle planétaire la sous-estimation des forêts en zones sèches et qui justifie directement la problématique de cette thèse. La section 2.3 passe en revue les méthodes de télédétection applicables aux milieux arides. La section 2.4 traite des fondements théoriques de l'optimisation du plan d'échantillonnage de calibration. La section 2.5 expose les concepts clés des dynamiques des écosystèmes xérophiles. La section 2.6, enfin, identifie les lacunes spécifiques au sud-ouest de Madagascar qui motivent les quatre articles de cette thèse.

2.1. Cartographie des végétations arides : défis et enjeux globaux

La cartographie de la végétation par télédétection constitue aujourd'hui le principal outil de surveillance des ressources forestières à l'échelle mondiale. Cependant, les résultats produits pour les écosystèmes arides accusent des biais systématiques qui reflètent la difficulté à adapter des méthodes conçues pour les forêts tropicales humides à des formations ligneuses ouvertes, structurellement et spectralement très différentes (Abel et al., 2023; Bastin et al., 2017). Ces biais ne sont pas anecdotiques : ils ont des implications directes pour les politiques de conservation, les inventaires nationaux de gaz à effet de serre, et la gestion des ressources ligneuses dans les pays les plus arides (UNCCD, 2024).

2.1.1 Biais des produits globaux dans les milieux ouverts

Les produits globaux de cartographie de la végétation les plus utilisés — Global Forest Cover Change (GFCC, Sexton et al., 2013), Global Forest Watch (GFW, Hansen et al., 2013), Copernicus Global Land Service (CGLS-LC100, Buchhorn et al., 2020) et Dynamic World (DW, Brown et al., 2022) — sont produits à partir d'algorithmes d'apprentissage automatique entraînés sur des bases de données de référence majoritairement issues de forêts tropicales humides et de forêts tempérées. Cette asymétrie de représentation dans les données d'entraînement introduit un biais systématique de sous-estimation du couvert dans les milieux ouverts, documenté à l'échelle planétaire par Bastin et al., (2017) et confirmé à l'échelle régionale africaine par (Abel et al., 2023).

Dans les écosystèmes arides, ce biais est amplifié par trois effets conjugués. Premièrement, la réflectance du sol nu domine le signal spectral en saison sèche, lorsque les espèces caducifoliées ont perdu leurs feuilles et que les espèces succulentes

ont réduit leur surface photosynthétique active. Les pixels Landsat (30 m) intègrent alors un mélange spectral sol-végétation qui, pour les formations ouvertes, se rapproche davantage des classes non-forestières que des classes arborées des référentiels de calibration (Bastin et al., 2017; De Marzo et al., 2021). Deuxièmement, la résolution spatiale de 30 m est inadaptée à la détection des arbres isolés ou des arbustes de faible diamètre de couronne, qui constituent une fraction significative du couvert ligneux dans les fourrés xérophiiles (Brandt et al., 2020). Troisièmement, la phénologie caducifoliée de nombreuses espèces de forêts sèches (*Commiphora*, *Euphorbia* à feuilles caduques) fait varier l'indice de végétation entre 0,1 et 0,6 selon la saison, rendant toute classification mono-date fortement dépendante de la période d'acquisition de l'image (Abel et al., 2023).

2.1.2 Définitions du couvert forestier et hétérogénéité inter-produits

L'absence de définition universelle du terme « forêt » constitue l'une des sources les plus profondes de discordance entre les produits globaux. Sexton et al., (2015) ont montré que les estimations de la superficie forestière mondiale varient de 32,1 à 41,4 millions de km² selon la définition retenue, soit un écart de près de 30 %. Cet écart est directement lié aux seuils de couvert utilisés pour délimiter la classe forêt : les définitions institutionnelles varient de 10 % (FAO/FRA) à 30 % (IUCN) selon les organisations, et certains produits n'appliquent aucun seuil défini a priori mais laissent l'algorithme de classification déterminer les frontières de classe (Brown et al., 2022).

Dans les milieux ouverts, le choix du seuil est décisif. Un seuil de 30 % exclut la majorité des fourrés xérophiiles africains et malgaches, dont le couvert arboré est structurellement inférieur à cette valeur (Bastin et al., 2017). Un seuil de 10 % intègre ces formations mais introduit des confusions avec les savanes herbeuses à arbre isolé. La distinction entre arbres et arbustes — que les produits optiques à 30 m sont incapables d'opérer de manière fiable — accentue cette incertitude : dans les fourrés à *Didiereaceae* et *Euphorbiaceae* du sud-ouest malgache, les individus ligneux atteignent rarement 5 m de hauteur, les plaçant dans une zone-frontière entre les définitions arborées et arbustives (Abel et al., 2023; Brinkmann et al., 2014).

Dans cette thèse, la dégradation est définie comme une diminution du couvert ligneux observée dans le temps au sein d'une même unité spatiale, et non comme un simple écart à un seuil de référence universel. Le seuil de 10 % n'est donc pas mobilisé ici comme indicateur de dégradation en soi, mais comme une borne inférieure permettant de circonscrire les formations xérophiiles structurellement ouvertes : toute variation observée autour de ce seuil doit être interprétée avec prudence, une partie pouvant relever de l'incertitude de mesure propre aux produits globaux à 30 m plutôt que d'un changement écologique réel.

Au-delà de cette frontière de hauteur, le statut botanique des *Didiereaceae* et *Euphorbiaceae* pose un problème proprement spectral : leurs tiges succulentes, gorgées d'eau et dépourvues de feuillage pendant une grande partie de l'année, présentent une signature de réflectance intermédiaire entre la végétation arborée dense

et le sol nu, distincte de celle du bois ligneux *sensu stricto*. Cette particularité invite à distinguer, au sein du couvert ligneux défini dans le Glossaire, le compartiment du couvert ligneux succulent (cf. Glossaire) plutôt que de le fondre indistinctement dans la catégorie générale.

Le choix d'un seuil de 10 % pour caractériser la dégradation (cf. Glossaire) ne relève pas d'une convention arbitraire : il répond à l'exigence méthodologique de dépasser la marge d'erreur instrumentale des capteurs employés, afin que tout signal détecté puisse être attribué sans ambiguïté à un changement écologique réel plutôt qu'au bruit de mesure. Bastin et al. (2017) ont montré que les estimations du couvert forestier en zone aride sont particulièrement sensibles aux erreurs de classification à proximité des seuils bas, ce qui renforce la nécessité de fixer un seuil opérationnel suffisamment éloigné de cette zone d'incertitude. La quantification précise de cette marge d'erreur (RMSE, biais absolu, RRMSE, pBias) pour les données mobilisées dans cette thèse est développée au Chapitre 4.

À cette difficulté liée à la végétation s'ajoute celle du sol nu, omniprésent dans les paysages arides et semi-arides. Le sol nu partage avec la végétation sénescente ou squelettique des signatures spectrales proches dans le visible et le proche infrarouge, ce qui complique la séparation entre couvert ligneux clairsemé et substrat minéral exposé, en particulier en saison sèche lorsque la phénologie caducifoliée accentue cette confusion. Les variations d'humidité, de texture et de couleur du sol — encore amplifiées par l'hétérogénéité géologique du sud-ouest malgache — introduisent un bruit spectral additionnel qui dégrade la précision des cartes de couvert produites par des algorithmes majoritairement calibrés et validés dans des contextes où le sol nu occupe une part marginale du paysage.

Ces divergences de définition se traduisent par des désaccords cartographiques substantiels, documentés à plusieurs échelles. Sexton et al. (2015) ont montré que les estimations de la superficie forestière mondiale varient de près de 30 % selon le produit utilisé, en raison de seuils de définition incompatibles. Dans quelle mesure ces désaccords affectent spécifiquement les quatre produits retenus dans cette thèse dans le contexte des végétations xérophiles de Toliara II — c'est précisément l'objet de l'évaluation conduite au Chapitre 6.

Cette hétérogénéité de définitions a des implications pratiques directes : deux questionnaires utilisant deux produits différents pour surveiller le couvert d'une même zone aride peuvent aboutir à des diagnostics opposés sur l'état ou la tendance de cette végétation. La compréhension de ces sources de divergence est un prérequis indispensable à tout travail de cartographie comparative, ce qui motive directement les analyses du Chapitre 6.

2.1.3 Implications pour les zones arides tropicales

Les zones arides tropicales, qui couvrent environ 45 % de la surface émergée (Právělie et al., 2019), concentrent certaines des forêts les plus menacées et les moins documentées. Elles abritent plus de deux milliards de personnes directement dépendantes des ressources ligneuses pour le bois-énergie et l'alimentation animale (Reynolds et al., 2007; UNCCD, 2024). La sous-estimation du couvert dans ces

biomes a des conséquences qui dépassent la précision cartographique : elle conduit à des sous-estimations du stock de carbone terrestre, à des biais dans les modèles de dynamique climatique globale, et à une invisibilisation de ces écosystèmes dans les mécanismes internationaux de financement de la conservation tels que REDD+ (Bastin et al., 2017; Brandt et al., 2020).

À l'échelle de Madagascar, les estimations nationales de la couverture forestière produites entre 1996 et 2015 ont systématiquement exclu les formations à canopée ouverte, réduisant ainsi les fourrés xérophiles à des zones non-forestières dans les bilans officiels (IEFN, 1996; MEFT et al., 2009; ONE et al., 2013). Cette exclusion représente une perte d'information sur plusieurs millions d'hectares d'écosystèmes à forte valeur de conservation, dont les fourrés à Didiereaceae du Domaine du Sud constituent un cas emblématique (Razanaka, 2004).

2.2. L'étendue mondiale des forêts en zones sèches

L'article de Bastin et al., (2017), publié dans *Science*, constitue un point d'inflexion dans la compréhension de la couverture forestière mondiale. En révélant l'ampleur de la sous-estimation des forêts dans les biomes secs, ce travail a posé les bases d'une remise en cause méthodologique dont les implications locales, à l'échelle du sud-ouest de Madagascar, structurent directement la problématique de cette thèse. Cette section en développe les résultats, la méthode, et les implications.

2.2.1 Une sous-estimation systématique à l'échelle planétaire

Bastin et al., (2017) ont estimé la superficie des forêts dans les biomes secs à partir d'une démarche de photo-interprétation systématique portant sur 213 000 placettes réparties dans l'ensemble des biomes arides et semi-arides du monde. Leur résultat principal est que les forêts en zones sèches couvrent 1 079 millions d'hectares, soit environ 40 % de plus que les estimations antérieures les plus larges, qui plafonnaient à 500-550 millions d'hectares (FAO, 2010; Hansen et al., 2013). Cette différence correspond à 1,3 milliard d'hectares non répertoriés dans les inventaires globaux précédents — une superficie supérieure à celle du Canada entier.

Cette sous-estimation n'est pas aléatoire : elle est systématiquement concentrée dans les biomes secs à végétation ouverte, précisément là où les algorithmes entraînés sur des forêts fermées perdent leur sensibilité. En Afrique subsaharienne, les nouveaux chiffres atteignent jusqu'au double des estimations antérieures. La région méditerranéenne, l'Asie centrale, l'Amérique du Sud (Caatinga, Chaco), et le pourtour australien (Pindan) figurent également parmi les zones où l'écart est le plus marqué. La distribution de la densité d'arbres par biome, telle qu'établie par Bastin et al., (2017), souligne que les zones les plus sous-estimées correspondent aux régions à faible densité de données de terrain de calibration — un biais méthodologique que leur approche vise précisément à corriger.

À l'échelle de Madagascar, cette sous-estimation se manifeste par l'absence des fourrés xérophiles du sud-ouest dans les estimations nationales antérieures. Brinkmann et al., (2014) et Randriamalala et al., (2015), ont montré, à partir de

données de terrain, que les formations ligneuses de cette région sont systématiquement déclassées dans les cartes issues de classifications binaires à 30 m, en raison de la confusion spectrale sol-végétation en saison sèche.

2.2.2 Méthodologie de référence : photo-interprétation

La force de la démarche de Bastin et al., (2017) réside dans son ancrage sur l'observation directe d'images à très haute résolution, plutôt que sur la classification automatique d'images multispectrales à 30 m. Les 213 000 placettes carrées (1 ha) ont été distribuées de manière systématique dans les biomes secs de l'ensemble des continents. Pour chaque placette, des photo-interprètes formés ont estimé le couvert arboré (en pourcentage) à partir d'images Google Earth, Bing, et d'autres sources à très haute résolution disponible gratuitement en ligne.

Ce protocole a été mis en œuvre via la plateforme Open Foris Collect Earth, développée par la FAO, qui structure l'interface de photo-interprétation et assure la traçabilité des données collectées. Collect Earth Online (CEO), version dérivée développée ultérieurement par Google et l'Université de Clark, reprend ce même paradigme et constitue l'outil utilisé pour la calibration des modèles dans les Chapitres 4 et 5 de cette thèse (Bey et al., 2016). Le lien méthodologique entre l'approche de Bastin et al., (2017) et les travaux de cette thèse est donc direct : la démarche de calibration par photo-interprétation sur images à très haute résolution constitue une réponse opérationnelle aux biais des classifications automatiques documentés dans leur article.

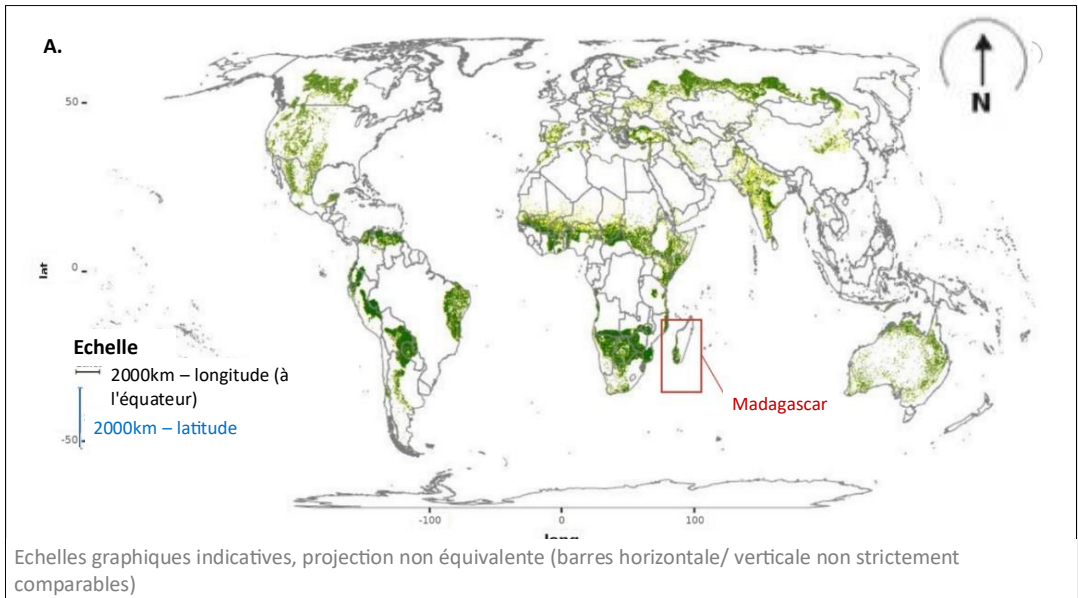
La robustesse de l'approche a été assurée par la participation de plus de 200 opérateurs locaux, géographiquement distribués, réduisant les biais d'interprétation liés à la méconnaissance de la flore locale. Cette dimension participative constitue un atout supplémentaire dans des écosystèmes comme les fourrés xérophiles de Madagascar, où la distinction entre un *Alluaudia* sans feuilles et un sol calcaire nu est difficile à opérer sans connaissances floristiques locales.

2.2.3 Distribution géographique mondiale des forêts sèches

L'analyse de Bastin et al., (2017) permet de localiser avec précision les grands ensembles de forêts sèches sous-estimées à l'échelle mondiale (Figure 2.1). L'Afrique subsaharienne concentre la plus grande superficie de forêts sèches non répertoriées, notamment dans la zone soudano-sahélienne (Sahel, bassin du lac Tchad), les forêts claires de miombo d'Afrique australe et orientale, et les forêts de Mopane d'Afrique du Sud. Le pourtour méditerranéen, en particulier le Maghreb et la péninsule ibérique, accuse également un déficit important. En Amérique du Sud, la Caatinga brésilienne et le Chaco argentin et paraguayen représentent des millions d'hectares absents des estimations globales précédentes. En Asie, les forêts claires de l'Asie centrale (Kazakhstan, Iran) et du sous-continent indien (forêts de teck et forêts sèches du Deccan) sont concernées.

Madagascar occupe une position particulière dans ce contexte : l'île figure parmi les pays où l'écart entre estimations antérieures et nouvelles estimations est proportionnellement le plus élevé pour les formations ouvertes (Bastin et al., 2017;

Figure 2.1.A). Les cartes B montre que le couvert forestier en zone aride à Madagascar se concentra essentiellement dans la moitié sud de l'île, conformément à la délimitation UNCC/CBD des zones arides. Le Domaine du Sud — qui correspond approximativement à la Région Atsimo Andrefana et aux régions limitrophes — est l'épicentre de cette sous-estimation à l'échelle de l'île, en raison de la quasi-absence de couvert herbacé sous la strate ligneuse et de la structure aphyllé en saison sèche. Le zoom sur la zone d'étude (Figure 2.1B, encadré en rouge) révèle une couverture forestière dense et continue à l'ouest, qui se fragmente progressivement vers l'est, illustrant le gradient d'aridité caractéristique du District de Toliara II. Cette spécificité floristique et spatiale place le District de Toliara II comme un cas d'application paradigmatique des conclusions de Bastin et al., (2017).



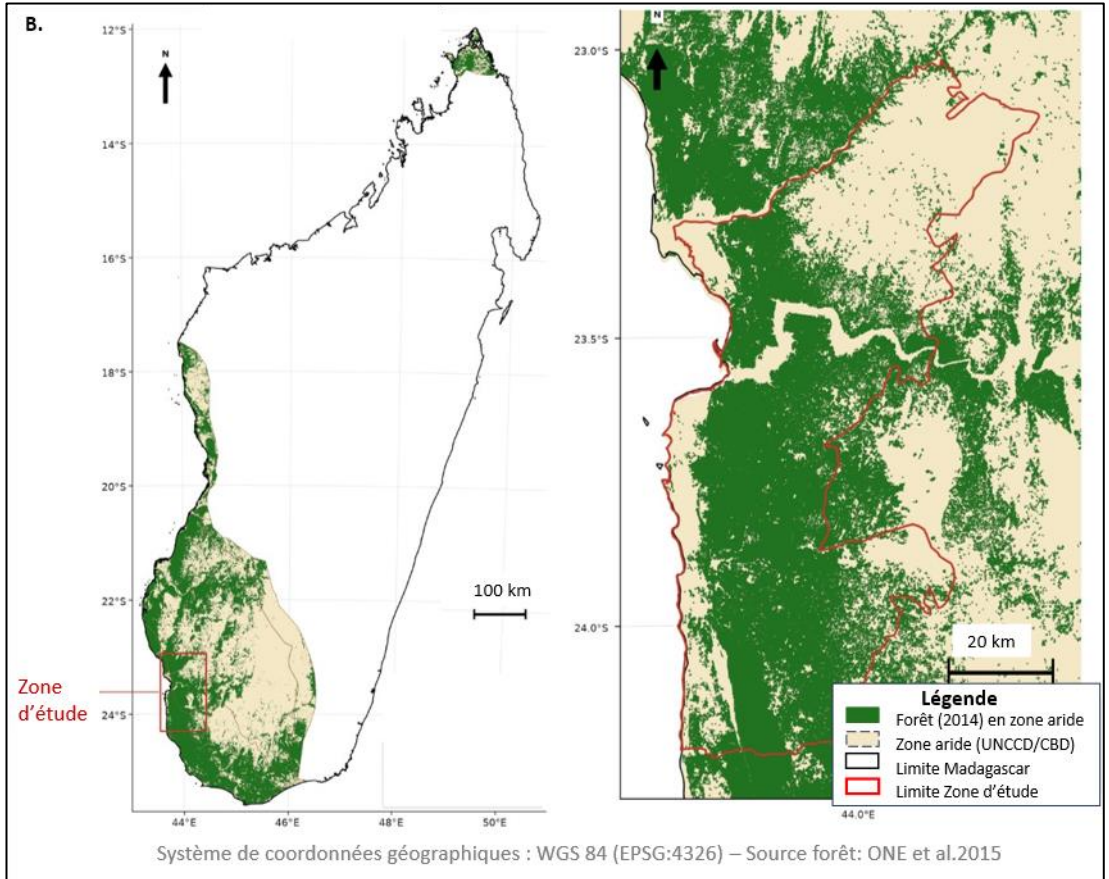


Figure 2-1 : Distribution mondiale des forêts (vert) et des formations non forestières à couvert arboré $\geq 10\%$ (jaune) dans les zones arides (Bastin et al., 2017), avec localisation de de la zone d'étude. (A) Carte mondiale, avec localisation de Madagascar (encadré rouge). (B) Couvert forestier (2014) au sein des zones arides de Madagascar (UNCCD/CBD) et zoom sur la zone d'étude du District de Toliara II, montrant le même couvert forestier

2.2.4 Implications cartographiques : pourquoi les produits globaux échouent en zones arides

Les résultats de Bastin et al., (2017) permettent d'identifier quatre mécanismes d'échec des produits globaux en zones sèches, qui constituent autant de défis méthodologiques auxquels cette thèse répond.

Le premier mécanisme est lié aux seuils de définition du couvert forestier. L'application d'un seuil de 10 % ou 30 % de couvert pour définir la forêt exclut mécaniquement une fraction importante des forêts ouvertes africaines et malgaches. Dans les biomes secs, le spectre du couvert s'étend de 5 à 70 % selon la formation et la saison, sans rupture nette qui permettrait de discriminer forêt et non-forêt sur la base d'un seuil fixe. L'approche continue de la couverture — modéliser le couvert

comme une variable quantitative plutôt que comme une classe — constitue la réponse méthodologique à ce problème, et c’est le choix central des Chapitres 4 et 6 de cette thèse.

Le deuxième mécanisme est la confusion spectrale. Dans les biomes secs, la réflectance du sol nu représente 60 à 90 % de la surface d’un pixel Landsat de 30 m en saison sèche, même dans des formations ligneuses dont le couvert réel est de 30 à 40 % (Bastin et al., 2017). Les algorithmes de classification binaire interprètent alors ces pixels comme non-forestiers, car leur signature spectrale est davantage proche des classes sol nu ou savane que des classes arborées des données d’entraînement.

Le troisième mécanisme est la résolution spatiale insuffisante. Les arbres hors forêt — isolés ou en petits groupes — représentent une fraction significative du couvert ligneux des zones arides, en particulier dans les zones de transition entre fourrés fermés et espaces agricoles. Brandt et al., (2020) ont montré, à partir de l’analyse d’images très haute résolution par apprentissage profond, que l’Afrique subsaharienne contient plus de 10 milliards d’arbres en dehors des forêts fermées, dont la moitié se situe dans les zones sèches. Ces arbres sont quasi-totalement absents des produits globaux à 30 m.

Le quatrième mécanisme est la phénologie caducifoliée. Les classifieurs entraînés sur des images d’une seule date ne capturent qu’un instantané de la phénologie, or dans les forêts sèches, l’indice de végétation normalisé (NDVI) d’un même peuplement peut varier de 0,1 à 0,5 selon la saison. L’utilisation de séries temporelles intégrant la saisonnalité constitue une réponse à cette limite (DeVries et al., 2015; Gorelick et al., 2017).

2.3. Méthodes de télédétection appliquées aux zones arides

Les méthodes de télédétection disponibles pour la cartographie des végétations arides ont connu des progrès considérables depuis les années 2010, portés par la démocratisation des archives satellitaires, le développement des plateformes de calcul distribué, et l’essor des algorithmes d’apprentissage automatique. Leur application effective dans les milieux ouverts reste néanmoins conditionnée par des choix méthodologiques que les sections suivantes examinent.

2.3.1 Classification supervisée et *Random Forest*

La classification supervisée consiste à assigner chaque pixel ou objet de l’image à une classe prédéfinie sur la base d’une fonction discriminante apprise à partir d’un jeu de données de référence (zones d’entraînement). Parmi les algorithmes disponibles, *Random Forest* (RF) s’est imposé comme la méthode de référence pour la cartographie des écosystèmes tropicaux (Belgiu and Drăguț, 2016; Breiman, 2001). RF est un ensemble d’arbres de décision construits sur des sous-échantillons aléatoires des données d’entraînement ; la prédiction finale est obtenue par vote majoritaire (classification) ou moyenne (régression) de l’ensemble des arbres.

Plusieurs propriétés font de RF un choix adapté aux milieux arides. Sa capacité à

intégrer des prédicteurs de nature et d'échelle différentes (indices spectraux multi-dates, variables topographiques, indicateurs de texture) permet de contrebalancer partiellement la confusion spectrale entre sol et végétation par l'introduction de variables structurelles. Sa robustesse aux données déséquilibrées est un atout dans les contextes où la classe arborée est minoritaire. Enfin, RF peut être utilisé en mode régression pour produire une estimation continue du couvert (variable quantitative), ce qui contourne la difficulté du choix d'un seuil de classification (Breiman, 2001).

Comparativement, l'algorithme de maximum de vraisemblance (Maximum Likelihood), largement utilisé dans les études malgaches antérieures (Brinkmann et al., 2014; J. Randriamalala et al., 2015), repose sur une hypothèse de distribution gaussienne des réflectances par classe. Cette hypothèse est rarement vérifiée dans les milieux hétérogènes à forte variabilité intra-classe. Le SVM présente une précision comparable au RF dans de nombreux contextes (Sheykhmousa et al., 2020), mais ses hyperparamètres sont plus difficiles à optimiser (Probst et al., 2019) et son comportement est moins transparent (Siemers and Bajorath, 2023).

Les réseaux de neurones convolutifs (CNN) et, plus largement, les méthodes de deep learning, représentent la frontière actuelle des méthodes de cartographie. (Guirado et al., 2020) ont montré leur supériorité pour estimer le couvert arboré dans les zones sèches à partir d'images à très haute résolution. Brandt et al. (2020, 2021) ont cartographié des arbres individuels dans le Sahel ouest-africain par CNN sur des images Maxar à 0,5 m de résolution. Ces méthodes nécessitent cependant des volumes de données de calibration et des ressources informatiques sans commune mesure avec les approches RF, et restent difficilement applicables dans des contextes opérationnels à faibles ressources comme le sud-ouest de Madagascar.

2.3.2 Séries temporelles Landsat et Google Earth Engine

L'archive Landsat, active depuis 1972, constitue la seule source de données satellitaires permettant une analyse continue du couvert végétal sur plusieurs décennies à l'échelle mondiale. Les capteurs Landsat 4-5 (TM), Landsat 7 (ETM+) et Landsat 8-9 (OLI) couvrent une plage spectrale de 0,45 à 2,35 μm avec une résolution spatiale de 30 m et un cycle de revisite de 16 jours. L'harmonisation des collections entre capteurs, réalisée via le protocole de correction de surface atmosphérique USGS (Collection 2) et les facteurs de normalisation inter-capteurs (Roy et al., 2016), permet de construire des séries temporelles continues couvrant des périodes de 30 à 40 ans.

La plateforme Google Earth Engine (GEE) a transformé l'utilisation pratique de ces archives en rendant accessible, à partir d'un navigateur internet, des pétaoctets de données satellitaires précalculées et des ressources de calcul distribué (Gorelick et al., 2017). GEE permet de construire des composites spectraux multi-dates, de calculer des indices spectraux sur l'ensemble de l'archive, et d'entraîner des modèles RF directement en nuage à l'échelle d'un pays ou d'un continent. Cette démocratisation de l'accès aux données et au calcul a rendu possibles des études comme celle de Bastin et al., (2017) et les travaux du Chapitre 8 de cette thèse, qui mobilisent 36 années d'images Landsat sur l'ensemble du District de Toliara II.

Dans les milieux arides, l'exploitation des séries temporelles apporte deux

avantages décisifs. D'une part, elle permet de construire des composites phénologiques intégrant plusieurs images de saisons différentes, capturant à la fois le signal de végétation en saison des pluies et la signature structurelle de la litière ou des tiges en saison sèche. D'autre part, elle permet de détecter des changements graduels de couvert — la dégradation lente par prélèvement de bois ou par feux récurrents — que les comparaisons biennales ou décennales sont incapables de révéler (DeVries et al., 2015).

2.3.3 Photo-interprétation à très haute résolution (Collect Earth Online)

La photo-interprétation sur images à très haute résolution constitue à la fois une méthode de production de données de calibration et, dans le cadre de Bastin et al., (2017), une méthode de cartographie à part entière. La plateforme Collect Earth Online (CEO), développée par Google et l'Université de Clark à partir du protocole Open Foris (Bey et al., 2016), permet à des opérateurs formés d'interpréter des placettes sur des mosaïques d'images Google Earth, Bing, et Planet Labs, disponibles à des résolutions de 0,5 à 3 m.

L'avantage principal de cette approche est de s'affranchir des confusions spectrales qui affectent les images à 30 m. À 0,5 m de résolution, la couronne individuelle d'un *Alluaudia procera* de 4 m de hauteur est identifiable, même en saison sèche, par sa forme colonnaire caractéristique. Cette identification visuelle, impossible à formaliser dans un modèle spectral, peut être réalisée par un opérateur local familier de la flore du sud-ouest malgache.

La limite de la méthode est son coût en temps humain qualifié et la subjectivité inhérente à toute photo-interprétation. Des protocoles de double lecture et de consensus inter-opérateurs permettent de quantifier et de réduire ces incertitudes (Olofsson et al., 2014). L'association de données CEO avec un algorithme RF sur images Landsat combine ainsi la précision de la photo-interprétation à très haute résolution avec la couverture spatiale des séries temporelles à moyenne résolution — c'est l'approche retenue dans le Chapitre 5.

2.3.4 Indices spectraux et harmonisation inter-capteurs

Les indices spectraux constituent des transformations normalisées des réflectances satellitaires qui amplifient le signal de la végétation par rapport au bruit de fond du sol. Le NDVI (*Normalized Difference Vegetation Index*, Rouse et al., 1974) est l'indice le plus utilisé, mais sa saturation au-delà de couverts denses le rend peu sensible dans les milieux ouverts où les couverts sont inférieurs à 40 %. Le NBR (*Normalized Burn Ratio*) est sensible aux feux récents et à la dégradation post-incendie. Le NDWI (*Normalized Difference Water Index*) capture l'état hydrique de la végétation et est particulièrement utile pour distinguer les espèces succulentes des espèces caducifoliées en saison sèche (Gao, 1996).

Dans les fourrés xérophiles, Abel et al., (2023) ont montré que les combinaisons spectrales intégrant le moyen infrarouge (bandes SWIR 1 et 2 de Landsat) sont plus discriminantes que le NDVI seul pour distinguer la végétation ligneuse du sol calcaire.

La texture spectrale et la variation temporelle des indices (amplitude saisonnière du NDVI, pente interannuelle) constituent des prédicteurs complémentaires pertinents pour les modèles RF appliqués aux milieux arides (DeVries et al., 2015).

L'harmonisation inter-capteurs est un prérequis de toute analyse de série temporelle long terme. Le passage de Landsat 7 (ETM+, 1999) à Landsat 8 (OLI, 2013) introduit des décalages radiométriques documentés qui peuvent fausser les tendances de couvert si l'harmonisation n'est pas appliquée. Roy et al., (2016) ont développé des facteurs de normalisation linéaire basés sur des comparaisons de scènes de nadir, disponibles dans la Collection 2 USGS et largement adoptés dans la littérature de cartographie des forêts sèches.

2.4. Recension des travaux antérieurs sur les méthodes utilisées dans la cartographie

2.4.1 Référence bibliographique

Recension des travaux antérieurs sur les méthodes utilisées dans la cartographie des écosystèmes forestiers des zones arides : cas du sud-ouest de Madagascar. Ramalason, Felana Nantenaina ; Olivia Lovanirina Rakotondrasoana ; Bastin, Jean-François. 2025 • In Comité scientifique du colloque (Ed.) LES CAHIERS DE L'ED GRND

Ce chapitre a été présenté lors qu'un colloque scientifique et publié dans Les Cahiers de l'ED GRND sous format d'un mini-article.

2.4.2 Abstract

The forest ecosystems of southwestern Madagascar exhibit significant diversity, comprising dry forests and various types of xerophytic shrublands. However, deforestation is escalating in this region, exacerbated by the fragmentation and complexity of these ecosystems, making aerial mapping challenging, particularly due to their resemblance to herbaceous vegetation. Existing estimates often underestimate vegetation in arid areas due to a lack of consideration for open vegetation. Mapping studies in this area are poorly documented and mainly rely on conventional remote sensing methods, utilizing medium-resolution images with a limited number of observation dates. Lasry et al. (2004) were among the few to employ higher-resolution imagery, but their estimates were deemed inaccurate, predicting complete deforestation within just 22 years, a scenario which did not occur despite the increase in anthropogenic pressures. This study synthesizes and identifies various approaches and images used to map arid zone vegetation at different geographic scales, providing avenues for improving ecosystem mapping in southwestern Madagascar.

Keywords: Arid zone, xerophytic shrublands, dry forests, remote sensing, open forests, Madagascar

2.4.3 Résumé

Les écosystèmes forestiers du sud-ouest de Madagascar présentent une grande diversité, avec des forêts sèches et divers types de fourrés xérophiles. Néanmoins, la déforestation s'intensifie dans cette région, exacerbée par la fragmentation et la complexité de ces écosystèmes, ce qui rend leur cartographie aérienne difficile, en particulier en raison de la similitude avec les végétations herbacées. Les estimations existantes sous-estiment souvent les végétations des zones arides, en raison du manque de considération pour les végétations ouvertes. Les études cartographiques dans cette zone sont peu documentées et se limitent principalement à des méthodes classiques de télédétection, utilisant des images de résolution moyenne avec un nombre restreint de dates d'observation. Les travaux de Lasry et al. (2004) ont été parmi les rares à utiliser des images de meilleure résolution, mais les estimations ont été jugées erronées, prédisant une déforestation complète des forêts en seulement 22 ans après, ce qui ne s'est pas produit malgré l'augmentation des pressions anthropiques. Cette étude a permis de synthétiser et d'identifier les différentes approches et images utilisées pour cartographier les végétations des zones arides à différentes échelles géographiques, offrant ainsi des pistes de recherche pour améliorer la cartographie des écosystèmes du sud-ouest de Madagascar.

Mots clés : Zone aride, fourrés xérophiles, forêts sèches, télédétection, forêts ouvertes, Madagascar

2.4.4 Introduction

Les écosystèmes forestiers du sud-ouest de Madagascar présentent une diversité remarquable, comprenant des forêts sèches et divers types de fourrés xérophiles, influencés par les conditions édaphiques et climatiques locales. Ces végétations revêtent une importance écologique majeure, caractérisée par un taux d'endémisme élevé, atteignant jusqu'à 95% selon (Razanaka, 2004), et une valeur socio-économique significative, en fournissant des ressources en bois essentielles pour les communautés locales. Cependant, la déforestation s'accélère dans cette région, avec un taux croissant atteignant 2% par an entre 2010 et 2017, selon (Vieilledent et al., 2020). Cette tendance est alimentée principalement par des activités anthropiques telles que l'agriculture sur brûlis, la production de charbon et le pâturage, exacerbées par la croissance démographique (Casse et al., 2004; Randriamalala et al., 2019). La cartographie de ces écosystèmes forestiers arides présente des défis, notamment en raison de leur fragmentation et de leur similitude visuelle avec les végétations herbacées lors de l'observation aérienne, comme souligné par (De Marzo et al., 2021).

Face à ces enjeux, cette étude vise à répondre à la question de recherche suivante : quelles sont les méthodes les plus appropriées pour cartographier de manière précise les écosystèmes forestiers du sud-ouest de Madagascar ? L'hypothèse de travail postule que les recherches cartographiques dans cette région ne sont pas suffisamment documentées malgré l'évolution de la télédétection, avec une prédominance de méthodes de télédétection classiques utilisant des images à faible résolution et des données temporelles limitées. De plus, les études nationales négligent souvent les

végétations ouvertes, malgré leur importance dans la composition globale de la couverture végétale. Ainsi, cette étude vise à examiner les études cartographiques existantes dans d'autres pays ainsi que celles réalisées à Madagascar, dans le but d'identifier les meilleures pratiques méthodologiques pour une cartographie plus précise des écosystèmes forestiers du sud-ouest de Madagascar.

2.4.5 Matériels et méthodes

La zone d'étude se trouve dans la Région Atsimo Andrefana, District de Toliara II, caractérisée par un climat subaride, avec une température moyenne de 24°C et des précipitations annuelles de 418 mm (Gaspard et al., 2018). La végétation comprend des forêts sèches au nord et des fourrés xérophiles au sud.

Notre étude a débuté par une recherche documentaire approfondie utilisant des mots-clés pertinents et des opérateurs booléens dans les moteurs de recherche Google Scholar et BASE (Bielefeld Academic Search Engine), conformément à la méthodologie de Pochet (2015). Cette recherche a permis de recueillir des travaux récents sur les méthodes et paramètres de télédétection, avec un accent particulier sur la cartographie des écosystèmes arides, couvrant les cinq dernières années à l'échelle mondiale et les dix dernières années pour le cas spécifique de Madagascar. Ensuite, nous avons comparé les études de cartographie effectuées dans d'autres pays et à l'échelle mondiale avec celles réalisées dans la zone d'étude, afin d'identifier les lacunes et d'orienter nos recherches pour réduire les incertitudes sur l'état et la situation des écosystèmes arides du sud-ouest de Madagascar.

2.4.6 Résultats

Diverses méthodes et algorithmes sont utilisés à la fois à l'échelle internationale et locale pour estimer les écosystèmes arides, notamment le Maximum de vraisemblance, le Support Vector Machine, la classification orientée objet, les techniques non-supervisées, semi-supervisées, la photo-interprétation, le deep-learning et les random forests. Nos résultats mettent en lumière une évolution significative des méthodes de télédétection des écosystèmes arides à l'échelle mondiale, tout en soulignant des lacunes au niveau local (cf. Tableau 5.1.). Bien que des techniques telles que les random forests et le deep-learning soient largement utilisées à l'échelle mondiale, elles n'ont pas encore été appliquées aux écosystèmes arides du sud-ouest de Madagascar. Les études locales se sont limitées à des analyses binaires (forêts/non-forêts) avec des images à une résolution de 30 mètres (Landsat), alors qu'il est maintenant possible d'utiliser des images à haute et très haute résolution pour une cartographie plus précise, réduisant ainsi les biais.

Notre analyse a révélé une négligence des végétations ouvertes dans les études locales, malgré leur importance dans la composition globale de la couverture végétale. Par exemple, l'étude de (Bastin et al., 2017) a mis en évidence une variation remarquable de la superficie des forêts dans les zones arides, passant de 105 millions d'hectares en 2000 à 542 millions d'hectares en 2010, grâce à une analyse plus approfondie du gradient de végétation avec des images à très haute résolution.

De plus, les études de (Lasry et al., 2004) ainsi que de (Ravonjimalala et al., 2017) se sont distinguées par l'utilisation d'images de meilleure résolution (image SPOT de 10 mètres) pour analyser la dynamique de la déforestation dans les forêts sèches de Mikea. Leurs résultats ont suggéré un taux de déforestation pouvant atteindre environ 40% sur une période de 15 ans, laissant entrevoir une déforestation complète de ces forêts en seulement 22 ans, soulevant ainsi des incertitudes sur la situation de ces écosystèmes.

Tableau 2-1: Synthèse des méthodes de télédétection des écosystèmes des zones arides au niveau international et local

<i>Méthodes/ Algorithmes</i>	Etudes au niveau international		Etudes locales				
	Images	Auteurs	Images	Types de végétation	Auteurs	Année d'étude	Résultats
Maximum de vraisemblance	Sentinel 2	<i>(Pande et al., 2021)</i>	Landsat	Fourrés xérophiles	<i>(Brinkmann et al., 2014)</i>	1973, 1993, 2001, 2013	45% en 40 ans (1,12%/an)
	Landsat	<i>((Laxmi Goparaju et al., 2019)</i>	Landsat	Forêts sèches	<i>(Ramalason, 2015)</i>	2000/2002, 2007/2008, 2013/2014	19% en 14ans soit 1,35%/an
Classification non supervisée (isodata)	Landsat	<i>(Bajirao et al., 2018)</i>					
Support Vector Machine (SVM)	Combinaison données Lidar et données optiques	<i>(Cao et al., 2018)</i>	Landsat	Fourrés xérophiles	<i>(Randriamalala et al., 2015)</i>	1989, 2001, 2010	1989-2001 (0,7 %/an) et 2001-2010 (1,9 %/an)
			SPOT	Forêts des Mikea	<i>(Ravonjimalala et al., 2017)</i>	1999, 2005, 2014	39,8 % en 15 ans soit 2,65%/an
Classification orienté objet	Landsat	<i>(Shirvani et al., 2020)</i>					
Photo-interprétation	Image à très haute résolution gratuit disponible en ligne	<i>(Bastin et al., 2017)</i>	SPOT	Forêts des Mikea	<i>(Lasry et al., 2004)</i>	1986, 1997, 1999, 2001	43,6% en 15 ans (2,9%/an)
		<i>(Martin-Ortega et al., 2017)</i>					
Classification semi-supervisée	Landsat	<i>(Teferi, 2018)</i>	Landsat	Forêts de Ranobe PK 32	<i>(Tovondrazane et al., 2020)</i>	1985, 1999, 2018	1985 à 2018 : 2,3%/an

<i>Deep-learning - CNN</i>	Données radar Sentinel 2 et données optiques	(Brandt and Stolle, 2021; Brandt et al., 2020)	-
	Image à résolution très élevée avec License	(Guirado et al., 2017)	
<i>Random forest</i>	Données Lidar et données optiques	(Cao et al., 2018)	-
	Sentinel 2	(Adjognon et al., 2019)	

2.4.7 Discussion

La cartographie des végétations du sud-ouest de Madagascar révèle plusieurs opportunités pour l'application de méthodes non encore utilisées, telles que *random forests* et *deep-learning*, ainsi que l'utilisation d'images à plus haute résolution. L'utilisation de gradients de végétation et d'images à très haute résolution permet une estimation plus précise de la couverture arborée, arbustive et boisée, y compris l'identification des arbres individuels. Des estimations de la couverture végétale à l'échelle mondiale ont été avancées par des chercheurs tels que (Brown et al., 2022; Buchhorn et al., 2020; Hansen et al., 2013; Sexton et al., 2013) ayant fait des études sur les différents gradients de végétation mais ont appliqués des méthodes similaires aux études locales induisant à des incertitudes et nécessite encore une étude encore plus approfondie. L'approche méthodologique de (Bastin et al., 2017) conduit à une meilleure compréhension de l'étendue des forêts, avec jusqu'à deux fois plus de forêts identifiées. De plus, l'analyse multi-dates est possible grâce à l'utilisation de Google Earth Engine (GEE), offrant ainsi une vision temporelle plus large des changements végétaux. Les modèles *random forests* (RF) se révèlent être des outils prometteurs par l'intégration de caractéristiques spectrales, spatiales et d'indices supplémentaires, améliorant ainsi la précision des classifications. (Senanayake et al., 2023) Les images à très haute résolution capturent des informations détaillées sur la faible couverture végétale et les interférences de fond (Ma et al., 2023). Ils permettent d'évaluer les changements de structure de la végétation après les épisodes de pluie et la relation entre la structure des arbustes pérennes et la densité du couvert végétal annuel (Abdullah et al., 2021).

2.4.8 Conclusion

Face à ces lacunes, il est essentiel d'explorer de nouvelles pistes de recherche pour améliorer la cartographie des écosystèmes arides du sud-ouest de Madagascar. L'utilisation de techniques telles que la photo-interprétation et le « *deep-learning* », sur la base d'images à haute résolution offre un potentiel prometteur pour améliorer la précision des cartes. De plus, l'adoption de méthodes de classification telles que l'algorithme « *random forest* » en combinaison avec des images à haute résolution, comme Sentinel ou des images captées par drone, représente une piste de recherche intéressante à explorer dans le domaine de la télédétection des végétations des zones arides. Ces avancées méthodologiques pourraient contribuer à une meilleure compréhension et gestion des écosystèmes du sud-ouest de Madagascar, en fournissant des données cartographiques plus précises et détaillées pour soutenir les efforts de conservation et de gestion durable. En outre, la collaboration avec des chercheurs internationaux travaillant sur des méthodes de télédétection avancées pourrait également ouvrir de nouvelles perspectives pour une cartographie plus précise des écosystèmes arides, en tenant compte des spécificités locales.

2.5. Optimisation du plan d'échantillonnage de calibration

La précision d'un modèle de cartographie par apprentissage automatique dépend de manière critique de la qualité et de la représentativité des données de calibration. En milieu aride, où les relevés de terrain sont coûteux et logistiquement contraints, la définition d'un plan d'échantillonnage optimal constitue un enjeu central — c'est l'objet du Chapitre 5.

2.5.1 Densité d'échantillonnage

La densité optimale d'échantillonnage dépend de la variabilité spatiale de la variable à modéliser et de la résolution spatiale du modèle. Dans les études de cartographie forestière à 30 m, la taille et la distribution des placettes influencent directement la précision des estimations selon la complexité de la végétation (Fassnacht et al., 2014; Mcroberts and Tomppo, 2007). Dans les milieux hétérogènes comme les fourrés xérophiles, la variabilité spatiale est élevée à courte distance — deux placettes séparées de 200 m peuvent avoir des couverts de 10 % et 60 % selon le substrat — ce qui plaide pour des densités d'échantillonnage élevées.

Cependant, l'augmentation de la densité d'échantillonnage a un coût logistique croissant dans des zones à piste difficile d'accès, et peut introduire de l'autocorrélation spatiale positive entre observations proches, biaisant positivement les métriques de validation croisée standard (Roberts et al., 2017). La validation par blocs spatiaux (spatial block cross-validation) permet de contrôler ce biais en s'assurant que les données de validation ne se trouvent pas dans le voisinage spatial des données d'entraînement.

2.5.2 Configuration spatiale : systématique versus stratifiée-aléatoire

La configuration spatiale d'un plan d'échantillonnage influence la représentativité des données vis-à-vis de l'hétérogénéité du milieu. Un plan systématique (grille régulière) maximise la dispersion spatiale et assure la représentation de tous les types de milieu, mais peut sous-représenter les zones de transition et les gradients. Un plan stratifié-aléatoire, dans lequel la stratification est fondée sur des covariables accessibles par télédétection (strates de couvert, types de substrat, éloignement à la côte), permet d'augmenter la représentation des classes minoritaires tout en maintenant une couverture spatiale satisfaisante (Cochran, 1977; Olofsson et al., 2014).

Dans les fourrés xérophiles du District de Toliara II, l'hétérogénéité biophysique — substrat calcaire versus sableux, gradient de hauteur intérieur-côte, gradient pluviométrique nord-sud — fournit des bases naturelles de stratification, conformes aux recommandations de Cochran, (1977) pour les milieux à forte variance spatiale. Le Chapitre 7 exploite ces gradients pour évaluer différentes configurations d'échantillonnage.

Ces deux approches sont illustrées dans la Figure 2.2, qui présente le protocole

comparatif mis en œuvre dans cette étude : une configuration **dispersée** (grille régulière, 1 placette par cellule) maximisant la dispersion spatiale, et une configuration **clusterisée** (point central aléatoire avec n voisins les plus proches selon la distance euclidienne). Les deux stratégies sont testées sur 13 tailles d'échantillon ($n = 10$ à 505 placettes), avec 100 itérations par taille et 4 configurations de données satellitaires (S1, S2, S1+S2, AlphaEarth).

2.5.3 Interaction entre précision et coût logistique

Un plan d'échantillonnage idéal maximise la précision du modèle tout en minimisant le nombre de placettes terrain à réaliser, puisque chaque placette terrain représente plusieurs jours de travail en milieu aride isolé. Ce compromis entre précision et coût a été formalisé par Congalton and Green, (2019) à travers la notion de matrice de confusion pondérée, et appliqué à la cartographie des forêts tropicales par Olofsson et al., (2014), qui proposent un cadre statistique permettant de déterminer le nombre minimal de placettes par classe pour atteindre un intervalle de confiance donné sur les estimations de superficie.

Dans les milieux arides où la classe arborée est structurellement peu contrastée spectralement avec le sol, la précision du modèle tend à plafonner après un certain nombre de placettes, et l'ajout d'observations supplémentaires ne compense plus le coût de collecte. L'identification de ce seuil de rentabilité marginale est d'une importance directe pour les acteurs de terrain qui doivent allouer des ressources de monitoring limitées dans les régions les plus fragiles — c'est l'objectif du Chapitre 5.

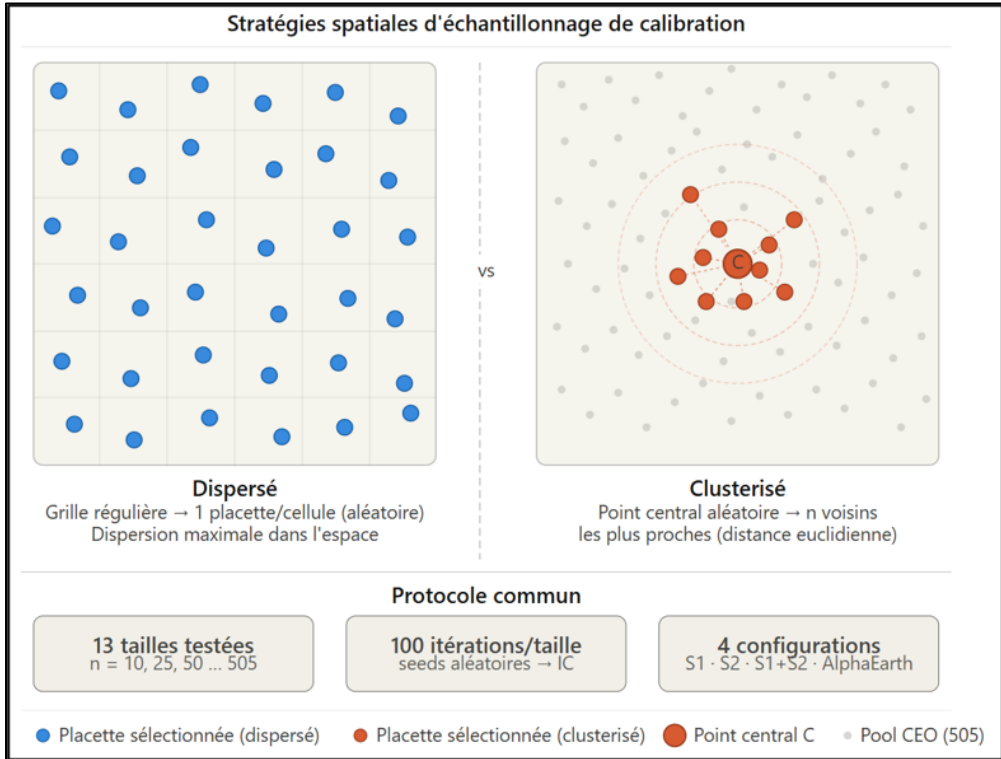


Figure 2-2: Illustration des deux stratégies spatiales d'échantillonnage de calibration comparées dans l'étude.

2.6. Dynamiques des écosystèmes xérophiiles : concepts clés

La compréhension des dynamiques temporelles du couvert ligneux dans les zones arides requiert un cadre conceptuel distinct de celui utilisé pour les forêts tropicales humides. Dans les milieux arides, les écosystèmes opèrent près de seuils biophysiques critiques, où des perturbations modérées mais récurrentes peuvent déclencher des transitions vers des états alternatifs. Cette section expose les concepts clés mobilisés dans l'analyse des dynamiques du Chapitre 8.

2.6.1 États de dégradation et résilience écologique

La résilience écologique est définie comme la capacité d'un écosystème à absorber des perturbations tout en maintenant ses fonctions et sa structure (Holling, 1973b; Walker et al., 2004). Dans les zones arides, la résilience est fortement contrainte par la disponibilité en eau : un écosystème soumis à une perturbation (feu, coupe) dans un contexte de déficit hydrique structurel récupère plus lentement que le même écosystème dans un milieu plus humide, car la repousse dépend d'une pluviométrie

incertaine et de la capacité de régénération végétative des souches (Bond and Van Wilgen, 1996).

Les trajectoires de changement écologique dans les milieux arides peuvent être réversibles (succession vers l'état initial après abandon des pressions) ou basculer vers des états alternatifs stables. Le concept d'état alternatif stable (Holling, 1973b; Scheffer et al., 2001) stipule que certains systèmes écologiques peuvent occuper plusieurs états d'équilibre distincts sous les mêmes conditions environnementales externes. Le passage d'un état riche en ligneux à un état dominé par la végétation herbacée ou le sol nu peut être déclenché par une perturbation dépassant un seuil critique — les feux récurrents constituant le mécanisme de bascule le plus documenté dans les savanes africaines (Hirota et al., 2011; Staver et al., 2011).

Dans les fourrés xérophiles du District de Toliara II, la déforestation par l'agriculture sur-brûlis ou *hatsake* conduit à une dégradation localement irréversible à court terme, car la banque de graines est détruite par le feu et les conditions de germination sont défavorables dans les premières années suivant la coupe. En revanche, la dégradation par production de charbon — qui ne détruit pas la totalité de la strate ligneuse — peut être suivie d'une récupération partielle si la pression est réduite (J. R. Randriamalala et al., 2015). La quantification de ces trajectoires à l'échelle de la série temporelle Landsat constitue l'objectif du Chapitre 8.

Ces trajectoires non linéaires interrogent par ailleurs les limites de la classification binaire forêt/non-forêt couramment employée en télédétection. Un pixel Landsat de 30 m passant de 60 % à 15 % de couvert ligneux reste classé « forêt » dans la plupart des schémas binaires usuels tant qu'un seuil arbitraire (souvent fixé autour de 10 à 20 %) n'est pas franchi, alors même qu'une telle perte de couvert peut déjà signaler l'amorce d'un basculement vers un état dégradé difficilement réversible (Hirota et al., 2011; Staver et al., 2011). Dans les paysages « en mosaïque » du sud-ouest malgache, où se côtoient à fine échelle fourrés denses, formations clairsemées et sols nus, cette approche binaire masque ainsi précisément les dynamiques de seuil — les *tipping points* (cf. Glossaire) — qu'il s'agit de détecter. C'est cette limite qui justifie le recours, dans cette thèse, à une caractérisation continue du couvert plutôt qu'à une classification catégorielle.

2.6.2 Régimes de feu en zone aride

Le feu est à la fois un facteur écologique naturel et une perturbation anthropique majeure dans les milieux arides africains. Sa fréquence, son intensité et sa saisonnalité déterminent la structure et la composition des formations ligneuses à des échelles temporelles de décennies à siècles (Archibald et al., 2013; Bond and Keeley, 2005). Les feux de faible fréquence (inférieurs à une fois par décennie) favorisent le maintien de la strate ligneuse en permettant la régénération entre deux événements. À l'inverse, les feux récurrents à haute fréquence exercent une pression de mortalité sur les régénérations, poussant progressivement l'écosystème vers des états herbacés ou arbustifs dégradés (Staver et al., 2011; Wigley et al., 2009).

Dans les savanes africaines, la réponse des ligneux aux feux suit une relation dose-réponse avec un seuil critique au-delà duquel le couvert ligneux décline de manière

accentuée (Bond and Van Wilgen, 1996; Smit et al., 2010). Ce phénomène d’hystérésis écologique — dans lequel le chemin de retour vers l’état initial diffère du chemin de dégradation — est bien documenté pour les interactions feu-ligneux. Dans le nord-est du District de Toliara II, où les feux pastoraux affectent la savane à *Heteropogon contortus* sur substrat calcaire, Rakotoarimanana and Grouzis, (2006) ont documenté des effets différenciés selon la fréquence et la saison d’occurrence du feu.

Les données MODIS MCD64A1 (burned area product, Giglio et al., (2018), disponibles depuis 2001 à résolution de 500 m, permettent de reconstituer les régimes de feu sur deux décennies. Malgré leur résolution limitée, ces données constituent le seul jeu de données de feu spatialement exhaustif sur la zone d’étude et sont mobilisées dans le Chapitre 8 pour analyser la relation entre récurrence des feux et évolution du couvert ligneux.

2.6.3 Pression anthropique et conversion d’usage des terres

La pression anthropique sur les écosystèmes ligneux s'exerce selon des gradients spatiaux structurés par la distance aux routes, aux villages et aux marchés énergétiques (Geist and Lambin, 2002; Laurance et al., 2002). Dans les zones arides, où les ressources ligneuses sont à la fois rares et fortement sollicitées, ces gradients sont particulièrement marqués. La théorie de l’accès différentiel aux ressources prédit une dégradation du couvert en concentrique autour des sources de demande, avec une intensité décroissant avec la distance (Laurance et al., 2002).

Dans le District de Toliara II, la proximité à la ville de Toliara et aux routes nationales constitue le principal déterminant de la pression de carbonisation, comme le montrent les travaux de Randriamalala et al., (2019, 2015) et Brinkmann et al., (2014). La croissance démographique du district (2,9 %/an selon INSTAT, 2020) accroît mécaniquement la frontière d’exploitation en repoussant les zones d’approvisionnement vers des secteurs de plus en plus éloignés de la ville.

La distinction entre déforestation (perte nette de couvert par conversion permanente) et dégradation (réduction du couvert sans changement de classe d’occupation du sol) est conceptuellement importante dans les milieux ouverts. Dans les bilans forestiers nationaux, seule la déforestation est généralement comptabilisée, car les classifications binaires sont insensibles à la dégradation graduelle (Brinkmann et al., 2014; Hansen et al., 2013). L’approche continue développée dans le Chapitre 8 permet précisément de quantifier la dégradation en distinguant la perte de qualité de couvert de la perte de surface, révélant une dégradation structurelle importante absente des classifications conventionnelles.

2.7. Cas du sud-ouest de Madagascar : lacunes identifiées

La synthèse des sections précédentes permet d’identifier trois lacunes cumulatives spécifiques au sud-ouest de Madagascar, qui structurent directement la problématique et les hypothèses de cette thèse. Ces lacunes découlent du croisement entre les limites

des produits globaux documentées dans la littérature internationale et les caractéristiques écologiques particulières des fourrés xérophiles du Domaine du Sud, telles qu'elles sont décrites par Brinkmann et al., (2014), Randriamalala et al., (2015) et Razanaka, 2004.

2.7.1 Carences en données de cartographie historique précise

Les études cartographiques disponibles pour le District de Toliara II reposent majoritairement sur des classifications binaires forêt/non-forêt appliquées à des images Landsat de 30 m sur un nombre limité de dates (deux à quatre dates, généralement espacées de dix ans). Cette approche a deux limites majeures. D'une part, elle exclut par construction les formations à canopée ouverte (couvert inférieur à 30-40 %), qui représentent une fraction significative des fourrés xérophiles. D'autre part, elle est incapable de distinguer les différents états de dégradation d'une même formation, traitant un fourré ouvert dégradé et un fourré dense intact comme deux points indiscernables dans l'espace spectral (Brinkmann et al., 2014; J. Randriamalala et al., 2015).

Cette carence documentaire a des conséquences directes sur la gestion des ressources forestières : les plans d'aménagement forestiers et les rapports nationaux de déforestation soumis à la Convention-cadre des Nations Unies sur les changements climatiques (CCNUCC) ne comptabilisent pas les formations ouvertes, produisant des bilans de déforestation structurellement sous-estimés pour les zones arides. Les mécanismes expliqués dans ce chapitre — biais des produits globaux, confusion spectrale, définitions incompatibles du couvert — permettent de comprendre pourquoi ces lacunes sont systématiques plutôt qu'accidentelles.

2.7.2 Absence d'analyse systématique des dynamiques en longue série

Aucune étude disponible dans la littérature n'a quantifié la dynamique temporelle continue du couvert ligneux du District de Toliara II sur l'ensemble de la période d'observation Landsat disponible (depuis 1989). Les études existantes se limitent à des comparaisons entre deux ou trois dates, ce qui ne permet pas de distinguer les tendances long terme des fluctuations interannuelles liées à la variabilité climatique, ni d'identifier les phases d'accélération ou de ralentissement de la dégradation (Brinkmann et al., 2014; Randriamalala et al., 2019).

L'absence de série temporelle continue empêche également de quantifier les relations entre, d'une part, les pressions anthropiques — fréquence des feux et proximité aux habitations — et, d'autre part, la perte de couvert ligneux, et de déterminer si ces relations suivent une dynamique linéaire ou si elles présentent des seuils au-delà desquels la dégradation s'accélère de manière irréversible. Le Chapitre 8 comble cette lacune en mobilisant 36 années de données Landsat harmonisées sous Google Earth Engine.

2.7.3 Besoin d'optimisation des protocoles de calibration

Les modèles RF de cartographie du couvert nécessitent des données de calibration

terrain suffisamment denses et spatialement représentatives pour produire des estimations précises. Dans le District de Toliara II, les coûts logistiques des campagnes terrain — accessibilité limitée, distances importantes, infrastructure routière déficiente — contraignent sévèrement le nombre de placettes réalisables. Il n'existe pas, dans la littérature disponible sur les milieux arides de Madagascar, d'étude systématique du compromis entre densité d'échantillonnage, configuration spatiale et précision du modèle dans ce type de milieu.

Cette lacune a des implications pratiques directes pour les gestionnaires qui souhaitent mettre en place des systèmes de surveillance continue du couvert : sans connaître le nombre minimal de placettes permettant d'atteindre une précision acceptable, il est impossible d'allouer les ressources de monitoring de manière rationnelle. Le cadre théorique de l'échantillonnage stratifié et du compromis précision-coût, exposé en section 2.4, fournit la base méthodologique pour répondre à ce besoin dans le Chapitre 6.

2.8. Synthèse du Chapitre 2

Ce chapitre a présenté le cadre théorique et conceptuel de cette thèse. Il a montré que la cartographie des végétations arides par télédétection demeure un défi, en raison notamment des difficultés à distinguer la végétation du sol nu, de la résolution parfois insuffisante des capteurs et des variations saisonnières de la végétation. Ces limites peuvent entraîner des erreurs importantes dans l'estimation du couvert végétal.

Dans le sud-ouest de Madagascar, ces difficultés sont accentuées par les caractéristiques particulières des fourrés xérophiiles. La structure ouverte de la végétation, la présence d'espèces perdant leurs feuilles durant la saison sèche et les signatures spectrales spécifiques de certaines familles dominantes rendent la détection de la végétation plus complexe que dans les écosystèmes forestiers humides.

Pour répondre à ces défis, cette thèse combine la photo-interprétation d'images à très haute résolution avec des méthodes de modélisation basées sur les séries temporelles Landsat et l'algorithme Random Forest. Cette approche vise à améliorer la caractérisation de la végétation ligneuse et le suivi de son évolution dans le temps.

Le chapitre a également souligné que la dégradation des écosystèmes arides est souvent progressive et difficile à détecter à partir d'une simple distinction entre zones dégradées et non dégradées. Une approche basée sur des variables continues apparaît donc plus adaptée pour quantifier les changements de végétation et mieux comprendre les dynamiques en cours.

Ces éléments mettent en évidence plusieurs besoins de connaissance dans le sud-ouest de Madagascar : améliorer la cartographie de la végétation ligneuse, documenter son évolution à travers le temps et optimiser les stratégies d'échantillonnage utilisées pour produire ces informations. Ces besoins constituent le point de départ des travaux présentés dans les Chapitres 5, 6, et 7.

Chapitre 3

Description de la zone d'étude

Ce chapitre présente les caractéristiques biophysiques et socio-économiques du district de Toliara II, qui constitue le cadre géographique de cette thèse. Les zones d'analyse varient selon les chapitres, en fonction des objectifs propres à chaque étude. Ainsi, le Chapitre 5, consacré à la recension des méthodes cartographiques antérieures, adopte une approche régionale englobant l'ensemble des écosystèmes forestiers arides et semi-arides du sud-ouest de Madagascar, afin d'identifier les principales lacunes méthodologiques à une échelle élargie. En revanche, les Chapitres 6 (évaluation des produits globaux de cartographie), 7 (optimisation du plan d'échantillonnage) et 8 (analyse des dynamiques temporelles du couvert ligneux sur 36 ans) se concentrent sur la partie sud du district, située au sud du fleuve de Fiherenana (23°00'–24°30' S, 43°30'–44°30' E ; 4 826 km²), désignée ci-après comme la zone d'étude effective. Ce choix s'explique par la disponibilité des données de terrain, la représentativité d'un gradient complet de formations xérophiles typiques des conditions les plus arides du district, ainsi que par l'importance de cette zone pour les enjeux de conservation nécessitant des informations cartographiques fiables.

3.1. Localisation géographique et délimitation de la zone d'étude

Le District de Toliara II est situé dans la Région Atsimo Andrefana, dans le sud-ouest de Madagascar (province de Toliara). Il couvre une superficie administrative d'environ 910 000 ha (Raoliarivelo et al., 2010) et est délimité au nord par la ville de Toliara et le District de Toliara I, à l'est par les districts de Betioky-Sud et d'Ampanihy, au sud par le district de Betioky-Sud, et à l'ouest par le canal du Mozambique. Les communes rurales incluses dans la zone d'étude sont : Ambohimahavelona, Ambolofoty, Anakao, Andranohinaly, Andranovory (partie sud), Antanimena Onilahy, Beheloka, Behompy (partie sud), Betsinjaka, Manorofify, Miary Ambohibola, Mitsinjo Betanimena, Saint Augustin et Soalara Sud.

La zone d'étude effective des Chapitres 6 à 8 couvre 4 826 km², au sud du fleuve de Fiherenana (Figure 3.1). Les altitudes y sont comprises entre le niveau de la mer et 598 m (moyenne : 162 m ; écart-type : 129 m ; Copernicus DEM GLO-30), avec un gradient progressif de la côte vers les plateaux calcaires de l'intérieur.

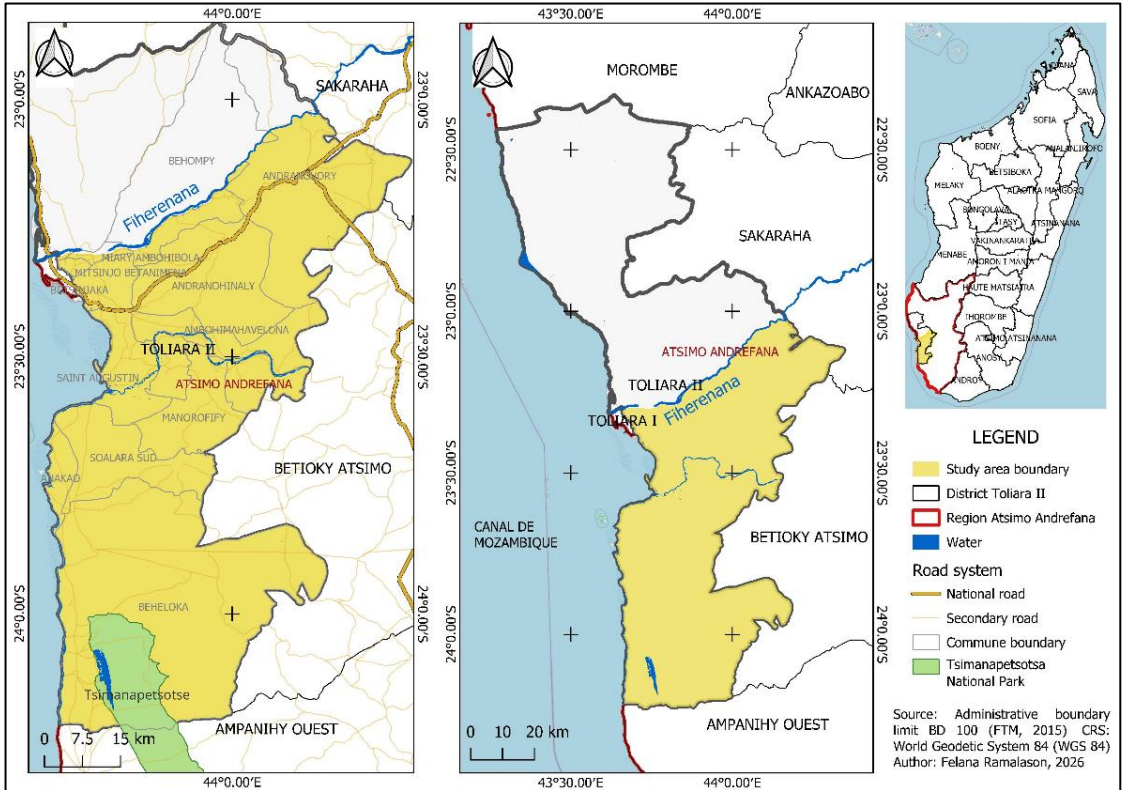


Figure 3-1: Localisation géographique du District de Toliara II et de la zone d'étude. À gauche : zone d'étude effective (en jaune) au sein du District de Toliara II, avec le réseau routier et les limites communales. À droite : situation du district dans la Région Atsimo Andrefana et positionnement régional. En médaillon : localisation dans Madagascar

3.2. Cadre bioclimatique

3.2.1 Caractérisation de l'aridité

Le District de Toliara II présente un climat aride à semi-aride, caractérisé par un indice d'aridité (IA) moyen de 0,21 (intervalle : 0,13–0,36 ; classification UNCCD ; Trabucco and Zomer, 2019; UNCCD, 2024). Selon cette classification, environ 59 % de la superficie du district sont classés arides (IA < 0,20), principalement dans les secteurs méridionaux du district, tandis que 41 % relèvent de la catégorie semi-aride (IA : 0,20–0,50), concentrés dans les secteurs septentrionaux (Tableau 3.1 ; Figure 3.2C). Aucune zone hyper-aride ou sub-humide n'est détectée au sein du district. Ce district figure parmi les plus secs de Madagascar, à l'opposé des forêts humides de la côte orientale.

L'hétérogénéité climatique est prononcée à l'intérieur du district : les précipitations annuelles varient de moins de 400 mm dans les secteurs méridionaux (Gaspard et al., 2018) à 600–1 000 mm dans les secteurs septentrionaux (Leprun et al., 2009), soit un

rapport supérieur à deux sur une distance d'environ 150 km selon un axe nord–sud (Figure 3.2A). La variabilité interne de l'IA confirme ce gradient : la zone aride présente une distribution resserrée ($\sigma = 0,019$), tandis que la zone semi-aride est nettement plus hétérogène ($\sigma = 0,037$), traduisant un gradient pluviométrique progressif vers l'intérieur des terres. Cette variabilité spatiale conditionne directement la structure et la composition des formations végétales, en générant un gradient continu entre des formations forestières denses au nord et des fourrés xérophiles ouverts au sud (Gaspard et al., 2018; Raselimanana et al., 2012; Razanaka, 2004).

Sur la zone d'étude effective, la précipitation annuelle moyenne est de 466 mm (intervalle : 274–770 mm ; CHIRPS, Funk et al., 2015) et la température annuelle moyenne de 24,8 °C (ERA5-Land, Hersbach et al., 2020) (Figure 2.2B). Des maximums diurnes supérieurs à 38 °C sont enregistrés en saison chaude (novembre–mars). La combinaison d'une évapotranspiration potentielle élevée et de précipitations limitées génère un déficit hydrique structurel sur sept à neuf mois par an, qui conditionne les adaptations morphologiques de la végétation décrites en section 2.5.

Tableau 3-1 : Statistiques de l'indice d'aridité (IA) par classe climatique au sein du District de Toliara II, calculées à partir du raster à résolution de 1 km (EPSG:32738 ; 4 747 pixels valides). L'IA est défini comme le rapport précipitations annuelles / évapotranspiration potentielle annuelle (P/ETP). La classification suit la nomenclature UNCCD (Trabucco and Zomer, 2019 ; UNCCD, 2024).

Zone	Seuil IA	% zone	IA min	IA moyen	IA max	Écart-type
Aride	0.05 – 0.20	59.1 %	0.122	0.167	0.200	0.019
Semi-aride	0.20 – 0.50	40.9 %	0.200	0.259	0.341	0.037

3.2.2 Saisonnalité des précipitations

Le régime pluviométrique du District de Toliara II est fortement saisonnier (Figure 3.2). La quasi-totalité des précipitations est concentrée entre décembre et mars, avec une moyenne mensuelle d'environ 90 mm en janvier et inférieure à 4 mm en août. Sept mois consécutifs reçoivent moins de 30 mm. Cette sécheresse saisonnière prolongée, combinée à une évapotranspiration potentielle élevée ($P = 365$ mm/an ; $T = 24,5$ °C), génère un déficit hydrique structurel qui gouverne la phénologie des végétations ligneuses de la zone (Razanaka, 2004).

Les analyses de tendance long terme indiquent une réduction progressive et une variabilité interannuelle accrue des précipitations dans le sud-ouest malgache (IPCC, 2023). La région est également exposée aux effets indirects des cyclones tropicaux touchant la côte orientale de Madagascar, qui se traduisent occasionnellement par des épisodes de pluies intenses modifiant temporairement le bilan hydrique local.

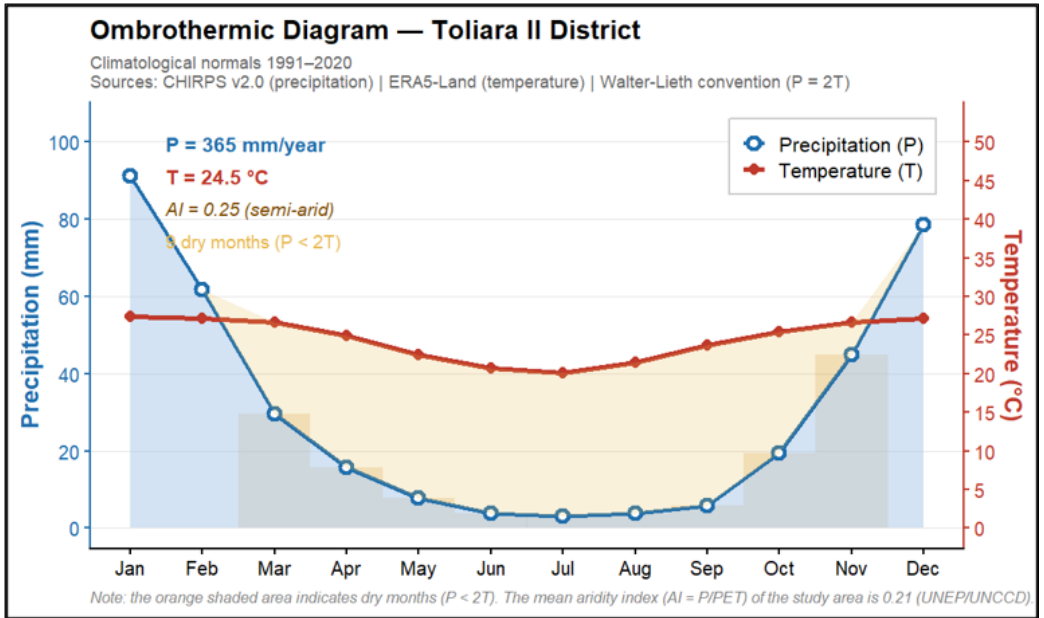
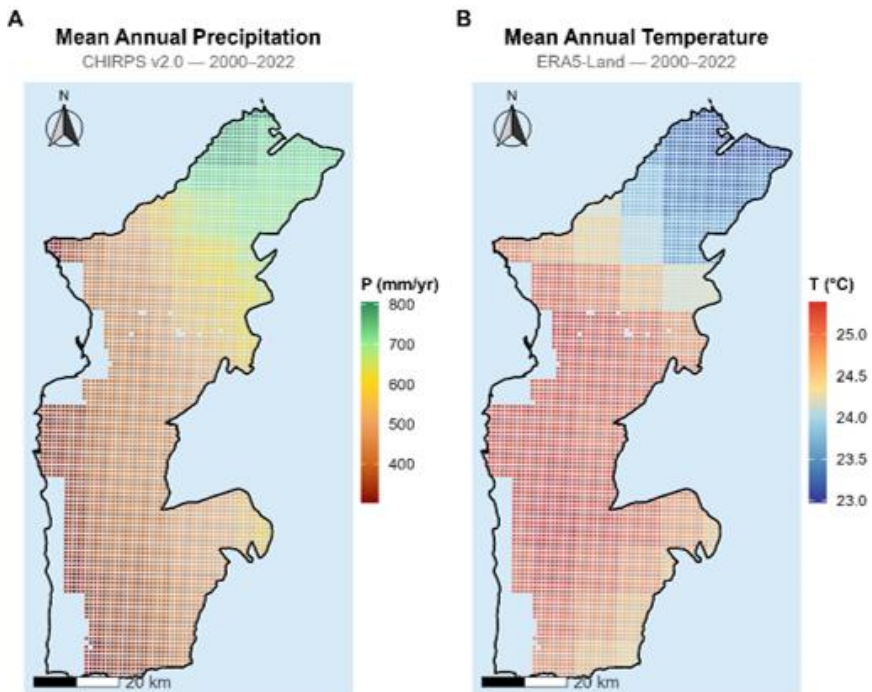


Figure 3-2 : Diagramme ombrothermique du District de Toliara II (normales climatologiques 1991–2020)



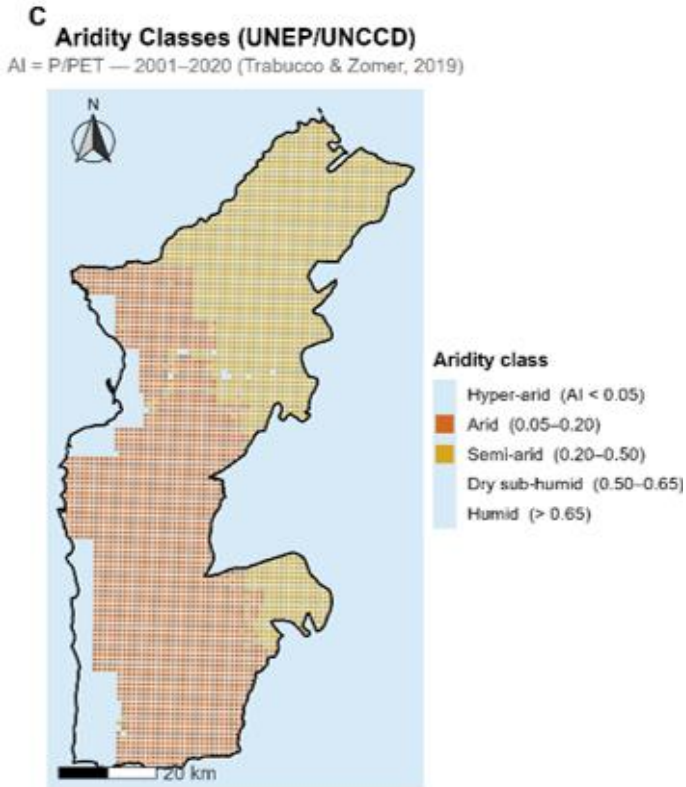


Figure 3-3: Caractérisation climatique du District de Toliara II : (A) précipitation moyenne annuelle ; (B) température moyenne annuelle ; (C) classes d'aridité (UNEP/UNCCD).

Sources : CHIRPS v2.0, ERA5-Land, Trabucco & Zomer (2019).

3.3. Relief et hydrographie

3.3.1 Relief

Le relief du District de Toliara II est globalement peu marqué, avec des altitudes comprises entre le niveau de la mer et 598 m (altitude moyenne : 162 m, écart-type : 129 m ; Copernicus DEM GLO-30). Le terrain s'abaisse progressivement du nord-est, où se concentrent les reliefs les plus élevés et accidentés, vers le littoral sud-ouest, avec une transition entre des plateaux calcaires à l'est et de larges plaines sableuses côtières à l'ouest (Figure 3.4). Les plateaux calcaires, composés de dalles karstiques peu profondes, supportent les fourrés xérophiles ouverts à dominante de *Didiereaceae* et *Euphorbiaceae*. La frange littorale est occupée par des sables blancs carbonatés qui conditionnent à leur tour la distribution des formations xérophiles denses, tandis que les sables roux ferrugineux prédominent dans les secteurs de l'intérieur au nord du district (Leprun et al., 2009; Raselimanana et al., 2012).

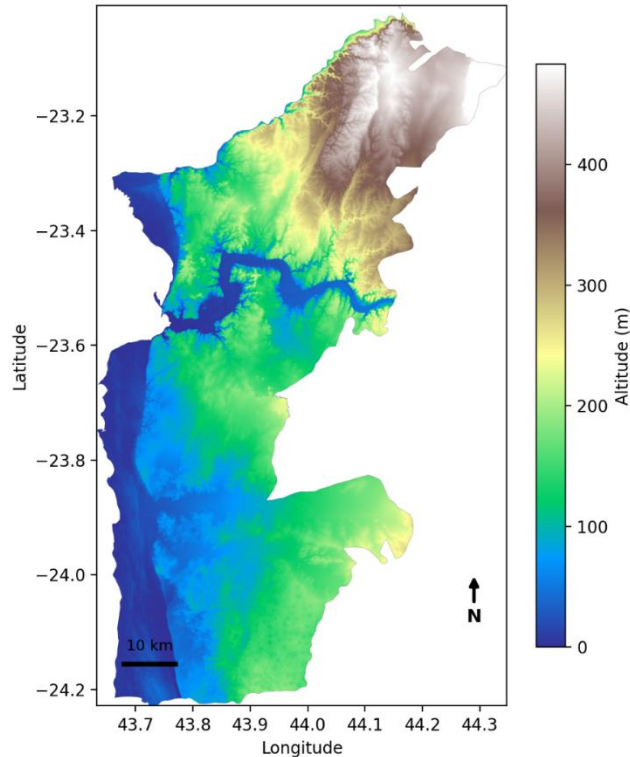


Figure 3-4 : Modèle numérique de terrain (MNT) du District de Toliara II. Le relief est globalement bas et plat sur la majeure partie du district (0-150 m), à l'exception d'un massif plus élevé (jusqu'à plus de 400 m) au nord-est, en bordure du district. Source : Copernicus DEM GLO-30 (ESA), via Google Earth Engine.

3.3.2 Hydrographie

Le réseau hydrographique du district est dominé par des cours d'eau à écoulement intermittent, dont l'activité est strictement corrélée à la saisonnalité pluviométrique (MERIT Hydro, Yamazaki et al., 2019). Le fleuve de Fiherenana délimite la bordure septentrionale de la zone d'étude effective (Figure 3.1), tandis que l'Onilahy borde la partie méridionale du district. Les nappes phréatiques sont généralement profondes et peu accessibles aux végétaux ligneux à enracinement superficiel, renforçant la dépendance des écosystèmes xérophiiles aux seules précipitations.

3.4. Géologie et pédologie

3.4.1 Substrats géologiques

Le substratum géologique du District de Toliara II est dominé par des formations sédimentaires d'âge secondaire à tertiaire, comprenant des calcarénites, des calcaires éocènes et des grès du Crétacé (Leprun et al., 2009). Ces formations conditionnent

directement la structure et la composition des formations végétales en déterminant la capacité de rétention en eau des sols. Les calcaires compacts du plateau karstique, particulièrement représentés dans le secteur d'Andranohinaly, sont associés à des fourrés xérophiles très ouverts à arbustes bas et contournés sur dalle affleurante (Figure 3.4, photos 4 et 5). Les formations sédimentaires côtières, à dominante sableuse, supportent quant à elles une végétation basse et éparse à *Euphorbia stenoclada* caractéristique de la frange littorale (Figure 3.4, photo 6). Cette correspondance entre nature du substratum et physionomie de la végétation est confirmée par les relevés de terrain : la carte ci-dessous (Figure 3.5) localise les principaux types de substrats observés — dalle calcaire, sable blanc et sable roux — et illustre la cohérence spatiale entre l'affleurement rocheux du plateau karstique et les zones de fourrés xérophiles d'une part, et les substrats sableux côtiers et la végétation littorale d'autre part.

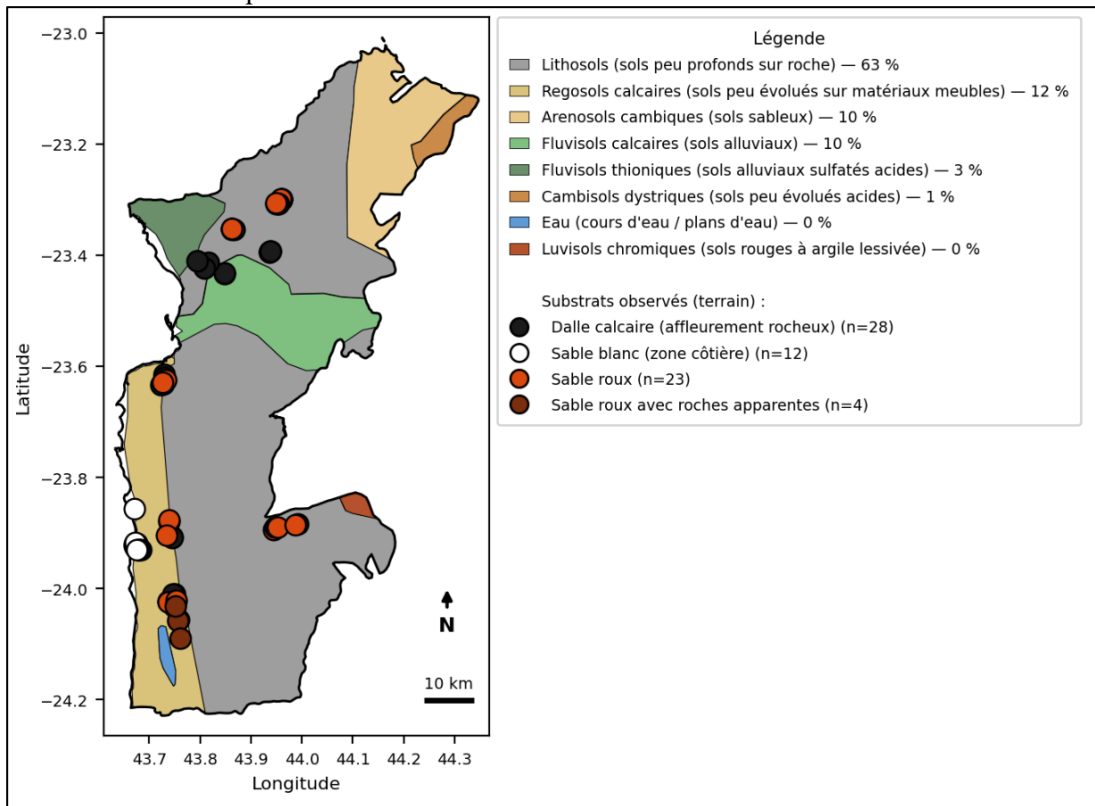


Figure 3-5 : Groupes de sols (FAO-UNESCO) et substrats observés sur le terrain — District de Toliara II. Le district est dominé par les Lithosols (63 %), avec des Regosols calcaires, Arenosols et Fluvisols en complément. Les 67 points superposés montrent les substrats relevés sur le terrain par drone (2024-2025) : dalle calcaire, sable blanc, sable roux et sable roux avec roches apparentes. Source : Carte mondiale des sols FAO-UNESCO (1974), via GeoServer resiliencemada.gov.mg ; relevés de terrain par drone, 2024-2025

3.4.2 Types de sols et relation végétation–sol

La variabilité pédologique du District est directement liée aux substrats géologiques et à la topographie. Quatre grands types de sols structurent la distribution spatiale des formations végétales (Randriamalala et al., 2019, 2016) (Figure 3.4).

(1) Les sables blancs côtiers (sols sableux calcaires) : associés à une végétation basse et éparse à *Euphorbia stenoclada*, prédominants dans la zone littorale et semi-littorale (Figure 3.4, photo 6).

(2) Les sols sableux roux ferrugineux : typiques des secteurs intérieurs, qui supportent des formations arbustives à arborées à couvert relativement fermé en saison humide, avec une défeuillaison marquée en saison sèche, dominées par *Commiphora spp.* et *Euphorbia spp.*, ainsi que des fourrés xérophiles arbustifs à *Alluaudia procera* (Figure 3.4, photos 1, 2 et 3).

(3) Les sols calcaires compacts du plateau (calcisols) : associés à des fourrés xérophiles très ouverts, à arbustes bas et contournés sur dalle calcaire affleurante, dominés par *Alluaudia procera* et *Euphorbia stenoclada*, dont la composition floristique diffère significativement de celle des formations sur substrats sableux (Randriamalala et al., 2019) (Figure 3.4, photos 4 et 5).

(4) Les sols alluviaux des bas-fonds, le long des fleuves Onilahy et Fiherenana : à nappe alluviale peu profonde, ils supportent des cultures irriguées et, localement, des formations ligneuses azonales (Masezamana et al., 2013).

Cette correspondance étroite entre composition floristique et type de sol constitue un paramètre structurant pour l'interprétation des performances des produits globaux de cartographie de la végétation analysés dans le Chapitre 3, en particulier pour les formations xérophiles ouvertes des substrats calcaires.

Cette hétérogénéité pédologique a une incidence directe sur la réflectance de surface et, par conséquent, sur la robustesse des produits cartographiques globaux. Comme indiqué au Chapitre 2 (section 2.2), le sol nu domine le signal spectral en saison sèche dans les écosystèmes arides ; cet effet est particulièrement marqué pour les formations sur calcisols et sur sables côtiers, où le couvert ligneux est naturellement clairsemé et où la fraction de sol nu visible reste élevée même en saison humide, contrairement aux formations plus fermées des sols sableux roux ferrugineux. L'échelle à laquelle s'exprime cette variabilité pédologique est elle-même multiple : à l'échelle macro (plusieurs kilomètres), elle correspond à la distribution des quatre grands substrats géologiques décrits ci-dessus ; à l'échelle micro (parcelles de l'ordre de 10 m × 10 m), elle reflète des variations plus fines de profondeur de sol, de pierrosité et de topographie locale qui modulent la structure et la composition de la végétation au sein même d'un même type de sol (Randriamalala et al., 2019). Cette structuration à plusieurs échelles constitue une source potentielle de confusion spectrale pour les classifications fondées sur des pixels de résolution kilométrique à décamétrique, et justifie une lecture prudente des cartes de couvert ligneux dans les secteurs à dominante calcaire ou sableuse côtière.



Figure 3-6: Diversité des formations végétales du District de Toliara II selon les substrats pédologiques. 1. Formation arbustive à arborée sur sol ferrugineux en saison humide, couvert relativement fermé (est). 2. Même formation en saison sèche, illustrant la défeuillaison marquée des ligneux. 3. Fourré xérophile à *Alluaudia procera* sur sable roux en saison humide. 4 et 5. Fourrés xérophiles très ouverts sur dalle calcaire affleurante, à arbustes bas et contournés. 6. Végétation basse et éparse à *Euphorbia stenoclada* sur sable blanc côtier

3.5. Végétation

3.5.1 Gradient structural est–ouest et types de formations

La végétation du District de Toliara II s'organise selon deux gradients structuraux conjugués (Figure 3.5). Un gradient nord–sud, dicté par la pluviométrie, oppose les formations forestières denses du nord (précipitations 600–1 000 mm/an) aux fourrés xérophiles ouverts du sud (moins de 400 mm/an ; Gaspard et al., 2018; Leprun et al., 2009). Un gradient est–ouest, lié à la nature des substrats pédologiques, détermine la hauteur et la composition des formations ligneuses : depuis les forêts sèches sur sols ferrugineux à l'est, pouvant atteindre 8–12 m de hauteur, jusqu'aux végétations basses et ouvertes à *Euphorbia stenoclada* sur sables blancs côtiers à l'ouest, réduites à des arbustes de 2–4 m (Figure 2.3 ; Raselimanana et al., 2012; Razanaka, 2004). Ces deux gradients conjugués génèrent une mosaïque végétale complexe dont la cartographie par télédétection constitue l'objet central des Chapitres 3 à 5.

Les forêts denses sèches caducifoliées et les forêts de transition avec des vestiges de baobabs occupent principalement la partie orientale du district, sur substrats ferrugineux à sols roux (Figure 2.3). La strate arborée, dominée par *Commiphora spp.* et *Euphorbia spp.*, peut atteindre 8 à 12 m de hauteur. Ces formations présentent une phénologie caducifoliée marquée, avec une défeuillaison quasi-complète en saison sèche qui réduit fortement leur signal spectral en vert et proche infrarouge, compliquant leur discrimination par télédétection optique (Noromiarilanto et al., 2016).

En progressant vers l'ouest, les forêts sèches cèdent la place aux fourrés xérophiles arbustifs à *Alluaudia procera* sur sables roux, puis aux fourrés ouverts sur dalle calcaire (Figure 2.3). La strate arbustive, dense à 2–6 m de hauteur, présente un feuillage réduit ou absent dont la signature spectrale en saison sèche est proche de celle du sol nu (Abel et al., 2023). La frange côtière, sur sables blancs carbonatés en contact avec le Canal du Mozambique, est occupée par une végétation basse et ouverte à *Euphorbia stenoclada* (Figure 2.3 ; Morat, 1969; Razanaka, 2004). Ces formations xérophiles constituent la forme d'adaptation extrême des végétations ligneuses aux conditions d'aridité sévère et représentent le principal objet d'étude des Chapitres 4, 5 et 6 de cette thèse.

3.5.2 Composition floristique et endémisme remarquable

Le taux d'endémisme spécifique des écosystèmes du sud-ouest malgache est exceptionnel : il atteint 95 % au niveau de l'espèce et 48 % au niveau du genre (Razanaka, 2004). Trois familles botaniques concentrent l'essentiel de la diversité des fourrés xérophiles et présentent des adaptations morphologiques spécifiques qui compliquent leur détection par télédétection optique.

- Didiereaceae : *Didiera madagascariensis*, *Alluaudia procera*, *Alluaudia ascendens*, *Alluaudia dumosa* — espèces exclusivement endémiques de Madagascar, à tiges columellaires succulentes et quasi-aphylles (Figure 2.3).

- Euphorbiaceae : *Euphorbia stenoclada*, *Euphorbia laro* — espèces à tiges photosynthétiques charnues, à latex blanc, dont la signature spectrale est quasi-

identique au sol nu en saison sèche (Figure 2.3; Abel et al., 2023).

• Fabaceae (Leguminosae) : *Acacia spp.*, *Baudouinia spp.*, *Colvillea racemosa* — espèces plus abondantes dans les fourrés semi-décidus intérieurs, sur substrats sablo-ferrugineux.

Ces adaptations morphologiques — microphyllie, aphyllie, spinescence et succulence — ont pour conséquence directe une réduction de l'indice de surface foliaire et une similarité de signature spectrale avec le sol nu en saison sèche, ce qui constitue l'un des défis méthodologiques centraux traités dans le Chapitre 3 (Abel et al., 2023; Gaspard et al., 2018).

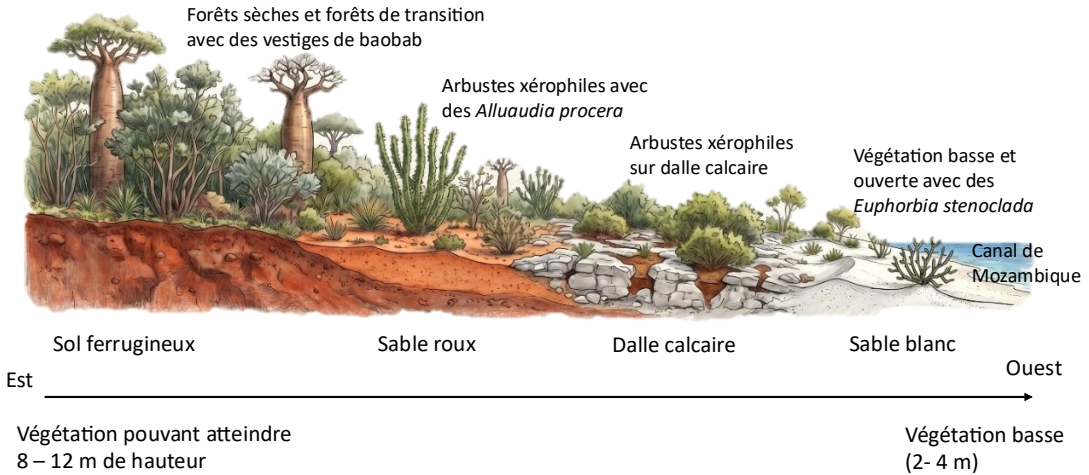


Figure 3-7: Transect schématique est-ouest illustrant le gradient structural et floristique de la végétation du District de Toliara II en fonction des substrats pédologiques

3.5.3 Valeur de conservation de la faune

Les écosystèmes xérophiiles du District de Toliara II abritent une faune vertébrée d'exception. Les espèces les plus remarquables recensées dans les forêts sèches du nord du district comprennent (Raselimanana et al., 2012; Razanaka, 2004) :

- Lémuriens diurnes : *Propithecus verreauxi* (sifaka de Verreaux, classé Endangered IUCN), *Lemur catta* (maki, Endangered IUCN).
- Lémuriens nocturnes : *Lepilemur ruficaudatus*, *Microcebus coquereli*.
- Grand prédateur endémique : *Cryptoprocta ferox* (fosa), plus grand mammifère carnivore endémique de Madagascar, Vulnérable IUCN.
- Mammifères : *Potamochoerus larvatus* (phacochère malgache), hérissons endémiques (*Tenrec ecaudatus*, *Echinops telfairi*).
- Avifaune endémique : *Coua ruficeps*, *C. cursor*, *C. verreauxi*, *Newtonia archboldi*, *Falcula palliata* ; *Lophotibis cristata* (quasi-menacée IUCN ; Raselimanana et al., 2012).

La présence de ces espèces, dont la survie est directement conditionnée par le maintien du couvert ligneux, confère à la surveillance de la végétation xérophile une valeur de conservation qui dépasse le cadre strictement forestier. La quantification de la dynamique du couvert ligneux (Chapitre 8) répond ainsi à un enjeu de conservation identifié par les listes rouges UICN (Lewin et al., 2024).

3.6. Contexte socio-économique et démographique

Le District de Toliara II comptait environ 391 939 habitants en 2020 selon les projections de l'INSTAT fondées sur le recensement national de 2018, avec une population à majorité féminine (51,1 %) et majoritairement rurale (INSTAT, 2020). La taille moyenne des ménages ruraux dans le sud-ouest malgache est estimée à 5,2 personnes par ménage (INSTAT, 2020). La croissance démographique, estimée à 2,9 % par an à l'échelle nationale (INSTAT, 2020), entraîne une augmentation continue de la demande en ressources naturelles, notamment en bois-énergie et en terres agricoles, constituant un facteur de pression additionnel sur les formations xérophiles. La densité de population est plus élevée autour de la ville de Toliara et des axes routiers principaux, qui sont également les zones où la pression sur les ressources ligneuses est la plus marquée (Raoliarivelo et al., 2010).

L'économie locale repose sur trois activités principales, toutes fortement dépendantes des écosystèmes ligneux : l'agriculture pluviale à base de maïs et de manioc, l'élevage extensif de petits ruminants, et la production de charbon de bois (Hänke et al., 2017; Raoliarivelo et al., 2010). La ville de Toliara, dont la population est en croissance rapide, génère une demande en charbon estimée à plus de 43 000 tonnes par an (Randriamalala et al., 2016), demande qui est couverte en quasi-totalité par l'exploitation des fourrés xérophiles environnants. Cette demande urbaine croissante, couplée à l'absence d'accès généralisé à l'électricité dans les zones rurales, explique la persistance de la carbonisation artisanale comme stratégie de revenu pour les ménages ruraux malgré ses effets reconnus sur le couvert ligneux (Casse et al., 2004; Hänke et al., 2017).

3.7. Pressions anthropiques sur les écosystèmes xérophiles

3.7.1 Agriculture sur brûlis (hatsake)

L'agriculture sur brûlis, localement dénommée hatsake dans le sud-ouest de Madagascar, constitue la principale cause directe de déforestation dans le district (Carrière et al., 2022; Casse et al., 2004). Elle consiste en l'abattage et le brûlage de la végétation ligneuse pour mettre en culture des parcelles destinées principalement au maïs, abandonnées après deux à trois années d'exploitation en raison de l'épuisement rapide des sols sableux (Casse et al., 2004) (Figure 3.6A). Dans le district de Toliara II, Raoliarivelo et al., (2010) ont documenté l'expansion de cette pratique en lien direct avec la croissance démographique et l'absence d'alternatives agricoles sur les substrats calcaires et sableux. La régénération des fourrés xérophiles après brûlis est lente en conditions arides, de l'ordre de plusieurs décennies, ce qui

rend la perte de couvert localement irréversible à l'échelle d'observation de cette thèse (Noromiarilanto et al., 2016; Randriamalala et al., 2019).

L'amplitude et la distribution spatiale de la déforestation par hatsake dans la zone d'étude effective sont quantifiées dans le Chapitre 8 à partir d'une série temporelle Landsat couvrant 36 ans.

3.7.2 Production de charbon de bois

La production de charbon de bois représente la deuxième menace majeure sur le couvert ligneux du district (Figure 3.6B). Randriamalala et al., (2016) estiment que la ville de Toliara consomme plus de 43 000 tonnes de charbon par an, produit quasi-exclusivement par carbonisation artisanale dans les fourrés xérophiles environnants, principalement à base d'*Euphorbia* et d'*Alluaudia* (Hänke et al., 2017). Cette exploitation est le fait de ménages ruraux dont la production de charbon constitue souvent la principale source de revenus monétaires (Casse et al., 2004; Hänke et al., 2017). Contrairement à l'agriculture sur brûlis, la carbonisation n'implique pas nécessairement une destruction totale de la strate ligneuse mais engendre une dégradation progressive du couvert par réduction sélective des individus les plus grands, affectant la structure et la densité des formations sans aboutir systématiquement à une conversion de l'usage du sol (Brinkmann et al., 2014; J. R. Randriamalala et al., 2015).

3.7.3 Feux de pâturage et pratiques pastorales

L'élevage dans le District de Toliara II est essentiellement extensif, dominé par des troupeaux de caprins et d'ovins qui pâturent librement dans les formations xérophiles (Figure 3.6C). Il est important de noter que le pâturage de petits ruminants (chèvres et moutons) ne constitue pas actuellement une menace directe sur le couvert ligneux des fourrés xérophiles dans cette région. En effet, les chèvres du sud-ouest malgache sont des brouteuses sélectives qui exploitent une diversité remarquable d'espèces végétales (163 espèces recensées dans la zone d'étude) sans exercer de pression directionnelle sur la composition de la communauté végétale ligneuse (Hester et al., 2006; Ranaivoharivelo et al., 2026). Ranaivoharivelo et al. (2026) ont montré, dans les communes de Soalara-Sud et d'Andranohinaly (District de Toliara II), que le comportement alimentaire des caprins dans les fourrés xérophiles est guidé principalement par la qualité nutritionnelle et l'accessibilité des fourrages, avec une flexibilité saisonnière importante, sans que la pression de pâturage ne se traduise par une dégradation directe du couvert ligneux. La consommation directe de la strate ligneuse reste donc marginale en dehors de situations de surpâturage localisé (Hänke et al., 2017; Randriamalala et al., 2022).

En revanche, la pratique des feux de pâturage — mises à feu délibérées pour stimuler la repousse de l'herbe et faciliter l'accès aux parcours — constitue une menace potentielle sur les formations ligneuses dont les effets méritent attention. Cette pratique est surtout documentée dans le secteur nord-est de la zone d'étude, où la végétation est caractérisée par des formations herbacées et arbustives de type savane à *Heteropogon contortus* sur plateau calcaire, plus inflammables que les fourrés

xérophiles sensu stricto de la zone côtière (Noromiarilanto et al., 2016; Rakotoarimanana and Grouzis, 2006). Dans ces secteurs, des éleveurs allument délibérément des feux pour favoriser la repousse de l'herbe et améliorer l'accès aux zones de parcours (Scales, 2014). Les effets à long terme de ces feux pastoraux sur la dynamique du couvert ligneux dans la zone d'étude sont analysés de manière quantitative au Chapitre 8.

Dans les systèmes arides africains comparables, les feux pastoraux ont été identifiés comme un vecteur important de dégradation du couvert ligneux (Bond and Keeley, 2005; Rakotoarimanana and Grouzis, 2006; Smit et al., 2010; Wigley et al., 2009). Dans les fourrés xérophiles malgaches, l'effet des feux sur la végétation ligneuse dépend fortement du type de formation concerné : les formations herbacées et arbustives du plateau calcaire nord-est sont plus vulnérables aux perturbations par le feu que les fourrés côtiers denses dominés par les Euphorbiaceae et Didiereaceae (Randriamalala et al., 2019).

A. Agriculture sur-brûlis



C. Production de charbon de bois



B. Feux de pâturage



Figure 3-8: Illustrations des pressions anthropiques sur les fourrés xérophiles du District de Toliara II.

3.7.4 Gouvernance forestière et crises politiques

La gestion des ressources forestières dans le sud-ouest de Madagascar s'inscrit dans un contexte de gouvernance institutionnelle fragile. Les crises politiques successives de 2002 et 2009 ont conduit à un affaiblissement récurrent des capacités de contrôle de l'administration forestière, avec pour conséquence des augmentations ponctuelles des pressions d'exploitation sur les formations ligneuses (Scales, 2014; Valencia, 2014).

Le système de gestion communautaire des ressources forestières (GELOSE/VOI), mis en place à partir de 1996, a été largement affecté par ces crises institutionnelles. Les transferts de gestion aux communautés locales, qui constituaient un levier potentiel de conservation, ont souvent été neutralisés par le retrait du soutien institutionnel en période de transition politique (Scales, 2014). Cette fragilité structurelle accroît le risque de déforestation dans les périodes d'instabilité, en autorisant de fait des pratiques d'exploitation non contrôlées (De Marzo et al., 2021; Valencia, 2014).

3.8. Historique des études cartographiques antérieures dans la zone

Avant les travaux de cette thèse, les études cartographiques des formations ligneuses du District de Toliara II étaient peu nombreuses et souffraient de limitations méthodologiques structurelles. Les principaux travaux disponibles sont les suivants.

Lasry et al., (2004), à partir d'images SPOT, ont évalué la déforestation des forêts de Mikea et extrapolent un taux de déforestation observé entre 1986 et 2001 pour prédire la disparition totale de ces forêts à l'horizon 2003–2010. Ces projections illustrent les biais induits par les classifications spectrales classiques appliquées aux végétations xérophiles, dont la signature spectrale en saison sèche se confond avec le sol nu (Abel et al., 2023).

Brinkmann et al., (2014) ont produit des cartographies binaires forêt/non-forêt sur images Landsat pour plusieurs localités du sud-ouest (Beheloka, Betioky-Sud, Itampolo), en utilisant une classification par maximum de vraisemblance. Cette approche ne capture ni la diversité structurale des formations ni la dégradation graduelle du couvert.

Randriamalala et al., (2015) ont cartographié les dynamiques spatiales des fourrés xérophiles de Betioky-Sud et Soalara-Sud sur trois dates Landsat (1990, 2000, 2010), produisant des matrices de transitions post-forestières à l'échelle des communes. Si cette étude constitue l'analyse diachronique la plus ciblée sur les fourrés xérophiles du District de Toliara II, elle demeure limitée à trois dates de classification catégorielle, sans variable continue de couvert, et ne couvre pas l'ensemble de la période d'observation Landsat disponible depuis 1989. Ces estimations de couvert sous-évaluées pour les formations ouvertes sont cohérentes avec le biais systématique identifié par Bastin et al., (2017) pour les zones arides tropicales.

Ces lacunes cartographiques constituent la motivation empirique directe des Chapitres 3 à 6 de cette thèse. Elles illustrent à l'échelle locale le déficit de représentation des végétations ligneuses arides dans les produits globaux, documenté à l'échelle planétaire par Bastin et al., (2017).

3.9. Justification du choix de la zone d'étude et pertinence pour les objectifs de la thèse

Le District de Toliara II constitue un cas d'étude adapté aux objectifs de cette thèse pour trois raisons complémentaires. Premièrement, l'écosystème présente une hétérogénéité biophysique marquée : le gradient de précipitations nord-sud et le gradient structural intérieur-côte génèrent un continuum de formations ligneuses — des forêts denses aux fourrés arbustifs très ouverts — sur une superficie contenue, ce qui offre un cadre idéal pour tester les méthodes de cartographie par télédétection sur une large gamme de couverts. Les particularités spectrales des espèces dominantes — Didiereaceae et Euphorbiaceae à tiges succulentes, aphylls ou microphylls — dont la signature en saison sèche se confond avec le sol nu (Abel et al., 2023), amplifient

la difficulté de détection et rendent ce terrain scientifiquement exigeant. Deuxièmement, les pressions anthropiques (hatsake, carbonisation, feux pastoraux) sont actives, diversifiées et quantifiables sur la période couverte par les archives Landsat (1989–2025), ce qui permet d'analyser des dynamiques temporelles réelles dans le Chapitre 8. Troisièmement, les conditions bioclimatiques ($IA = 0,21$), la structure clairsemée des formations et les pressions d'usage sont représentatives d'autres régions arides tropicales d'Afrique et d'Asie du Sud (Bastin et al., 2017; Brandt et al., 2020), ce qui confère aux méthodes développées une transférabilité potentielle au-delà de Madagascar.

Chapitre 4

Méthodologie générale

Ce chapitre présente le cadre méthodologique commun aux quatre articles de cette thèse. Il décrit de manière systématique les données mobilisées, les protocoles de collecte, les méthodes de modélisation, et les procédures de validation, afin de donner au lecteur les clés nécessaires à la compréhension des chapitres analytiques qui suivent. Chaque article étant centré sur une question spécifique, certaines étapes méthodologiques sont partagées entre plusieurs études ; la section 4.9 précise l'articulation entre chaque article, ses hypothèses, ses données, et les méthodes qui lui sont propres.

4.1. Vue d'ensemble de la démarche

La démarche scientifique adoptée dans cette thèse repose sur trois enchaînements logiques. Le premier vise à établir un état de l'art des méthodes de cartographie existantes appliquées aux écosystèmes arides malgaches, pour en identifier les lacunes et les biais (chapitre 5). Le deuxième vise à développer et optimiser un cadre de cartographie du couvert ligneux continu adapté à ces milieux, en comparant des sources de données satellitaires et des stratégies d'échantillonnage de calibration (chapitre 6 et 7). Le troisième vise à analyser les dynamiques temporelles du couvert sur 36 ans et à en identifier les facteurs anthropiques explicatifs (chapitre 8).

Ces trois enchaînements mobilisent une chaîne méthodologique commune articulée autour de quatre éléments (Figure 4.1) : (1) des données de référence terrain et par photo-interprétation servant à la calibration et à la validation des modèles ; (2) des données prédictives ; (3) des modèles de régression Random Forest (RF) produisant des estimations continues du couvert ligneux ; (4) des analyses spatiales et statistiques portant sur les dynamiques de dégradation et leurs déterminants. La Figure 4.1 représente le schéma général de la démarche.

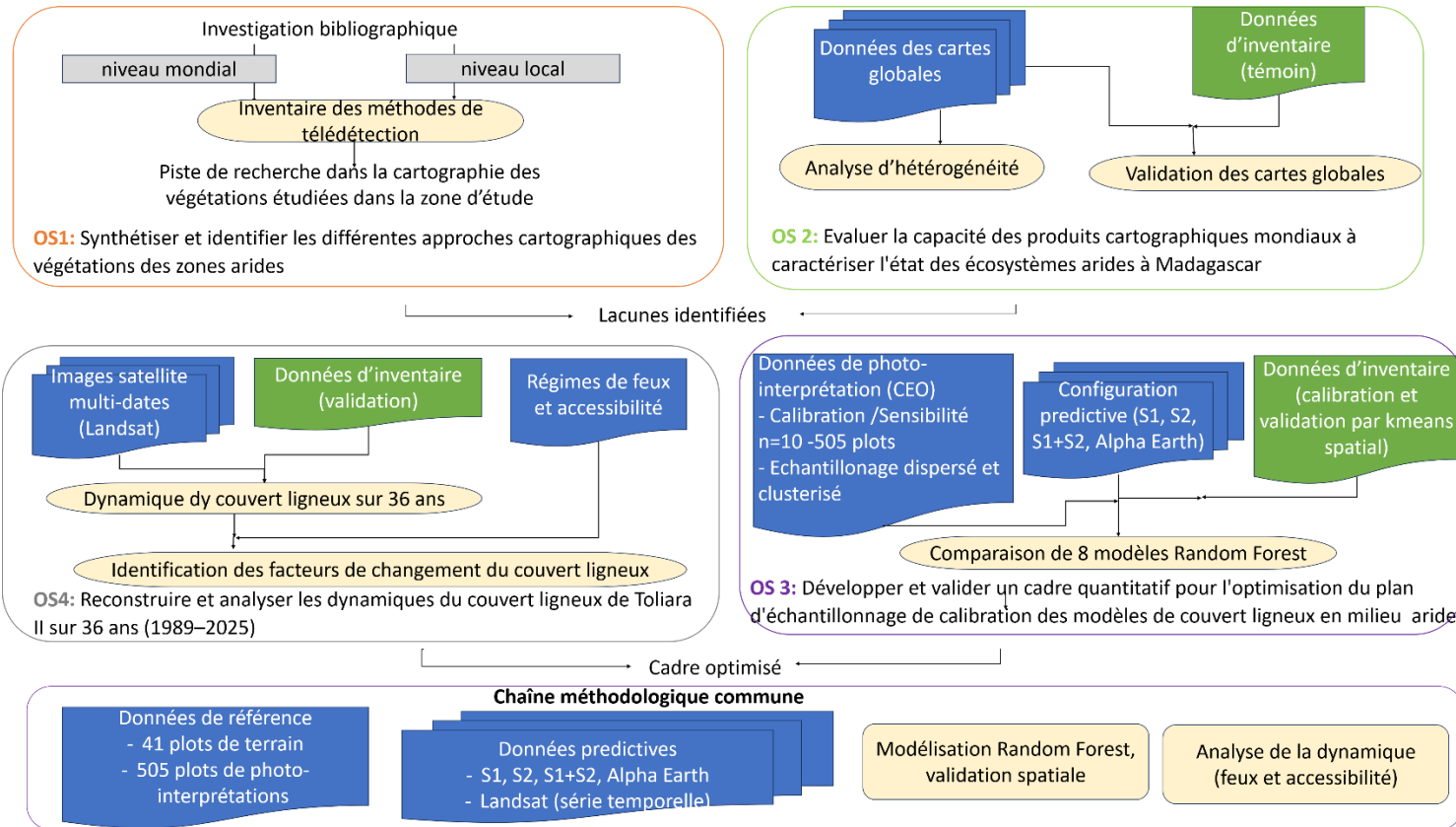


Figure 4-1 : Cadre méthodologique général de la thèse. Les quatre objectifs spécifiques (OS1–OS4) sont représentés par des cadres colorés organisés selon une progression logique : recension des méthodes (OS1), évaluation des produits globaux (OS2), optimisation de l'échantillonnage (OS3), et analyse des dynamiques long terme (OS4). La rangée inférieure présente les éléments méthodologiques communs aux quatre articles : données de référence, configurations prédictives, modélisation Random Forest et analyse des dynamiques.

4.2. Données de référence

Deux sources de données de référence ont été produites pour cette thèse : un inventaire terrain de 41 placettes — une placette désignant ici une unité d'échantillonnage de surface fixe (carrée ou circulaire, délimitée au sol ou sur l'image) au sein de laquelle le couvert ligneux est mesuré ou estimé — réalisé en avril-mai 2022, et un jeu de photo-interprétation de 505 placettes acquises entre 2021 et 2022 via la plateforme Collect Earth Online. La localisation des jeux de placettes au sein du district est représentée en figure 5.2. Ces deux jeux de données sont complémentaires : les placettes terrain constituent la référence indépendante de validation, tandis que les placettes photo-interprétées constituent la principale source de calibration des modèles Random Forest. Les caractéristiques respectives des deux jeux de données de référence sont résumées en Figure 4.2.

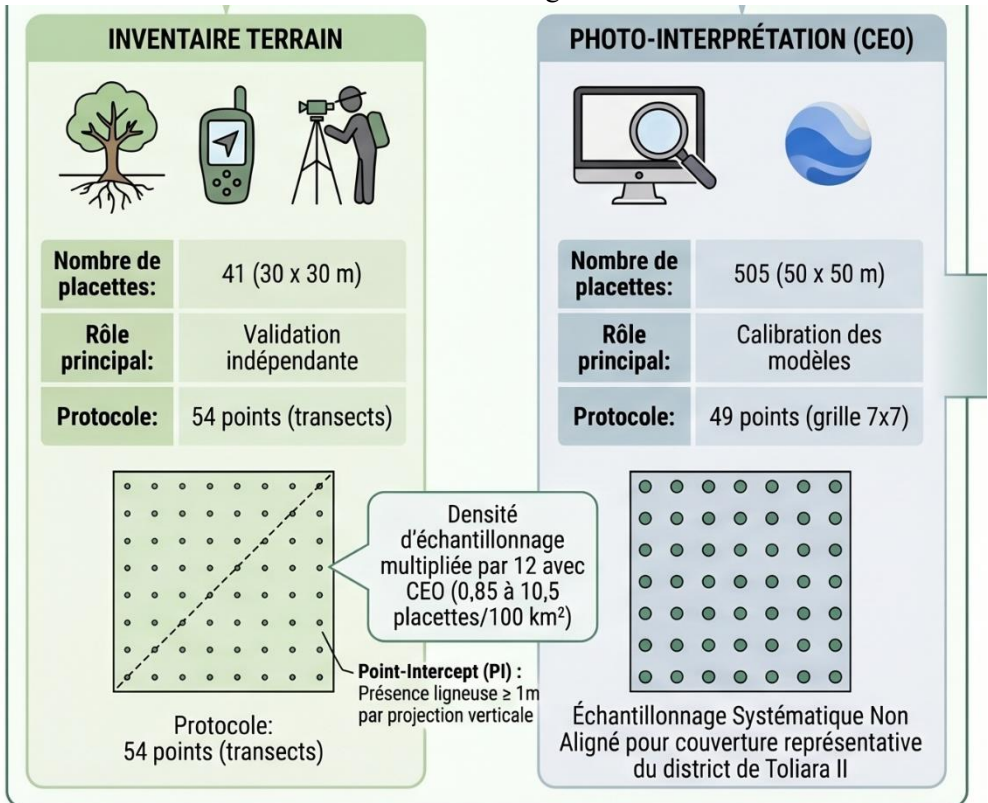


Figure 4-2 : Comparaison des deux sources de données de référence : inventaire terrain (41 placettes, validation) et photo-interprétation CEO (505 placettes, calibration).

4.2.1 Inventaires terrain

Les données de terrain ont été collectées entre avril et mai 2022, en coïncidence avec la période d'acquisition des données satellitaires Sentinel utilisées dans les

Chapitres 7 et 8. Quarante et une placettes de 30×30 m ont été établies selon un échantillonnage aléatoire stratifié, les strates étant définies par le croisement de cinq classes de couvert ligneux (à intervalles de 20 %) et de quatre types de formation végétale. Cette stratégie garantit la représentation du gradient complet de couvert, des formations ouvertes (< 10 %) aux fourrés denses (> 80 %). La sélection des sites a privilégié les zones d'homogénéité structurelle dans un rayon de 200 m, afin de limiter l'effet de la variabilité infra-pixel sur les comparaisons observé/prédict (Karl et al., 2017). La géolocalisation des placettes a été réalisée avec un Garmin GPSMAP 65s (précision $\pm 1,5$ m) et la délimitation au sol à l'aide d'un hypsomètre Vertex.

Le couvert ligneux a été mesuré par la méthode point-intercept (PI), adaptée du protocole du réseau australien TERN (Australian Terrestrial Ecosystem Research Network ; Karl et al., 2017; White et al., 2012). Au sein de chaque placette, 54 points de mesure ont été distribués le long de quatre transects orthogonaux espacés de 10 m, avec un point tous les 2 m. La présence de végétation ligneuse d'une hauteur supérieure ou égale à 1 m a été enregistrée à chaque point par projection verticale. Le couvert ligneux par placette a été calculé comme le rapport entre le nombre de points positifs et le nombre total de points observés.

La méthode point-intercept présente plusieurs avantages pour la cartographie par télédétection : elle fournit une estimation directe du couvert en projection verticale, compatible avec la grandeur physique mesurée par les capteurs optiques ; elle est reproductible et indépendante de l'opérateur lorsque le protocole est standardisé ; elle permet une comparaison directe avec les estimations des placettes photo-interprétées ($R^2 = 0,775$, RMSE = 0,125 entre les 41 co-placettes ; Chapitre 7). Ces 41 placettes ont servi de données de validation indépendante pour l'ensemble des modèles Random Forest développés aux Chapitres 7 et 8. Le dispositif de placette et le protocole de mesure sont schématisés en Figure S4.1 (Annexe S4).

4.2.2 Photo-interprétation

Un total de 505 placettes de 50×50 m a été établi par photo-interprétation d'orthophotographies à très haute résolution (2021-2022) via la plateforme Collect Earth Online (CEO ; Bey et al., 2016). CEO est une plateforme web développée par Google et Clark University qui permet l'interprétation structurée de placettes sur des mosaïques d'images Google Earth, Bing, et Planet Labs à des résolutions de 0,3 à 3 m.

Le plan d'échantillonnage suit un protocole systématique non aligné (« *unaligned systematic sampling* » ; Cochran, 1977), avec un tirage aléatoire d'un point par maille d'une grille régulière couvrant l'ensemble du district de Toliara II. La distribution spatiale des 505 placettes selon ce protocole (Stratified Systematic Unaligned Sampling, SSUS) est présentée en Figure S5.1 (Annexe S5). Des variables environnementales d'aide à l'interprétation — altitude (NASADEM), précipitations annuelles moyennes, indices NDVI, EVI et NDWI — étaient affichées comme couches contextuelles dans l'interface CEO pour guider l'évaluation visuelle (Colditz et al., 2011).

Au sein de chaque placette CEO, 49 points d'observation étaient disposés sur une grille 7×7 espacés d'environ 7 m. La présence ou l'absence de végétation ligneuse à chaque point a été déterminée par projection verticale, en appliquant des critères biophysiques identiques à ceux utilisés sur le terrain (hauteur minimale de 1 m). Le couvert ligneux de la placette a été calculé comme la proportion de points positifs. Cette approche augmente la densité d'échantillonnage spatiale d'un facteur 12 par rapport à l'inventaire terrain (10,5 vs 0,85 placettes/100 km²), tout en maintenant la compatibilité métrique avec les mesures au sol.

La validation interne de la méthode de photo-interprétation, réalisée par comparaison directe avec les 41 placettes terrain co-localisées ($R^2 = 0,775$; RMSE = 0,125 ; pente = 0,92 — détails en §4.2.1 et Chapitre 7), confirme la validité des estimations CEO comme données de calibration dans le contexte logistique contraint du sud-ouest malgache, où le coût d'un inventaire terrain exhaustif serait prohibitif. La distribution comparée des valeurs de couvert ligneux entre les deux jeux de données est illustrée en Figure S4.2 (Annexe S4), confirme l'absence de biais systématique entre les deux méthodes.

4.3. Données prédictives

4.3.1 Sentinel-1/2 et AlphaEarth Foundations

Dans le Chapitre 7, consacré à l'optimisation du plan d'échantillonnage, quatre configurations de données satellitaires ont été comparées pour l'année de référence 2022. Toutes les acquisitions couvrent la période avril 2022, coïncidant avec la collecte terrain.

Sentinel-1 (S1) : données SAR en bande C issues de la collection GEE COPERNICUS/S1_GRD (mode « Interferometric Wide Swath », polarisations VV et VH, résolution 10 m). Un total de 10 scènes ascendantes et descendantes a été composé en médiane mensuelle pour avril 2022. La correction du bruit thermique, la calibration radiométrique, et la correction de terrain sont incluses dans le pipeline standard GRD (Torres et al., 2012). L'intérêt du radar réside dans sa sensibilité partielle à la structure du couvert et à l'humidité de la végétation, indépendante des conditions d'éclairement solaire.

Sentinel-2 (S2) : données de réflectance de surface Level 2A issues de la collection COPERNICUS/S2_SR, après correction atmosphérique par le processeur ESA Sen2Cor. Les bandes B2 (bleu, 490 nm), B3 (vert, 560 nm), B4 (rouge, 665 nm), B8 (PIR, 842 nm), B11 (SWIR, 1 610 nm) et B12 (SWIR, 2 190 nm) ont été rééchantillonnées à 10 m. Un total de 25 scènes sélectionnées sur un critère de couverture nuageuse inférieure à 20 % (CLOUDY_PIXEL_PERCENTAGE) ont été composées en médiane pour avril 2022 (Drusch et al., 2012).

Sentinel-1+2 (S1+S2) : configuration combinée intégrant la sensibilité structurelle du radar (S1) et la discrimination spectrale optique (S2), selon l'approche proposée par Acharki et al. (2021) pour les milieux arides.

AlphaEarth Foundations : modèle géospatial d'embedding générant des représentations de 64 bandes capturant les trajectoires spatio-temporelles annuelles

d'un pixel (Brown et al., 2025). Chaque embedding intègre des indices spectraux (NDVI, EVI, NDWI), des métriques de texture issues d'images optiques et radar (Sentinel-2 L1C, Landsat 8/9, Sentinel-1 GRD), des variables topographiques (pente, exposition, courbure depuis le Copernicus DEM GLO-30) et des indices climatiques spatiaux (ERA5-Land). Cette configuration offre la représentation prédictive la plus complexe et la plus intégrée des quatre.

4.3.2 Série temporelle Landsat multi-capteurs 1989-2025

L'analyse des dynamiques à long terme (Article 4, chapitre 8) s'appuie sur une série temporelle Landsat couvrant la période 1989-2025, extraite des produits Collection 2 Level-2 via Google Earth Engine (Cardille et al., 2024; Gorelick et al., 2017). Ces produits fournissent la réflectance de surface après correction atmosphérique par les algorithmes LEDAPS pour Landsat 4-7 (Masek et al., 2006) et LaSRC pour Landsat 8-9 (Vermote et al., 2016), dans le cadre de l'archive Collection 2 produite par l'USGS (Crawford et al., 2023). Les données ont été acquises successivement à partir de Landsat 5 TM (1989-2011), Landsat 7 ETM+ (2012), Landsat 8 OLI (2013-2021), et Landsat 9 OLI-2 (2022-2025). L'inventaire complet des 1 527 scènes utilisées est fourni en annexe (Tableau S1 du Chapitre 8).

La fenêtre temporelle d'acquisition annuelle a été fixée à janvier-juin, correspondant à la transition saison des pluies / saison sèche. Cette période présente deux avantages décisifs dans le contexte des fourrés xérophiles : les espèces décidues conservent encore une fraction significative de leur feuillage, maximisant le signal spectral de la biomasse ligneuse ; la couverture nuageuse est réduite par rapport à la pleine saison des pluies (décembre-mars), augmentant le nombre de scènes utilisables. Des composites médians annuels ont été générés à partir de toutes les scènes disponibles dans cette fenêtre, après élimination des nuages, ombres et pixels saturés via les masques de qualité Collection 2 (bandes QA_PIXEL et QA_RADSAT).

Quatre années ont présenté une couverture nuageuse excessive dans la fenêtre standard (1992, 2000, 2009, 2010) : pour ces années, le composite médian a été calculé à partir des scènes disponibles sur les deux années adjacentes (année-1 et année+1), constituant un pool de trois ans. Cette procédure maintient la cohérence phénologique tout en maximisant la disponibilité de scènes, conformément aux approches de composition multi-temporelle appliquées dans des études Landsat long terme confrontant des lacunes d'images (Hermosilla et al., 2015; White et al., 2012). Les années 2002 et 2003 n'ont pu être traitées en raison d'un nombre insuffisant de scènes utilisables après filtrage nuageux, combiné à la défaillance du Scan Line Corrector de Landsat 7 ETM+ en mai 2003 (Markham et al., 2004). Pour les analyses nécessitant une série temporelle continue, les valeurs de ces deux années ont été estimées par interpolation linéaire entre 2001 et 2004.

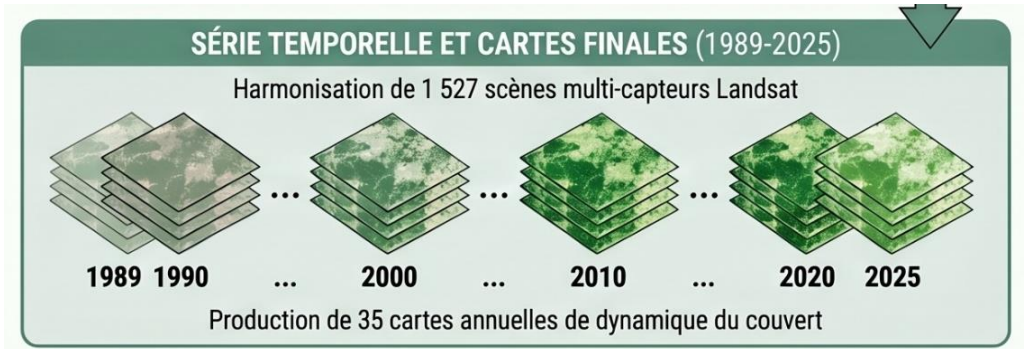


Figure 4-3 : Construction de la série temporelle Landsat harmonisée (1989-2025) à partir de 1 527 scènes multi-capteurs, produisant 35 cartes annuelles de couvert ligneux.

4.3.3 Variables spectrales et temporelles

Dix indices spectraux ont été calculés sur chaque composite annuel Landsat pour caractériser les dimensions structurelles et fonctionnelles de la végétation : NDVI (Rouse et al., 1974), NDWI (Gao, 1996), SAVI (Huete, 1988), EVI (Huete et al., 2002), NBR (Key and Benson, 2006), NDBI (Zha et al., 2003), ratio NIR/Red (Jordan, 1969), ratio SWIR/NIR, Luminosité et Verdure du Tasseled Cap (Crist and Cicone, 1984). Ces indices complètent les sept bandes de réflectance de surface Landsat (B1-B7 pour Landsat 8/9), constituant un espace prédictif de 17 variables (Tableau 4.1).

L'harmonisation inter-capteurs constitue un préalable indispensable à la construction d'une série temporelle couvrant plusieurs générations de capteurs. La différence principale entre Landsat 5/7 et Landsat 8/9 tient à l'absence de la bande SWIR2 (2,09-2,35 μm) sur les capteurs TM et ETM+. Pour assurer la cohérence spectrale, un pseudo-SWIR2 a été généré pour Landsat 5 et 7 en multipliant la bande SWIR1 (1,55-1,75 μm) par un facteur d'ajustement empirique de 0,95 pour approximer le SWIR2 de Landsat 8/9 (Roy et al., 2016). Cette procédure produit un espace prédictif homogène de 17 variables sur l'ensemble de la série 1989-2025.

Tableau 4-1 : Variables spectrales et temporelles utilisées dans les modèles Random Forest (série Landsat 1989-2025).

N°	Variable	Formule / source	Dimension captée
1-7	Bandes B1-B7 (réflectance de surface)	Collection 2 USGS	Signal brut multispectral
8	NDVI	$(\text{NIR}-\text{R})/(\text{NIR}+\text{R})$ — (Rouse et al., 1974)	Activité photosynthétique
9	NDWI	$(\text{NIR}-\text{SWIR1})/(\text{NIR}+\text{SWIR1})$ — (Gao, 1996)	État hydrique
10	SAVI	$[(\text{NIR}-\text{R})/(\text{NIR}+\text{R}+\text{L})] \times (1+\text{L})$, $\text{L}=0,5$ — (Huete, 1988)	Couvert avec correction sol
11	EVI	$2,5 \times (\text{NIR}-\text{R})/(\text{NIR}+6\text{R}-7,5\text{B}+1)$ — (Huete et al., 2002)	Couvert dense, moins saturé
12	NBR	$(\text{NIR}-\text{SWIR2})/(\text{NIR}+\text{SWIR2})$ — (Key and Benson, 2006)	Feux, dégradation post-incendie
13	NDBI	$(\text{SWIR1}-\text{NIR})/(\text{SWIR1}+\text{NIR})$ — (Zha et al., 2003)	Surfaces bâties, sol nu
14	Ratio NIR/Red	NIR/Red — (Jordan, 1969)	Biomasse relative
15	Ratio SWIR/NIR	$\text{SWIR1}/\text{NIR}$	Stress hydrique
16	Tasseled Cap Luminosité	Combinaison lin. pondérée — (Crist and Cicone, 1984)	Réflectance générale
17	Tasseled Cap Verdure	Combinaison lin. pondérée — (Crist and Cicone, 1984)	Végétation photosynthétique

4.4. Modélisation Random Forest

4.4.1 Paramètres d'entraînement

L'estimation du couvert ligneux a été réalisée par l'algorithme Random Forest (RF) utilisé en mode régression, implémenté sous Google Earth Engine (Belgiu and Drăguț, 2016; Breiman, 2001). RF constitue un ensemble d'arbres de décision construits sur des sous-échantillons aléatoires des données d'entraînement ; la prédiction finale est obtenue comme la moyenne des prédictions de l'ensemble des arbres. Le choix du mode régression permet de produire une estimation continue du couvert dans l'intervalle [0-100 %], évitant le recours à un seuil de classification qui exclut mécaniquement les formations ouvertes.

Le RF peut être mis en œuvre selon deux modes principaux. En mode classification, l'algorithme attribue chaque pixel à une classe discrète (par exemple présence/absence de couvert ligneux, ou classes de densité), ce qui facilite l'interprétation thématique mais impose un seuillage arbitraire et peut masquer les gradients de couvert continu, en particulier dans les formations ouvertes ou hétérogènes. En mode régression, l'algorithme prédit une valeur continue (ici, le pourcentage de couvert ligneux), ce qui préserve la variabilité fine du gradient mais rend l'évaluation et l'interprétation des erreurs plus complexes (RMSE plutôt que matrice de confusion). Le mode régression a été retenu dans cette thèse en raison de l'hypothèse d'un paysage « patchy », c'est-à-dire structuré par une mosaïque de formations végétales aux limites diffuses et aux densités de couvert continûment variables, pour lequel un seuillage en classes discrètes risquerait de gommer des transitions écologiquement significatives (voir également Chapitre 7, section sur l'architecture Random Forest, pour la discussion appliquée à la comparaison Sentinel/AlphaEarth).

La variable réponse (couvert ligneux, exprimé en proportion) a été transformée en racine carrée avant l'entraînement, afin de stabiliser la variance et d'améliorer les performances prédictives aux extrémités du gradient de couvert (0-10 % et 80-100 %). Les prédictions ont été retransformées par mise au carré et contraintes à l'intervalle [0-100 %]. Les hyperparamètres retenus sont détaillés dans l'Article 2, Chapitre 8.

Le modèle calibré sur l'année de référence 2022 a été appliqué à l'ensemble de la série temporelle Landsat (1989-2025), produisant 35 cartes annuelles de couvert ligneux à 30 m de résolution spatiale. Cette hypothèse de stationnarité du modèle repose sur le fait que les relations spectrales entre le signal Landsat et le couvert ligneux sont physiquement stables à travers le temps pour des capteurs harmonisés, les variations interannuelles du couvert étant captées par la variation du signal spectral lui-même.

Pour le chapitre 7 (comparaison Sentinel/AlphaEarth), les hyperparamètres ont été optimisés par recherche sur grille (grid search) : $n_{tree} = 500$, $m_{try} = \sqrt{p}$ (où p est le nombre de prédicteurs), $n_{odesize} = 5$, conformément aux recommandations de (Probst et al., 2019). Huit modèles indépendants ont été entraînés (quatre configurations satellitaires \times deux sources de calibration : terrain vs. photo-interprétation) pour permettre la comparaison systématique de l'effet de la source de calibration et de la configuration spectrale sur la précision des modèles.

4.4.2 Importance des variables

L'importance relative de chaque variable prédictrice a été évaluée par la métrique Mean Decrease in MSE (diminution moyenne de l'erreur quadratique moyenne lors de la permutation d'une variable ; Breiman, 2001). Cette métrique permet d'identifier les bandes et indices spectraux les plus discriminants pour la prédiction du couvert ligneux dans les fourrés xérophiles, et de justifier le choix des variables retenues dans le modèle final. L'analyse de l'importance des variables contribue également à l'interprétation écologique des modèles : une importance élevée du SWIR et du NBR,

par exemple, indique que la structure ligneuse en saison sèche et les cicatrices de feu constituent des signaux informatifs, ce qui est cohérent avec la physique des milieux arides (Abel et al., 2023).

4.5. Validation des modèles : performance et validation spatiale

La performance des modèles a été évaluée par deux approches complémentaires, permettant de distinguer la précision locale (validation externe) et la généralisation spatiale (blockCV).

La validation externe a été réalisée par régression entre les prédictions du modèle et les observations des 41 placettes terrain indépendantes, calculant le R^2 , la RMSE, la MAE, et le biais. Ces 41 placettes n'ont jamais été utilisées dans la calibration des modèles, garantissant leur indépendance. Pour le modèle Landsat (Chapitre 8), les résultats de la validation externe indiquent $R^2 = 0,75$, $RMSE = 13,74 \%$, et un biais négligeable de $-0,21 \%$.

La validation spatiale par blocs (blockCV ; Valavi et al., 2018) a été appliquée aux 505 placettes photo-interprétées. Des blocs spatiaux de 10 km de côté ont été définis, et le jeu de données a été partitionné en 5 plis (folds). Cette approche évalue la capacité de généralisation géographique du modèle en s'assurant que les données d'entraînement et de test sont spatialement séparées, fournissant une estimation plus réaliste de la performance en dehors des zones échantillonnées (Roberts et al., 2017). Pour l'article 4, chapitre 8, le blockCV donne $R^2 = 0,64$ (blocs de 10 km), révélant un gradient de performance entre zones bien représentées et zones hétérogènes distantes des placettes de calibration. Pour le chapitre 3 (modèle Sentinel-1+2 avec photo-interprétation), le blockCV indique $R^2 = 0,88$ ($RMSE = 0,11$).

Pour l'article 3, chapitre 7, la validation spatiale sur les 41 placettes terrain a été implémentée par partitionnement en cinq plis géographiques via k-means (à partir des coordonnées des placettes, $k = 5$, $nstart = 25$), maximisant la séparation géographique entre plis. L'absence d'autocorrélation spatiale résiduelle entre plis a été vérifiée par le test de Moran ($I = -0,025$, $p = 0,5$), confirmant l'indépendance spatiale des plis.

4.6. Analyse des dynamiques temporelles

4.6.1 Classification des états de dégradation du couvert

Les cartes continues de couvert ligneux ont été reclassifiées en cinq états discrets représentant un gradient de dégradation écologique (Tableau 8.3, chapitre 8). Cette classification permet d'analyser l'évolution temporelle des superficies occupées par chaque état entre 1989 et 2025, et de visualiser les transitions de classes. Les bornes de classe ont été définies sur la base de l'état écologique des peuplements observés sur le terrain et validées par photo-interprétation drone sur les 41 placettes terrain.

La classe « formations non ligneuses » (0-10 %) correspond au seuil minimum de couvert définissant les formations ouvertes selon (Bastin et al., 2017; Brandt et al., 2020). Ce seuil de 10 % a également été utilisé pour calculer les taux de changement

du couvert ligneux selon la méthode FAO standardisée (Puyravaud, 2003), définissant la superficie ligneuse totale comme l'ensemble des pixels dépassant ce seuil.

4.6.2 Tendances 1989-2025

Les taux annuels de changement du couvert ligneux ont été calculés par la méthode standardisée FAO (Puyravaud, 2003) : $r = (1/t_2 - t_1) \times \ln(A_2/A_1)$, où A_1 et A_2 représentent les superficies de couvert ligneux aux dates t_1 et t_2 . Cette formule est analytiquement équivalente à un taux de croissance exponentiel continu, et évite le biais introduit par la méthode linéaire simple pour des périodes longues. Les taux ont été calculés pour la période complète 1989-2025 et pour quatre sous-périodes correspondant aux phases temporelles identifiées dans la série : 1989-2001, 2001-2013, 2013-2019 et 2019-2025.

L'évolution proportionnelle des superficies occupées par chaque classe de dégradation a été représentée sous forme d'une série temporelle de proportions de surface (area proportion plot), permettant de visualiser les transferts entre classes de couvert au cours du temps. Cette représentation intègre l'ensemble du gradient de couvert dans un même graphique, et révèle des tendances de réduction du couvert dense et d'expansion des classes dégradées qui sont invisibles dans les analyses de déforestation binaire conventionnelles.

4.6.3 Identification des points de changement

L'identification des points de rupture dans la série temporelle du couvert moyen a été réalisée par régression polynomiale de degré 2 (régression parabolique) appliquée à l'ensemble de la série 1989-2025. Cette approche permet d'identifier la date d'inflexion (point de minimum ou de maximum de la parabole), interprétable comme le tournant d'une phase de dégradation vers une phase de récupération potentielle, ou inversement. La robustesse de ces points d'inflexion a été vérifiée visuellement et par comparaison avec les tendances de couvert par classe.

Des analyses de tendance pixel par pixel ont été réalisées par régression linéaire de la pente du couvert (1989-2025), permettant de cartographier les zones de dégradation progressive (pente négative significative) et de récupération (pente positive significative) à l'échelle du district. Cette cartographie des tendances a été croisée avec la distance aux habitations et la fréquence des feux pour identifier des gradients spatiaux de pression.

4.7. Analyse des facteurs explicatifs

4.7.1 Régimes de feux — MODIS MCD64A1

Les superficies brûlées ont été cartographiées annuellement à partir du produit MODIS MCD64A1 v6.1, disponible à 500 m de résolution depuis 2001 (Giglio et al., 2018). Ce produit fournit la date de brûlage de chaque pixel détecté comme brûlé (champ BurnDate) ; un pixel a été codé présence de feu (BurnDate > 0) ou absence pour chaque année. La période d'analyse du régime de feux couvre 2000-2025,

correspondant à la disponibilité du produit. Avant 2000, aucun produit satellitaire de superficie brûlée ne combine résolution spatiale et couverture temporelle mensuelle comparables, rendant toute analyse cohérente du régime de feux impossible pour la période antérieure.

Trois variables ont été calculées pour chaque pixel : (1) la fréquence cumulative de feux (nombre d'années de feu sur 2000-2025), (2) le temps depuis le dernier feu (YSF, années depuis le dernier événement détecté), et (3) le statut brûlé/non-brûlé (au moins un feu détecté vs. aucun sur la période). Les associations entre ces variables de régime de feux et le changement de couvert ligneux ont été quantifiées par trois approches complémentaires : un test de Mann-Whitney comparant les zones brûlées et non brûlées, des corrélations de Spearman entre fréquence de feux, YSF, et changement de couvert, et une régression polynomiale de degré 2 pour identifier le point d'inflexion de la relation fréquence-changement de couvert. Un échantillon aléatoire de 10 000 points a été tiré dans l'ensemble du district comme compromis entre représentativité spatiale et faisabilité computationnelle ; après exclusion des points à données manquantes, 1 364 points ont été retenus pour l'analyse.

4.7.2 Gradient de distance aux habitations — Google Open Buildings

La localisation des bâtiments a été extraite de la base de données Google Open Buildings v3 (Sirko et al., 2021), extraite via Google Earth Engine. Cette base de données, construite par détection de bâtiments sur images satellitaires à haute résolution, constitue la source la plus exhaustive et à jour pour cartographier la présence humaine en milieu rural malgache, où les données cadastrales officielles sont absentes ou obsolètes.

Dix zones tampons concentriques de 1 à 10 km (par pas de 1 km) ont été délimitées autour de chaque bâtiment, définissant dix anneaux de distance caractérisant le gradient d'influence humaine sur le couvert ligneux. Dans chaque anneau, la superficie ligneuse (ha , seuil $> 10\%$) a été calculée pour 1989 et 2025, et le taux annuel de changement de couvert a été dérivé par la méthode Puyravaud, (2003) sur ces superficies. Des corrélations de Spearman ($n = 10$ zones) ont quantifié les associations entre distance aux bâtiments et dynamiques du couvert ligneux.

Cette approche par anneaux concentriques opérationnalise le modèle théorique de de Laurance et al., (2002) décrit au Chapitre 3 : elle permet de tester empiriquement l'hypothèse d'une dégradation décroissante avec la distance aux sources de demande, et d'en quantifier le gradient à l'échelle du district.

4.8. Articulation entre les articles

Le Tableau 4.2 présente la matrice de correspondance entre les quatre articles de la thèse, leurs hypothèses de recherche, les données mobilisées, et les méthodes spécifiques à chacun. Cette matrice permet de visualiser la cohérence de l'architecture d'ensemble : chaque article répond à une lacune identifiée au Chapitre 3, et les résultats de l'article précédent informent le suivant — Le chapitre 5 définit les lacunes

méthodologiques ; Les chapitres 6 et 7 développent et optimisent les solutions ; Le chapitre 8 applique le cadre optimisé à l'analyse long terme.

Tableau 4-2: Matrice d'articulation des quatre articles : hypothèses, données et méthodes.

Article	Titre court	Hypothèse principale	Données clés	Méthodes spécifiques
1	Recension des travaux antérieurs (acte de colloque)	H1.1 : Les méthodes antérieures n'identifient pas les écosystèmes ouverts xérophiles du SO Madagascar	Revue bibliographique, produits globaux (GFCC, GFW, CGLS, DW)	Analyse comparative qualitative, synthèse critique
2	Potentiel des cartes globales dans les écosystèmes arides	H 1.2 : Les produits globaux sous-estiment systématiquement le couvert dans les milieux xérophiles malgaches	41 placettes terrain, 505 placettes CEO, GFCC, GFW, CGLS-LC100, DW	Comparaison multi-produits, régression RF, matrice de confusion, ANOVA
3	Importance du plan d'échantillonnage	H 2.1 : La configuration spatiale de l'échantillonnage détermine la précision indépendamment de la complexité du capteur	41 placettes terrain, 505 placettes CEO, S1, S2, S1+S2, AlphaEarth	Comparaison 8 modèles RF, sensibilité à la densité (n = 10-505), blockCV
4	Dynamiques 36 ans (Landsat 1989-2025)	H 2.2 : La dégradation suit des effets-seuil non linéaires liés à la fréquence des feux et la proximité aux habitations	1 527 scènes Landsat, 505 CEO, 41 terrain, MODIS MCD64A1, Google Open Buildings	RF régression 35 cartes annuelles, pente pixel, méthode Puyravaud, Spearman, régression poly deg 2

4.9. Synthèse du chapitre 4

Ce chapitre a décrit l'ensemble du dispositif méthodologique déployé dans cette thèse. La chaîne méthodologique repose sur trois piliers : des données de référence à double source (41 placettes terrain PI et 505 placettes CEO), une série temporelle Landsat de 36 ans (1 527 scènes, 17 variables harmonisées), et une modélisation par forêt aléatoire en régression produisant des cartes continues de couvert ligneux validées par deux approches complémentaires (validation externe $R^2 = 0,75$; blockCV $R^2 = 0,64$). L'analyse des dynamiques mobilise la méthode FAO (Puyravaud, 2003) pour les taux de changement, et les corrélations de Spearman avec les régimes de feux (MODIS MCD64A1) et la proximité aux habitations (Google Open Buildings v3) pour identifier les facteurs explicatifs. L'articulation entre les trois articles suit une logique d'escalade de complexité et de résolution : de l'évaluation des produits globaux (Chapitre 5) à l'optimisation de l'échantillonnage pour la modélisation RF (Chapitre 6), jusqu'à l'application à la dynamique long terme par série temporelle Landsat (Chapitre 7).

Chapitre 5

Potential of global vegetation maps in capturing xerophytic vegetation cover: insights from Madagascar's arid ecosystems

5.1. Référence bibliographique

Ramalason, F.N., Rakotondrasoa, O.L., Linden, A.V., Renard, G., Ratsimba, H.R., Bogaert, J., Bastin, J.-F., 2025b. Potential of global vegetation maps in capturing xerophytic vegetation cover: insights from Madagascar's arid ecosystems. *Journal of Arid Environments* 229, 105406. <https://doi.org/10.1016/j.jaridenv.2025.105406>

5.2. Abstract

Global vegetation maps are essential tools for ecosystem monitoring, yet they often fail to accurately represent vegetation in arid regions. This study evaluates the performance of four global products—GFCC, GFW, CGLS-LC100, and DW—in the xerophytic shrublands of southwestern Madagascar, a region undergoing rapid deforestation. Satellite-derived estimates were compared with field data collected in 2022 from 41 plots (900m² each) using the point-intercept method. Results revealed substantial discrepancies among products, with a coefficient of variation reaching 78.94% for tree cover, primarily due to differences in vegetation definitions and classification methods. Aggregating tree and shrub cover into a single woody cover layer significantly reduced this variability by 58.19%. Among the products tested, DW provided the most accurate estimates of woody cover ($R^2 = 0.70$), while CGLS-LC100 was the most responsive to floristic composition. Product biases varied by vegetation type: DW and CGLS-LC100 tended to overestimate tree cover in structurally developed plots, while GFCC and GFW underestimated it in multi-stemmed thickets. Shrub cover was consistently underestimated across all products. In this context of active forest loss, overestimations may reflect past landscape conditions, while underestimations expose structural limitations. These findings highlight the need for mapping strategies that better capture the complexity of dryland vegetation.

Keywords: Global Forest Cover Change (GFCC), Global Forest Watch (GFW), Copernicus Global Land Cover (CGLS-LC100), Dynamic World (DW), xerophytic shrubs, arid zone.

5.3. Introduction

Arid ecosystems, which cover approximately 45% of the Earth's land surface, face increasing threats from both climate change and human activities (Právělie et al., 2019). Among these anthropogenic pressures, deforestation accelerates desertification and soil degradation, leading to a decline in ecosystem services (IPCC, 2023). Despite these challenges, arid ecosystems play a crucial role in mitigating climate change by sequestering carbon and influencing atmospheric dynamics through the reflection and absorption of solar radiation (Zhang et al., 2005). Furthermore, these ecosystems support about 50% of the global livestock population, thereby contributing significantly to land and water resource management (Ramond et al., 2022). However, despite their ecological and socioeconomic importance, arid regions are often overlooked in conservation efforts compared to other ecosystems. Only 12% of drylands are formally protected, in contrast to 21% of non-drylands (Lewin et al., 2024). Moreover, forest vegetation in arid ecosystems has received less attention than in rainforests or temperate regions (Abel et al., 2023). This underrepresentation has

led to a notable scarcity of field data particularly on vegetation structure and canopy composition which is essential for the accurate assessment of land cover (Zhang-Zheng et al., 2024). The lack of such ground-based information limits the reliability of remote sensing-derived vegetation maps and poses significant challenges for monitoring and effective ecosystem management (Bastin et al., 2017).

In this regard, remote sensing has emerged as a particularly relevant technique for overcoming the lack of field data and acquiring information in inaccessible and understudied areas, thereby improving the quality of maps representing these ecosystems (Lebzak, 2022). Nevertheless, uncertainties persist, particularly regarding the accuracy of maps produced through remote sensing (Bastin et al., 2017). These uncertainties often stem from the lack of standardized definitions for terms such as "forest". Sexton et al., (2015) highlighted substantial discrepancies in forest area estimates across different remote sensing methods, ranging from $32.1 \times 10^6 \text{ km}^2$ to $41.4 \times 10^6 \text{ km}^2$, due to ambiguities in the definition of "forest".

For arid regions, the mapping of forest landscapes presents specific challenges. The complex vegetation structures, which include drought-adaptive features such as leaf deciduousness and the presence of succulent plants, along with degradation factors like deforestation and ecosystem fragmentation, introduce significant biases in mapping efforts (De Marzo et al., 2021). Additionally, the open nature of many arid areas complicates remote sensing measurements, as the reflectance of individual trees can merge with bare soil and shadows (Bastin et al., 2017; Brandt et al., 2020). The insufficient spatial resolution of satellite images, which fails to detect isolated trees, further limits the accuracy of vegetation quantification, leading to cartographic biases (Brandt et al., 2020; Ma et al., 2023).

In this context, Madagascar, and more specifically the arid zones of the southwestern part of the island, represents a particularly relevant study model. These regions host dry forests and xerophytic shrublands characterized by a high diversity of vegetation gradients as well as notable differences in canopy closure and vegetation height (Hänke et al., 2017; Randriamalala et al., 2022). However, these ecosystems are severely threatened by human activities such as slash-and-burn agriculture, charcoal production, and goat grazing, leading to progressive fragmentation of the landscapes. These dynamics make the region particularly sensitive to temporal discrepancies, with an expected tendency for older satellite data to overestimate vegetation cover compared to more recent ground observations. Although previous studies have been conducted locally, they have primarily relied on binary forest/non-forest classifications using medium-resolution imagery, which fails to capture the floristic and structural complexity of these vegetation types (Brinkmann et al., 2014; J. R. Randriamalala et al., 2015)

Global reference products such as GFCC, GFW, DW, and CGLS_LC100 provide vegetation cover data at a global scale (Sexton et al., 2013; Buchhorn et al., 2020; Hansen et al., 2013; Brown et al., 2022). However, their accuracy in arid ecosystems remains uncertain due to the structural and morphological diversity of these environments. To address this gap, this study assesses the relevance of these products in a dryland ecosystem. Instead of generating new maps, the analysis adopts an end-

user perspective by comparing existing datasets both against each other and with independent field inventories. A vegetation typology based on dominant plant families and structural characteristics (height, diameter, and stem density) is employed to explore sources of disagreement. Two hypotheses are tested: (i) global products display substantial internal heterogeneity, limiting their interchangeability in arid environments; and (ii) estimation discrepancies are partly explained by vegetation composition and structure, with certain formations being more accurately captured depending on the product. The findings aim to inform the critical use of global vegetation datasets in ecologically complex and underrepresented dryland regions.

5.4. Materials and methods

5.4.1 Study area

The study area is in the Atsimo Andrefana region of southwestern Madagascar, specifically in the Toliara II district, south of the Fiherenana River. It spans between 23°00 and 24°30 S and 43°30 and 44°30 E, covering approximately 482,592 hectares (Figure 2). The region has a semi-arid climate with an average temperature of 24.8°C and annual rainfall of 350 mm (Gaspard et al., 2018)

Rainfall varies significantly, with less than 400 mm annually in the south (Gaspard et al., 2018) and between 600 and 1000 mm in the north (Leprun et al., 2009). The local economy relies heavily on natural resources, with charcoal production from thorny shrublands and extensive goat grazing being the dominant activities. These practices are deeply interconnected with the forest ecosystem and contribute significantly to landscape transformation (Hänke et al., 2017). The study focused on southern Toliara II due to security concerns in the north and the importance of the typical arid ecosystems in the southern part.

5.4.2 Data collection

The study used global vegetation cover maps available on the Google Earth Engine platform (Gorelick et al., 2017) and field inventory data. Vegetation was characterized by assessing tree cover, shrub cover, and wooded cover both on the ground and from global maps. Tree cover refers to the land area covered by tree canopies (Hansen et al., 2013), shrub cover to the extent of shrubs and/or bushy vegetation (Cherif et al., 2024), and wood cover to the combined area of trees and shrubs (FAO, 2020b). In this context, trees are defined as woody vegetation over 5 meters tall, and shrubs as woody vegetation up to 5 meters (Hansen et al., 2013; Sexton et al., 2013) (Figure 6.1.) (Figure A1). Each inventory plot measured 30 meters per side.



Trees (over 5 meters in height) with shrubs
(under 5 meters in height)

Figure 5-1: Example of mixed woody vegetation structure showing tall trees (>5 m) and understory shrubs (<5 m) in a dry forest formation of southwestern Madagascar

5.4.2.1 Technical description of global vegetation cover datasets

This study uses four global vegetation cover datasets (Tableau 6.1): (Sexton et al., 2013; Buchhorn et al., 2020; Hansen et al., 2013; Brown et al., 2022). Sexton's map is the official NASA reference in the Global Forest Cover Change (GFCC) tool, Hansen's map is used by Global Forest Watch (GFW), Buchhorn's map is a product of the European Space Agency's Copernicus Global Land Service (CGLS-LC100), and Brown's map is from the Dynamic World (DW) collection. Maps from 2019 were used to maintain temporal consistency, except for Sexton's map from 2015, the most recent available data. This temporal lag introduces an overestimation bias in 2015 due to high deforestation rates in the region (Vieilledent et al., 2018).

Tableau 5-1: Comparison of GFW, GFCC, CGLS-LC100, and DW products based on key parameters relevant to global vegetation mapping. The table summarizes definitions, data sources, resolution, seasonality, availability, and methods as described by the original developers. No models were trained by the authors; this is a descriptive comparison of existing datasets

	(Sexton et al., 2013)	(Hansen et al., 2013)	(Buchhorn et al., 2020)	(Brown et al., 2022)
Products	Global Forest Cover Change (GFCC)	Global Forest Watch (GFW)	Copernicus Global Land Cover (CGLS-LC100)	Dynamic world Collection (DW)
Definition Tree cover, Shrub cover	Percentage of horizontal area covered by vegetation over 5 meters in height, projected onto the ground in each 30 meter pixel	All vegetation over 5 meters height	Proportion of land cover estimates for land cover types (trees, shrubs)	Proportion of area covered by trees/shrubs
Images/ layers used	MODIS Landsat	Landsat	Landsat, Sentinel-1 et Sentinel-2, données des capteurs PROBA V	Sentinel-2
Image resolution	250 m 30 m	30m	10m, 30m, 100m	10m
Season of images used	All seasons	Growing season is more suitable for mapping	All seasons	All seasons
Year with available data	2000, 2005, 2010, 2015	2000 - 2021	2015-2019	2015- présent
Training data	Overlay of MODIS image tree cover with Landsat reflectance layer	Interpretation of very high-resolution (Quick bird) with existing tree cover layers (Landsat, MODIS)	Collected via Sentinel 2 grid (10 x10m) through manual classification using Google Maps and Microsoft Bing images (GEO-WIKI)	Semi-supervised classification of Sentinel 2 et annotation based on annotation on high-resolution satellite images (Google Maps)
Methods/ Algorithms	CubistTM regression algorithm tree	Analysis of Landsat data via Google Earth Engine platform combining public data and reflectance value data analysis – Tree Decision Analysis	Random Forest classification, regression of data by biome (tree, shrub) combined with ecological layers	Deep-learning - training a convolutional neural network CNN)

5.4.2.2 Field data

Field data were collected from April to May 2022 using a stratified random sampling design to capture gradients in vegetation cover, aridity, and soil types. A preliminary wood cover map, based on 505 photointerpreted 2500 m² plots (Figure

6.2) using Collect Earth Online (Bey et al., 2016), was generated via a Random Forest model trained on Sentinel-2 data (Breiman, 2001). This map guided the selection of 41 field plots (30 × 30 m), ensuring structural homogeneity within 200 m and representing the diversity of local vegetation. In each plot, tree, shrub, and total woody cover were estimated using the point-intercept method developed by the Terrestrial Ecosystem Research Network (TERN), Australia, with 54 points along four transects (Karl et al., 2017; White et al., 2012). In parallel, full inventories were carried out to record floristic composition and structural attributes (height, diameter, stem density). These combined datasets supported both accuracy assessment and the analysis of error patterns by vegetation type.

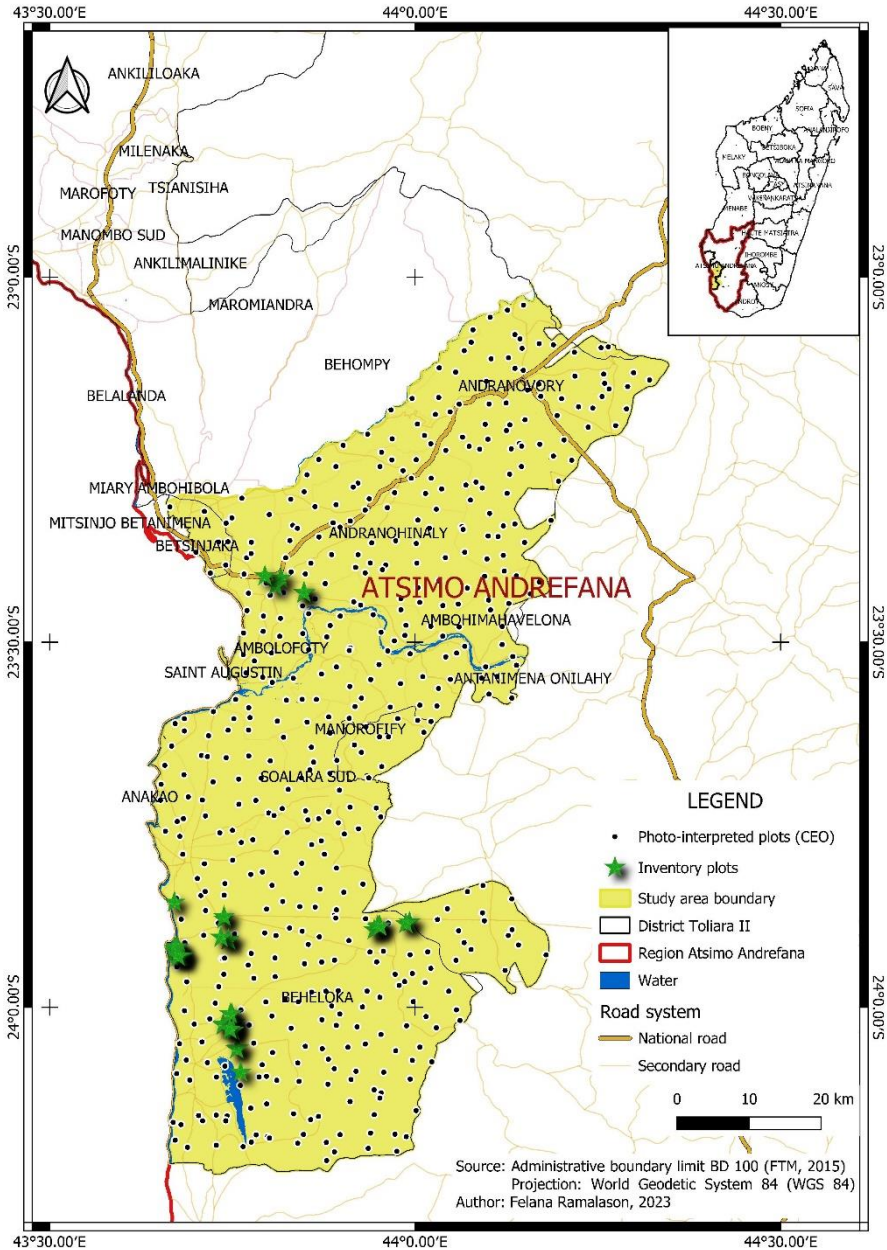


Figure 5-2: Map showing the location of the study area in the Toliara II District, southwestern Madagascar. The red line indicates the regional boundary, and the black line outlines the district (FTM, 2015). The study area is highlighted in yellow. Black dots represent 505 photo-interpreted plots (2500 m² each) collected using Collect Earth Online (CEO) in 2021, which were used to generate a reference map for stratified plot selection. Green stars indicate the 41 inventory plots selected along a wood cover gradient and surveyed during the 2022 field campaign (April–May).

5.4.3 Analytical framework

The overall analytical workflow step by step is summarized in Figure 6.3.

5.4.3.1 Assessment of inter-product heterogeneity

To assess the internal consistency of global vegetation datasets, we compared estimates of tree cover (TC), shrub cover (SC), and wood cover (WC) across four products. Prior to statistical tests, data normality (Shapiro-Wilk test) and variance homogeneity (Bartlett test) were verified. When assumptions were met, one-way ANOVA followed by Tukey's HSD test was applied to identify significant differences among products (Streiner, 2018).

Spatial heterogeneity was further assessed by computing pixel-wise coefficient of variation maps and histograms (Nuzzo, 2019) (Figure 4), where the coefficient of variation was calculated as the standard deviation divided by the mean value across products, expressed as a percentage (Zhang et al., 2005). This relative measure of dispersion allows for meaningful comparison across pixels with different mean values. Higher coefficient of variation values indicates greater relative inconsistency among estimates from different products.

In addition, vegetation cover was reclassified into "forest," "other wooded land," and "non-forest" using FAO thresholds (FAO, 2020b). Corresponding surface areas were compared between products using chi-square tests to evaluate thematic consistency (Ermakov, 2023).

5.4.3.2 Comparison with data field

To evaluate product accuracy, vegetation estimates were compared to field measurements from 41 plots. For each plot, values of TC, SC, and WC extracted from the maps were compared to field data using simple linear regression. Model performance (Piñeiro et al., 2008) was assessed using the slope, intercept, coefficient of determination (R^2), and mean absolute error relative to the mean of observed values (MAEr), providing a synthetic assessment of estimation accuracy and bias (Nzabarinda et al., 2022).

To detect systematic biases, relative errors were plotted against field-observed values, enabling the identification of overestimation or underestimation trends across vegetation density gradients (Randrianasolo et al., 2019).

To assess whether estimation errors varied according to vegetation characteristics, plots were grouped into two typologies: (i) **Floristic composition**, based on a Principal Component Analysis (PCA) of the relative abundance of dominant plant families, preceded by a **Hellinger transformation** to accommodate the compositional nature of the data and reduce the influence of rare or absent taxa (Borcard et al., 2011); and (ii) **Vegetation structure**, derived from three key descriptors: vegetation height, stem diameter, and stem density (Palla et al., 2011).

For each product, the significance of error differences between the two vegetation types was tested. When assumptions of normality and homoscedasticity were met, ANOVA was used; otherwise, a Wilcoxon rank-sum test was applied. Results were visualized using boxplots of absolute relative errors per vegetation group

5.4.3.3 Statistical analysis

Statistical tests were chosen based on residual diagnostics: ANOVA and Tukey’s HSD for inter-product comparisons, Chi-square for FAO class surface comparisons, Linear regression for field comparison, and ANOVA or Wilcoxon tests for vegetation-type effects on error patterns. A significance level of $\alpha = 0.05$ was used throughout. The combination of parametric and non-parametric approaches ensured robust interpretation across methods and hypotheses.

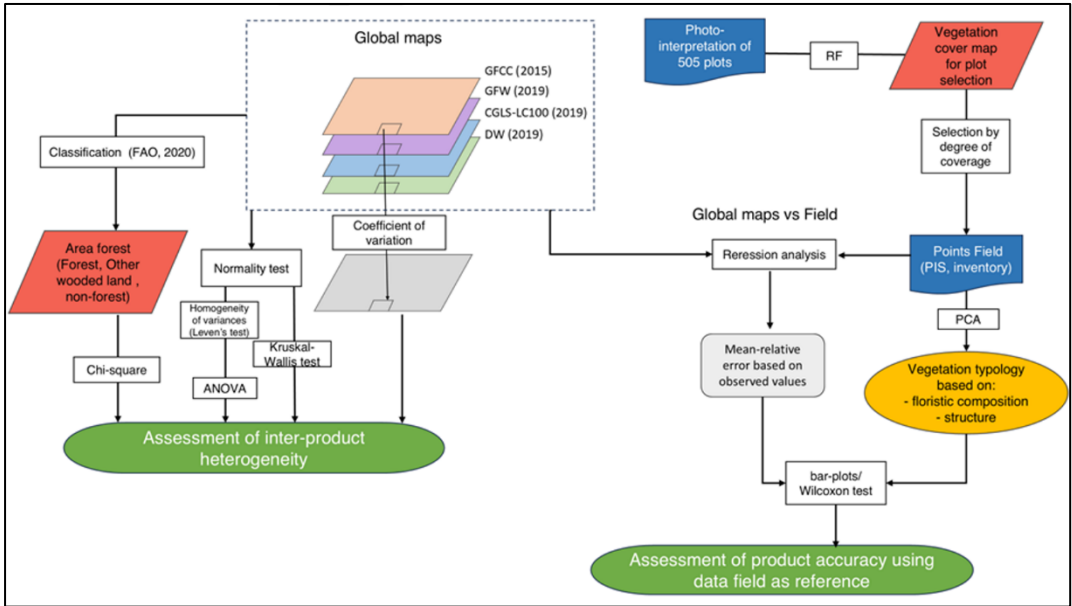


Figure 5-3: Flowchart of the methodological framework. Left: Assessment of inter-product heterogeneity, including FAO-based classification, chi-square tests for forest area comparison, ANOVA/ Kruskal-Wallis test, and relative standard deviation maps. Right: Comparison with field data. A vegetation cover map from 505 photo-interpreted plots (RF model) guided the selection of 41 inventory plots (PIS method). Accuracy was assessed via regression and mean-relative error. Plots were grouped by floristic and structural typologies (PCA), and error differences were tested using ANOVA or Wilcoxon tests. ANOVA = Analysis of Variance; RF = Random Forest; PIS = Point Intercept Sampling; MAEr = Mean Absolute Error relative to observed mean.

5.4.4 Methodological limitations related to seasonal and temporal inconsistencies

This study has three main limitations. First, the analysis is constrained by the minimum mapping unit (MMU), defined by the spatial resolution of the global products. Although DW and CGLS-LC100 offer resolutions as fine as 10 meters, all datasets were resampled to 30 meters to ensure comparability with GFW and GFCC (Tableau 6.1). While this standardization is necessary, it may limit the ability to detect

fine-scale vegetation structures, particularly in fragmented or sparsely vegetated landscapes.

The second concerns data seasonality: only the DW and GFCC products incorporate seasonal composites, whereas GFW and CGLS-LC100 rely on annual averages. However, in this region characterized by an extended dry season lasting up to nine months (Chanteau, 2006), and where field data were collected in April–May—a period during which most species are already defoliated—the impact of seasonal mismatch is considered limited (Diédhiou et al., 2020).

The third limitation is temporal: the GFCC product is only available up to 2015, and CGLS-LC100 up to 2019. The year 2019 was therefore selected as the reference period to maximize overlap across products. Additionally, GFW is based on a static tree cover map from the year 2000, updated annually using a loss mask (Hansen et al., 2013). In the context of ongoing deforestation, a moderate overestimation of vegetation cover relative to 2022 field data can thus be interpreted as an indicator of product reliability.

5.5. Results

5.5.1 Heterogeneity of existing remote sensing products

5.5.1.1 Variation of vegetation cover

Remote sensing products exhibit substantial variability in vegetation cover estimates (Figure B1). While GFCC and GFW report only tree cover, CGLS-LC100 and DW include both tree and shrub components. Tree cover estimates are relatively consistent across GFCC, GFW, and CGLS-LC100 (5.51–7.67%), but markedly higher in DW (27.28%). For woody cover, DW also reports much higher values (52.30%) compared to CGLS-LC100 (25.24%) (Tableau B1). The coefficient of variation (CV) indicates high heterogeneity for tree cover (78.94%), moderate for shrub cover (63.08%), and low for total woody cover (20.75%), with significant differences (p -value $< 2.2e-16$) (Tableau A1), particularly in coastal areas (Figure 6.4). Notably, combining tree and shrub layers into a single woody cover category substantially reduces heterogeneity (CV reduction of 58.19%) (Figure 6.4), suggesting that discrepancies primarily stem from methodological differences in stratification rather than in the detection of woody vegetation itself. These findings highlight that although classification approaches vary in distinguishing trees from shrubs, they converge more consistently in identifying overall woody-covered areas.

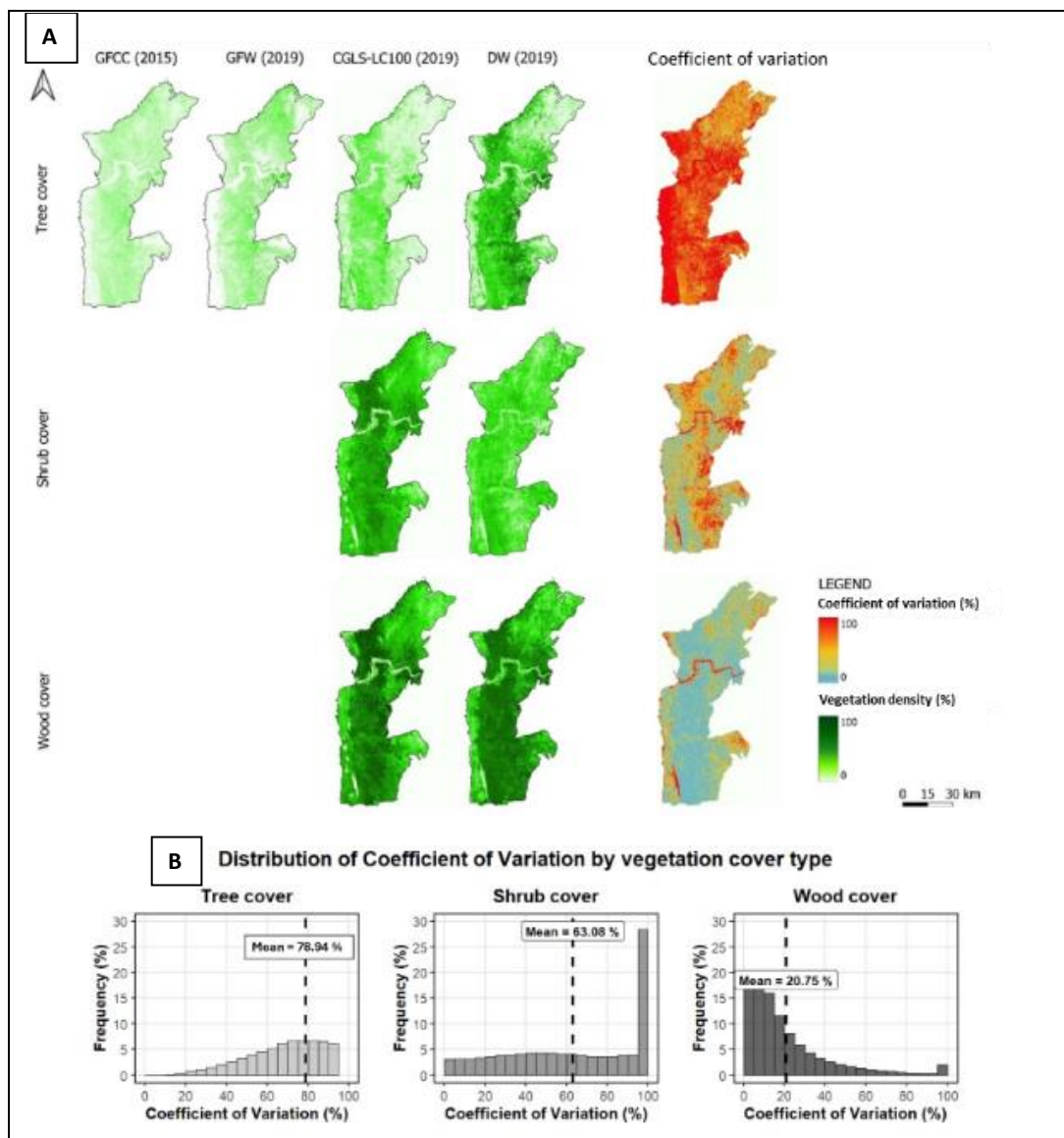


Figure 5-4: Vegetation cover estimates and variability in the Toliara II District (Madagascar) from four global products: GFCC (2015), GFW (2019), CGLS-LC100 (2019), and DW (2019). Panel A: Maps of tree, shrub, and total woody cover (%), along with their coefficients of variation (CV). Green shades represent vegetation density; red, yellow–blue gradients show spatial variability (CV, %), with red indicating high variation. Panel B: CV distributions for each cover type. Tree cover shows the highest variability (mean CV = 78.94%), followed by shrub cover (63.08%) and woody cover (20.75%). CV was calculated as $(\text{standard deviation} / \text{mean}) \times 100$ across products.

5.5.1.2 Comparison of forest area

Forest cover estimates differ widely across global products. GFCC (2015) reports the lowest forest area (12.54%), while GFW, CGLS-LC100, and DW estimate 37.41%, 46.96%, and 72.22%, respectively (Figure 6.5). Although GFW and CGLS-LC100 provide similar overall proportions, their spatial classifications diverge, with GFW often assigning forest to areas labeled as other wooded land or non-forest by CGLS-LC100. DW presents the highest forest extent, suggesting broader classification criteria. Coastal areas display marked discrepancies, being classified as non-forest by GFCC and GFW, but as forest by CGLS-LC100 and DW. A chi-square test reveals significant differences in forest category distributions among the products ($p < 2.2e-16$) (Tableau B2), underscoring inconsistencies in classification approaches. These results indicate that the vegetation cover classifications are not homogeneous across the examined data sources.

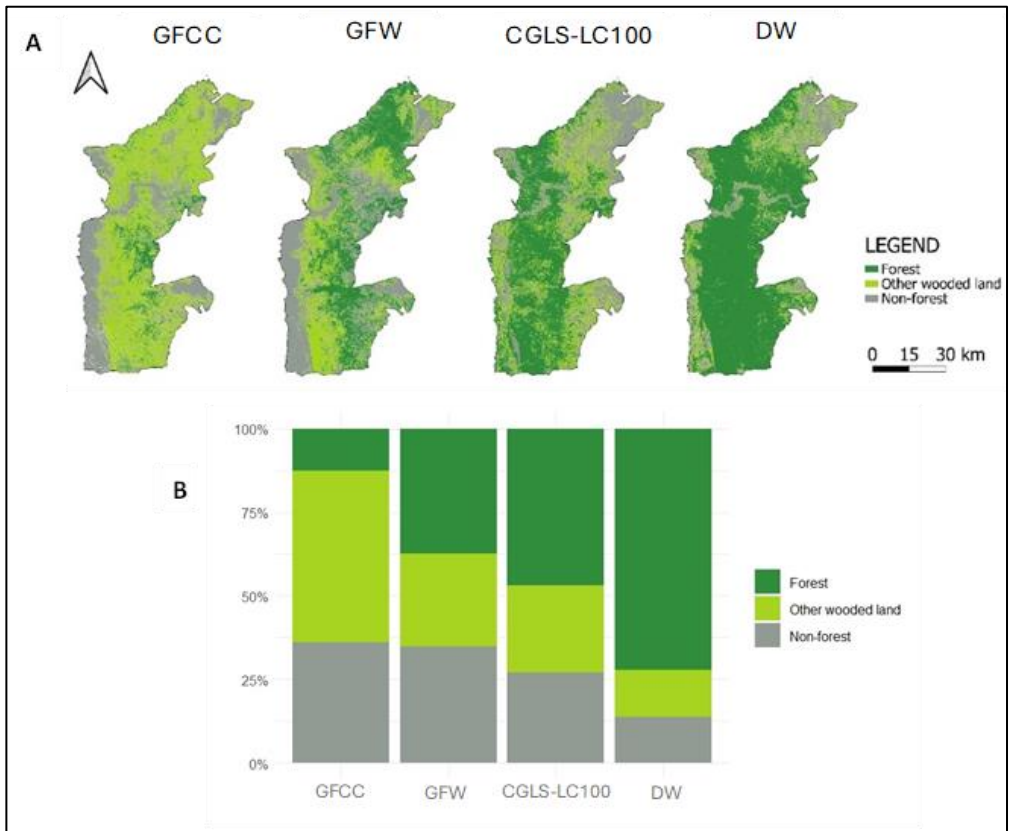


Figure 5-5: Heterogeneity in forest area estimates across 4 global vegetation products GFCC (2015), GFW (2019), DW (2019), and CGLS-LC100 (2019) in Toliara II District, Madagascar. Panel (A) shows the spatial distribution of forested areas as defined by FAO (2020), while panel (B) highlights discrepancies in total forest area among the datasets. In a context of active deforestation, older datasets such as GFCC (2015) would be expected to report substantially

higher forest cover than more recent products. However, this is not observed, raising concerns about the internal consistency of these globally referenced datasets.

5.5.2 Comparison of products estimates with field data

5.5.2.1 Analysis of Relative Errors and Bias Identification

The comparison with 2022 field data reveals substantial variability in the accuracy of vegetation cover estimates across products. For tree cover, GFW and DW perform best ($R^2 = 0.52$), followed by GFCC ($R^2 = 0.45$) and CGLS-LC100 ($R^2 = 0.29$). However, CGLS-LC100 shows the highest relative error (MAEr = 0.91) (Figure 6.6). Regression slopes indicate that GFCC and GFW tend to underestimate tree cover, especially at higher values (slopes = 1.64 and 1.27), while DW and CGLS-LC100 show overestimation. For shrub cover, both DW and CGLS-LC100 display low correlation with field data ($R^2 = 0.05$ and 0.08) despite moderate MAEr values (~ 0.36). The slopes below 1 and high intercepts (DW: 0.45 / 44.45; CGLS-LC100: 0.66 / 29.89) indicate a systematic underestimation. In contrast, wood cover estimates are more accurate (Tableau C1). DW shows the highest agreement ($R^2 = 0.70$, MAEr = 0.19), followed by CGLS-LC100 ($R^2 = 0.35$, MAEr = 0.29) (Figure 6.6). Although both slightly underestimate wood cover, they perform better than for individual vegetation strata. Overall, DW provides the most consistent estimates, particularly for wood cover, while shrub cover remains poorly captured by all products.

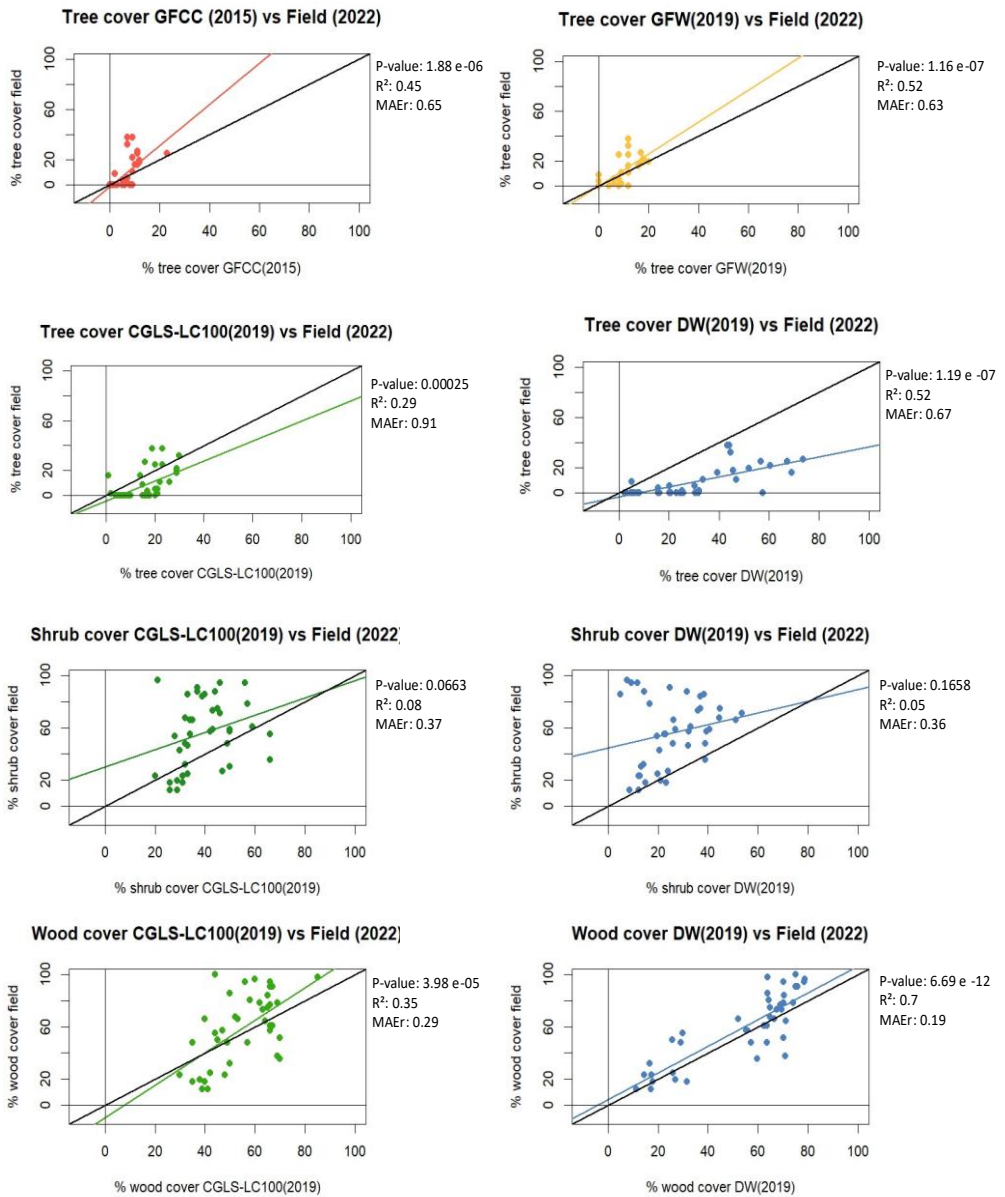


Figure 5-6 : Linear regression plots comparing tree cover, shrub cover, and wood cover estimates from global remote sensing products to 2022 field data across the Toliara II District, Madagascar. Each panel shows the product-specific regression line (in color) and the 1:1 reference line (in black), representing perfect agreement with field data. The position and slope of the regression lines highlight trends of over- or underestimation for each vegetation type. Statistical indicators (R², p-value, MAEr) summarize the strength and accuracy of the relationship between remote sensing estimates and field observations.

Relative error analysis reveals distinct biases across products and vegetation types. For tree cover (Figure D1), GFW and GFCC overestimate low values (below 5%), with errors above 100%, decreasing as observed values increase. DW shows extreme overestimation below 10%, sometimes exceeding 150%, but errors are reduced at higher cover. CGLS-LC100 maintains low errors below 10% yet underestimates medium to high values.

For shrub cover (Figure D1), both DW and CGLS-LC100 exhibit increasing underestimation as shrub cover increases. DW underestimates values above 50%, with errors nearing -150%. CGLS-LC100 follows a similar trend with slightly lower error magnitude.

For wood cover (Figure D1), errors are less pronounced. DW underestimates consistently, especially above 60%. CGLS-LC100 overestimates low values (below 20%), but error decreases progressively and approaches zero as cover increases.

These results suggest that all products are less accurate at low vegetation densities, with DW and CGLS-LC100 offering more consistent estimates for total wood cover than for individual strata.

5.5.2.2 Influence of vegetation typology on model errors

Two main clusters were identified based on floristic composition (Figure 6.7A) (Figure 6.8). Cluster 1 (30 plots) is floristically diverse and dominated by families such as *Burseraceae*, *Fabaceae*, *Anacardiaceae*, and *Hernandiaceae*. Cluster 2 (11 plots) is characterized by succulent and shrubby species, notably from *Asteraceae*, *Euphorbiaceae*, and *Zygophyllaceae*. Structural clustering shows that Cluster 1 corresponds to low-stature xerophytic thickets with thin, multi-stemmed trunks, while Cluster 2 includes more developed vegetation with thicker stems and slightly greater height (Figure 6.7B) (Figure 6.8).

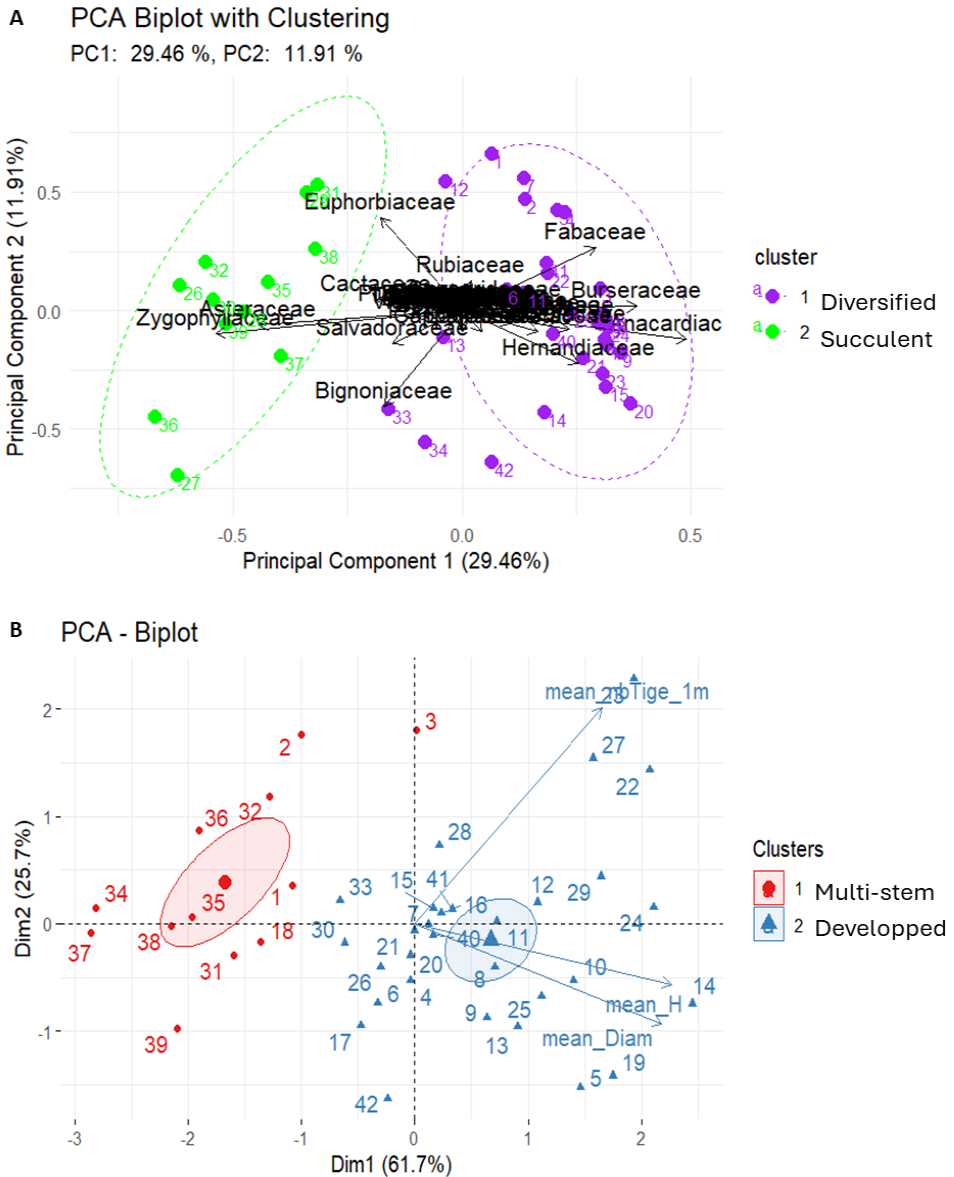


Figure 5-7 : Principal Component Analysis (PCA) biplots showing clustering of 42 vegetation plots in the Toliara II District, Madagascar. A. PCA biplot based on Hellinger-transformed species composition, with clusters defined by floristic composition. Cluster 1 (purple) represents plots dominated by dry forest species, and Cluster 2 (green) includes plots characteristic of succulent thickets. B. PCA biplot based on structural attributes (mean height, diameter, and stem count), with clusters defined by vegetation structure. Cluster 1 (red) includes multi-stemmed, open formations; Cluster 2 (blue) corresponds to denser, woody formations. Arrows indicate the main variables influencing axis separation





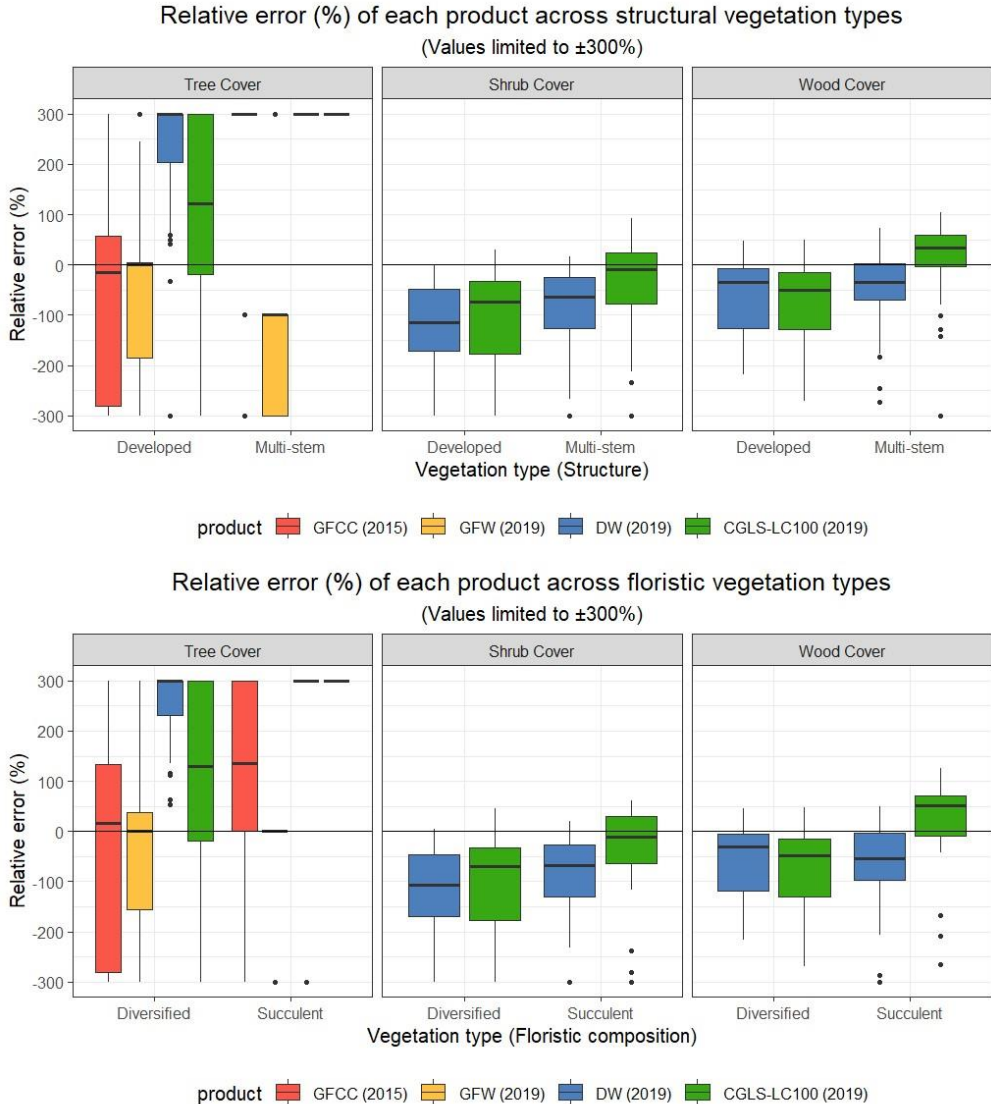
	Cluster 1 - Structure	Cluster 2 - Structure
Cluster 1 – Floristic composition		
	High morphological diversity, small-sized individuals	High morphological diversity, but with better-developed individuals
Cluster 2 -Floristic composition		
	Small-sized succulents and shrubs under severe arid conditions	Better-developed succulents and shrubs under more favorable conditions

Figure 5-8 : Illustrations of the diversity of xerophytic ecosystems captured by drone in August 2023. These images showcase different types of xerophytic vegetation in the study area, highlighting the variation in species and ecosystem structure within the region.

The analysis of relative errors highlights significant differences across vegetation types in southwestern Madagascar, a region undergoing active deforestation. In this context, overestimations by satellite products (GFCC 2015, GFW, DW, and CGLS-LC100 2019) relative to 2022 field data are considered favorable. For tree cover, CGLS-LC100 shows the highest sensitivity to floristic composition ($p < 0.001$) (Tableau D2), with marked overestimation in succulent-dominated formations, as expected (Figure 9). GFCC and DW also show significant differences ($p < 0.05$) (Tableau D2), while GFW remains nonsignificant. Regarding wood cover, only CGLS-LC100 significantly distinguishes between floristic types ($p < 0.001$), overestimating in succulent areas and underestimating in floristically diverse ones. For shrub cover, CGLS-LC100 is again significant ($p < 0.01$) (Tableau D2), with slight overestimation in succulent formations (Figure 6.9). In terms of vegetation structure, all products except GFW are significantly influenced in the case of tree cover ($p < 0.001$) (Tableau D2), with more positive errors in structurally developed formations compared to multi-stemmed types. CGLS-LC100 also shows strong

structural sensitivity for wood and shrub layers. Overall, CGLS-LC100 aligns most consistently with observed deforestation trends, providing favorable overestimations in the expected formations. DW performs adequately for tree cover but is less effective for other layers. GFW shows limited responsiveness to structural and floristic



heterogeneity, reducing its reliability in dry, variable ecosystems (Figure 6.9).

Figure 5-9: Relative errors (%) in tree, shrub, and wood cover across vegetation typologies in the Toliara II District, Madagascar. (A) Errors by structural types (developed vs. multi-stem). (B) Errors by floristic composition (diversified vs. succulent). Boxplots compare deviations between field data and estimate from four global products (GFCC 2015, GFW 2019, DW 2019, CGLS-LC100 2019), with relative errors limited to $\pm 300\%$. The results

highlight that some global datasets better capture specific vegetation types, emphasizing the importance of selecting context-appropriate products for ecological analyses.

5.6. Discussion

5.6.1 Inter-product heterogeneity causes and implications for arid ecosystems

The analysis supports the first hypothesis: global vegetation products show high internal heterogeneity, especially in arid environments, which limits their interchangeability. Tree cover estimates vary widely between products—ranging from 5.5% (GFCC, GFW) to 27.3% (DW) — reflecting differences in definitions and classification methods. GFCC and GFW apply a minimum height threshold of 5 meters (Hansen et al., 2013; Sexton et al., 2013) excluding shorter woody vegetation common in drylands. In contrast, DW and CGLS-LC100 include both trees and shrubs without strict height thresholds, resulting in broader coverage. These definitional differences explain the underrepresentation of low-stature, multi-stemmed vegetation in certain products. Prior studies (Wang et al., 2022) have also highlighted such inconsistencies between land cover datasets. However, wood cover estimates are more stable across products. This consistency suggests that combining tree and shrub layers into a single woody category reduces classification discrepancies and better reflects vegetation patterns in arid landscapes (Bastin et al., 2017).

5.6.2 Systematic biases in vegetation cover estimation across global products

Field-based analysis reveals that all global products tend to underestimate vegetation cover, particularly in low-density areas typical of arid ecosystems. This confirms the observation that current products are not calibrated to capture sparse and heterogeneous vegetation. GFCC and GFW consistently underestimate tree cover in plots with low canopy height and density, in agreement with similar results from African savannas and drylands (Adzhar et al., 2022) and other studies (Fagan, 2020). DW, while more inclusive, tends to overestimate tree cover in open landscapes, possibly due to spectral confusion between trees, shrubs, and grasses. This misclassification has also been reported in other dry regions (Los et al., 2012). CGLS-LC100 performs relatively well for low tree cover (<10%) but underestimates higher values, a pattern consistent with observations in semi-arid China (Wang et al., 2022). For shrub cover, both DW and CGLS-LC100 show systematic underestimation, particularly as cover increases. (Tian et al., 2017) reported similar findings, where shrub-rich drylands were poorly captured by global datasets.

5.6.3 Ecological factors affecting model performance: insights from vegetation typology

In southwestern Madagascar, a region undergoing rapid deforestation, this study demonstrates that the accuracy of global remote sensing products is strongly influenced by vegetation typology—namely floristic composition (succulent vs. diverse formations) and structural characteristics (multi-stemmed vs. developed canopies). Since the evaluated products (GFCC 2015, GFW 2019, DW 2019, CGLS-LC100 2019) predate field data collected in 2022, an overestimation of vegetation cover is considered both expected and desirable, as it reflects pre-deforestation conditions.

Among the tested products, CGLS-LC100 (2019) showed the highest sensitivity to both floristic and structural contrasts. It overestimates cover in succulent-dominated areas, as anticipated, and underestimates in floristically diverse formations, which are more likely degraded, indicating a reliable response to ecological variability. These findings are consistent with (Buchhorn et al., 2020), who reported strong performance of CGLS-LC100 in homogeneous vegetation types. DW (2019) also overestimates tree cover in diverse plots but shows limited robustness in shrub and wood layers, particularly in structurally complex environments. In contrast, GFW (2019) demonstrates weak sensitivity to vegetation heterogeneity, a limitation also noted by Sexton et al., (2015), rendering it less suitable for fine-scale typological assessments. DW and CGLS-LC100 exhibit higher errors in floristically diverse assemblages and structurally developed plots, consistent with observations from other structurally complex arid ecosystems (Los et al., 2012).

These patterns align with previous research in semi-arid and dry tropical systems. Studies by (Bachantourian et al., 2022) reveal that global land cover datasets often misrepresent xerophytic vegetation due to training models biased toward mesic or moist tropical forests. Similarly, Torbick et al., (2005) found that structurally complex dry forests tend to be underestimated by global products. This highlights the need for model calibration using locally relevant vegetation typologies, incorporating both floristic and structural attributes. As emphasized by Los et al., (2012); Schneider et al., (2017), integrating functional vegetation traits into model development is critical for improving remote sensing reliability in dynamic, degraded ecosystems.

5.6.4 Resolution, seasonality and deforestation: interpreting product limitations

The findings of this study reveal significant discrepancies in the accuracy of global vegetation products depending on the floristic and structural typology of vegetation in southwestern Madagascar. However, these outcomes must be interpreted considering several key methodological limitations.

Spatial resolution: All products were harmonized to a spatial resolution of 30 meters—even those originally available at 10 m (DW, CGLS-LC100). While this standardization was necessary to enable comparison, it likely smoothed out fine-scale structural contrasts, particularly between multi-stemmed and developed formations.

This spatial resampling reduces the ability to detect ecological fragmentation and micro-heterogeneity typical of xerophytic shrublands. Recent studies highlight that high-resolution data (<10 m) are essential for capturing structural complexity in heterogeneous dry ecosystems (Wang et al., 2022).

Seasonality: Although some products (e.g., GFW, CGLS-LC100) are based on annual averages, the field data were collected in April–May, during the peak of the dry season when most woody species are leafless. In this phenological context, optical sensors tend to underestimate vegetation cover. This discrepancy likely explains the marked underestimation observed in floristically diverse and structurally developed (intact) formations. Similar issues have been documented in seasonal tropical environments where annual composites fail to reflect intra-annual vegetation dynamics (Pelletier et al., 2019).

Deforestation context: Conversely, succulent- and multi-stem-dominated formations—typically associated with ecosystem degradation and colonization by pioneer species (e.g., *Euphorbiaceae*)—tended to be overestimated. Given the study region's ongoing deforestation, and the fact that satellite products date from 2015–2019 while field data are from 2022, this overestimation is consistent with expected landscape trajectories and should not be interpreted as a model error.

Together, these findings suggest that products like CGLS-LC100 and DW may capture some ecological transitions but perform poorly in intact formations. This limits their use for monitoring degradation and reinforces the need to integrate local vegetation typologies into model calibration, as recommended by Adzhar et al., (2021; Schneider et al., (2017).

5.7. Conclusion

This study evaluated the performance of global vegetation cover products (GFCC, GFW, CGLS-LC100, DW) in the dryland ecosystems of southwestern Madagascar. Results confirm substantial heterogeneity among these products (CV=78.94% for tree cover), primarily due to methodological differences. Combining tree and shrub layers considerably reduces this variability (CV reduction of 58.19%). From an end-user perspective, the analysis reveals that: (1) total woody cover estimates are generally more reliable than individual strata, with DW offering the best accuracy ($R^2=0.70$, MAEr=0.19), (2) CGLS-LC100 shows significant sensitivity ($p<0.001$) to succulent-dominated formations (*Euphorbiaceae*, *Asteraceae*) and developed vegetation structures, and (3) all products exhibit systematic errors in low-height, multi-stemmed xerophytic formations characteristic of arid zones. These errors may compromise sustainable management efforts by neglecting the ecological importance of complex vegetation formations. To address these challenges, it is essential to develop algorithms adapted to xerophytic ecosystems, integrate complementary technologies (radar, drone, LiDAR), establish long-term monitoring plots, and adopt classification schemes that better reflect the floristic and structural diversity of arid environments.

Chapitre 6

The importance of the sampling design in mapping woody cover in arid ecosystems

6.1. Référence bibliographique

Ramalason, F.N., Rakotondrasoa, O.L., Vander Linden, A., Renard, G., Randriamalala, J.R., Vereecken, N.J., Razakamiarmanana, A., Ranaivoharivelo, M.F., Raholinarivo, S., Rakoto Ratsimba, H., Bogaert, J., Bastin, J.-F., 2026. The importance of the sampling design in mapping woody cover in arid ecosystems. *GIScience & Remote Sensing* 63, 2658305. <https://doi.org/10.1080/15481603.2026.2658305>

6.2. Abstract

Accurate calibration data remains a major constraint in ecological remote sensing, particularly in arid ecosystems where sparse and heterogeneous woody vegetation is challenging to detect. Despite advances in sensor technology and modelling algorithms, optimal calibration sampling design has received limited attention, with current approaches varying unsystematically (35–1,000 plots/1,000 km²) without empirical benchmarks. Here, we develop and evaluate a quantitative framework for optimizing sampling strategies in remote-sensing models of woody cover through systematic comparison of (i) calibration data source (field surveys versus photointerpretation), (ii) spatial configuration (clustered versus dispersed), and (iii) sampling density. This integrated approach enables isolating the relative contributions of each design component to mapping accuracy — a critical gap in current remote sensing methodology.

We applied this framework to Sentinel-1, Sentinel-2, combined Sentinel-1+2, and AlphaEarth Foundations across Madagascar's arid southwest, validating predictions through spatial cross-validation using independent field plots. Photointerpretation substantially outperformed field-based calibration under typical arid-zone constraints ($R^2=0.88$, RMSE=0.11 versus $R^2=0.46–0.66$, RMSE=0.17–0.21). Performance saturated at 20.7 - 41.4 dispersed calibration plots per 1,000 km² across all predictors, beyond which gains became marginal. Dispersed strategies required half as many samples as clustered designs to achieve comparable accuracy, demonstrating that spatial distribution outweighs sample size. Once adequate sampling density and distribution were achieved, Sentinel-1+2 and AlphaEarth performed similarly ($R^2\approx 0.85–0.86$), indicating that sampling design outweighs predictor complexity.

Our framework provides empirically derived operational thresholds (minimum: 15.5 plots/1,000 km²; optimal: 20.7 -41.4 plots/1,000 km²) and a transferable methodology for determining optimal calibration densities in heterogeneous ecosystems where logistical constraints limit field sampling.

Keywords: sampling design, photointerpretation, remote sensing, arid ecosystems, woody vegetation.

6.3. Introduction

Arid and semi-arid ecosystems cover 40.6% of Earth's land surface and support 2.3 billion people (UNCCD, 2024), providing critical ecosystem services including wood

production, carbon sequestration, and biodiversity conservation (D'Odorico et al., 2013; Feldman et al., 2024). Woody vegetation structures these ecosystems by influencing hydrological cycles, soil stability, and wildlife habitats (Eldridge et al., 2011). However, accelerated degradation driven by anthropogenic pressures — defined here as human-induced disturbances including land clearing, overgrazing, resource extraction and agricultural expansion (Reynolds et al., 2007)— and climate change makes monitoring essential for conservation and sustainable management (Feng and Fu, 2013).

Traditional field inventories provide accurate data but present major limitations: prohibitive costs, logistical constraints, and limited spatial coverage restrict large-scale application (McRoberts et al., 2018). Satellite remote sensing enables broad-scale monitoring but encounters specific challenges in arid environments. Sparse and heterogeneous vegetation, low biomass, short phenological cycles, and spectral confusion between senescent vegetation and bare soil complicate detection (Abel et al., 2019; Brandt et al., 2016). Adaptive forms including deciduousness, leaf reduction, and waxy coatings further disrupt spectral detection. Low canopy density creates additional interpretation difficulties, as tree spectral signatures easily confound with exposed soil or shadows (Bastin et al., 2017; Brandt et al., 2020). Landscape fragmentation and deforestation further accentuate these challenges (De Marzo et al., 2021).

Recent advances in very high-resolution mapping have substantially improved our capacity to characterize woody vegetation at fine spatial scales. Global tree density estimation has been achieved at sub-kilometer resolution (Crowther et al., 2015), while sub-pixel decomposition methods including pixel-swapping and localized partial unmixing enable tree counting at landscape scale from medium-resolution imagery (Pouliot et al., 2018). Drone-based canopy mapping further extends these capabilities to centimeter-level resolution in semi-arid shrublands (Pérez-Luque et al., 2022). Deep learning applied to submetre imagery has enabled wall-to-wall mapping of over 1.8 billion individual tree crowns across dryland Africa (Brandt et al., 2020). Integrating Sentinel-1 (radar) and Sentinel-2 (optical) data offers improved detection through complementarity: radar ensures canopy penetration and moisture sensitivity, while optical data provide detailed spectral information (Achariki et al., 2021; Dostálová et al., 2014). Multi-temporal remote sensing analyses have further demonstrated the value of combining optical and spatial data for monitoring forest cover changes at regional scale (Valjarević et al., 2018). However, effectiveness depends critically on calibration data quality and representativeness (Ploton et al., 2020).

A critical methodological gap persists: despite widespread recognition that calibration sampling design influences model accuracy, current approaches vary unsystematically. Published studies report sampling densities ranging from 35 to 1,000 plots per 1,000 km² and plot sizes from 0.5 to 1 hectare, reflecting no consensus on optimal design (Bastin et al., 2017; Goldblatt et al., 2017; Van Passel et al., 2020). This uncertainty generates inefficiencies: some projects oversample at the expense of spatial coverage, while others undersample, compromising accuracy. Consequently,

global woody cover maps present systematic biases in arid ecosystems (Ramalason et al., 2025b). Critically, no study has systematically quantified the relative contributions of calibration data source, spatial configuration, and sampling density to mapping accuracy across multiple sensor types—preventing development of evidence-based sampling guidelines.

Photointerpretation of very high-resolution imagery emerges as a promising intermediate solution between costly field campaigns and potentially biased global products (Pengra et al., 2007). This approach exploits human expertise to visually interpret complex vegetation patterns, enabling production of more numerous and representative samples while maintaining acceptable accuracy. Field plots provide high measurement precision but limited spatial coverage under typical logistical constraints in arid regions, while photointerpretation enables production of more numerous and spatially representative samples while maintaining acceptable accuracy. Concurrently, new pre-processed satellite datasets offer opportunities to improve ecological mapping. AlphaEarth Foundations, for example, combines spectral, textural, and topographic data through deep learning in standardized, directly exploitable formats (Brown et al., 2025).

Here, we address this methodological gap by developing an integrated quantitative framework for optimizing calibration sampling design. Our approach advances beyond previous studies in three ways. First, we systematically isolate and quantify the independent effects of three calibration design components—data source (field versus photointerpretation), spatial configuration (clustered versus dispersed), and sampling density (10–505 plots)—through factorial experimental design with 100 iterations per configuration. Second, we evaluate these effects across four satellite predictor sets with varying complexity (Sentinel-1, Sentinel-2, Sentinel-1+2, AlphaEarth Foundations) to determine whether optimal sampling strategies depend on predictor sophistication. Third, we implement rigorous spatial cross-validation using independent field plots, ensuring performance estimates reflect true predictive capacity rather than overfitting to spatially autocorrelated data (Ploton et al., 2020; Valavi et al., 2018).

We apply this framework to Madagascar's arid southwest, encompassing dry forests and xerophytic thickets with high variability in vegetation structure and intense anthropogenic pressures, including deforestation linked to charcoal production, fuelwood extraction, and construction timber harvesting, as well as slash-and-burn agriculture (*tavy*) (Hänke et al., 2017; Randriamalala et al., 2022). This environmental heterogeneity provides a suitable testbed for evaluating sampling strategies across broad ecological gradients characteristic of global arid ecosystems.

Our objectives are to: (1) quantify how sampling design influences model accuracy relative to predictor complexity through systematic comparison; (2) identify minimal and optimal densities of calibration samples using multi-criteria convergence analysis; and (3) provide empirically derived operational thresholds and a transferable framework for designing calibration datasets under logistical and financial constraints. We test three hypotheses: (1) photo-interpreted samples produce more accurate maps than field inventories through better spatial representativeness; (2)

performance increases with sample size up to an identifiable saturation threshold beyond which marginal gains diminish; and (3) multi-sensor integration improves accuracy progressively from single-sensor approaches to multi-dimensional composites, but only when calibration design is adequate.

This framework provides the first empirically derived quantitative guidelines for calibration sampling in arid ecosystem remote sensing, directly addressing the current lack of standardized approaches. The resulting operational thresholds and comparative methodology are transferable to other heterogeneous ecosystems where logistical constraints limit field sampling, offering practical guidance for balancing mapping accuracy against resource investment.

6.4. Materials and methods

6.4.1 Study area

We applied our framework to the Atsimo Andrefana region of southwestern Madagascar (4,826 km²; 23°00'-24°30'S, 43°30'-44°30'E). The region experiences an arid to semi-arid climate (mean aridity index AI = 0.21; range: 0.13–0.36, following the UNCCD classification; (Trabucco and Zomer, 2019; UNCCD, 2024; Valjarević et al., 2025) with mean annual precipitation of 466 mm (range: 274–770 mm) and a mean annual temperature of 24.8°C (2000–2022: CHIRPS Funk et al., 2015; ERA5-Land, Hersbach et al., 2020). The southern portions of the region are classified as arid (AI < 0.20) while most of the zone is semi-arid (AI 0.20–0.50). Rainfall is strongly seasonal, concentrated between December and March (January mean: 151 mm; August mean: 4 mm), with seven months receiving less than 30 mm per month. The hydrographic network is dominated by seasonal rivers with intermittent flow (MERIT Hydro; Yamazaki et al., 2019)). Topographic relief is moderate, with elevations ranging from near sea level to 598 m (mean: 162 m; SD: 129 m; Copernicus DEM GLO-30).

The region is increasingly affected by intensifying agricultural droughts and heat waves, threatening food security and endemic biodiversity. Although geographically less exposed to tropical cyclones than Madagascar's eastern coast, the southwest remains vulnerable to their indirect impacts — strong winds and intense rainfall episodes — in a context of high social vulnerability. Climate projections indicate a long-term trend toward reduced and more variable annual rainfall, increasing physiological stress on woody vegetation (IPCC, 2023). These climatic pressures underscore the urgency of robust and cost-effective woody cover monitoring frameworks for this region.

Vegetation comprises two co-existing structurally heterogeneous formations along an east-west gradient: (1) coastal xerophytic thickets, characterised by open structure and dominance of succulent species from endemic families Didiereaceae (*Didiera madagascariensis*, *Alluaudia procera*) and Euphorbiaceae (*Euphorbia stenoclada*); and (2) inland spiny forest, denser and more semi-deciduous, with a greater contribution of Fabaceae. This environmental heterogeneity provides a suitable testbed for evaluating sampling strategies across broad gradients.

6.4.2 *Satellite data acquisition and preprocessing*

6.4.2.1 Multi-sensor configurations

We used four satellite configurations to enable comparative assessment of predictors with variable complexity.

Sentinel-1 (S1): C-band SAR data were obtained from the Google Earth Engine (GEE) collection COPERNICUS/S1_GRD (Ground Range Detected, Level-1, Interferometric Wide Swath mode, VV and VH polarizations, 10 m resolution), offering partial canopy penetration and sensitivity to vegetation structure and moisture (Torres et al., 2012). A total of 10 scenes from both ascending and descending orbits were acquired over the study area (S1.2 Tableau S1)

Sentinel-2 (S2): Level 2A surface reflectance data were obtained from the GEE collection COPERNICUS/S2_SR, atmospherically corrected by the ESA Sen2Cor processor. Bands B2 (blue, 490 nm), B3 (green, 560 nm), B4 (red, 665 nm), B8 (NIR, 842 nm), B11 (SWIR, 1,610 nm), and B12 (SWIR, 2,190 nm) were included, resampled to 10 m resolution. NDVI $[(B8-B4)/(B8+B4)]$ complemented spectral bands (Drusch et al., 2012). A total of 25 scenes were selected based on a cloud cover threshold of $< 20\%$ (S1.2 Tableau S2)

Sentinel-1+2 (S1+S2): Combined configuration integrating radar structural sensitivity with optical spectral discrimination (Acharki et al., 2021)

AlphaEarth Foundations: Geospatial embedding model generating 64-band embeddings representing annual space-time trajectories (Brown et al., 2025). Each pixel embedding integrates spectral indices (NDVI, EVI, NDWI), textural metrics from optical and radar imagery (Sentinel-2 L1C, Landsat 8/9, and Sentinel-1 GRD imagery), topographic variables (elevation, slope, aspect, curvature from Copernicus DEM GLO-30), and spatialized climatic indices (derived from ERA5-Land), capturing ecosystem multidimensional complexity in analysis-ready format.

6.4.2.2 Temporal acquisition and preprocessing

All imagery was acquired during April 2022, corresponding to field data collection (S1.1 Sentinel-1 GRD data (VV and VH polarizations, Interferometric Wide Swath mode) were obtained from the COPERNICUS/S1_GRD GEE collection and composited as a monthly median for April 2022 (S1.1, S1.3). Thermal noise removal, radiometric calibration, and terrain correction are included in the standard GRD pipeline. Sentinel-2 surface reflectance imagery was obtained from COPERNICUS/S2_SR, filtered for cloud cover $< 20\%$ (CLOUDY_PIXEL_PERCENTAGE), and composited using a pixel-wise median over April 2022 acquisitions (S1.1, S1.3). AlphaEarth annual embeddings (64 bands, 2022) were obtained from GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL and mosaicked across the study area. All datasets were clipped to the study area and processed at 10 m resolution. Spectral and radar values were extracted at plot centroids.

6.4.3 Calibration data sources

6.4.3.1 Field data

Field sampling (April–May 2022) used stratified random design. We established 41 plots (30×30 m) distributed proportionally across strata defined by combining five woody cover classes (20% intervals) with four vegetation formations. Each site was selected to ensure structural homogeneity within a 200 m radius. Plots were geolocated using a Garmin GPSMAP 65s (positional accuracy ± 1.5 m) and delineated using a Vertex hypsometer, with plots systematically located in homogeneous vegetation patches to minimise pixel-boundary effects.

Woody cover was measured using the Point-Intercept method adapted from the Australian Terrestrial Ecosystem Research Network protocol (Karl et al., 2017; White et al., 2012). Within each plot, 54 points were systematically distributed along four orthogonal transects spaced 10 m apart, with points positioned every 2 m. Woody presence (≥ 1 m height) was recorded at each point by vertical projection. Woody cover was calculated as the proportion of positive points (S4.1, Figure S4.2).

6.4.3.2 Photo-interpreted data

We established 505 plots (50×50 m) through photo-interpretation of very high-resolution orthophotographs (2021–2022) using Collect Earth Online (Bey et al., 2016).

Sampling followed a stratified systematic unaligned design to capture the full woody cover gradient (S5.1, Figure S5.1). Strata were defined using a systematic grid with one random point per cell. Environmental variables including topography (elevation from NASADEM), climatic gradients (mean annual precipitation), and vegetation indices (NDVI, EVI, NDWI) were displayed as auxiliary contextual layers in Collect Earth Online to support woody cover visual assessment during photo-interpretation (Colditz et al., 2011).

Each plot contained 49 observation points arranged in a 7×7 grid with approximately 7 m spacing (Figure 7.1, S5.2). Woody vegetation presence or absence at each point was determined through vertical photointerpretation using identical biophysical criteria as field data. This approach multiplied spatial sampling density by 12-fold while maintaining metric compatibility with field measurements.

Direct comparison between photo-interpreted and field-measured woody cover values for the 41 co-located field plots yields $R^2 = 0.775$, $RMSE = 0.125$, $MAE = 0.100$, and a mean relative error of 23.7% (regression slope = 0.92, bias = +0.042), confirming acceptable accuracy of photo-interpreted estimates relative to field measurements and supporting their use as calibration input under the logistical constraints typical of arid regions.

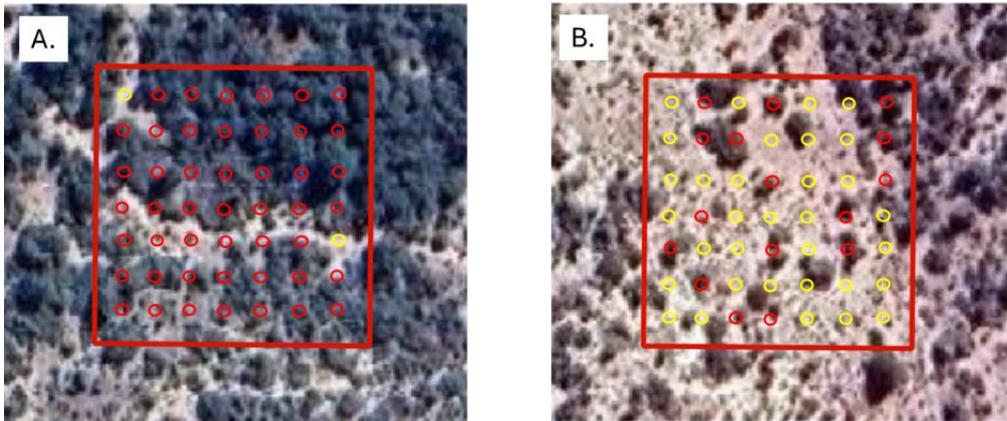


Figure 6-1: Photointerpretation of woody cover using very high-resolution orthophotographs. Examples of (A) dense woody cover ($47/49=0.96$) and (B) sparse woody cover ($16/49=0.33$). Each 50×50 m plot (red square) contains 49 observation points in a 7×7 grid. Red circles indicate woody vegetation presence; yellow circles indicate absence. A total of 505 plots were photo-interpreted from 2021–2022 orthophotographs using Collect Earth Online (Bey et al., 2016), following environmental stratification across topographic, climatic, vegetation, and edaphic gradients.

6.4.4 Comparative modeling and validation

Our analytical framework compared two calibration approaches across four satellite data sources using Random Forest algorithms (Figure 2). This design enabled systematic evaluation of how calibration data type, sample size, spatial distribution, and satellite data complexity influence woody cover mapping accuracy.

6.4.4.1 Random Forest architecture

We developed eight Random Forest models independently: four calibrated exclusively with 41 field plots (one per satellite configuration) and four utilizing all 505 photo-interpreted plots (one per configuration). In this supervised regression framework, each calibration plot provided a labeled observation — a field-measured or photo-interpreted wood cover value — paired with predictor values extracted as the mean of all pixels within each plot. For the Sentinel-1+2 configuration, predictors consisted of spectral and radar bands; for the AlphaEarth configuration, predictors were 64-dimensional embeddings derived from a deep learning model. Final models were then applied pixel-by-pixel across the entire study area to generate continuous wood cover prediction maps (0–100%). Random Forest was selected for its robustness against overfitting, ability to handle non-linear relationships, and resistance to correlated variables (Belgiu and Drăguț, 2016). The supervised learning process comprised three steps: (1) extraction of spectral or embedding predictor values per calibration polygon, calculated as the mean of all pixels within each plot footprint (30×30 m for field plots; 50×50 m for photo-interpreted plots); (2) supervised Random Forest model fitting on (predictor, observed woody cover) pairs, with hyperparameter optimisation via grid search ($n_{tree} = 500$, $m_{try} = \sqrt{p}$ where p is the number of

predictors, nodesize = 5; Probst et al., 2019); (3) pixel-by-pixel spatial prediction across the entire study area to generate continuous woody cover maps (0–100%), following recommendations from Probst et al. (2019) (S3.1, S3.2, S3.3, Tableau S3.1). The response variable underwent square-root transformation for variance stabilization before modeling, with back-transformation for evaluation (S3.3, S6.1). Variable importance was assessed using Mean Decrease in MSE, a metric inherent to the Random Forest algorithm (Breiman, 2001) ; detailed results are provided in Supplementary Figure S3.3.

6.4.4.2 Validation protocols

We implemented spatial cross-validation by partitioning the 41 field plots into five mutually exclusive geographic blocks through k-means clustering of coordinates ($k=5$, $nstart=25$), maximizing geographic separation between folds (Valavi et al., 2018) (S2.1, S2.3, Figure S2.1). Spatial independence was verified using Moran's I test ($I=-0.025$, $p=0.5$), confirming absence of residual spatial autocorrelation between folds (S2.2). These spatial folds, defined from field plots, were consistently applied to evaluate both field-based and photointerpretation-based models, ensuring methodological consistency for direct performance comparison.

6.4.5 Spatial and statistical analyses

Model performance was assessed through three complementary metrics. First, we quantified prediction accuracy using spatial cross-validation on the 41 field plots, calculating RMSE, relative error, and R^2 (S7, S6.1). Second, we examined distributions of predicted values across the study area. Third, we quantified spatial uncertainty through coefficient of variation across the five folds for each pixel (Nuzzo, 2019; Zhang et al., 2005) (S7.4). Systematic comparative analysis between satellite configurations enabled identifying relative influence of predictive data complexity versus calibration quality and density.

6.4.6 Sensitivity analysis to sampling density

6.4.6.1 Experimental design

To determine optimal photo-interpreted sampling density thresholds, we tested thirteen sample sizes: $n \in \{10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 505\}$ plots}, designed to capture the transition from minimal to complete sampling, with densification around expected transition zones (Mellor et al., 2015).

For each sample size below 505, we performed 100 iterations with different random seeds to quantify sampling variability and estimate performance confidence intervals (Olofsson et al., 2014) (S6.2). This analysis was repeated for each satellite configuration.

6.4.6.2 Contrasting spatial strategies

We compared two spatial distribution strategies. **Dispersed sampling** overlaid a regular grid and randomly selected one plot per cell, maximizing environmental representativeness. **Clustered sampling** randomly selected one central point then

selected nearest plots based on Euclidean distance, simulating logistical constraints typical of conservation programs (Wang et al., 2005). Each combination of sample size and spatial strategy was systematically tested with all satellite configurations.

6.4.7 Evaluation of metrics and critical threshold identification

6.4.7.1 Performance metrics

We used three complementary metrics: R^2 quantified variance explained, RMSE measured average prediction error magnitude, and mean relative error normalized errors relative to observed values, facilitating comparisons across cover ranges. For clarity, RMSE is reported here in absolute units (percentage points of woody cover on the 0–100% scale used throughout, see Figure 4.7), while the relative error metric corresponds to the relative RMSE (RRMSE), i.e. RMSE normalized by the mean observed cover value. A signed bias term is reported separately where relevant (e.g. as the regression slope/intercept of the validation against photo-interpreted reference data), rather than as part of these three primary accuracy metrics.

6.4.7.2 Optimal threshold identification

Sampling thresholds were identified based on model performance and stability, evaluated through repeated calibrations for each sample size. Following the uncertainty-aware framework of Olofsson et al. (2014), the analysis accounts for performance variability to ensure statistically comparable estimates across configurations.

The minimal threshold corresponds to the smallest sample size meeting acceptable criteria: $R^2 \geq 0.75$, RMSE efficiency $\geq 85\%$, and coefficients of variation below 15% for both metrics. The optimal threshold identifies the point where additional plots yield marginal gains, determined using stricter criteria: $R^2 \geq 0.84$, RMSE efficiency $\geq 92\%$, coefficients of variation below 12%, and average RMSE improvement below 2% across subsequent sample sizes. Thresholds were identified independently for each satellite configuration and spatial design. Sample sizes were normalized as sampling density (plots per 1000 km²) relative to the study area (4,826 km²).

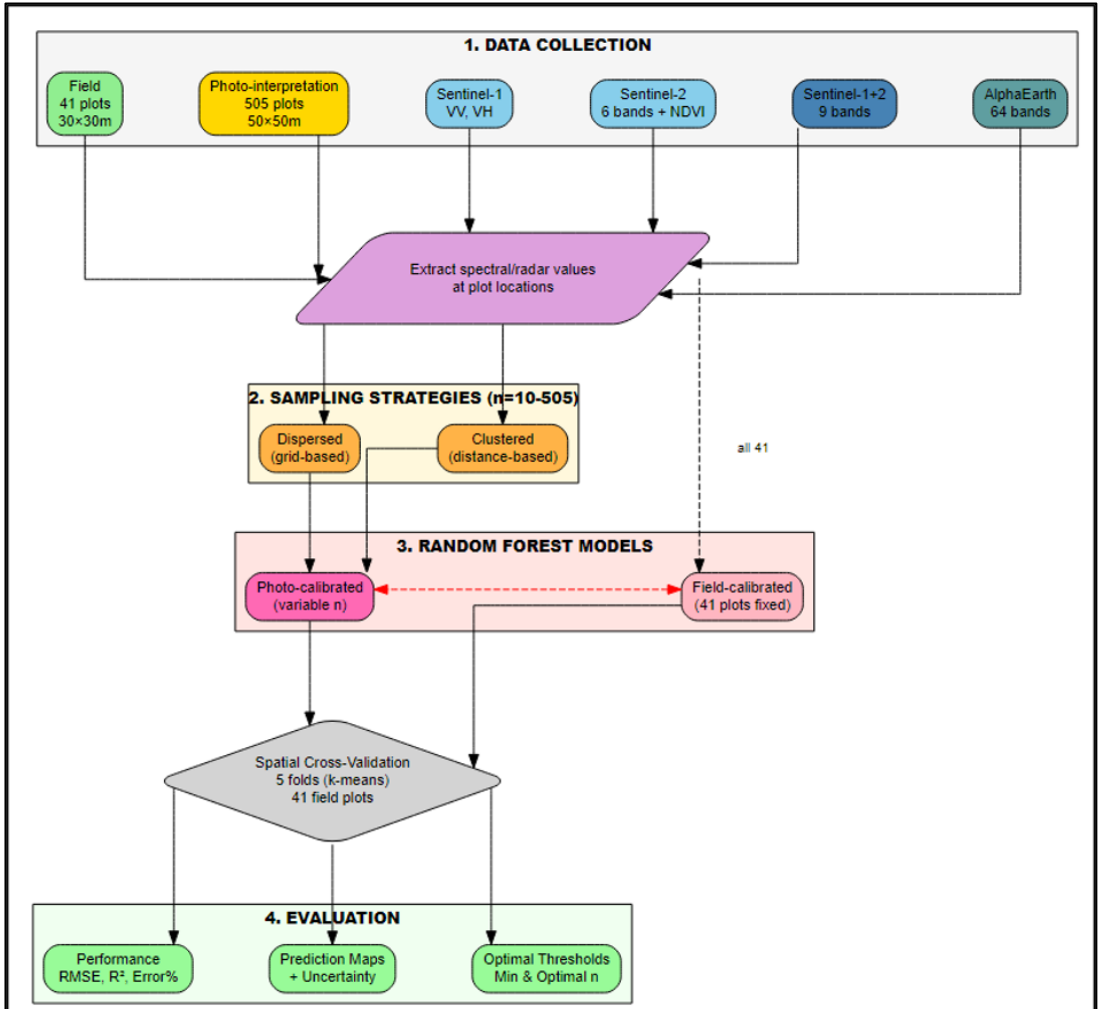


Figure 6-2: Methodological framework. The workflow integrates multi-source data collection (field plots, photointerpretation, and four satellite configurations: Sentinel-1(COPERNICUS/S1_GRD), Sentinel-2 (COPERNICUS/S2), Sentinel-1+2, and AlphaEarth), two sampling strategies (dispersed vs. clustered with n=10-505), and two calibration sources (photo-calibrated with variable n and field-calibrated with 41 plots). Models are validated using spatial cross-validation where the 41 field plots are partitioned into 5 folds (k-means) and evaluated based on performance metrics (RMSE, R², Error%), prediction maps with uncertainty, and optimal thresholds and sample sizes.

6.5. Results

6.5.1 Spatial representativeness outweighs measurement precision

Direct comparison between calibration approaches—field measurement precision (41 plots) versus photointerpretation spatial density (505 plots)—revealed substantial performance differences when validated against identical independent field data (Figure 3). Although each photo-interpreted observation relies on visual assessment rather than direct field measurement, their 12-fold greater spatial density (10.5 versus 0.85 plots/100 km²) consistently produced superior models across all satellite configurations. Using Sentinel-1+2, photointerpretation achieved $R^2 = 0.88$ (RMSE =

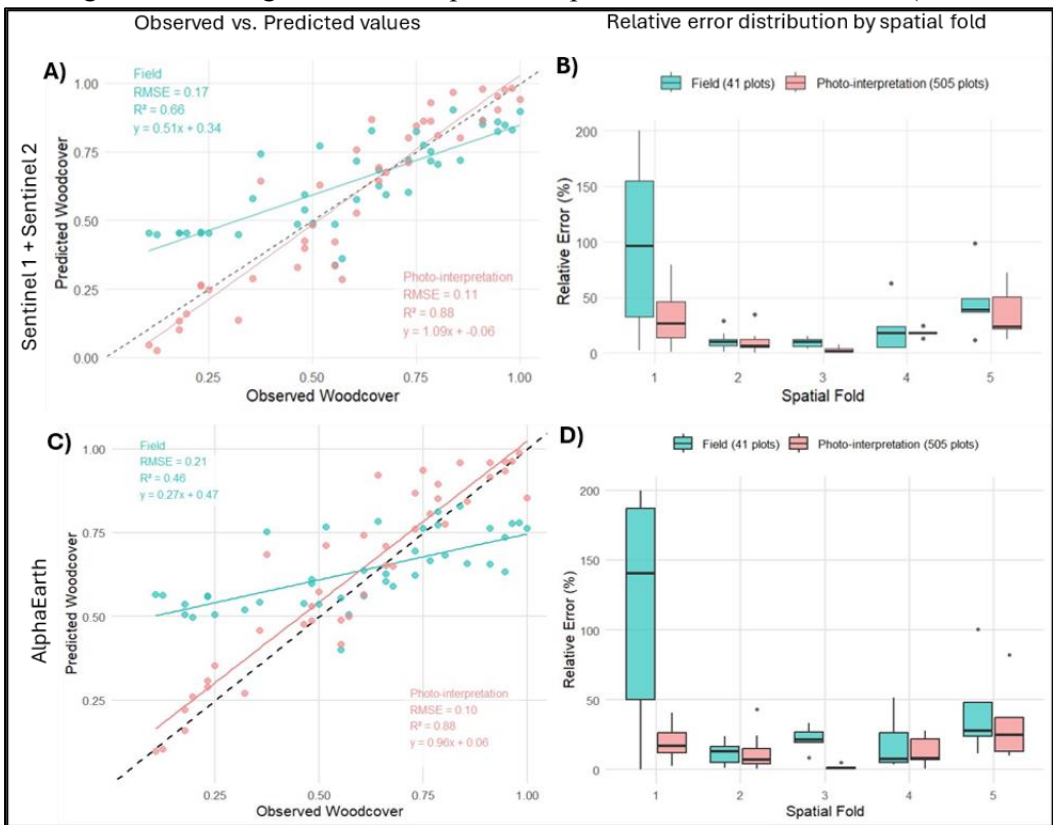


Figure 6-3: Performance comparison between field and photointerpretation calibration for mapping woody cover with Sentinel-1+2 (A, B) and AlphaEarth (C, D). (A, C) Observed vs. predicted scatter plots demonstrate photointerpretation superiority (505 plots: $RMSE = 0.11$, $R^2 = 0.88$ for Sentinel-1+2; $RMSE = 0.10$, $R^2 = 0.88$ for AlphaEarth) over field calibration alone (41 plots: $RMSE = 0.17$, $R^2 = 0.66$ for Sentinel-1+2; $RMSE = 0.21$, $R^2 = 0.46$ for AlphaEarth). (B, D) Relative error distribution by spatial validation blocks reveals reduced and stable variability for photointerpretation calibration versus field approaches for both satellite data sources.

0.11) compared to $R^2 = 0.66$ (RMSE = 0.17) for field calibration—a 35% error reduction (Figure 3A). This performance gap widened with AlphaEarth: Photointerpretation maintained $R^2 = 0.88$ (RMSE = 0.10) while field calibration declined to $R^2 = 0.46$ (RMSE = 0.21) (Figure 3C).

The relationship between observed and predicted values differed markedly between approaches. Photointerpretation produced near-linear relationships with regression slopes of 1.09 (Sentinel-1+2) and 0.96 (AlphaEarth), approximating ideal 1:1 correspondence (Figure 3A, C). Field calibration yielded slopes of only 0.51 (Sentinel-1+2) and 0.27 (AlphaEarth), indicating systematic underestimation of woody cover. This compression became particularly acute above 60% cover, where field-calibrated predictions plateaued regardless of observed values.

Cross-validation revealed contrasting spatial stability patterns. Field calibration exhibited high variability among spatial folds, with median relative errors ranging from 20% to 150% across blocks (Figure 3B, D). This instability intensified with predictor complexity: AlphaEarth produced wider error distributions (10–190%) than Sentinel-1+2 (20–150%), suggesting insufficient sample size to constrain high-dimensional predictor space. Photointerpretation maintained stable median relative errors of 15–35% across all spatial blocks and satellite configurations, demonstrating robust spatial generalization with 505 calibration plots.

6.5.2 Calibration density determines spatial pattern fidelity

Predicted woody cover maps differed qualitatively between calibration approaches (Figure 4). Field-calibrated maps (41 plots) displayed spatially homogeneous patterns with most predictions concentrated between 40–60% woody cover. Distribution histograms showed unimodal peaks centered at 50%, failing to reproduce the bimodal distribution observed in field data (primary mode: 20–30%; secondary mode: 60–70%) (Figure 7.4E, G). This compression toward intermediate values occurred with both Sentinel-1+2 (Figure 7.4A, E) and AlphaEarth (Figure 7.4C, G). Prediction uncertainty was high (30–40%) and spatially concentrated in coastal zones, expanding more broadly under AlphaEarth (Figure 7.4I, K).

Photo-interpretation-calibrated maps (505 plots) captured spatial heterogeneity through complex mosaics alternating between high-cover (70–90%) and low-cover (10–30%) zones (Figure 7.4B, D). Predicted distributions exhibited clear bimodality matching field observations, with primary peaks at 20–30% and secondary peaks at 60–70% (Figure 7.4F, H). This distributional fidelity persisted across both satellite configurations. Prediction uncertainty remained low (10–25%) and spatially uniform, with localized increases only in ecotonal transition zones (Figure 4J, L).

6.5.3 Sampling strategy and size control model convergence

Systematic variation of photo-interpretation sample size revealed distinct convergence trajectories by sampling strategy (Figure 7.5). Under dispersed sampling, performance metrics stabilized rapidly: R^2 exceeded 0.80 at 100–150 plots and plateaued at 0.88 beyond 200 plots for both Sentinel-1+2 and AlphaEarth. RMSE declined from 0.20 (10 plots) to 0.11–0.12 (>150 plots), while relative error decreased from 30% to 20% over the same range. Marginal gains beyond 200 plots were minimal (<2% R^2 improvement per 50-point increment).

Clustered sampling exhibited slower convergence with greater variability. Between 10 and 100 plots, R^2 fluctuated broadly (0.50–0.70) with large error bars indicating high sensitivity to specific plot locations. Performance approached dispersed sampling levels only at 400–500 plots ($R^2 = 0.85$ – 0.88 , RMSE = 0.11–0.12), requiring 2–2.5× more observations to achieve equivalent accuracy.

Satellite configuration comparisons showed differential data requirements. Sentinel-1 alone plateaued at $R^2 = 0.68$ regardless of sample size, confirming fundamental limitations of single-polarization radar for woody cover estimation. Sentinel-2 required 25–30% more plots than Sentinel-1+2 to reach comparable performance, demonstrating synergistic value of optical-radar fusion. AlphaEarth and Sentinel-1+2 converged to similar asymptotic performance ($R^2 = 0.88$, RMSE = 0.11) from 150 plots onward under dispersed sampling, indicating limited marginal benefit of foundation model embeddings once adequate sampling is achieved.

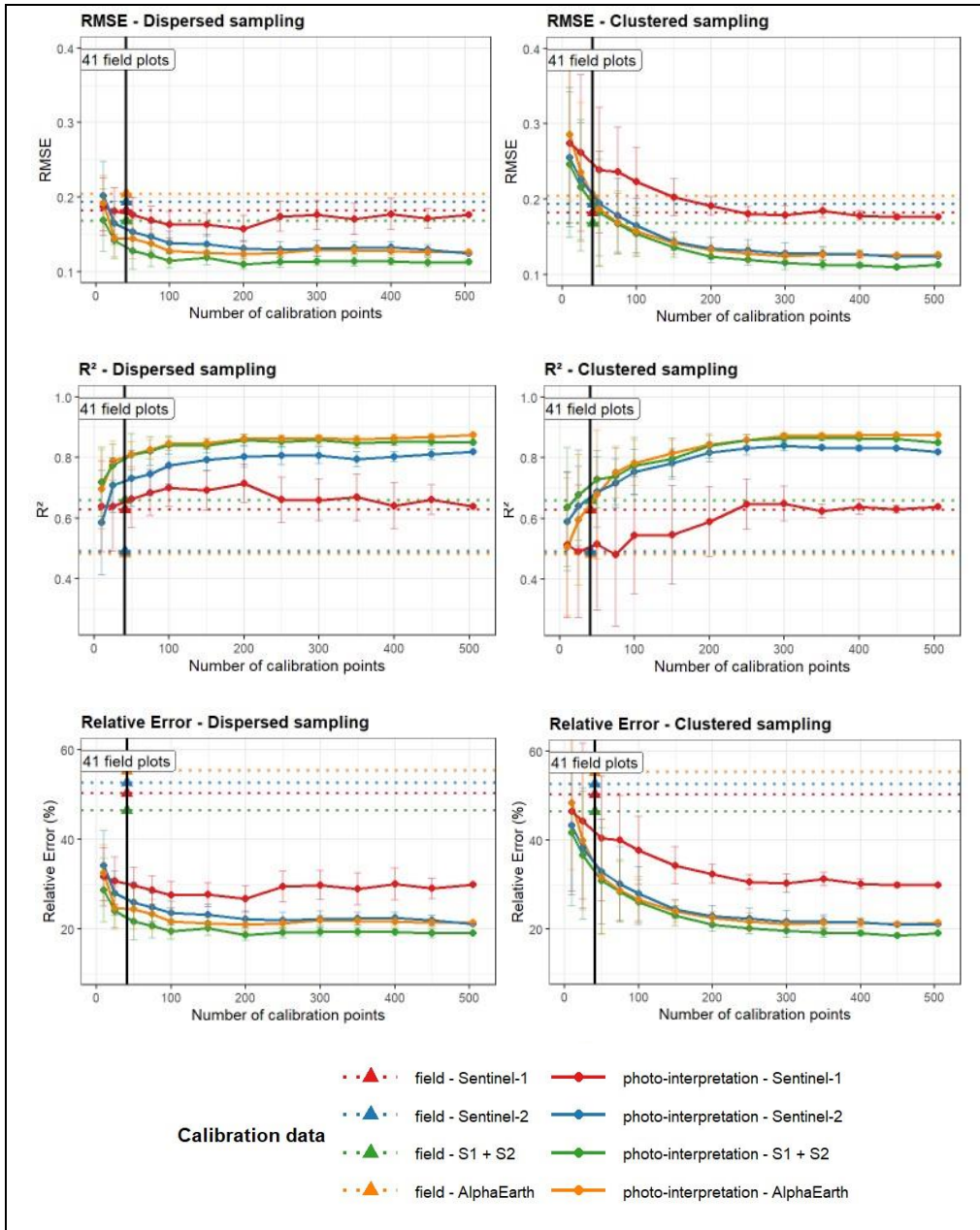


Figure 6-5: Performance metrics across photo interpretation sample sizes and sampling strategies. RMSE (top), R² (middle), and relative error (bottom) for dispersed (left) and clustered (right) sampling using Sentinel-1, Sentinel-2, Sentinel-1 + Sentinel-2, and AlphaEarth. Dashed lines with triangles: field-based calibration; solid lines with circles: photo interpretation-based calibration. Vertical line indicates baseline field calibration (41 plots). Error bars show standard deviations over 100 iterations.

6.5.4 Operational thresholds depend on spatial distribution

Quantitative threshold identification revealed consistent patterns across satellite configurations (Tableau 7.1). For dispersed sampling, minimum acceptable performance ($R^2 \geq 0.82$, $RMSE \leq 0.14$) emerged at 15.5 plots/1000 km² (75 plots). Optimal performance balancing accuracy and sampling effort ($R^2 \geq 0.85$, $RMSE \leq 0.11$) occurred at 20.7–41.4 plots/1000 km² (100–200 plots), beyond which cost-benefit ratios declined.

Clustered sampling required substantially higher densities. Minimum thresholds reached 31.1–41.4 plots/1000 km² (150–200 plots)—double that of dispersed sampling—while optimal thresholds extended to 51.8–62.2 plots/1000 km² (250–300 plots). These elevated requirements reflect spatial redundancy inherent to clustered designs, where proximate observations provide diminishing marginal information.

Satellite configuration had minimal impact on threshold locations: Sentinel-1+2 and AlphaEarth differed by only 1–3% R^2 at equivalent sample sizes, with threshold densities varying within 10 plots/1000 km². Notably, minimal photo-interpretation sampling (75 dispersed plots; $R^2 = 0.82$) surpassed conventional field calibration (41 plots; $R^2 = 0.46$ – 0.66), demonstrating superior efficiency for resource-constrained applications.

Tableau 6-1: Performance metrics across photo interpretation sample sizes and sampling strategies. RMSE (top), R^2 (middle), and relative error (bottom) for dispersed (left) and clustered (right) sampling using Sentinel-1, Sentinel-2, Sentinel-1 + Sentinel-2, and AlphaEarth. Dashed lines with triangles: field-based calibration; solid lines with circles: photo interpretation-based calibration. Vertical line indicates baseline field calibration (41 plots). Error bars show standard deviations over 100 iterations.

Distribution	Source	Threshold	Plots/ 1000 km ²	N plots	RMSE	R ²
Dispersed	Sentinel-1+2	Minimal	15.5	75	0,12	0,82
		Optimal	41.4	200	0,11	0,85
	AlphaEarth	Minimal	15.5	75	0,14	0,82
		Optimal	20.7	100	0,14	0,84
Clustered	Sentinel-1+2	Minimal	41.4	200	0,12	0,83
		Optimal	62.2	300	0,13	0,86
	AlphaEarth	Minimal	31.1	150	0,14	0,81
		Optimal	51.8	250	0,13	0,85

6.6. Discussion

6.6.1 Superiority of photointerpretation for model calibration

Models calibrated with 505 photo-interpreted plots achieved substantially higher performance ($R^2 = 0.88$, $RMSE = 0.11$) compared to those calibrated with 41 field plots ($R^2 = 0.46$ – 0.66 , $RMSE = 0.17$ – 0.21) using spatially independent clusters via k-means spatial clustering for validation, representing a 35–90% improvement depending on satellite configuration. The superior performance of photointerpretation reflects a fundamental calibration principle: spatial representativeness outweighs measurement precision for machine learning models. Sparse field plots (41) created environmental gaps forcing model extrapolation (slopes = 0.27 – 0.51), while dense photointerpretation (505) enabled interpolation (slopes = 0.96 – 1.09). For context, Dynamic World — a globally trained deep-learning land cover product — achieved $R^2 = 0.70$ on the same 41 independent field plots (Brown et al., 2022; Ramalason et al., 2025b) markedly lower than our locally calibrated models ($R^2 = 0.88$), demonstrating the substantial benefit of regionally tailored calibration. Prior to this study, woody cover mapping in southwestern Madagascar had been limited to discrete land cover classification rather than continuous cover estimation (Ramalason et al., 2025a), underscoring the novelty of the spatially continuous approach developed here.

Spatial representativeness translates into remarkable cross-validation stability: relative errors remain homogeneous (10–50%) regardless of geographic fold tested, while field calibration exhibits strong instability with errors ranging from 5% to 190% across folds, confirming results of Tsendbazar et al. (2021), who demonstrated that spatial representativeness outweighs raw sample quantity for large-scale mapping.

Direct validation of photo-interpreted against field-measured woody cover values for the 41 co-located plots confirms acceptable measurement accuracy ($R^2 = 0.77$, $RMSE = 0.12$, slope = 0.92 , bias = $+0.042$), with the slight positive bias consistent with the known tendency of VHR imagery to capture canopy extent more completely than the point-intercept method. This demonstrates that photo-interpreted data provides both adequate spatial coverage and acceptable individual measurement quality for model calibration.

Our results align with the growing use of photointerpretation for ecological mapping. Provencher-Nolet et al., (2014) achieved 85% accuracy for forest habitat classification using aerial photographs, while Pérez-Luque et al. (2022) demonstrated strong correlation ($R^2 = 0.81$) between drone-based estimates and field sampling in Mediterranean shrublands. Francis et al. (2023) showed that integrating drone and satellite data enables precise monitoring of vegetation changes at landscape scale. Advances in deep learning (Symeonakis et al., 2022) and increasing accessibility of cloud platforms (Guida-Johnson et al., 2021) further enhance operational potential.

These performances do not diminish the intrinsic value of field surveys, which provide high point-level accuracy and complementary ecological information (floristic composition, vertical structure). However, in vast arid landscapes where

logistical and financial constraints limit field sampling density (McRoberts et al., 2018), photointerpretation constitutes a pragmatic alternative, provided rigorous validation and quality control protocols are implemented.

6.6.2 Critical thresholds and spatial distribution of samples

The identification of critical sampling density thresholds represents a major methodological contribution for designing calibration campaigns. Three key findings emerge.

First, spatial distribution dominates: clustered sampling requires twice as many points as dispersed sampling to achieve equivalent performance. For spatially dispersed sampling, we identified a minimum threshold of 15.5 plots/1000 km² (75–100 plots) ensuring acceptable accuracy ($R^2 = 0.82$, $RMSE \leq 0.14$), and an optimal threshold of 20.7–41.4 plots/1000 km² (100–200 plots) maximizing cost-benefit ratio ($R^2 = 0.85$ – 0.86 , $RMSE = 0.11$ – 0.12). Beyond this threshold, performance gains become marginal (< 2% per 50 additional plots), suggesting saturation of predictive capacity. Clustered sampling requires substantially higher densities: 31.1–41.4 points/1000 km² (150–200 plots) for minimum threshold and 51.8–62.2 points/1000 km² (250–300 plots) for optimal threshold—systematically double that of dispersed sampling.

Second, AlphaEarth provides negligible advantage over Sentinel-1+2 (1–3% R^2 gain) when adequate calibration samples are available, demonstrating that sample design outweighs satellite data sophistication.

Third, minimal photointerpretation (75 dispersed points; $R^2 = 0.82$) outperforms 41-point field calibration ($R^2 = 0.46$ – 0.66), confirming superior cost-effectiveness for resource-limited monitoring programs. This difference persists regardless of satellite data source, demonstrating that spatial distribution constitutes the dominant factor affecting model performance. Even with minimal but dispersed photo-interpreted sampling, performance substantially exceeds that obtained with field plots, underscoring critical importance of geographic coverage.

These results complement the work of Mellor et al. (2015), who observed saturation of Random Forest performance beyond 200–300 samples for land cover mapping, and Qu et al. (2024), who identified performance plateaus for forest mapping from approximately 150 samples. Our contribution extends these observations by precisely quantifying thresholds for different spatial strategies and data configurations, providing directly transferable operational benchmarks.

Normalization as surface density (plots/1000 km²) facilitates extrapolation to other geographic contexts, although the thresholds identified here are specific to this biome and the Random Forest modelling algorithm. More fragmented or structurally complex ecosystems may require higher densities to capture environmental variability (Stehman and Foody, 2019). Conversely, more spatially uniform ecosystems are expected to reach performance saturation at lower densities. Nevertheless, these thresholds provide a solid foundation for dimensioning sampling efforts in operational monitoring programs, particularly for government agencies, NGOs, and local projects facing budget constraints.

6.6.3 Multi-sensor complementarity: gains conditional on sample adequacy

Comparative analysis of satellite data sources reveals nuanced results challenging our hypothesis of systematic progressive improvement with predictor complexity. Individual sensors showed distinct limitations: Sentinel-1 alone remained constrained regardless of sample size (R^2 max = 0.68), confirming that radar data, while valuable for capturing structural information including canopy height, density, and moisture, cannot fully represent spectral variability necessary for accurate woody cover estimation in heterogeneous arid landscapes. Sentinel-2 performed better but required 25–30% additional calibration plots compared to Sentinel-1+2 to achieve equivalent performance.

The Sentinel-1+2 combination demonstrated clear advantages with limited samples, achieving $R^2 = 0.66$ with only 41 field plots compared to lower values for individual sensors. This synergy—radar providing structural metrics and optical data capturing biochemical variability including chlorophyll and vegetation indices—proves particularly valuable in arid ecosystems where seasonal phenology and sparse vegetation complicate spectral interpretation, confirming findings by Frison et al. (2018) and Reiche et al. (2018) in tropical forests, and recent studies in arid zones (Karakizi et al., 2023).

However, AlphaEarth contradicted expected progressive improvement. With adequate photo-interpreted sampling (150–200 plots), both AlphaEarth and Sentinel-1+2 converged toward nearly identical performance ($R^2 = 0.88$, RMSE = 0.11), with only 1–2% marginal superiority for AlphaEarth. More strikingly, with limited field calibration (41 plots), AlphaEarth substantially underperformed ($R^2 = 0.46$ vs 0.66 for Sentinel-1+2).

This performance collapse illustrates the curse of dimensionality (Hawkins, 2004). AlphaEarth integrates dozens of variables including spectral indices, textural metrics, and topographic derivatives, creating a high-dimensional predictor space. With only 41 observations, Random Forest cannot stably estimate relationships between these numerous predictors and woody cover, resulting in overfitting where the model memorizes training data rather than learning generalizable patterns. Conversely, Sentinel-1+2 combines fewer fundamental variables including radar polarizations, spectral bands, and NDVI, requiring less calibration data for robust parameter estimation.

Variable importance analysis further revealed that VH backscatter dominated Sentinel-1+2 predictions under both calibration sources, while AlphaEarth showed a more diffuse importance distribution under photointerpretation, reflecting broader predictor exploitation with larger training sets (Figure S3.3)

These findings demonstrate that sample design—not satellite data sophistication—primarily determines mapping accuracy in resource-constrained contexts. Investing in spatially representative photo-interpreted samples combined with freely accessible dual-sensor data (Sentinel-1+2) proves more effective than limited field calibration

paired with complex multi-dimensional composites, confirming observations by (Foody, 2009; Mellor et al., 2015) that sample size and spatial distribution outweigh predictor complexity.

6.6.4 Limitations and methodological perspectives

Five principal limitations warrant explicit acknowledgment.

First, photointerpretation relies on visual judgment by human operators, introducing inter-observer variability that can affect result consistency. In this study, two operators followed a standardized protocol, but no formal inter-rater reliability assessment was conducted. Insufficient image resolution further amplifies this uncertainty in open and sparse woody formations where woody-herbaceous discrimination becomes ambiguous. Future studies should systematically cross-validate subsamples across multiple operators to quantify and minimize these biases.

Second, the limited number of field plots (41) constrains external validation of photo-interpreted models. Direct comparison between photo-interpreted and field-measured woody cover values for the 41 co-located plots ($R^2 = 0.775$, $RMSE = 0.125$, $slope = 0.92$, $bias = +0.042$) partially addresses this limitation by quantifying the accuracy of photo-interpreted data as calibration input; a fully independent validation set of larger size would nonetheless further strengthen conclusions. However, we implemented rigorous spatial cross-validation via k-means spatial clustering, which accounts for spatial autocorrelation by ensuring that validation clusters are geographically distinct from training clusters. This approach is fundamental, as Ploton et al. (2020) revealed poor predictive performance for large-scale ecological models when spatial autocorrelation is ignored. Recent methodological developments, such as the CAST package (Meyer et al., 2024) and recommendations by Stock, (2025), reinforce the necessity of these approaches for robust evaluation of model spatial generalization.

Third, the calibration density thresholds identified here (15.5–41.4 plots per 1,000 km²) are specific to the Malagasy spiny forest biome and to the Random Forest algorithm, whose predictive performance varies with ecosystem context (Belgiu and Drăguț, 2016). Structurally heterogeneous ecosystems will require higher calibration densities, whereas spatially homogeneous ecosystems are expected to reach performance saturation at lower densities (Fassnacht et al., 2014). Furthermore, alternative algorithms — notably Support Vector Machines or deep learning approaches potentially better suited to exploiting the high-dimensional embedding space of AlphaEarth — may yield different threshold values (Fassnacht et al., 2016). Validating this framework with alternative algorithms therefore constitutes a priority for future research.

6.6.5 Implications for sustainable management of arid ecosystems

These results have direct applications for sustainable management of woody resources. Fine-scale mapping of woody cover enables identification of declining areas (deforestation, overgrazing), remnant vegetation cores requiring priority conservation, and areas favorable for natural or assisted regeneration. (Brandt et al., 2020) revealed an unexpected number of trees in Sahel drylands using high-resolution imagery, substantially revising carbon stock estimates. Sengani et al. (2023) highlighted the importance of woody cover mapping for early detection of degradation in semi-arid regions.

The proposed methodology integrates with Land Degradation Neutrality (LDN) objectives, providing spatially explicit information to guide restoration actions (Bastin et al., 2019), track international commitments, and justify investments. The accessibility of platforms like Collect Earth Online facilitates implementation of participatory protocols where local communities can contribute to data collection, strengthening local ownership and sustainability of monitoring systems (Saah et al., 2019).

6.6.6 Transferability and operational recommendations

Our sampling optimization framework is directly transferable to global arid and semi-arid ecosystems with heterogeneous woody vegetation across all cover density levels. This framework was specifically developed within the Malagasy spiny forest biome, a structurally coherent ecological unit; applying it to coarser geographic units would dilute vegetation heterogeneity and reduce threshold precision, whereas finer ecological subdivisions would improve stratification efficiency. Implementation requires: (i) Sentinel-1+2 combined data or AlphaEarth Foundations as predictors, (ii) photointerpretation via platforms like Collect Earth Online, and (iii) dispersed sampling representing all vegetation classes and cover types present in the landscape.

The quantitative thresholds identified (20.7–41.4 dispersed plots per 1,000 km² for optimal accuracy; 15.5 plots per 1,000 km² as minimal configuration) provide indicative reference benchmarks applicable to savannas, Sahelian zones, and xerophytic formations across continents, though independent recalibration is recommended for structurally distinct ecosystems or when using alternative modelling algorithms (Belgiu and Drăguț, 2016; Fassnacht et al., 2016). The comparative approach constitutes a standardized reproducible protocol applicable across environmental contexts. Compatibility with free Sentinel data and open-source photointerpretation platforms removes technological and financial barriers for developing countries. Managers can adapt sampling densities according to budget and local heterogeneity. For landscapes with greater structural complexity, we recommend increasing target sampling densities and adopting systematic stratified random sampling to ensure representation of all dominant ecosystem types (Fassnacht et al., 2014; Stehman and Foody, 2019). Conversely, spatially homogeneous ecosystems are

expected to reach performance saturation at lower densities (Mellor et al., 2015; Qu et al., 2024).

Finally, although AlphaEarth embeddings implicitly encode climatic signals, their individual contributions remain uninterpretable, making explicit precipitation predictors a valuable and transparent complement. Direct integration of freely available high-resolution precipitation data (e.g., CHIRPS, available in GEE at 5.5 km resolution) into Sentinel-based configurations therefore represents a straightforward improvement recommended for future applications, particularly where strong precipitation gradients drive vegetation heterogeneity.

6.7. Conclusion

This study demonstrates that very high-resolution photointerpretation provides a robust operational alternative to field inventories for calibrating woody cover models in arid ecosystems. Models calibrated with 505 photo-interpreted plots achieved substantially higher accuracy ($R^2 = 0.88$) compared to 41 field plots ($R^2 = 0.46\text{--}0.66$), confirming superiority of spatial representativeness over point-level precision.

These findings advance prior work in three directions. Our empirically derived thresholds (20.7–41.4 dispersed plots/1,000 km²) align with saturation patterns reported by Mellor et al. (2015) beyond 200–300 samples for land cover mapping and by Qu et al., (2024) from approximately 150 samples for soil property prediction, while providing spatially normalised, biome-specific benchmarks absent from earlier studies. Our demonstration that spatial representativeness outweighs measurement precision extends the conclusions of Tsendbazar et al.,(2021) — who showed this principle at the validation stage — to the calibration stage itself. Finally, the convergence of Sentinel-1+2 and AlphaEarth under adequate sampling reinforces Foody, (2009) and Mellor et al., (2015), now demonstrated across foundation model embeddings.

Collectively, these results make six methodological contributions: (1) the first quantitative framework explicitly isolating calibration source, spatial configuration, and sampling density; (2) empirical evidence that spatial representativeness outweighs point-level precision; (3) a systematic cross-sensor comparison across four predictor sets; (4) operational sampling-density thresholds expressed as plots per 1,000 km²; (5) demonstration that dispersed sampling achieves equivalent accuracy with half the samples of clustered designs; and (6) rigorous spatially blocked cross-validation ensuring reliable and generalizable performance estimates.

The identified thresholds provide directly transferable operational benchmarks for global arid ecosystems, implementable with Sentinel-1+2 or AlphaEarth predictors, dispersed photointerpretation via Collect Earth Online, and representative sampling across vegetation gradients. Compared to globally trained products such as Dynamic World ($R^2 = 0.70$ on the same independent field plots; Ramalason et al., 2025), locally calibrated models with adequate photointerpretation sampling ($R^2 = 0.88$) demonstrate the substantial benefit of regional calibration, consistent with the growing evidence that spatially representative training data matter more than model

complexity (Foody, 2009; Tsendbazar et al., 2021). Principal limitations — single-biome derivation, dependence on Random Forest, and absence of formal inter-observer reliability assessment — define clear future priorities: threshold validation across biomes, evaluation of alternative algorithms, integration of high-resolution precipitation data (CHIRPS), and exploration of multi-temporal and drone-based acquisition.

Chapitre 7

Spatio-temporal analysis of structural degradation of xerophytic thickets masked by limited net area loss: a 36-year Landsat assessment (1989–2025)

7.1. Référence bibliographique

F.N. Ramalason · O.L. Rakotondrasoa · A. Vander Linden · G. Renard · J.R. Randriamalala · N.J. Vereecken · A. Razakamiaramanana · I.H. Hantanirina · M.F. Ranaivoharivelo · S. Raholinarivo · K.R.S. Sambieni · H. De Lame · M.M. Mukenza · J. Bogaert · J.-F. Bastin, 2026. Structural degradation of xerophytic thickets masked by limited net area loss : a 36-year Landsat assessment (1989–2025), soumis à la revue Discover forests

7.2. Abstract

Binary forest/non-forest classifications systematically fail to detect gradual structural degradation in woody formations where continuous cover ranges widely (from sparse to dense). This monitoring gap limits evidence-based conservation of semi-arid dryland ecosystems that provide critical ecosystem services to millions of people.

We developed a continuous woody cover monitoring framework using 36 years of Landsat imagery (1989–2025) in xerophytic thickets of southwestern Madagascar. The model was calibrated on 505 photo-interpreted plots and independently validated on 41 field plots ($R^2 = 0.751$). Drivers of degradation—fire regime and settlement proximity—were analyzed using MODIS burned area data and satellite-derived building locations.

Net woody cover loss was modest (-3.2% over 36 years), yet structural transformation was profound: intact thickets declined from 62.8% to 29.0% of landscape area, while degraded thickets increased from 18.5% to 39.5%. Degradation accelerated nonlinearly beyond 8.9 cumulative fire events (median loss -18.0% vs. -0.1% for unburned areas). Woody cover loss declined monotonically with distance from settlements ($-0.166\% \text{ yr}^{-1}$ at 1 km to $-0.090\% \text{ yr}^{-1}$ at 10 km), revealing spatially explicit management zones.

Continuous cover monitoring reveals hidden degradation invisible to categorical maps, enabling conservation prioritization and fire management thresholds. The 8.9 fire-event threshold provides an operational target: areas below it retains recovery potential, while areas exceeding it require active restoration. This framework is directly transferable to tropical drylands where categorical products systematically fail to detect gradual structural change.

Key words: woody cover; xerophytic thickets; southwestern Madagascar; Landsat time series; fire regime; human proximity gradient.

7.3. Introduction

Dryland and semi-arid ecosystems cover approximately 40% of the Earth's land surface and support more than two billion people, making them a critical biome for global socio-ecological systems (Reynolds et al., 2007; UNCCD, 2024). They deliver essential ecosystem services including woody biomass production, carbon storage, support for endemic biodiversity, and natural forage for livestock (D'Odorico et al., 2013). Despite their importance, these ecosystems remain underrepresented in global vegetation monitoring systems, particularly regarding woody cover dynamics. Recent

work has shown that dryland woody vegetation is substantially more extensive than previously estimated, exposing biases inherent to conventional mapping approaches (Bastin et al., 2017). This underestimation has direct implications for degradation monitoring, as open formations remain systematically excluded from global forest inventories, preventing the detection of gradual cover loss in structurally open ecosystems.

Accurate monitoring of woody cover is essential for understanding degradation dynamics in dryland ecosystems and for supporting sustainable natural resource management. However, persistent methodological constraints hinder progress: strong phenological variability among deciduous species, structural heterogeneity of landscapes, and temporally sparse observations mask the nonlinear degradation and regeneration trajectories that characterize these systems (DeVries et al., 2015; Reynolds et al., 2007). These limitations are particularly acute across African drylands, where the majority of woody vegetation falls outside formal forest definitions (Brandt et al., 2020).

Current satellite-based assessments represent woody vegetation as a discrete land-cover class rather than as a continuous cover variable. This categorical approach systematically overlooks open woody formations and fails to detect the gradual, partial degradation processes that dominate vegetation change in xerophytic shrublands, where subtle declines in fractional cover may already reflect major ecological transitions without triggering categorical reclassification. Globally trained classifiers have demonstrated limited accuracy in structurally complex dryland environments (Ramalason et al., 2026), where spectral heterogeneity is high and training data from other biomes may not transfer well (Brown et al., 2022; Ramalason et al., 2025b).

Recent advances in remote sensing offer new opportunities to overcome these limitations. Access to multi-decadal Landsat archives through Google Earth Engine enables the analysis of dense, harmonized time series, opening the way for annual, large-scale woody cover monitoring (Amani et al., 2020; Gorelick et al., 2017). Machine learning methods, particularly Random Forest regression, allow continuous woody cover estimation by integrating spectral information related to vegetation structure, water status, and phenology (Belgiu and Drăguț, 2016; Breiman, 2001). High-resolution photointerpretation provides a reliable calibration data source in contexts where field surveys are costly and logistically constrained (Bey et al., 2016).

Dryland woody dynamics are strongly influenced by fire regimes and human pressure. Fires can promote regeneration at low frequency or drive conversion toward herbaceous systems at high frequency, while proximity to settlements creates resource extraction gradients that progressively reduce woody cover (Archibald et al., 2013; Laurance et al., 2002). Yet the joint quantification of these pressures and their threshold effects on woody cover dynamics remains limited by the lack of long-term, continuous monitoring data.

This study addresses this gap by exploiting 36 years of harmonized Landsat imagery (1989–2025). Two hypotheses are tested: (H1) woody cover has declined nonlinearly

over 36 years, with structural degradation affecting predominantly high-cover formations; (H2) woody cover loss increases nonlinearly with fire frequency and proximity to settlements, revealing threshold effects in the ecosystem response to disturbance. This study aims to: (1) quantify annual and periodic rates of woody cover change over 36 years (1989–2025); (2) characterize the temporal trajectory of woody cover at landscape scale; (3) evaluate the relationship between woody cover change and fire regime (cumulative frequency and time since last fire); (4) analyse the effect of distance from settlements on woody cover dynamics.

7.4. Materials and Methods

7.4.1 Study area

The study was conducted in the southern part of the Toliara II district (Atsimo-Andrefana region), southwestern Madagascar, covering 4,826 km² south of the Fiherenana River (23°00'–24°30' S, 43°30'–44°30' E). The climate is arid to semi-arid (mean aridity index AI = 0.21; range: 0.13–0.36, UNCCD classification; (Trabucco and Zomer, 2019; UNCCD, 2024)), with mean annual precipitation of 466 mm (range: 274–770 mm) and a mean annual temperature of 24.8°C (2000–2022; CHIRPS, (Funk et al., 2015); ERA5-Land, (Hersbach et al., 2020)). Southern sectors are classified as arid (AI < 0.20) while the majority of the zone is semi-arid (AI 0.20–0.50). Rainfall is strongly seasonal, concentrated between December and March (January mean: 151 mm; August mean: 4 mm), with seven months receiving less than 30 mm. The hydrographic network is dominated by seasonal rivers with intermittent flow (MERIT Hydro; (Yamazaki et al., 2019)). Topographic relief is moderate, with elevations ranging from near sea level to 598 m (mean: 162 m; SD: 129 m; Copernicus DEM GLO-30).

Vegetation comprises two structurally heterogeneous formations along an east-west gradient: (1) coastal xerophytic thickets, characterized by open structure and dominance of succulent species from endemic families Didiereaceae (*Didiera madagascariensis*, *Alluaudia procera*) and Euphorbiaceae (*Euphorbia stenoclada*); and (2) inland spiny thickets, denser and more semi-deciduous, with greater contribution of Fabaceae. Population density is higher in the northern sector, along the national road RN7 corridor north of the Onilahy River. The main drivers of woody cover loss are charcoal production and slash-and-burn agriculture (hatsaky) (Brinkmann et al., 2014; Randriamalala et al., 2021, 2019). In the north-eastern sector, pasture fires are set during the dry season to stimulate grass regrowth in savanna areas (Brinkmann et al., 2014), with goats subsequently grazing on post-fire herbaceous regrowth. Extensive goat herding is otherwise widespread across the study area, grazing in xerophytic thickets without deliberate burning (Randriamalala et al., 2019). This ecological and human heterogeneity makes the study area a suitable testbed for evaluating woody cover dynamics across broad gradients.

7.4.2 Satellite data acquisition and preprocessing

7.4.2.1 Multi-sensor Landsat time series

The analysis draws on a multi-sensor Landsat time series covering 1989–2025, extracted from Collection 2 Level-2 products via Google Earth Engine (Cardille et al., 2024; Gorelick et al., 2017). These products provide surface reflectance following atmospheric correction by LEDAPS (Landsat 5 and 7) and LaSRC (Landsat 8 and 9) algorithms. Data were acquired successively from Landsat 5 TM (1989–2011), Landsat 7 ETM+ (2012 only, used as transition year), Landsat 8 OLI (2013–2021), and Landsat 9 OLI-2 (2022–2025).

The annual acquisition window spans January to June, corresponding to the transition between the wet and dry seasons. This period coincides with favorable phenology for woody cover observation: deciduous species retain most of their foliage, maximizing the spectral signal of woody biomass while limiting confusion with seasonal herbaceous vegetation.

For four years with excessive cloud cover within the standard January–June window (1992, 2000, 2009, and 2010), the median composite was calculated from all scenes available across the two adjacent years (year–1 and year+1), constituting a three-year pool. This approach maintains phenological consistency while maximizing scene availability. Similar multi-temporal compositing procedures have been applied in long-term Landsat studies facing image gaps (Hermosilla et al., 2015; White et al., 2014).

The years 2002 and 2003 could not be processed due to insufficient usable scenes after cloud filtering, compounded by the Scan Line Corrector failure of Landsat 7 ETM+ in May 2003 (Markham et al., 2004). No woody cover raster maps were therefore produced for these two years. For analyses requiring a continuous time series — class proportion plots and trend calculations — values for 2002 and 2003 were estimated by linear interpolation between 2001 and 2004 using the `na.approx` function from the `zoo` package (Zeileis and Grothendieck, 2005). Both years were excluded from all spatial analyses and deforestation rate calculations. The complete inventory of 1,527 scenes used is provided in Table S1.

7.4.2.2 Image preprocessing

Landsat images were processed through a standardized preprocessing chain including elimination of clouds, shadows, and saturated pixels using Collection 2 quality masks (QA_PIXEL and QA_RADSAT bands). Annual median composites were generated from all scenes available within the January–June window. This approach minimizes residual atmospheric effects and inter-scene variability while preserving the spectral signal of woody vegetation during the wet-to-dry season transition.

7.4.2.3 Spectral indices and inter-sensor harmonization

Ten spectral indices were computed to characterize structural and functional dimensions of vegetation: NDVI (Rouse et al., 1974), NDWI (Gao, 1996), SAVI

(Huete, 1988), EVI (Huete et al., 2002), NBR (Key and Benson, 2006), NDBI (Zha et al., 2003), NIR/Red ratio (Jordan, 1969), SWIR/NIR ratio, and the Brightness and Greenness components of the Tasseled Cap transformation (Crist and Cicone, 1984), capturing photosynthetic activity, water status, cover structure, and overall reflectance.

To ensure temporal consistency across sensor generations, a pseudo-band was generated for Landsat 5 and 7 by multiplying the SWIR1 band (band 5, 1.55–1.75 μm) by an empirical adjustment factor of 0.95 to approximate the SWIR2 band (band 6, 2.09–2.35 μm) present on Landsat 8 and 9 (Roy et al., 2016). This adjustment produced a homogeneous predictor space of 17 variables (7 spectral bands + 10 indices) across the full 1989–2025 time series.

7.4.3 Calibration and spatial cross-validation data

A total of 505 plots of 50 \times 50 m were established by photointerpretation of very high-resolution imagery acquired between 2021 and 2022 using the Collect Earth Online platform (Bey et al., 2016), following the sampling design and photointerpretation protocol described in Ramalason et al., (2026). These plots served two distinct purposes: calibration of the Random Forest model for the 2022 reference year, and spatial block cross-validation (blockCV) to assess the geographic generalization capacity of the model across the study area (Valavi et al., 2018).

Sampling followed a systematic grid with one randomly placed point per cell, designed to cover the full woody cover gradient. Environmental variables — topography (elevation from NASADEM), climatic gradient (mean annual precipitation), and vegetation indices (NDVI, EVI, NDWI) — were displayed as auxiliary contextual layers in Collect Earth Online to guide visual assessment during photointerpretation (Colditz et al., 2011).

Within each plot, 49 observation points were arranged on a regular 7 \times 7 grid (approximately 7-m spacing). Presence or absence of woody vegetation was assessed at each point by vertical interpretation using biophysical criteria identical to those applied in field surveys. Woody cover per plot was calculated as the proportion of points classified as woody.

7.4.4 Field validation data

Field data collected between April and May 2022 served as independent external validation of the model. Stratified random sampling guided the establishment of 41 plots of 30 \times 30 m, distributed proportionally across strata defined by five woody cover classes (20% intervals) and four vegetation formation types. Sites were selected to ensure structural homogeneity within a 200-m radius, limiting the effect of sub-pixel spatial variability on predicted-observed comparisons.

Woody cover was measured using the point intercept method following the Australian Terrestrial Ecosystem Research Network protocol (Karl et al., 2017; White et al., 2012). Within each plot, 54 sampling points were placed along four orthogonal transects spaced 10 m apart, with one point every 2 m. Presence of woody vegetation

taller than 1 m was recorded by vertical projection. Woody cover was calculated as the ratio of positive points to total points observed.

7.4.5 Random Forest woody cover modelling

Woody cover estimation was performed using the Random Forest algorithm (Belgiu and Drăguț, 2016; Breiman, 2001) implemented in regression mode via Google Earth Engine. Model calibration drew on 505 photo-interpreted plots for the 2022 reference year. The predictor space included seven Landsat surface reflectance bands and ten derived spectral indices, totalling 17 variables.

The response variable was square-root transformed prior to training to stabilize variance and improve predictive performance at extreme values of the cover gradient. Predictions were back-transformed by squaring and constrained to the [0–100%] interval. Retained hyperparameters were: 750 trees, 7 candidate variables per node, minimum leaf size of 7, bootstrap fraction of 0.632, and random seed fixed at 123.

Model performance was assessed through two complementary approaches. External validation relied on regression between predictions and observations from 41 independent field plots, computing R^2 , RMSE, MAE and bias. Spatial block cross-validation (blockCV; Valavi et al., 2018) was applied to 505 photo-interpreted plots, defining 10-km spatial blocks and 5 folds. This approach tests geographic generalization capacity by ensuring that training and test data are spatially separated, providing a more realistic performance estimate outside sampled areas.

The model calibrated on 2022 was subsequently applied to the full Landsat time series (1989–2025), producing 35 annual woody cover maps at 30-m spatial resolution.

7.4.6 Woody cover change analysis






Annual rates of woody cover change were calculated using the standardized FAO method (Puyravaud, 2003). Woody surface area was defined by a minimum threshold of 10 % cover, corresponding to the boundary between non-woody formations (0–10 %) and woody formations (> 10 %), consistent with definitions used for open dryland formations (Bastin et al., 2017; Brandt et al., 2020). Rates were computed for the full 1989–2025 period and for four sub-periods corresponding to the temporal phases identified in the time series: 1989–2001, 2001–2013, 2013–2019, and 2019–2025.

7.4.7 Classification of ecological degradation states

Continuous woody cover maps were reclassified into five discrete ecological states representing a degradation gradient (Tableau 8.1). This classification enables analysis of the temporal evolution of areas occupied by each state between 1989 and 2025 and quantification of cover class transitions over the study period.

Tableau 7-1: Woody cover classification system along an ecological degradation gradient.

Continuous woody cover maps were reclassified into five discrete classes representing distinct ecological states, validated against field measurements and very high-resolution satellite image photointerpretation across 41 plots

Cover class	Interval (%)	Ecological state	Characterization	Field photo
Non-woody areas	0–10	Herbaceous formations	Bare soil, herbaceous vegetation	
Very degraded cover	10–40	Sparse thicket	Scattered shrubs	
Moderately degraded cover	40–60	Open anthropized thicket	Heterogeneous stand	
Degraded thickets	60–80	Dense secondary thicket	Advanced regeneration	
Intact thickets	80–100	Intact xerophytic thicket	Original structure preserved	

7.4.8 Anthropogenic drivers of woody cover dynamics

Spearman correlation analyses were conducted to quantify associations between anthropogenic disturbances and woody cover dynamics.

7.4.8.1 Fire regimes

Burned areas were mapped annually from the MODIS MCD64A1 v6.1 product at 500-m resolution (Giglio et al., 2018; Bastin et al., 2025), extracted via Google Earth Engine. The fire regime analysis covers 2000–2025, corresponding to product availability. Before 2000, no satellite burned area product combines comparable spatial resolution with monthly temporal coverage, making coherent fire analysis impossible for the earlier period. Each pixel was coded as fire presence or absence (criterion: $\text{BurnDate} > 0$). Three variables were computed: cumulative fire frequency (number of fire years per pixel over 2000–2025), time since last fire (YSF, years since last detected fire event), and burned/unburned status (at least one fire detected versus none over the period).

A random sample of 10,000 points was drawn across the study area as a compromise between spatial representativeness and computational feasibility. For each point, cumulative fire frequency, YSF, and woody cover change (2000–2025) were extracted. After excluding points with incomplete data, 1,364 points were retained for analysis. The high exclusion rate primarily reflects the limited spatial overlap between burned pixels and valid woody cover observations, as fire events are spatially concentrated in the northeastern sector. Fire impacts were assessed through three complementary approaches: a Mann–Whitney test comparing burned and unburned areas, Spearman correlations between fire frequency, YSF, and woody cover change, and a degree-2 polynomial regression to identify the inflection point of the frequency–cover change relationship.

7.4.8.2 Proximity to built-up areas

Building locations were derived from the Google Open Buildings v3 database (Sirko et al., 2021), extracted via Google Earth Engine. This database identifies buildings detected by high-resolution satellite imagery and represents the most comprehensive and up-to-date source for mapping human presence in rural Madagascar. Concentric buffer zones from 1 to 10 km (1-km steps) were delineated around built-up area centroids derived from aggregated building footprints, defining ten distance rings to characterize the human influence gradient on woody cover. Within each ring, woody surface area (ha, threshold $>10\%$) was calculated for 1989 and 2025, and the annual rate of woody cover change was derived using the FAO method (Puyravaud, 2003) on these surfaces. Spearman correlations ($n = 10$ zones) quantified associations between distance to buildings and woody cover dynamics.

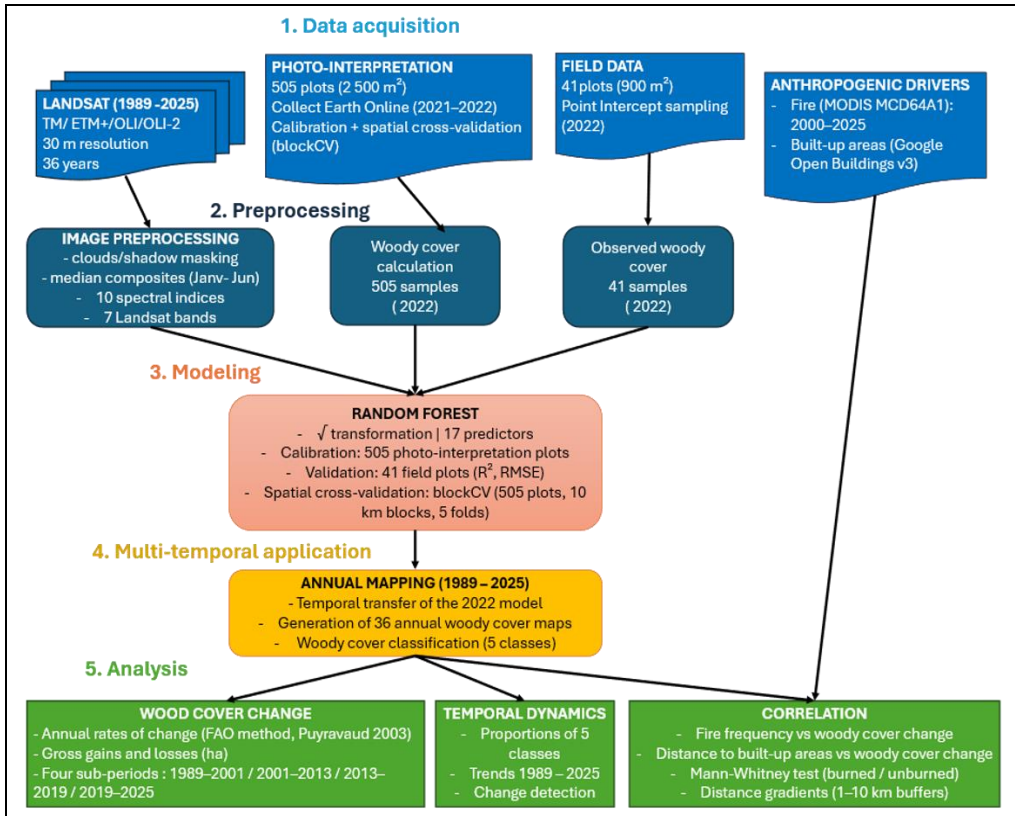


Figure 7-1: Methodological workflow for woody cover mapping and degradation analysis. The approach combines photointerpretation (n=505) and field validation (n=41) to calibrate a Random Forest model subsequently applied to the full Landsat time series (1989-2025). Temporal dynamics of woody cover are analyzed in relation to fire regimes and proximity to human built up areas.

7.5. Results

7.5.1 Woody cover modelling and mapping

External validation on 41 independent field plots indicated strong predictive performance: $R^2 = 0.751$, $RMSE = 13.74\%$, $MAE = 11.01\%$, and $bias = -0.21\%$, indicating negligible systematic error. Spatial block cross-validation (blockCV, 10-km blocks, 5 folds) on 505 photo-interpreted plots yielded $R^2 = 0.794$, $RMSE = 16.50\%$, $MAE = 12.20\%$, and $bias = -3.74\%$. Inter-fold R^2 ranged from 0.706 to 0.870 and $RMSE$ from 13.48% to 19.19%, indicating stable performance across spatial blocks. Predicted vs. observed scatterplots for both validation approaches are provided in Fig. S1. These results are consistent with the performance range reported for the same study area using Sentinel-1+2 and photointerpretation calibration ($R^2 = 0.85-0.88$; Ramalason et al., 2026).

The model was subsequently applied to the full Landsat time series (1989–2025), producing 35 annual woody cover maps at 30-m resolution, reclassified into five ecological degradation states (Tableau 8.1)

7.5.2 Temporal phases of woody cover loss and recovery

Over the full 1989–2025 period, total woody surface area declined from 477,109 ha to 461,924 ha, a net loss of 15,185 ha (-3.2%), corresponding to an annual rate of $-0.090\% \text{ yr}^{-1}$ (Tableau 8.2). Four distinct phases characterize this trajectory. A slight expansion was recorded during 1989–2001 ($+0.055\% \text{ yr}^{-1}$, +3,134 ha), consistent with the visual stability of intact thicket proportions observed in Fig. 3A. Accelerated loss followed during 2001–2013 ($-0.376\% \text{ yr}^{-1}$, -21,214 ha), representing the most intensive phase of woody cover decline over the study period. A continued but attenuated decline persisted during 2013–2019 ($-0.200\% \text{ yr}^{-1}$, -5,466 ha), before partial recovery emerged over 2019–2025 ($+0.304\% \text{ yr}^{-1}$, +8,362 ha). A robust (Theil–Sen) regression fitted on the annual values within each phase indicates that none of the four trends differs significantly from zero at the 95% confidence level, and a Chow test applied at the 2001, 2013 and 2019 boundaries does not detect statistically significant structural breaks ($p = 0.24, 0.085$ and 0.13 , respectively).

To assess whether this net loss could reflect measurement uncertainty around the 10 % threshold rather than genuine degradation, a Monte Carlo sensitivity analysis was performed: independent Gaussian noise ($\text{sd} = \text{RMSE} = 13.74\%$, from the external validation reported above) was added to each pixel's predicted cover, clipped to $[0, 100]$, reclassified at the 10 % threshold, and the resulting 1989–2025 loss recomputed over 1,000 simulations. The observed loss (15,185 ha) falls outside the 95 % envelope of losses expected under measurement noise alone (13,610–13,748 ha; mean 13,680 ha), indicating that the decline reflects a genuine ecological signal rather than an artefact of threshold-crossing noise near the 10 % boundary. This procedure assumes spatially independent pixel-level errors and therefore likely underestimates the true uncertainty, since remote-sensing classification errors are typically spatially autocorrelated; the resulting confidence intervals should be interpreted as a conservative lower bound on uncertainty rather than an exhaustive measure of it.

Tableau 7-2: Annual rates of woody cover change by phase (1989–2025). A₁ and A₂ denote woody surface area (ha) at the beginning and end of each phase, respectively.

Woody surface area corresponds to pixels with woody cover exceeding 10%.

Annual rates were calculated following Puyravaud (2003): $r = (1/\Delta t) \times \ln(A_2/A_1) \times 100$. Years 2002–2003 were excluded due to insufficient Landsat scene availability (Scan Line Corrector failure).

Phase	Period	Surface A ₁ (ha)	Surface A ₂ (ha)	Net change (ha)	Net change (%)	Annual rate (% yr ⁻¹)
Expansion	1989–2001	477 109	480 243	+3 134	+0.66	+0.055
Accelerated loss	2001–2013	480 243	459 028	-21 214	-4.42	-0.376
Continued decline	2013–2019	459 028	453 562	-5 466	-1.19	-0.200
Partial recovery	2019–2025	453 562	461 924	+8 362	+1.84	+0.304
Overall	1989–2025	477 109	461 924	-15 185	-3.18	-0.090

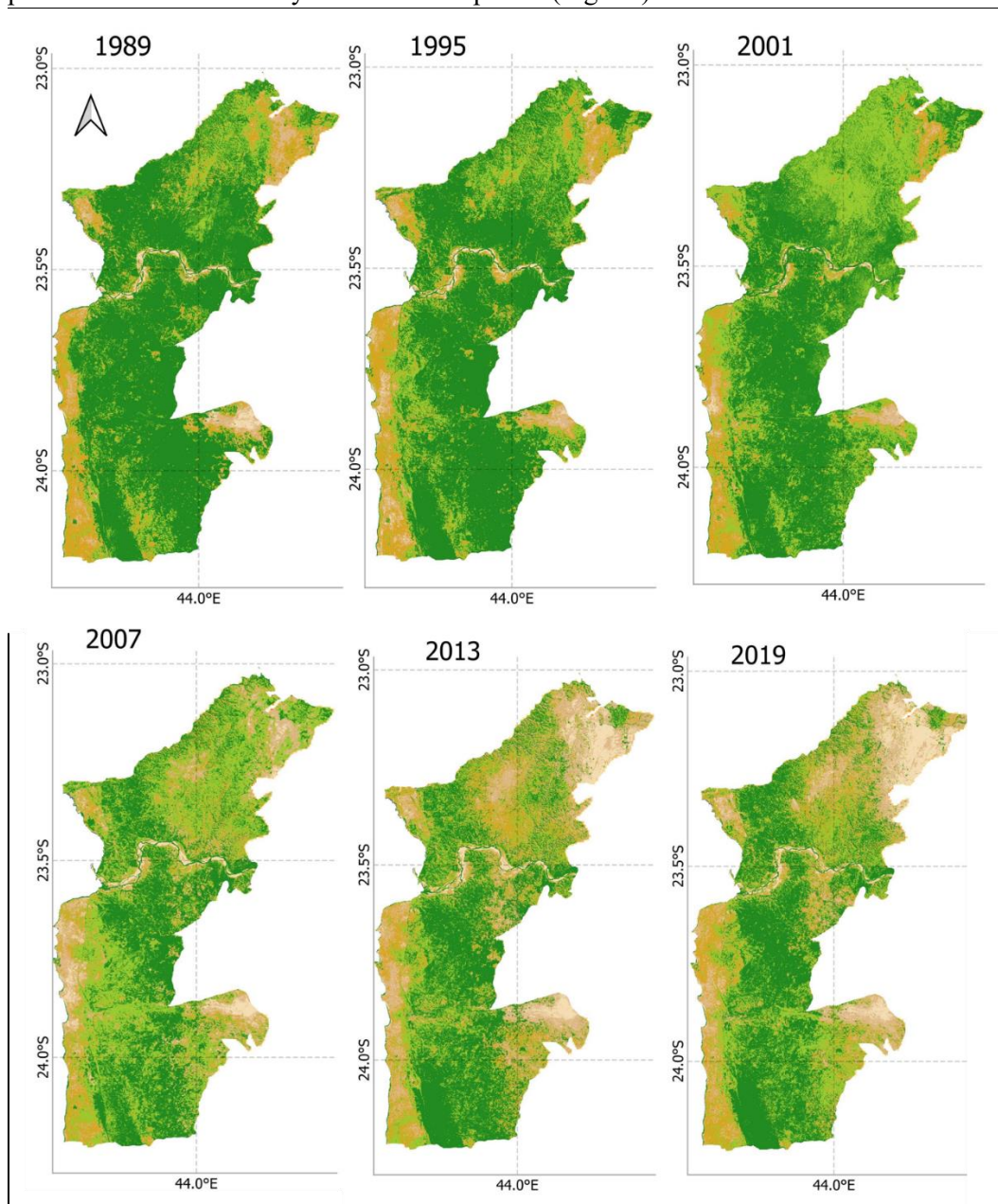
7.5.3 Long-term woody cover dynamics

Diachronic mapping reveals substantial landscape transformation over 36 years (Fig. 8.2, 8.3). Intact thickets (> 80 % cover) declined from 62.8 % of total surface area in 1989 to a minimum of approximately 10 % around 2011, before partially recovering to 29.0 % by 2025 (Fig. 8.3A). Degradation was spatially concentrated in the northeastern sector during 1989–2013, with a similar dynamic emerging in the eastern and southeastern sectors in more recent years, while the southwestern sector remained comparatively less affected (Fig. 8.2).

Degraded thickets (60–80 % cover) increased from 18.5 % to 39.5 %, becoming the dominant landscape component after 2013. Moderately degraded cover (40–60 %) peaked around 2011–2013 before partially declining. Non-woody areas and very degraded cover remained minor throughout, together representing less than 20 % of total area (Fig. 8.3A, 8.3B).

The most pronounced changes occurred during 2001–2013, when intact thickets lost 43 percentage points while degraded classes simultaneously expanded — a shift clearly visible in the crossing of the two dominant curves in Fig. 8.3A. The net change map (Δ woody cover 1989–2025) confirms dominant losses in the northeast and localized gains in central and southern sectors (Fig. 8.2). A progressive slowdown followed 2013, with partial recovery visible from 2019 onward (Fig. 8.3A, 8.3B). Cover class transition analysis (2000–2025) further confirms the structural nature of this degradation: 41.5 % of pixels classified as intact thickets (> 80 %) in 2000

transitioned to degraded thickets (60–80 %) by 2025, while 51.6 % of non-woody pixels showed no recovery over the same period (Fig. S2).



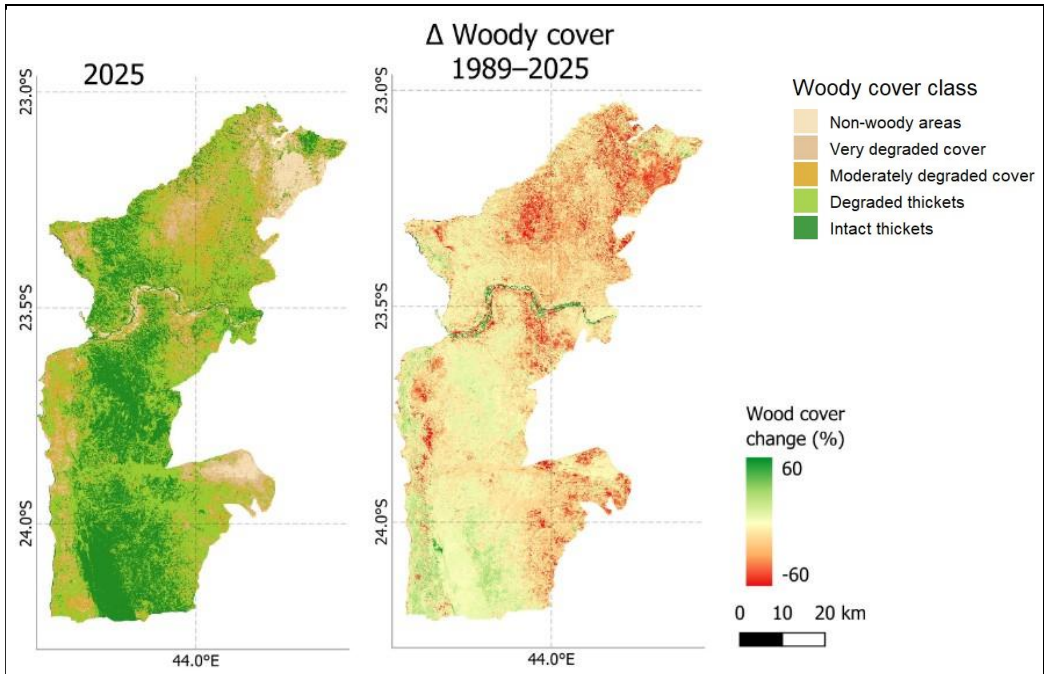
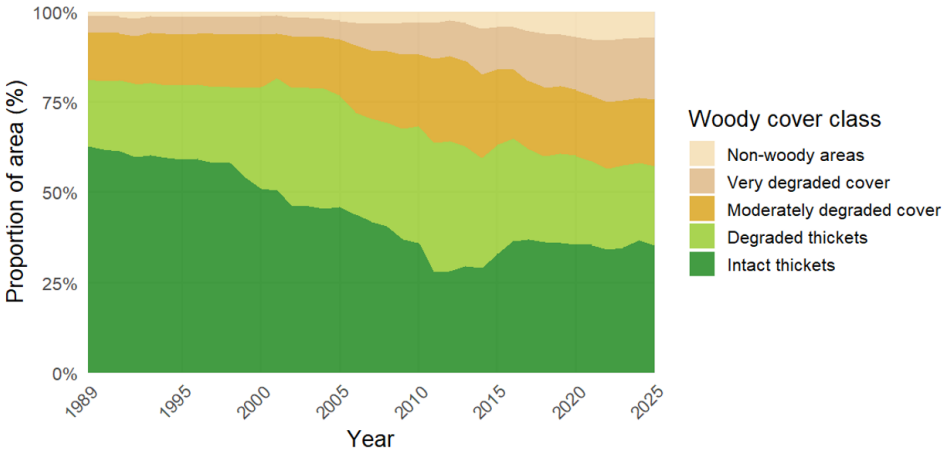


Figure 7-2 : Spatial distribution of woody cover classes at six selected time steps and net woody cover change (1989–2025). Classification maps illustrate the progressive expansion of degraded classes (beige to orange tones) and the contraction of intact woody thickets (dark green) over the 36-year period, most pronounced in the northern and eastern parts of the study area. The change map (Δ woody cover, 1989–2025) quantifies pixel-level variation in woody cover percentage; red tones indicate loss and green tones indicate gain

A. Trends in woody cover class proportions at six-year intervals (1989–2025)



B. Heatmap of Woody Cover Distribution

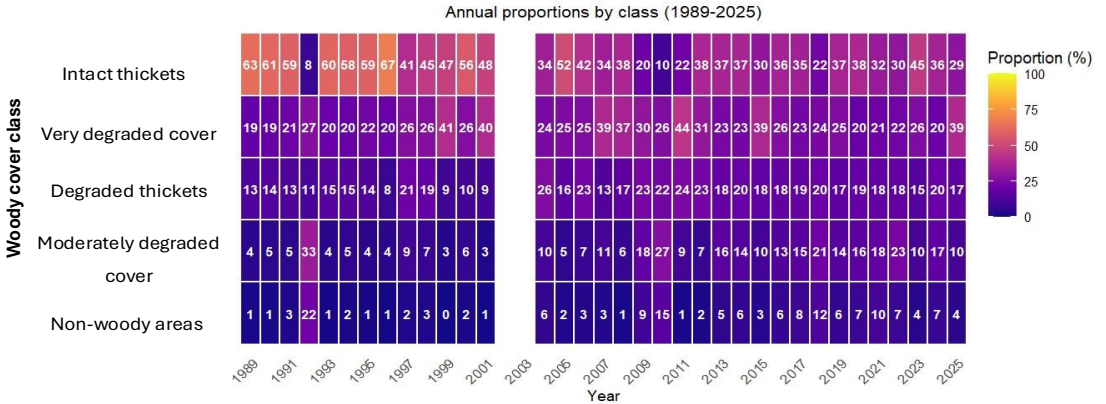


Figure 7-3 : Temporal dynamics of woody cover class proportions (1989–2025). (A) Six-year interval trends in the relative contribution of five woody cover classes to total landscape area. (B) Heatmap of annual proportions by class; cell values indicate percentage of total area. Both panels show a marked decline of intact thickets between 2001 and 2013, followed by partial stabilization. The 2003 data gap reflects insufficient image availability. Values for 1992, 2000, 2009–2010, and post-2022 may be affected by phenological bias due to image compositing outside the standard January–June window.

7.5.4 Impacts of fire regimes on woody cover

Mapping of cumulative fire frequency (2000–2025) reveals a marked spatial concentration of fires in the northeastern sector of the study area, with maximum frequency reaching 20 events per pixel, while the southern and western sectors remained largely unburned (Fig. 8.4A). This spatial pattern closely mirrors the distribution of woody cover losses, with the highest-frequency fire zones corresponding to the areas of most severe cover decline (Fig. 8.4B).

Comparison between burned and unburned pixels confirms a strong fire-driven cover loss signal (Fig. 8.5B). Burned pixels ($n = 4,112$) showed a median woody cover loss of -18.0% versus -0.1% for unburned pixels ($n = 10,888$), a highly significant difference (Mann–Whitney $W = 10,327,294$; $p < 2.2 \times 10^{-16}$; Cohen's $d = -0.967$). Significant spatial autocorrelation in residuals (Moran's $I = 0.330$; $p < 2.2 \times 10^{-16}$) indicates that the spatial clustering of fire events likely contributes to the strength of this association.

Using a second-degree polynomial regression to model the frequency–cover change relationship, the analysis identifies an inflection point at 8.9 cumulative fire events beyond which woody cover loss accelerates in a nonlinear manner, indicating that the resilience of xerophytic thickets is exceeded (Fig. 8.5A; $R^2 = 0.153$; $n = 15,000$; $p < 2.2 \times 10^{-16}$). Expressed as a fire-return interval rather than as a cumulative event count, this threshold over the 25-year fire record (2000–2025) corresponds approximately to a limiting frequency of one fire every three years, beyond which the resilience capacity of the vegetation is progressively exhausted. Below this threshold, losses remained moderate and relatively stable across fire frequency classes. Spearman correlations confirmed a negative dose–response relationship between fire frequency and woody cover change ($\rho = -0.124$; $p = 4.1 \times 10^{-6}$), and a positive association between time since last fire and woody cover change ($\rho = +0.158$; $p = 4.3 \times 10^{-9}$).

Post-fire recovery trajectories indicate slow and partial regeneration (Fig. 8.5C). Pixels burned within the past 0–6 years showed median losses between -21.0% and -24.6% , with no detectable short-term recovery signal. A significant reduction in losses was only observable beyond 10 fire-free years (median = -12.4% ; Kruskal–Wallis $\chi^2 = 1,155.8$; $df = 4$; $p < 2.2 \times 10^{-16}$). Linear regression estimated a recovery rate of $+0.49\%$ cover per fire-free year, though the low explanatory power of this relationship ($R^2 = 0.042$) reflects the high variability in individual pixel trajectories (Fig. 8.5C).

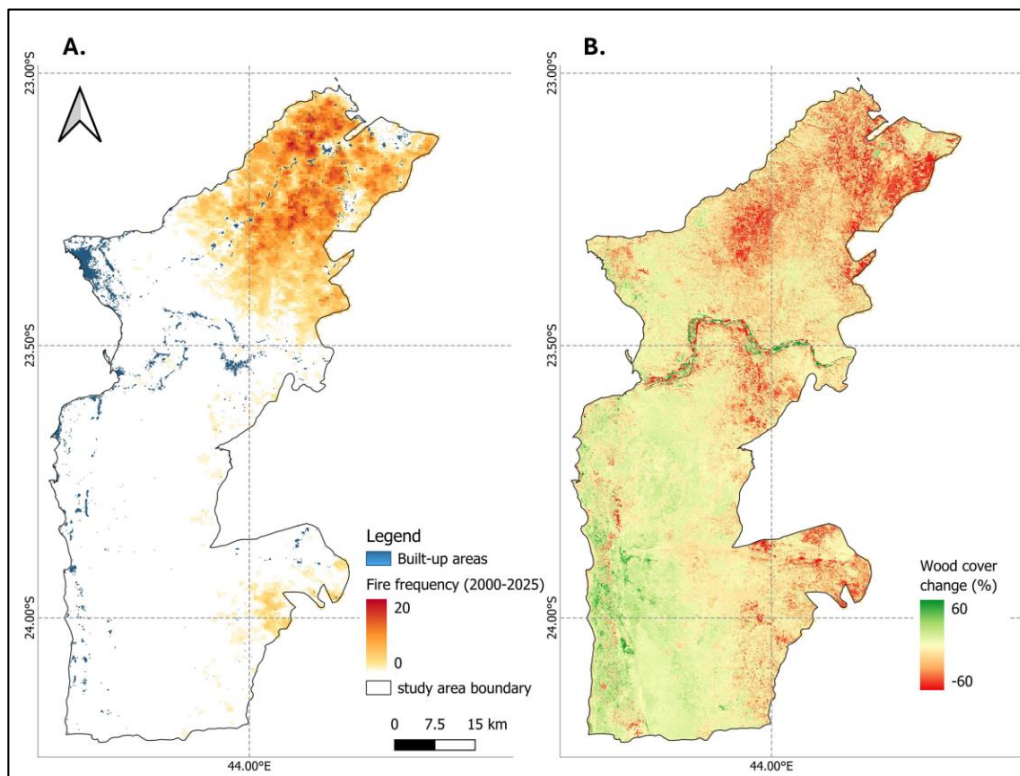


Figure 7-4: Spatial patterns of fire frequency and woody cover change (2000–2025). (A) Cumulative fire frequency derived from the MODIS MCD64A1 product (Bastin et al., 2025; Giglio et al., 2018); built-up areas (Google Open Buildings v3) are shown in blue. (B) Net woody cover change (2000–2025); red tones indicate loss and green tones indicate gain. The spatial co-occurrence of high fire frequency zones (northern and central-eastern sectors) and areas of severe woody cover loss confirms the spatial association between fire regime and vegetation degradation.

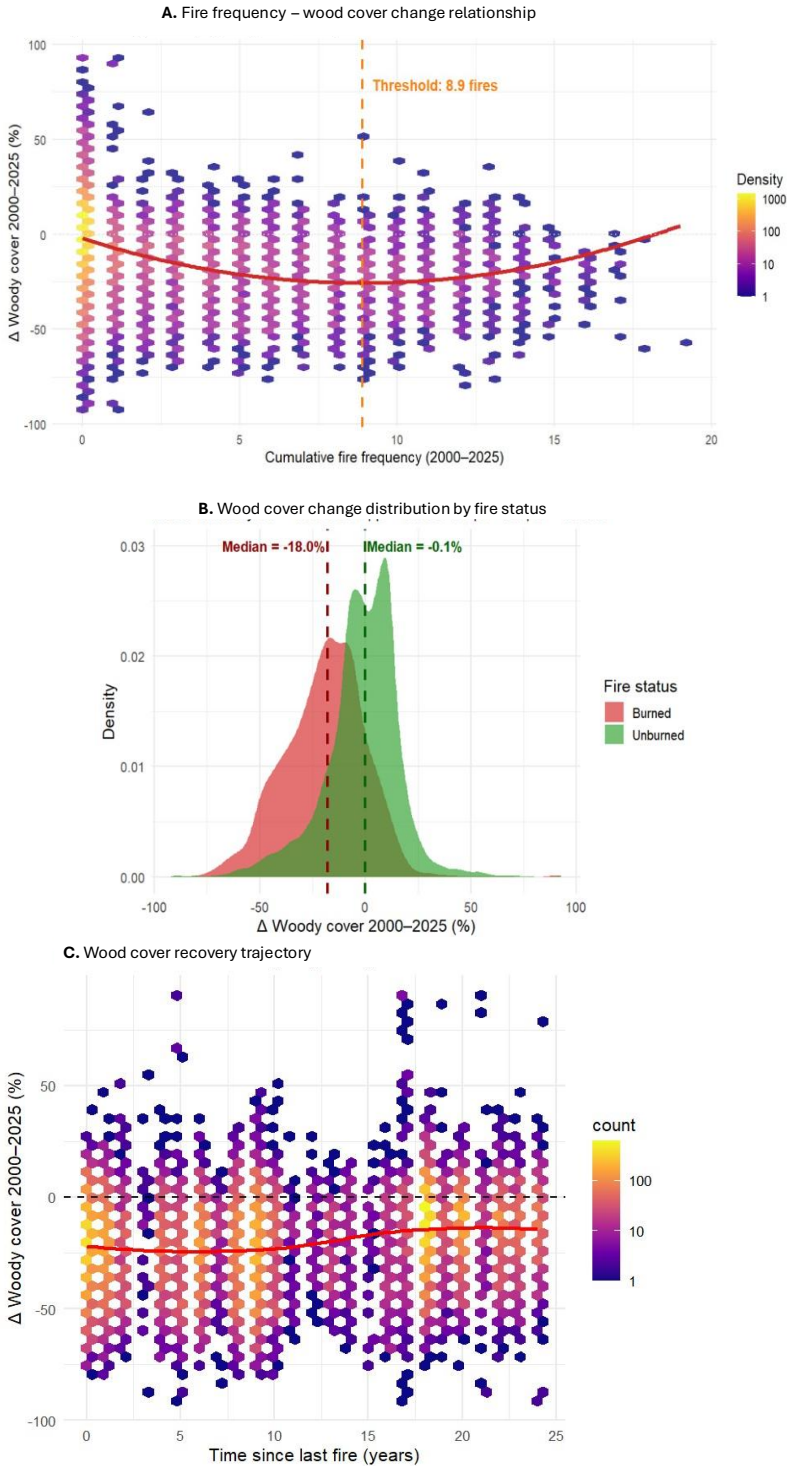


Figure 7-5: Relationship between fire regime and woody cover change (2000–2025). (A) Hexagonal density plot of cumulative fire frequency versus Δ woody cover; the polynomial regression curve (red) identifies a threshold at 8.9 fire events, corresponding to the inflection point beyond which degradation accelerates nonlinearly. (B) Density distribution of Δ woody cover for burned (median = -18.0%) and unburned (median = -0.1%) pixels, illustrating the magnitude of fire-driven loss. (C) Woody cover recovery trajectory as a function of time since last fire (years); the regression line shows no significant recovery trend over the observation period.

7.5.5 Distance gradient from built-up areas

The concentric buffer analysis reveals a monotonically decreasing gradient in woody cover loss with increasing distance from built-up areas (Fig. 6, Table 3). Annual rates of woody cover change declined from $-0.166\% \text{ yr}^{-1}$ at 1 km to $-0.102\% \text{ yr}^{-1}$ at 5 km and $-0.090\% \text{ yr}^{-1}$ at 10 km, while net losses followed the same pattern, from 5.81 % at 1 km to 3.18 % at 10 km over 1989–2025.

Mean woody cover increased rapidly from approximately 57 % at 1 km to 65 % at 4–5 km, then stabilised toward an asymptote of approximately 67 % beyond 6–7 km, as illustrated by the fitted LOESS curve (Fig. 6A). All buffer zones fall within the degraded (60–80 %) and moderately degraded (40–60 %) cover classes, with the very degraded class (10–40 %) restricted to the intensive zone (≤ 2 km), confirming that no intact thickets ($> 80\%$) persist within 10 km of built-up areas (Fig. 6B). This profile is consistent with a nonlinearly decreasing proximity effect, characteristic of woody resource extraction concentrated within a daily access radius. Three intensity zones characterize this gradient: intensive (≤ 2 km, $-0.161\% \text{ yr}^{-1}$), moderate (3–6 km, $-0.115\% \text{ yr}^{-1}$), and peripheral (> 6 km, $-0.092\% \text{ yr}^{-1}$), confirming a progressive decrease in disturbance intensity with distance from settlements (Table 3). Spearman correlation confirms this gradient quantitatively: net woody cover change (1989–2025) was strongly and positively correlated with distance from built-up areas ($\rho = +0.985$; $p = 2.3 \times 10^{-7}$).

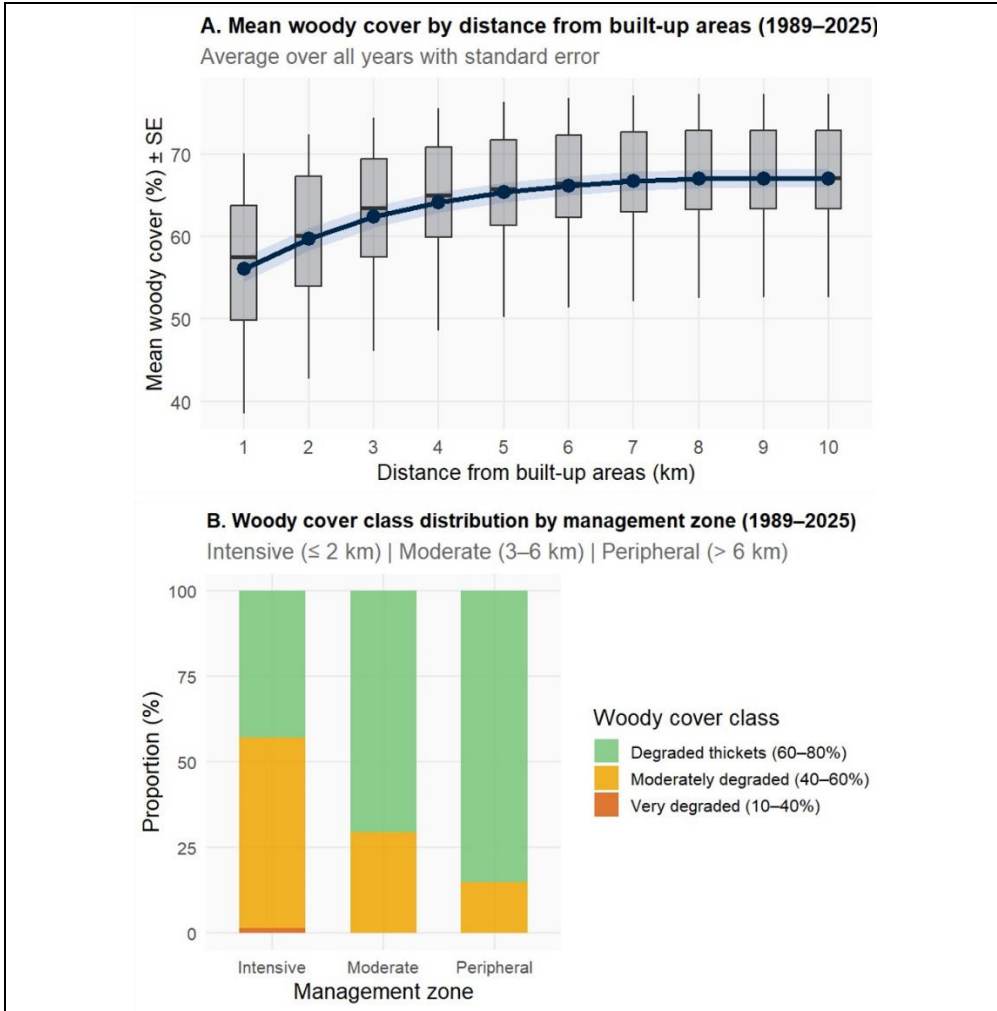


Figure 7-6: Woody cover distribution by distance from built-up areas (1989–2025). A. Mean woody cover (%) averaged over all years within each 1-km concentric buffer zone; error bars indicate \pm standard error across years. The fitted curve (LOESS regression, shaded band = 95 % confidence interval) shows a rapid increase from 57 % at 1 km to approximately 67 % beyond 6–7 km. Annual rates of woody cover change (FAO method) range from -0.166 % yr^{-1} at 1 km to -0.090 % yr^{-1} at 10 km (Table 3). B. Proportion of woody cover classes per management zone averaged over 1989–2025. No intact thickets (> 80 %) were detected within 10 km of built-up areas. Built-up area locations derived from Google Open Buildings v3 (Sirko et al., 2021).

Tableau 7-3: Net woody cover change by distance class from built-up areas (1989–2025). Annual deforestation rate calculated using the FAO method (Puyravaud, 2003). Intensity zones are defined as follows: intensive (≤ 2 km), moderate (3–6 km), peripheral (>6 km).

Distance (km)	Zone	Surface 1989 (ha)	Surface 2025 (ha)	Loss (ha)	Loss (%)	Rate (% yr ⁻¹)
1	Intensive	142,550	134,268	8,282	5.81	-0.166
2	Intensive	256,008	242,003	14,006	5.47	-0.156
3	Moderate	333,687	318,466	15,221	4.56	-0.130
4	Moderate	383,656	368,447	15,209	3.96	-0.112
5	Moderate	419,675	404,491	15,184	3.62	-0.102
6	Moderate	445,425	430,241	15,184	3.41	-0.096
7	Peripheral	464,799	449,615	15,185	3.27	-0.092
8	Peripheral	474,500	459,315	15,185	3.20	-0.090
9	Peripheral	476,987	461,802	15,185	3.18	-0.090
10	Peripheral	477,109	461,924	15,185	3.18	-0.090

7.6. Discussion

7.6.1 Nonlinear degradation trajectories over 36 years

The 36-year annual reconstruction revealed four distinct phases of woody cover dynamics that would remain undetected using conventional bi-temporal approaches (DeVries et al., 2015; Kennedy et al., 2010). The slight expansion recorded during 1989–2001 ($+0.055$ % yr⁻¹) suggests relative landscape stability during this period, consistent with limited anthropogenic pressure documented across southern Madagascar before the political crises of the 2000s (Vieilledent et al., 2018). The subsequent phase of accelerated loss during 2001–2013 (-0.376 % yr⁻¹) represents the most intensive degradation episode of the study period, temporally coinciding with Madagascar's political crises of 2002 and 2009 — periods previously associated with weakened forest governance and accelerated deforestation (ONE et al., 2015; Vieilledent et al., 2018), though our observational framework does not allow formal causal attribution. The attenuation of losses during 2013–2019 (-0.200 % yr⁻¹) and the partial recovery observed over 2019–2025 ($+0.304$ % yr⁻¹) suggest a gradual reduction in disturbance pressure, possibly linked to reduced fire activity in certain sectors, consistent with the recovery trajectories identified in section 3.4.

The principal finding of this trajectory analysis is that a limited net area loss (−3.2 % over 36 years) conceals a profound structural transformation: intact thickets declined from 62.8 % to 29.0 % of the landscape while degraded thickets increased from 18.5 % to 39.5 %. This transition from intact to degraded thickets is consistent with nonlinear state changes described in disturbance-prone dryland ecosystems, where gradual cover loss can trigger abrupt structural reorganization without proportional area change (Archibald et al., 2013). Cover class transition analysis further confirms this structural shift: 41.5 % of pixels classified as intact thickets in 2000 transitioned to degraded thickets by 2025, while 51.6 % of non-woody pixels showed no recovery over the same period (Fig. S2), suggesting limited reversibility in the most degraded sectors. However, it should be acknowledged that low woody cover values do not necessarily indicate anthropogenic degradation. In the study area, calcareous substrates (calcsols) support structurally open formations with large emergent individuals and high floristic richness that may exhibit cover values below 60 % under undisturbed conditions (Randriamalala et al., 2019). Vertical structure and floristic composition are therefore complementary indicators that continuous cover monitoring alone cannot fully capture (Maestre et al., 2016). This divergence between net area loss and structural degradation is precisely the dimension of ecosystem change that binary forest/non-forest classifications cannot capture, confirming the analytical value of continuous cover monitoring in dryland systems (Bastin et al., 2017; Brandt et al., 2020). The observed global loss rate (−0.090 % yr^{−1}) is substantially lower than the national deforestation rate for Madagascar (1.1 % yr^{−1} during 2010–2014; Vieilledent et al., 2018), which aggregates all forest types (moist, dry, spiny, and mangroves) at the national scale. Differences in spatial scope, forest definition, and methodology therefore limit direct comparison.

7.6.2 Fire as driver of structural degradation

The spatial concentration of fires in the northern and northeastern sectors primarily reflects two anthropogenic drivers: slash-and-burn agriculture (*hatsaky*) and charcoal production. The Belomotse Plateau (National Road RN7) was the main charcoal production site supplying Toliara during the 1990s, with woody resources so depleted that shrub roots were excavated for charcoal biomass (Randriamalala et al., 2021). Slash-and-burn agriculture is the primary proximate driver of deforestation in southwestern Madagascar (Brinkmann et al., 2014; Randriamalala et al., 2019). In the northern sector, pasture fires are also deliberately set during the dry season to stimulate grass regrowth in savanna areas (Brinkmann et al., 2014), with goats subsequently grazing on post-fire herbaceous regrowth. The strong contrast between burned and unburned pixels — median losses of −18.0 % versus −0.1 % (Mann–Whitney, $p < 2.2 \times 10^{-16}$; Cohen's $d = -0.967$) — quantifies the cumulative impact of these practices on woody cover at landscape scale (Smit et al., 2010; Wigley et al., 2009).

Fire frequency was highest in the degraded thickets class (60–80 % cover), with 39.4 % of pixels in this class having burned at least once over 2000–2025, compared to 21.2 % for intact thickets (> 80 %). This suggests that partially degraded formations

constitute an active degradation front where fire risk is disproportionately concentrated, rather than being uniformly distributed across the woody cover gradient.

The analysis identifies a threshold at 8.9 cumulative fire events beyond which woody cover loss accelerates in a nonlinear manner, indicating that the resilience of xerophytic thickets is exceeded (Aleman et al., 2016; Archibald et al., 2013). Below this threshold, losses remained moderate, suggesting that xerophytic thickets retain resilience to low-frequency fire disturbance, consistent with the adaptive traits of dominant succulent taxa from Didiereaceae and Euphorbiaceae families that characterize the spiny forest ecosystem of southwestern Madagascar (Antonelli et al., 2022). Beyond it, repeated burning progressively drives the system toward an alternative stable state dominated by herbaceous vegetation (Hirota et al., 2011; Staver et al., 2011).

Post-fire recovery was slow and incomplete, with no regeneration signal detectable within the first 10 fire-free years and a recovery rate of only $+0.49\% \text{ yr}^{-1}$ beyond this interval ($R^2 = 0.042$). This reflects the slow growth of shrub species (Gaspard et al., 2018), poor natural regeneration after clearing (Randriamalala et al., 2019), and water-limited recruitment dynamics in arid ecosystems (Maestre et al., 2016; Reynolds et al., 2007). Where fire recurrence reached up to 20 events over 25 years, it locally exceeds natural regeneration capacity, raising concerns about irreversible structural degradation in the most heavily burned areas (Bond and Keeley, 2005; Veldman et al., 2015).

7.6.3 Human proximity gradient and resource-use zonation

Annual woody cover change rates declined monotonically from $-0.166\% \text{ yr}^{-1}$ at 1 km to $-0.090\% \text{ yr}^{-1}$ at 10 km, with mean woody cover increasing from approximately 57 % to 67 % over the same distance range before stabilizing beyond 5–6 km (Fig. 6, Table 3). In the study area, this gradient reflects the combined effect of three distance-dependent anthropogenic pressures: charcoal production and slash-and-burn agriculture (*hatsaky*), most intense in the northern sector along the national road RN7 (Brinkmann et al., 2014; Randriamalala et al., 2021, 2019), and goat grazing, mainly concentrated within a 4–6 km radius around settlements. This pattern is consistent with concentric resource-use models in which extraction intensity declines nonlinearly with distance from settlements (Chomitz and Gray, 1996; Mertens and Lambin, 2000). Spearman correlation confirms this gradient quantitatively: net woody cover change (1989–2025) was strongly and positively correlated with distance from built-up areas ($\rho = +0.985$; $p = 2.3 \times 10^{-7}$; Table 5). The three intensity zones — intensive (≤ 2 km, $-0.161\% \text{ yr}^{-1}$), moderate (3–5 km, $-0.115\% \text{ yr}^{-1}$), and peripheral (6–10 km, $-0.092\% \text{ yr}^{-1}$) — provide a spatially explicit framework for conservation prioritization.

The cover class distribution by management zone further confirms this gradient: the intensive zone (≤ 2 km) concentrates the highest proportion of very degraded cover (10–40 %), while the peripheral zone (> 6 km) is dominated by degraded thickets (60–80 %) with virtually no very degraded pixels (Fig. 6B). Notably, intact thickets (> 80 %) are absent from all buffer zones within 10 km of built-up areas, indicating that

human proximity has effectively eliminated the highest-quality habitat class from the proximate landscape.

Goat pastoral lands are mainly located between 4 and 6 km from settlements depending on the season, corresponding to the moderate intensity zone (-0.115% yr^{-1}). Within this zone, fire frequency remained below the 8.9 cumulative event threshold and woody cover showed stable or slightly increasing trends, suggesting that low-intensity small ruminant grazing does not necessarily drive directional woody cover loss (Brandt et al., 2017; Dardel et al., 2014; Vetter, 2005). Beyond 6 km, the reduction of direct anthropogenic pressures contributes to the stabilization of woody cover observed in the peripheral zone (-0.092% yr^{-1}). Disentangling the respective contributions of rainfall variability, fire frequency, and livestock pressure to these local dynamics requires finer-resolution land use data. Fire frequency and proximity to settlements are themselves not fully independent — pastoral burning is concentrated near villages, where herders use fire to stimulate pasture regrowth — yet their effects on woody cover are analytically distinguishable here, as the fire-frequency threshold (8.9 events) and the distance gradient ($\rho = +0.985$) emerge from separate statistical tests applied to overlapping but not identical spatial subsets. A joint interaction-effects model (e.g., a generalized additive model with a fire frequency \times distance tensor term) would be needed to formally quantify any synergy between these two drivers, but this requires pixel-level covariate data beyond the scope of the present analysis.

7.6.4 Methodological advances and transferability

The continuous Random Forest approach developed here addresses a fundamental limitation of existing global monitoring products. The Tropical Moist Forest map (TMF; Vancutsem et al., 2021), designed for closed-canopy humid forests, applies canopy cover thresholds that systematically exclude open dryland formations: the majority of woody cover in the study area — including intact xerophytic thickets with 60–80 % cover — falls below TMF detection thresholds and is classified as non-forest. By estimating woody cover as a continuous variable across the full 0–100 % range, the present framework captures the gradual degradation spectrum that binary products cannot resolve, filling a monitoring gap documented across African drylands (Bastin et al., 2017; Brandt et al., 2020).

The performance advantage of locally calibrated models over globally generalized products is further illustrated by a comparison with Dynamic World (Brown et al., 2022), a globally trained deep-learning land cover classifier that yielded $R^2 = 0.70$ on the same 41 independent field plots (Ramalason et al., 2025b) — approximately 7 percentage points below the locally calibrated Random Forest developed here ($R^2 = 0.751$). This gap underscores the limitations of globally uniform training datasets in structurally heterogeneous dryland ecosystems, where spectral responses vary substantially across formation types and degradation gradients (Ramalason et al., 2026). Spatial block cross-validation (blockCV; $R^2 = 0.794$) further confirms that model performance is maintained under realistic geographic extrapolation conditions, a key requirement for reliable multi-decadal temporal transfer (Valavi et al., 2018).

The Collect Earth Online photo-interpretation workflow provides a scalable and cost-effective calibration alternative in regions where field access is logistically constrained (Bey et al., 2016; Ramalason et al., 2026). The 505 photo-interpreted plots covered the full woody cover gradient and enabled robust spatial generalization across structurally heterogeneous formations, with optimal sampling density empirically confirmed at 20.7–41.4 dispersed plots per 1,000 km² for this study area — the density applied in the present study. This approach is directly transferable to other data-scarce dryland regions facing similar monitoring challenges.

This study demonstrates that annual, multi-decadal woody cover monitoring is operationally feasible in tropical drylands using Landsat imagery processed in Google Earth Engine (Gorelick et al., 2017).

7.6.5 Methodological limitations and perspectives

Several limitations should be acknowledged. The absence of data for 2002–2003 and the adjustment of acquisition windows in cloud-prone years (1992, 2000, 2009, 2010) may introduce temporal discontinuities and potential phenological artefacts, limiting interpretation of fine interannual fluctuations (Markham et al., 2004). The 30-m Landsat resolution may underestimate woody cover in highly open or fine-structured formations; future integration of Sentinel-2 or radar data could improve structural sensitivity and reduce interannual phenological noise (Reiche et al., 2018). Ramalason et al., (2026) demonstrated on the same study area that Sentinel-1+2 fusion achieved $R^2 = 0.85\text{--}0.88$ with adequate photointerpretation calibration, confirming the substantial gain in structural sensitivity offered by optical-radar complementarity over single-sensor Landsat approaches (Reiche et al., 2018).

Statistical analyses establish robust associations between woody cover dynamics, fire regime, and human proximity, but do not formally demonstrate causality. Future work should integrate population density as an explicit explanatory variable, given that human pressure gradients are known to scale nonlinearly with demographic density in dryland contexts (Brinkmann et al., 2014; Laurance et al., 2002). Additional variables — including livestock density, interannual rainfall variability, and charcoal extraction intensity — would further disentangle their respective contributions to observed cover dynamics (Noromiarilanto et al., 2016; Sedano et al., 2016).

The continuous cover framework developed here captures cover quantity but not cover quality. Future integration of canopy height data (e.g., GEDI lidar; (Dubayah et al., 2020) and floristic composition surveys would enable distinction between structurally open formations on calcareous substrates — which may be naturally open — and anthropogenically degraded formations, strengthening the ecological interpretation of cover dynamics.

The identified thresholds — 8.9 cumulative fire events and the 5–6 km stabilization distance — should be considered context-specific and require regional recalibration before transfer to other dryland systems (Archibald et al., 2013).

7.6.6 Implications for management and conservation

The fire threshold of 8.9 cumulative events provides a practical prioritization criterion: areas approaching this threshold retain passive recovery potential, whereas sectors already exceeding it likely require active restoration. In the northern sector, where pasture fires are deliberately set in savanna areas to stimulate grass regrowth (Brinkmann et al., 2014), agreements with local herding communities on fire frequency and timing could help maintain fire occurrence below the critical threshold and reduce degradation pressure in this zone (Crocker et al., 2023; Phelps et al., 2022). More broadly, long-term conservation of xerophytic thickets requires reducing community dependence on forest resources — primarily charcoal production and slash-and-burn agriculture — through viable alternative livelihoods. Improving goat husbandry and promoting income-generating activities that outcompete forest exploitation are more effective long-term strategies than restoration alone (Randriamalala et al., 2019).

The distance gradient supports spatially differentiated management. Within the intensive zone (≤ 2 km), community fuelwood plantations or alternative energy sources could reduce pressure on natural woody cover. Within the moderate zone (3–6 km), coordinated fire management and grazing rotation could help maintain woody cover above degradation thresholds. Beyond 6 km, passive conservation measures may suffice to preserve remaining intact thickets. The absence of intact thickets within 10 km of built-up areas underscores the urgency of extending conservation efforts beyond the immediate settlement periphery, where passive recovery potential may still exist.

The continuous cover framework provides a quantitative baseline against which future conservation interventions can be evaluated — a capability that binary land cover classifications cannot offer (Bastin et al., 2017; Brandt et al., 2020).

7.7. Conclusion

This study demonstrates that annual, multi-decadal woody cover monitoring is operationally feasible in tropical drylands by combining harmonized Landsat time series with locally calibrated Random Forest regression ($R^2 = 0.75$; RMSE = 13.74 %; blockCV $R^2 = 0.79$). The continuous cover framework provides a quantitative baseline for monitoring conservation interventions — a capability that binary forest/non-forest classifications cannot offer.

The principal finding is paradoxical: a limited net area loss ($-15,185$ ha; -3.2 %; -0.090 % yr^{-1}) masks profound structural degradation invisible. Intact thickets declined from 62.8 % to 29.0 % of the landscape while degraded thickets increased from 18.5 % to 39.5 %, with 41.5 % of intact thicket pixels transitioning to degraded thickets between 2000 and 2025 and 51.6 % of non-woody pixels showing no recovery. This structural transformation unfolded across four distinct phases: slight expansion (1989–2001, $+0.055$ % yr^{-1}), accelerated loss (2001–2013, -0.376 % yr^{-1}), continued decline (2013–2019, -0.200 % yr^{-1}), dynamics invisible to conventional bi-temporal assessments.

These phase-level rates should nonetheless be interpreted with caution. A robust (Theil–Sen) regression fitted on the annual values within each phase indicates that none of the four trends differs significantly from zero at the 95% confidence level, and a Chow test applied at the 2001, 2013 and 2019 boundaries does not detect statistically significant structural breaks ($p = 0.24, 0.085$ and 0.13 , respectively). An unsupervised breakpoint search across the full 1989–2025 series instead converges on a single dominant transition around 2016. The four-phase decomposition adopted here therefore remains ecologically motivated — anchored in documented disturbance events such as the political crises of 2002 and 2009 — but should be read as a qualitative description of the trajectory rather than as a set of statistically independent segments, given the limited number of annual observations available per phase ($n = 7–13$).

Among the anthropogenic drivers analyzed, fire recurrence and proximity to settlements emerged as the main drivers of structural degradation. Two quantifiable thresholds provide immediate operational guidance. (1) Fire: Degradation accelerates nonlinearly beyond 8.9 cumulative fire events — equivalent to one fire every three years over the 25-year record — with burned areas losing -18.0% cover versus -0.1% unburned. Recovery requires >10 fire-free years. Managers should prioritize agreements with local communities to bring fire frequency down to one fire every five years or less at a given site, a target that keeps plots within the partial-resilience zone identified above. (2) Distance from built-up areas: Woody cover loss declines monotonically from $-0.166\% \text{ yr}^{-1}$ at 1 km to $-0.090\% \text{ yr}^{-1}$ at 10 km, defining three management zones (intensive ≤ 2 km, moderate 3–6 km, peripheral >6 km). Within each zone, spatially differentiated strategies—from alternative energy sources (intensive) to fire management and grazing rotation (moderate) to passive conservation (peripheral)—can maintain woody cover above degradation thresholds.

Long-term conservation requires reducing community dependence on forest resources through alternative livelihoods. The continuous cover framework provides a quantitative baseline for monitoring these interventions. The identified thresholds (8.9 fire events, >10 fire-free years, built-up area gradient) offer immediately actionable guidance for conservation planning in tropical drylands with similar disturbance regimes.

Chapitre 8

Discussion générale

8.1. D'un diagnostic local à une approche intégrée

La surveillance des écosystèmes arides par télédétection souffre d'un déficit structurel bien documenté à l'échelle mondiale. Les outils de cartographie développés au cours des trente dernières années ont été conçus et calibrés principalement pour les forêts denses tropicales et tempérées, où le signal spectral de la végétation est fort et contrasté. Dans les milieux arides, la végétation ligneuse est souvent ouverte, basse et mélangée à du sol nu, ce qui rend sa détection difficile avec les mêmes algorithmes (Bastin et al., 2017). Cette inadéquation se traduit par des estimations très variables selon les produits disponibles et par une sous-estimation systématique des surfaces boisées en zones sèches, documentée à l'échelle planétaire (Brandt et al., 2020; Fagan, 2020). Les conséquences ne sont pas seulement scientifiques : les zones arides représentent environ 40 % de la surface terrestre et hébergent plus de deux milliards de personnes, dont les moyens de subsistance dépendent directement des ressources ligneuses (Reynolds et al., 2007; UNCCD, 2024). Une surveillance imprécise de leur couvert végétal se traduit par des décisions de gestion et de conservation fondées sur des données erronées.

Dans ce contexte, le sud-ouest de Madagascar constitue un cas d'étude particulièrement révélateur. Les fourrés xérophiles du District de Toliara II hébergent une biodiversité végétale exceptionnelle — avec un taux d'endémisme spécifique supérieur à 95 % — mais leur couvert ligneux est structurellement difficile à quantifier par télédétection. Les tiges succulentes des *Didiereaceae* et des *Euphorbiaceae* aphylls, qui dominent ces formations, présentent une réflectance proche de celle du sol nu en saison sèche, ce qui dérouté les algorithmes de détection entraînés sur d'autres biomes (Brinkmann et al., 2014; J. R. Randriamalala et al., 2015). À cette contrainte spectrale s'ajoutent des contraintes logistiques — l'éloignement et l'enclavement de nombreuses zones — et un déficit de données de référence locales, qui ont longtemps limité le développement de méthodes de cartographie adaptées. Le Chapitre 2 (§2.4) de cette thèse documente précisément cette lacune : la littérature cartographique sur le sud-ouest de Madagascar repose majoritairement sur des classifications binaires à faible résolution qui ne permettent pas de représenter les états intermédiaires de dégradation.

C'est à partir de ce double constat — inadéquation des méthodes globales et absence de méthodes locales validées — que les trois chapitres empiriques de cette thèse ont été construits en séquence logique. Le Chapitre 5 mesure l'ampleur des désaccords entre les produits globaux disponibles et les données de terrain indépendantes, fournissant ainsi une base de comparaison objective. Le Chapitre 6 développe et valide un protocole de calibration locale fondé sur la photo-interprétation via Collect Earth Online, en identifiant les seuils d'échantillonnage à partir desquels un modèle Random Forest atteint une précision satisfaisante et stable. Le Chapitre 7 applique ce protocole à une série temporelle Landsat de 36 ans pour reconstruire les dynamiques du couvert ligneux et analyser les facteurs explicatifs. Ensemble, ces trois chapitres forment une chaîne méthodologique dont chaque maillon répond à une limite identifiée par le précédent, en progressant du diagnostic vers la solution, puis vers

l'application. Ils répondent à deux hypothèses principales et leurs sous-hypothèses formulées dans l'introduction générale : l'inadéquation structurelle des méthodes cartographiques existantes pour les végétations xérophiles (H1.1), l'hétérogénéité des produits globaux disponibles et son lien avec la composition floristique (H1.2), la primauté du plan d'échantillonnage de calibration sur la sophistication des capteurs (H2.1), et la non-linéarité des dynamiques temporelles du couvert ligneux et le rôle des perturbations anthropiques (H2.2). La discussion qui suit évalue dans quelle mesure ces hypothèses sont confirmées, nuancées ou dépassées par les résultats obtenus.

8.2. État de la zone d'étude et enjeux pour les acteurs concernés

Les résultats présentés dans les chapitres précédents dessinent un état des lieux préoccupant du couvert ligneux xérophile du District de Toliara II. Les dynamiques observées sur trente-six ans révèlent une dégradation dominante, ponctuée d'une phase de récupération partielle entre 2019 et 2025, et un rôle prépondérant des feux récurrents dans l'érosion du couvert. Au-delà de leur portée scientifique, ces constats ont des conséquences directes pour un large éventail d'acteurs. Sur le plan écologique, la dégradation continue de formations végétales à très fort taux d'endémisme menace un patrimoine biologique difficilement substituable. Sur le plan socio-économique, les communautés locales dont les moyens de subsistance dépendent des ressources ligneuses (bois de chauffe, pâturage, produits non ligneux) sont directement exposées à l'appauvrissement de cette ressource, tandis que les agriculteurs pratiquant le brûlage comme outil agro-pastoral se trouvent au cœur des dynamiques de dégradation identifiées. Pour les décideurs politiques et les gestionnaires d'aires protégées, ces résultats fournissent une base empirique inédite pour cibler les interventions de conservation là où elles seront les plus efficaces. C'est dans cette perspective que les implications pratiques de ce travail sont développées ci-après, à destination de ces différents acteurs.

8.2.1 Zonation spatiale des priorités à partir des cartes de dynamiques

Les cartes de dynamiques produites au Chapitre 7 ne sont pas seulement un résultat scientifique — elles constituent un outil de planification directement utilisable par les gestionnaires des aires protégées et les services forestiers régionaux. En croisant les trajectoires temporelles du couvert avec le gradient de pression humaine et la fréquence des feux, il est possible de délimiter trois types de zones aux priorités d'intervention distinctes, sans recourir à des données supplémentaires. Cette zonation est représentée spatialement en Figure X, qui superpose les trois catégories de zones aux clusters d'habitations et au réseau routier au centre du district : le gradient du rouge (fronts de dégradation active, à proximité immédiate des habitations) au jaune (zones de récupération potentielle) puis au vert (noyaux résilients, en périphérie)

illustre la diminution de la pression de dégradation avec l'éloignement aux habitations, conformément au modèle de Laurance et al. (2002).

Les noyaux résilients correspondent aux formations qui ont maintenu un couvert élevé (> 60 %) sur l'ensemble de la période 1989–2025, malgré les pressions documentées. Ces zones, situées principalement au-delà de 6 km des habitations dans le secteur sud-ouest du district, représentent les réservoirs de biodiversité et de régénération les mieux préservés. Leur protection prioritaire est justifiée non seulement par leur état actuel, mais aussi par leur rôle potentiel de source de graines et de refuges faunistiques pour la régénération des zones dégradées environnantes. Dans le cadre des transferts de gestion VOI et des plans d'aménagement des aires protégées existantes, ces zones devraient bénéficier du statut de mise en défens stricte, avec un suivi annuel du couvert par le protocole CEO développé au Chapitre 6.

Les fronts de dégradation active désignent les zones situées à moins de 5 km des habitations et ayant perdu plus de 15 % de couvert entre 2013 et 2025. C'est dans ces zones que la pression de carbonisation et de prélèvement de bois est la plus intense, et que les interventions urgentes sont les plus nécessaires. Les résultats du Chapitre 7 indiquent que ces zones ne peuvent pas être restaurées par la seule mise en défens : la pression de prélèvement se déplace vers des zones plus éloignées si les besoins énergétiques des ménages ne sont pas satisfaits par d'autres moyens. Les actions les plus efficaces dans ces zones sont donc celles qui agissent simultanément sur la demande en bois-énergie — promotion des foyers améliorés, développement de plantations communautaires d'espèces à croissance rapide à proximité des villages — et sur l'offre alternative d'énergie (Baylis et al., 2017; Hänke et al., 2017). Le choix des espèces doit être adapté aux conditions édaphiques locales, qui varient fortement au sein du district : sur le plateau calcaire (sols minces, climat hyper-aride), *Prosopis juliflora*, *Acacia crassicaarpa* et le tsinefo (*Ziziphus spina-christi*) sont les mieux adaptés, l'Eucalyptus et l'*Acacia mangium* y étant peu performants en raison de la faible profondeur de sol et de la sécheresse extrême ; le tsinefo présente par ailleurs un intérêt fourrager pour l'élevage caprin, valorisé par le projet AGRIFO, qui appuie des éleveurs pilotes dans l'alimentation des femelles caprines gestantes et allaitantes et des petits ruminants; sur le littoral et les dunes sableuses, *Acacia crassicaarpa* et le filao (*Casuarina equisetifolia*, à croissance rapide, recherché pour le charbonnage et les poteaux) sont recommandés, le kily (*Tamarindus indica*) pouvant également réussir en bordure de cuvettes sableuses si la nappe phréatique reste accessible ; dans les vallées alluviales (Fiherenana, Onilahy), les conditions plus fertiles permettent *Eucalyptus camaldulensis*, *Acacia mangium* et le neem (*Azadirachta indica*), le kily pouvant y être intégré en agroforesterie le long des parcelles agricoles.

Les zones de récupération potentielle correspondent aux parcelles ayant connu une perte significative entre 2001 et 2013, suivie d'une stabilisation ou d'un léger rebond depuis 2019. Ces zones se situent principalement dans la bande intermédiaire entre 3 et 6 km des habitations, là où la pression d'exploitation est modérée mais où la végétation conserve encore une capacité de régénération. Si la pression des feux peut y être maintenue en dessous du seuil critique de 8,9 événements cumulés, ces zones sont candidates à une régénération naturelle assistée, moins coûteuse que la

restauration active et plus durable sur le long terme (Chazdon, R.L and Guariguata, M.R., 2016; Strassburg et al., 2020).

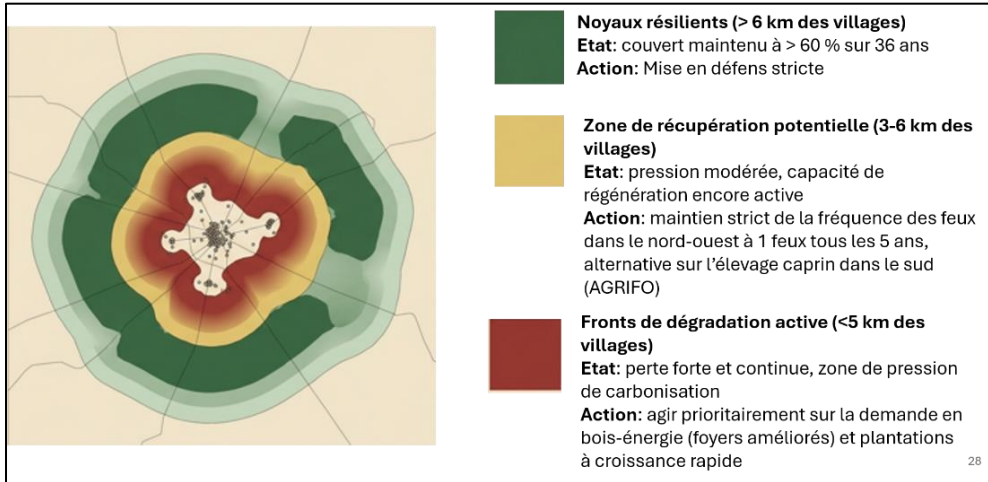


Figure 8-1 : Zonation spatiale des priorités d'intervention dérivée des cartes de dynamiques du couvert ligneux (1989-2025) et du gradient de distance aux habitations. Le centre regroupe les clusters d'habitations et le réseau routier ; le gradient du rouge (fronts de dégradation active) au jaune (zones de récupération potentielle) puis au vert (noyaux résilients) traduit la diminution de la pression de dégradation avec l'éloignement aux habitations.

Le cadre de zonation présenté ci-dessus repose sur deux pressions mesurables par télédétection — la proximité aux habitations et la fréquence des feux — qui n'épuisent pas l'ensemble des déterminants de la dégradation identifiés en introduction. L'agriculture sur brûlis (hatsake), désignée au Chapitre 2 comme la principale cause directe de déforestation dans le district, n'est pas individualisée dans la zonation spatiale proposée : ses emprises se confondent largement avec les fronts de dégradation active à proximité des habitations, dont elle constitue probablement le principal moteur, mais l'indicateur de distance ne permet pas de distinguer une parcelle convertie par hatsake d'une parcelle dégradée par carbonisation ou par pâturage. Une cartographie opérationnelle de l'hatsake nécessiterait un suivi des conversions du couvert ligneux vers un usage agricole identifiable — signatures spectrales de parcelles cultivées, calendriers culturaux —, ce qui dépasse le cadre méthodologique du présent travail, centré sur la dynamique du couvert ligneux plutôt que sur l'usage du sol qui lui succède. Sur le plan des leviers d'intervention, agir sur l'hatsake suppose des mesures qui sortent du périmètre des outils satellitaires développés ici — sécurisation du foncier, intensification ou diversification agricole, développement de filières alternatives au défrichage — et qui relèvent davantage de l'aménagement du territoire et des politiques agricoles que du suivi environnemental. Le dispositif de délimitation spatiale proposé ne se substitue donc pas à une politique de maîtrise de l'hatsake, mais peut l'appuyer en signalant les fronts où la conversion agricole progresse le plus rapidement, dès lors que cette information

satellitaire est croisée avec des données parcellaires ou des enquêtes de terrain non mobilisées dans cette thèse.

8.2.2 La maîtrise des feux comme levier d'intervention prioritaire

Le seuil de 8,9 feux cumulés identifié au Chapitre 7 fournit un critère opérationnel concret pour orienter les interventions de prévention. Les zones ayant accumulé entre 7 et 9 événements depuis 2000 représentent une zone de vigilance active : la végétation y est encore dans un état de résilience partielle, et des actions ciblées de réduction des feux peuvent encore prévenir le basculement vers un état fortement dégradé. Ces zones peuvent être identifiées et actualisées annuellement à partir du produit MODIS MCD64A1 dans Google Earth Engine, sans investissement logistique supplémentaire. Exprimé en intervalle de retour plutôt qu'en nombre d'événements cumulés — une formulation plus directement actionnable pour les communautés et les gestionnaires — ce seuil de 8,9 feux sur 25 ans correspond approximativement à une fréquence limite d'un feu tous les trois ans : au-delà, la capacité de résilience de la végétation s'épuise progressivement. La maîtrise effective du régime de feu, ramenant la fréquence à un feu tous les cinq ans ou moins, permet de maintenir les parcelles dans la zone de résilience partielle et constitue l'objectif pratique des conventions de brûlage contrôlé proposées ci-dessous.

La gestion des feux dans ce contexte ne peut pas reposer sur une interdiction généralisée du brûlage, qui se heurterait aux pratiques pastorales profondément ancrées dans les communautés d'éleveurs du district. Elle doit plutôt s'appuyer sur des conventions négociées avec ces communautés, définissant des calendriers de brûlage contrôlé, des zones de mise en défens rotatives et des corridors de non-brûlage autour des noyaux résilients identifiés. Ce type d'approche participative, documenté avec succès dans d'autres contextes africains (Archibald, 2016; Croker et al., 2023), est également cohérent avec les mécanismes de transfert de gestion VOI déjà en place dans plusieurs communes du district. Les cartes de fréquence des feux produites au Chapitre 7 constituent un support visuel directement utilisable pour ces négociations communautaires, en montrant concrètement les zones les plus affectées et les tendances temporelles sur 25 ans.

8.2.3 Un dispositif de monitoring actualisable à faible coût

Le protocole développé dans cette thèse repose exclusivement sur des données et des outils librement accessibles : les archives Landsat Collection 2 distribuées via Google Earth Engine (Gorelick et al., 2017), la plateforme de photo-interprétation Collect Earth Online (Bey et al., 2016), et l'algorithme Random Forest implémenté dans l'environnement R (Breiman, 2001). Cette combinaison ne nécessite ni équipements satellitaires propriétaires, ni logiciels commerciaux, ni infrastructure de calcul locale — des contraintes qui limitent fréquemment l'adoption des méthodes de télédétection avancées dans les pays à ressources limitées (Amani et al., 2020). Une session complète de recalibration du modèle, incluant la photo-interprétation de 500

placettes CEO et la production d'une carte de couvert annuelle, a été réalisée par deux opérateurs formés en trois à quatre semaines, ce qui est compatible avec les capacités opérationnelles des services déconcentrés de l'environnement à Toliara.

Un renouvellement de cette calibration tous les trois à cinq ans, couplé à une mise à jour annuelle des cartes de tendance à partir des composites Landsat, permettrait à l'Office National pour l'Environnement (ONE) et aux gestionnaires des aires protégées du District de Toliara II de disposer d'indicateurs quantitatifs actualisés du couvert ligneux. Ce type de dispositif s'inscrit dans les recommandations Olofsson et al., (2014) sur les bonnes pratiques d'estimation des changements de surface et dans le cadre du système de mesure, notification et vérification (MRV) requis par les mécanismes REDD+ auxquels Madagascar est engagé. Il répond également à un besoin explicitement identifié par Saah et al., (2019), qui soulignent que les plateformes de photo-interprétation participative comme Collect Earth Online offrent une voie concrète pour impliquer les institutions nationales dans la production et la mise à jour de données de référence locales, renforçant ainsi la durabilité des systèmes de surveillance environnementale.

Une fois validé à l'échelle du District de Toliara II, ce dispositif pourrait être étendu aux autres districts du domaine subaride malgache — Ampanihy, Beloha et Tsihombe — en appliquant la même chaîne de traitement sans modification majeure du protocole. Cette extension produirait la première cartographie continue multi-décennale du couvert ligneux de l'ensemble du bush du sud, comblant un vide documentaire identifié dès le Chapitre 2 (§2.4) et fournissant une base quantitative pour les futurs plans nationaux de restauration du paysage forestier, conformément aux engagements de Madagascar dans le cadre du Défi de Bonn (objectif de restauration de 4 millions d'hectares d'ici 2030 ; Strassburg et al., 2020).

8.2.4 Valorisation des résultats par et pour les communautés locales

Au-delà de leur portée scientifique et institutionnelle, les résultats de cette thèse produisent plusieurs livrables susceptibles d'être valorisés à l'échelle des communautés locales du District de Toliara II, à condition que leur restitution soit adaptée aux réalités sociales et économiques de ces populations.

Les cartes de couvert par classes, produites annuellement pour chaque secteur du district, constituent le premier livrable mobilisable à l'échelle communautaire. Contrairement aux indices spectraux ou aux taux de déforestation exprimés en pourcentages annuels, les cinq classes de couvert — de non-boisé à fourré intact — correspondent à des réalités visuellement identifiables sur le terrain par les éleveurs et les agriculteurs. Une communauté VOI disposant d'une carte montrant que son fourré de référence est passé de la classe « intact » à la classe « dégradé » entre 2013 et 2025 dispose d'un argument cartographique concret pour négocier des droits d'accès différenciés ou pour justifier une demande de transfert de gestion auprès des autorités forestières régionales. DeVries et al., (2016) ont montré, dans le contexte des mécanismes REDD+, que les cartes participatives ancrées dans une série temporelle

satellitaire fournissent aux communautés un contexte historique qui renforce leur capacité à interpréter les changements de couvert et à participer aux décisions de gestion ; c'est précisément ce type d'ancrage que la reconstruction annuelle sur 36 ans développée dans le Chapitre 7 peut offrir aux VOI du district.

Le seuil de 8,9 feux cumulés identifié au Chapitre 7 constitue le deuxième résultat directement traduisible en règle de gestion communautaire. Exprimé simplement, il signifie qu'au-delà d'environ neuf feux sur une parcelle depuis 2000, la végétation ne récupère plus sans intervention active. Combiné aux cartes de fréquence cumulée des feux, ce seuil permet à chaque communauté VOI de localiser les parcelles encore en deçà du seuil critique — celles où une réduction des brûlages peut encore être efficace — et de concentrer les efforts de convention locale sur ces zones à fort potentiel de résilience. Huy et al., (2024), dans une évaluation quasi-expérimentale menée dans les zones arides du Burkina Faso, ont montré que des interventions communautaires de sensibilisation fondées sur des données satellitaires de fréquence des feux réduisent effectivement l'incidence des feux dans les forêts protégées ; ce résultat indique que des données de feux spatialement explicites, restituées localement sous une forme compréhensible, peuvent modifier les comportements de brûlage à l'échelle communautaire.

La phase de récupération partielle identifiée entre 2019 et 2025 (+0,304 %/an) est probablement le résultat le plus porteur de sens pour les communautés. Elle montre, de manière empirique et spatialement explicitée, que les fourrés xérophiles récupèrent là où la pression des feux a diminué. C'est un argument de terrain en faveur des pratiques de gestion communautaire des feux : là où des conventions locales ont été respectées, la forêt répond. Restituer ce résultat aux communautés concernées — en localisant précisément les zones de rebond sur leurs propres terroirs — contribue à renforcer la légitimité des efforts de conservation engagés. Randriamiharisoa et al., (2024), dans une étude menée dans six aires protégées de Madagascar incluant des formations épineuses du sud, ont montré que des membres de communautés locales formés au monitoring produisent des données comparables à celles de scientifiques professionnels ; ce résultat, obtenu dans un contexte proche de celui du District de Toliara II, atteste que l'implication communautaire dans le suivi des ressources naturelles est réalisable à Madagascar lorsque la formation et l'accompagnement sont assurés.

Enfin, le protocole CEO développé dans cette thèse a été conçu pour minimiser les contraintes techniques d'accès : il repose exclusivement sur des outils gratuits accessibles en ligne et ne nécessite pas d'infrastructure de calcul locale. À terme, il pourrait être appris par des opérateurs formés au sein des services déconcentrés de l'ONE, réduisant la dépendance vis-à-vis d'experts extérieurs pour les mises à jour périodiques des cartes de couvert. Cette perspective reste néanmoins conditionnée, dans le contexte malgache, à des investissements préalables en formation, en équipements et en connectivité. Aubert et al., (2012), dans une analyse des transferts de gestion forestière à Madagascar, documentent précisément ces contraintes institutionnelles et soulignent que la durabilité des dispositifs de surveillance communautaire dépend d'un cadre d'accompagnement structuré impliquant

conjointement les services déconcentrés de l'État, les communautés et les coordinateurs de projets ; c'est précisément ce modèle que le projet AGRIFO s'efforce de mettre en place, conformément aux recommandations de Saah et al., (2019) sur le renforcement des capacités institutionnelles locales dans les programmes de surveillance environnementale.

8.3. Contribution aux Objectifs de Développement Durable

Les résultats de cette thèse s'inscrivent dans plusieurs Objectifs de Développement Durable. L'ODD 15 (Vie terrestre) est directement concerné : le protocole de suivi du couvert ligneux développé ici (Chapitres 5 à 7) fournit un indicateur quantitatif et reproductible de dégradation des fourrés xérophiiles, écosystème à fort taux d'endémisme non couvert par les produits cartographiques globaux existants. L'identification d'un seuil de fréquence de feu (Chapitre 7) et d'un gradient de pression lié à la proximité aux habitations relie ce travail à l'ODD 13 (Lutte contre les changements climatiques), dans la mesure où la résistance des fourrés xérophiiles aux perturbations contribue à la séquestration de carbone en milieu aride. La caractérisation des pressions sur des ressources ligneuses dont dépendent directement les moyens de subsistance locaux (bois de chauffage, pâturage) relie également ces résultats aux ODD 1 et 2 (Élimination de la pauvreté ; Faim zéro). Enfin, le dispositif de monitoring à faible coût proposé (§9.5.3), reposant sur des outils librement accessibles et mobilisable par les services déconcentrés de l'ONE et les communautés VOI, s'inscrit dans l'ODD 17 (Partenariats pour la réalisation des objectifs), en particulier le renforcement des capacités institutionnelles locales de suivi environnemental.

8.4. Apports méthodologiques transversaux

8.4.1 La calibration locale comme réponse aux limites des produits globaux

Le Chapitre 5 montre que les quatre produits globaux évalués — GFCC, GFW, CGLS-LC100 et Dynamic World — ne divergent pas de façon aléatoire dans le District de Toliara II. Leur hétérogénéité est structurée : les écarts entre produits sont les plus importants dans les formations à dominance d'Euphorbiaceae et de Didiereaceae, c'est-à-dire précisément là où la signature spectrale de la végétation ligneuse est la plus atypique. Les produits qui appliquent un seuil de couvert élevé pour définir la forêt (GFW : 30 %) sous-estiment systématiquement le couvert des formations ouvertes, tandis que les produits à classification continue (CGLS-LC100, Dynamic World) présentent des biais différents selon la densité structurale. Ce résultat indique que les limites des produits globaux dans cette région ne sont pas dues à un manque de précision générale, mais à une inadéquation entre les caractéristiques spectrales des végétations xérophiiles et les bases d'entraînement utilisées pour construire ces produits, lesquelles sont dominées par des formations forestières de zones tempérées et tropicales humides (Brown et al., 2022; Tsendbazar et al., 2021).

Un résultat particulièrement instructif concerne Dynamic World : ce produit s'avère le plus proche de la réalité terrain pour la variable couvert ligneux agrégé (arbres + arbustes), mais diverge fortement pour le couvert arboré. Plusieurs hypothèses permettent d'expliquer ce paradoxe apparent. D'une part, l'architecture de classification de Dynamic World (Brown et al., 2022), fondée sur un modèle d'apprentissage profond entraîné à l'échelle mondiale sur des images Sentinel-2 à 10 m, capte efficacement le signal de végétation ligneuse dans sa globalité, sans distinguer les strates verticales. D'autre part, dans les fourrés xérophiles du sud-ouest malgache, la frontière entre arbres et arbustes est floue : les individus des familles Didiereaceae et Euphorbiaceae atteignent rarement 5 m, et leur port architectural — tiges succulentes dressées, canopée diffuse — ne correspond à aucune des catégories arborescentes présentes dans les bases d'entraînement mondiales. Le classificateur assigne alors ces individus à la classe arbustive plutôt qu'arborescente, sous-estimant systématiquement le couvert arborien tout en restituant correctement le couvert ligneux total. Ce phénomène de confusion arbre/arbuste lié à la structure ouverte et au port atypique des espèces succulentes constitue une source d'erreur qui n'est pas corrigable par la seule amélioration de la résolution spatiale, mais qui nécessite une calibration locale sur des données de référence représentant spécifiquement ces formations. Ce résultat rejoint les conclusions d'Agassounon et al., (2025), qui montrent dans leur revue systématique sur la télédétection en milieux arides et semi-arides que la précision des classifications décroît substantiellement dès que les catégories végétales deviennent plus détaillées, et que 70 % des études publiées se limitent à des distinctions grossières forêt/non-forêt précisément pour éviter cette dégradation de performance.

Face à ce constat, la valeur ajoutée de la calibration locale est mesurable et substantielle. Dynamic World, le produit global le plus précis sur les données de terrain indépendantes du District de Toliara II, atteint un R^2 de 0,70 (Chapitre 5). Le modèle Random Forest calibré localement sur 505 placettes photo-interprétées via Collect Earth Online atteint 0,88 pour le modèle Sentinel-1+2 (Chapitre 6) et 0,75 sur la série temporelle Landsat (Chapitre 7). Cet écart de 0,05 à 0,18 point selon les capteurs peut sembler modeste en valeur absolue, mais il est décisif pour la classe de couvert la plus importante sur le plan de la gestion : les états intermédiaires entre 40 % et 70 % de couvert, qui correspondent aux fourrés en cours de dégradation. C'est précisément dans cette plage que les produits globaux accumulent le plus d'erreurs, et que le modèle local, calibré sur un échantillon représentant l'ensemble du gradient écologique local, parvient à discriminer des états que les produits globaux confondent.

Cette supériorité de la calibration locale s'explique en grande partie par la nature des données d'entraînement utilisées pour construire les produits globaux. Ces produits sont entraînés sur des millions de points répartis à l'échelle planétaire, mais les végétations succulentes à tiges photosynthétiques des zones arides malgaches y sont très marginalement représentées. Un algorithme entraîné principalement sur des forêts à feuilles larges ou des savanes africaines ne dispose pas des exemples nécessaires pour apprendre à distinguer un fourré à *Alluaudia procera* dense d'un fourré dégradé à *Euphorbia stenoclada* clairsemée. La calibration locale sur Collect Earth Online

permet au contraire de constituer un échantillon qui couvre spécifiquement le gradient de couvert propre à ces formations, avec des placettes réparties sur l'ensemble du district et stratifiées sur les cinq classes de couvert (Chapitre 6). C'est cette représentativité locale, et non la sophistication de l'algorithme, qui explique le gain de précision observé, conformément aux conclusions de Foody, (2009) et de Mellor et al., (2015).

Ce résultat n'invalide pas l'utilité des produits globaux pour d'autres usages ou d'autres régions. Il conditionne en revanche leur emploi dans les milieux arides tropicaux à une étape préalable de validation locale. Utiliser un produit global sans vérification indépendante dans ce contexte, c'est prendre le risque de baser des décisions de conservation sur des estimations de couvert qui peuvent s'écarter de la réalité de 20 à 40 points de pourcentage dans les formations les plus dégradées. La recommandation qui découle de ce chapitre est donc opérationnelle : tout programme de suivi du couvert ligneux dans les zones arides tropicales devrait systématiquement inclure une étape de validation sur données locales indépendantes, avant d'utiliser un produit global comme indicateur de tendance (Olofsson et al., 2014; Tsendbazar et al., 2021).

Ces résultats permettent de répondre directement à H1.2 : l'hétérogénéité inter-produits est confirmée et substantielle ($CV = 78,94\%$ pour le couvert arboré), mais elle se réduit à $20,75\%$ lorsque l'on agrège arbres et arbustes en une variable unique de couvert ligneux. Ce résultat nuance H1.2 de façon importante : le problème n'est pas tant la détection de la végétation ligneuse en général que la capacité à distinguer les strates verticales dans des formations structurellement hybrides. Il ouvre une voie d'utilisation conditionnelle des produits globaux — en se limitant à la variable couvert ligneux agrégé et en procédant à une validation locale préalable — dans des contextes où la calibration locale n'est pas possible. H1.1 est également confirmée par ce résultat : les méthodes existantes documentées au Chapitre 2 (§2.4) ne sont pas seulement peu répandues dans cette région, elles sont structurellement inadaptées, comme le prouve l'échec des produits globaux à restituer fidèlement le couvert des formations succulentes.

Ce travail permet de préciser et de nuancer les apports de Bastin et al. (2017), dont la recension mondiale fondée sur 213 000 placettes a mis en évidence une sous-estimation systématique du couvert forestier en zones arides par les approches binaires classiques. Cette étude opère cependant à une échelle et avec une résolution qui ne permettent pas de quantifier la part de cette sous-estimation imputable au choix d'une variable binaire forêt/non-forêt plutôt qu'à la qualité intrinsèque des produits de télédétection. Les résultats du présent travail apportent un éclairage complémentaire à cette échelle plus fine : même le meilleur produit global testé sur le terrain indépendant du District de Toliara II (Dynamic World, $R^2 = 0,70$) reste nettement moins performant que le modèle calibré localement ($R^2 = 0,88$ en Sentinel-1+2, $0,75$ en série Landsat), et cet écart se concentre précisément dans la plage de couvert intermédiaire (40–70 %) qu'une approche binaire ne peut par construction pas représenter. La valeur ajoutée de cette thèse par rapport à Bastin et al. (2017) ne consiste donc pas à contester l'ampleur de la sous-estimation qu'ils documentent à

l'échelle planétaire, mais à montrer, à l'échelle d'un district, qu'une partie substantielle de cette sous-estimation provient du choix même d'une variable discrète, et qu'une variable continue de couvert calibrée localement permet de la réduire significativement tout en révélant une dynamique de dégradation invisible à une lecture binaire.

8.4.2 La distribution spatiale de l'échantillon prime sur la précision de mesure

Le Chapitre 6 apporte une démonstration quantitative d'un principe qui est souvent mentionné dans la littérature mais rarement testé de façon aussi systématique dans les milieux arides : la façon dont les placettes de calibration sont réparties dans l'espace est plus déterminante pour la précision finale de la carte que la précision avec laquelle chaque placette est mesurée individuellement. Ce résultat mérite d'être discuté en détail, car il va à l'encontre d'une intuition courante selon laquelle des mesures de terrain plus précises produisent nécessairement de meilleures cartes.

Les 41 placettes de terrain collectées en 2022 ont été mesurées avec un protocole rigoureux — la méthode point-intercept avec 54 points par placette — qui fournit une estimation directe et reproductible du couvert ligneux en projection verticale. Pourtant, les modèles calibrés sur ces placettes atteignent au mieux un R^2 de 0,66 (modèle Sentinel-1+2, Chapitre 6), nettement inférieur aux modèles calibrés sur les 505 placettes photo-interprétées ($R^2 = 0,88$ pour le même capteur). La raison principale de cet écart n'est pas la précision de mesure — les données terrain sont plus précises que les données photo-interprétées — mais la représentativité spatiale. Les 41 placettes terrain sont concentrées dans des zones accessibles par piste, ce qui introduit un biais systématique : les formations denses éloignées des pistes sont sous-représentées, et le modèle apprend à prédire le couvert principalement dans les zones facilement accessibles. À l'inverse, les 505 placettes CEO couvrent l'ensemble du gradient écologique du district, y compris les secteurs les plus reculés, ce qui donne au modèle une vue représentative de la diversité des situations présentes sur le terrain.

La saturation des performances identifiée entre 20,7 et 41,4 placettes dispersées par 1 000 km² fournit un repère pratique important. En dessous de ce seuil, chaque placette supplémentaire améliore sensiblement la précision du modèle. Au-delà, les gains marginaux deviennent négligeables et l'investissement en photo-interprétation n'est plus rentable. Ce seuil de saturation est cohérent avec les observations de Mellor et al., (2015) qui avaient documenté un phénomène similaire pour la cartographie de la végétation en Australie, et avec les travaux de Ploton et al., (2020), qui ont montré que l'autocorrélation spatiale des données d'entraînement est une source fréquente de surestimation des performances dans les modèles écologiques à grande échelle. La validation spatiale par blocs (blockCV, Valavi et al., 2018) utilisée dans les Chapitres 7 et 8 permet précisément d'éviter ce piège en s'assurant que les données de validation sont géographiquement distinctes des données d'entraînement.

Ce résultat a une implication directe pour la conception des futurs programmes de cartographie en zones arides : avant d'investir dans des capteurs plus sophistiqués ou

des algorithmes plus complexes, il est plus efficace de s'assurer que l'échantillon de calibration est spatialement représentatif. Un échantillon de 75 placettes photo-interprétées bien distribuées produit une carte plus précise qu'un échantillon de 41 placettes terrain mal distribuées, même si ces dernières sont mesurées avec beaucoup plus de soin. Ce principe, qui déplace la priorité de la précision de mesure vers la représentativité spatiale, constitue l'un des apports méthodologiques les plus directement transférables de cette thèse.

Ce résultat apporte une réponse précise à H2.1 : la primauté du plan d'échantillonnage sur la sophistication des capteurs est confirmée de façon quantitative. Avec 75 placettes photo-interprétées dispersées, un modèle surpasse systématiquement celui calibré sur 41 placettes terrain plus précises mais spatialement biaisées. H2.1 est cependant enrichie d'une nuance importante : cette primauté de l'échantillonnage n'est pas absolue. Le Chapitre 6 montre que lorsque l'échantillon est suffisamment représentatif — au-delà du seuil de saturation de 41,4 placettes par 1 000 km² — la sophistication du capteur retrouve un rôle : Sentinel-1+2 et AlphaEarth convergent alors vers des performances quasi identiques ($R^2 = 0,88$), tandis qu'avec un échantillon insuffisant, AlphaEarth sous-performe significativement en raison du surajustement lié à la haute dimensionnalité de ses prédicteurs. La relation entre taille d'échantillon et choix du capteur est donc interactive, et non hiérarchique.

8.4.3 La variable continue de couvert comme outil de détection précoce de la dégradation

La quasi-totalité des bilans nationaux de déforestation, y compris ceux produits pour Madagascar, repose sur une classification binaire qui distingue les surfaces boisées des surfaces non boisées selon un seuil de couvert minimal (généralement 10 % ou 30 % selon les conventions nationales). Cette approche a l'avantage de la simplicité et de la comparabilité internationale, mais elle présente une limite fondamentale dans les milieux où la dégradation progresse de façon graduelle : un pixel dont le couvert passe de 80 % à 40 % reste classé comme « boisé » dans les deux cas, et la perte écologique associée à cette transition est totalement invisible dans le bilan national. C'est précisément la situation qui prévaut dans les fourrés xérophiles du District de Toliara II, où la dégradation qualitative — le glissement progressif d'un état de fourré intact vers un état de fourré dégradé — est la forme dominante de changement, bien avant que la déforestation stricte ne soit détectable (FAO, 2020b; ONE et al., 2015).

Le Chapitre 7 fournit une démonstration empirique de l'ampleur de cette invisibilité. Sur les 36 années analysées, la perte nette de surface végétale dans le district ne dépasse pas -3,2 %, soit un taux annuel de -0,090 %, qui est inférieur aux taux de déforestation documentés pour les forêts humides malgaches (Vieilledent et al., 2018). Ce chiffre, s'il était seul, conduirait à conclure que les fourrés xérophiles de Toliara II sont relativement stables. Or, l'analyse en variable continue révèle une réalité très différente : la proportion de fourrés intacts à couvert élevé (> 80 %) a reculé de 62,8 % à 29,0 % du paysage entre 1989 et 2025, tandis que les fourrés dégradés (60–80 %) progressaient de 18,5 % à 39,5 %. Cette transformation structurelle profonde, qui

correspond à une perte massive de biodiversité et de stockage carbone, est entièrement invisible dans un bilan binaire. C'est la variable continue qui la rend détectable.

Cette démonstration rejoint et prolonge plusieurs travaux récents qui plaident pour une évolution des standards de surveillance vers des indicateurs de qualité du couvert plutôt que de simple présence ou absence. Wulder et al., (2022), dans leur bilan de cinquante ans d'archives Landsat, soulignent que la richesse de ces séries temporelles est précisément leur capacité à documenter des changements gradients qui échappent aux approches bi-temporelles classiques. DeVries et al., (2015) ont montré que les perturbations forestières subtiles dans les forêts tropicales de montagne ne sont détectables qu'avec des approches en variable continue sur séries temporelles denses. Dans le contexte des mécanismes REDD+, Harris et al., (2021) et Pearson et al., (2017) ont établi que la dégradation forestière représente une source d'émissions de carbone comparable à la déforestation stricte, mais qu'elle est systématiquement sous-comptabilisée dans les bilans nationaux faute d'outils de détection adaptés. Les résultats du Chapitre 7 illustrent concrètement ce problème à l'échelle d'un district malgache et fournissent un argument empirique solide en faveur de l'adoption de la variable continue comme standard dans les programmes nationaux de surveillance des zones arides.

8.5. Dynamiques du couvert ligneux sur 36 ans : résultats et interprétations

8.5.1 Le glissement vers des états dégradés : une dynamique invisible aux classifications binaires

L'analyse de la trajectoire temporelle du couvert ligneux sur 36 ans (Chapitre 7) conduit à un résultat qui semble paradoxal au premier abord : la surface totale couverte par la végétation ligneuse n'a diminué que de 3,2 % entre 1989 et 2025, mais la structure du paysage a été profondément transformée. Ce paradoxe disparaît dès lors qu'on abandonne la lecture binaire pour une lecture en classes de couvert continu. Les fourrés à couvert élevé — ceux dont la densité de végétation ligneuse dépasse 80 % — représentaient 62,8 % du district en 1989. En 2025, ils n'en représentent plus que 29,0 %. Cette perte de plus de la moitié des fourrés intacts ne s'est pas traduite par une déforestation au sens classique du terme : les pixels n'ont pas basculé de la catégorie « boisé » à la catégorie « non boisé ». Ils ont glissé progressivement vers des classes de couvert intermédiaires — fourrés dégradés entre 60 et 80 %, fourrés très dégradés entre 40 et 60 % — qui restent « boisés » selon les définitions conventionnelles mais qui sont écologiquement très différents de l'état initial.

Cette forme de dégradation qualitative sans perte quantitative majeure est caractéristique des milieux arides soumis à des pressions diffuses et récurrentes, comme les feux pastoraux et le prélèvement sélectif de bois-énergie. Elle diffère fondamentalement du profil de déforestation observé dans les forêts humides malgaches, où la conversion agricole efface rapidement et complètement la végétation sur des parcelles entières, produisant des taux de déforestation élevés et facilement

détectables (Vieilledent et al., 2018 : entre $-0,4\%$ et $-1,0\%$ /an selon les régions). Il convient cependant de souligner qu'un faible taux de recouvrement ligneux n'est pas nécessairement synonyme de dégradation anthropique dans ce contexte. Les formations sur substrats calcaires (calcisols) du district peuvent présenter, même à l'état naturel, des couverts structurellement ouverts avec de grands individus émergés et une richesse floristique remarquable, dont les valeurs de couvert peuvent être inférieures à 60% sans que cela reflète une quelconque pression anthropique (Randriamalala et al., 2019). La structure verticale de la végétation et sa composition floristique constituent donc des indicateurs complémentaires indispensables que la seule variable de couvert continu ne peut pas capturer (Maestre et al., 2016). Cette nuance invite à interpréter les cartes de couvert avec précaution dans les zones à substrat calcaire, où la structure naturellement ouverte des formations peut être confondue avec une dégradation induite. L'intégration future de données de hauteur de canopée (GEDI) et de composition floristique à partir de relevés terrain permettrait de lever cette ambiguïté dans les zones les plus concernées. Comparer directement les taux bruts de perte de surface entre les forêts humides et les fourrés xérophiiles serait donc trompeur : non seulement les valeurs ne sont pas comparables, mais la nature même du processus de dégradation est différente. Dans les fourrés xérophiiles, l'indicateur pertinent n'est pas le taux de déforestation mais le taux de transition entre classes de couvert, et c'est cet indicateur que la variable continue permet de calculer.

Ce profil de dégradation graduelle rejoint des observations faites dans d'autres milieux arides africains. Brandt et al., (2017) ont montré, à partir d'une analyse multi-décennale des savanes d'Afrique subsaharienne, que la pression humaine — prélèvement de bois, feux pastoraux, agriculture itinérante — compense en grande partie les gains de végétation ligneuse liés à l'augmentation du CO_2 atmosphérique, sans produire nécessairement une perte nette de surface. Le résultat net visible depuis l'espace est une relative stabilité de la surface boisée, qui masque une transformation structurelle profonde documentée uniquement par des approches en variable continue. Les résultats du Chapitre 7 s'inscrivent dans cette même logique et en fournissent une illustration localement précise, avec une résolution temporelle annuelle sur 36 ans qui permet de suivre les phases successives de la dégradation. Ce type d'analyse, encore rare pour les fourrés xérophiiles de Madagascar, constitue une contribution directe au débat sur les méthodes de surveillance des milieux arides tropicaux.

Ces résultats apportent une réponse partielle à H2.2 : la non-linéarité de la trajectoire temporelle est confirmée, avec quatre phases distinctes documentées sur 36 ans. La dégradation structurelle affectant préférentiellement les formations à couvert élevé — fourrés intacts passant de $62,8\%$ à $29,0\%$ — confirme la sous-hypothèse H2.2. Cependant, H2.2 est aussi dépassée sur un point : le résultat le plus significatif n'est pas la non-linéarité de la trajectoire en elle-même, mais le fait que cette non-linéarité est invisible aux classifications binaires. Une analyse fondée sur un taux de déforestation brut aurait conclu à une relative stabilité du paysage ($-3,2\%$ sur 36 ans), masquant une transformation structurelle profonde que seule la variable continue permet de révéler.

8.5.2 Les feux et la proximité aux habitations comme moteurs principaux de la dégradation

Les résultats du Chapitre 7 permettent d'aller au-delà de la simple description de la trajectoire temporelle du couvert ligneux pour identifier les facteurs qui expliquent cette trajectoire. Deux variables ressortent comme moteurs principaux de la perte de couvert : la fréquence cumulée des feux depuis 2000 et la distance aux zones habitées. Ces deux variables ne sont pas indépendantes — les feux pastoraux sont concentrés à proximité des villages, où les éleveurs pratiquent le brûlage pour stimuler la repousse des pâturages — mais leurs effets sur le couvert ligneux sont distincts et peuvent être analysés séparément. Leur identification repose sur deux analyses complémentaires : une analyse de régression polynomiale pour les feux et une analyse par tampons concentriques pour la distance aux habitations, les deux étant fondées sur des données satellitaires indépendantes du modèle de couvert (MODIS MCD64A1 pour les feux, Google Open Buildings v3 pour les habitations).

La relation entre fréquence des feux et perte de couvert ligneux n'est pas linéaire. En dessous de 8,9 événements cumulés depuis 2000, la perte de couvert reste modérée et relativement stable d'une classe de fréquence à l'autre. Au-delà de ce seuil, la perte s'accélère de façon marquée : les pixels ayant subi plus de 8,9 feux cumulés affichent une perte médiane de $-18,0\%$ contre $-0,1\%$ pour les pixels non brûlés (Mann-Whitney, $p < 2,2 \times 10^{-16}$; Cohen's $d = -0,967$). Ce comportement à seuil indique que les fourrés xérophiles disposent d'une certaine capacité de résistance aux feux à faible fréquence, mais que cette capacité s'épuise progressivement lorsque les perturbations se répètent à intervalles trop courts pour permettre une récupération complète. Ce type de réponse non linéaire est cohérent avec les modèles théoriques d'états alternatifs stables dans les écosystèmes soumis à des perturbations récurrentes (Hirota et al., 2011; Scheffer et al., 2001) : en dessous d'un seuil critique, le système absorbe la perturbation et revient à son état initial ; au-delà, il bascule vers un état dégradé dont le retour en arrière est lent et incertain. La lenteur de la récupération observée dans le Chapitre 7 — un signal positif n'est détectable qu'au-delà de dix années sans feu, avec un taux de reconstitution de seulement $+0,49\%/an$ — confirme que ce basculement est difficile à inverser sans intervention active (Archibald et al., 2013; Noromiarilanto et al., 2016).

La relation entre distance aux habitations et perte de couvert est, elle, remarquablement régulière. Le taux annuel de perte décroît de façon monotone de $-0,166\%/an$ à 1 km des habitations à $-0,090\%/an$ à 10 km, avec une corrélation de Spearman de $\rho = +0,985$ entre distance et perte nette sur l'ensemble de la période 1989–2025. Ce gradient se structure en trois zones d'intensité décroissante : une zone intensive jusqu'à 2 km, une zone modérée entre 3 et 5 km, et une zone périphérique entre 6 et 10 km. Cette structure est cohérente avec un modèle d'exploitation des ressources ligneuses centré sur un rayon d'accès quotidien à pied ou en charrette, classiquement documenté pour la production de charbon de bois dans les zones arides africaines (Laurance et al., 2002; Noromiarilanto et al., 2016). Le fait que les taux de perte restent significativement positifs même en zone périphérique, à 10 km des

habitations, indique que la pression d'extraction s'étend progressivement vers des zones de plus en plus éloignées à mesure que les ressources ligneuses proches des villages se raréfient — un phénomène de front de dégradation mobile bien documenté dans d'autres contextes africains (Vieilledent et al., 2018). Pris ensemble, ces deux résultats — seuil de feux à 8,9 événements cumulés et gradient de distance ($\rho = +0,985$) — fournissent une base empirique solide pour la zonation spatiale des interventions de conservation, développée dans la section 9.5. Ils confirment également la sous-hypothèse H2.2 : la perte de couvert ligneux croît de façon non linéaire avec la fréquence des feux et de façon monotone avec la proximité aux habitations, révélant des effets de seuil dans la réponse écosystémique aux perturbations anthropiques. H2.2 est ainsi pleinement confirmée, avec une précision quantitative qui n'était pas anticipée lors de la formulation initiale des hypothèses.

8.5.3 Une récupération partielle depuis 2019

La quatrième phase identifiée dans la trajectoire temporelle du couvert ligneux (2019–2025, +0,304 %/an) constitue le résultat le plus inattendu du Chapitre 7. Après trois décennies de déclin quasi continu, le couvert ligneux du District de Toliara II amorce un rebond partiel sur les six dernières années de la série. Ce résultat mérite une discussion prudente, car plusieurs explications sont plausibles et les données disponibles ne permettent pas de trancher définitivement entre elles.

Une première explication est climatique. La période 2019–2022 correspond à des anomalies positives de précipitations sur le sud-ouest de Madagascar, associées à la phase La Niña de l'oscillation ENSO (Funk et al., 2015; Hersbach et al., 2020). Dans les formations décidues à phénologie fortement couplée aux précipitations, une série d'années plus humides que la normale peut produire un signal spectral de végétation plus fort, interprété par le modèle comme une augmentation du couvert. Cette explication climatique est plausible mais partielle : elle ne rend pas compte de la persistance du signal positif au-delà de 2022, après le retour à des conditions pluviométriques plus proches de la normale. Par ailleurs, la variable réponse du modèle — le couvert ligneux en projection verticale — est moins sensible aux variations interannuelles de verdissement herbacé que des indices comme le NDVI, ce qui limite l'effet de confusion entre signal ligneux et signal herbacé.

Une deuxième explication est liée à une possible réduction de la pression des feux dans certains secteurs. Les données MODIS MCD64A1 indiquent une diminution de la superficie brûlée annuelle dans plusieurs cellules du secteur nord-est du district entre 2019 et 2025, sans qu'il soit possible d'en identifier clairement la cause — réduction des pratiques de brûlage pastoral, mise en place de conventions locales de gestion des feux, ou simple variabilité interannuelle. Si cette réduction est réelle et durable, elle est cohérente avec les taux de récupération post-feu estimés dans le Chapitre 7 (+0,49 %/an au-delà de dix années sans feu) : une diminution significative de la fréquence des feux depuis 2015 dans certaines zones aurait pu produire un signal de récupération détectable à partir de 2019–2020. Cette hypothèse mériterait d'être testée en croisant les cartes de fréquence des feux avec les cartes de tendance du

couvert à l'échelle des cellules individuelles, ce qui constitue une perspective directe pour les travaux futurs.

Une troisième explication, d'ordre méthodologique, ne peut pas être écartée. La précision du modèle Random Forest dans les classes de couvert élevé est limitée par la saturation spectrale du signal Landsat au-delà de 70 % de couvert. Dans les zones où le couvert résiduel est encore dense, de légères variations dans la composition des composites annuels — liées à la disponibilité des scènes, à la phénologie ou à la couverture nuageuse résiduelle — peuvent introduire un bruit dans les estimations de tendance. Cette incertitude est inhérente à toute analyse fondée sur des séries temporelles de données optiques à 30 m dans un environnement à forte variabilité saisonnière. Elle ne remet pas en cause les tendances de long terme identifiées sur 36 ans, mais invite à interpréter avec prudence les tendances sur des sous-périodes courtes comme celle de 2019–2025. L'intégration de données radar Sentinel-1, moins sensibles aux conditions atmosphériques et à la saturation spectrale, constituerait une voie d'amélioration directe pour affiner ce signal sur les périodes récentes.

8.6. Transférabilité du cadre méthodologique

8.6.1 Conditions de transférabilité et seuils opérationnels

Le protocole développé dans cette thèse — photo-interprétation via Collect Earth Online, calibration par Random Forest, validation spatiale par blocs, application sur série temporelle Landsat dans Google Earth Engine — a été conçu pour répondre à des contraintes spécifiques : végétation ligneuse ouverte à signature spectrale atypique, accès terrain limité, absence de données de calibration préexistantes, et besoins de suivi sur le long terme à coût modéré. Ces contraintes ne sont pas propres au District de Toliara II. Elles caractérisent une large part des zones arides tropicales à travers le monde, ce qui confère au cadre méthodologique développé ici une portée potentiellement plus large que le seul contexte malgache.

Le protocole CEO + Random Forest + blockCV élaboré dans cette thèse a été repris et appliqué dans des écosystèmes forestiers hétérogènes d'Afrique centrale par Mpanda Mukenza et al., (2026, 2025), qui l'ont adapté à la cartographie des types forestiers de la Province de Lualaba (RD Congo). Ce fait atteste que les principes méthodologiques établis ici sont opérationnels au-delà du contexte malgache. Mpanda Mukenza et al., (2025) ont en outre montré que la prise en compte de la phénologie foliaire — et spécifiquement de la saison sèche comme fenêtre d'acquisition optimale — améliore la discrimination des types forestiers dans ce contexte semi-aride (précision globale de 80,66 %), validant ainsi le principe de sélection saisonnière appliqué dans le Chapitre 7.

La transférabilité du protocole est par ailleurs confortée par Agassounon et al., (2025), dont la revue systématique de 121 études en milieux arides et semi-arides désigne la saisonnalité et la résolution spatiale comme les deux obstacles majeurs à la cartographie LUC — obstacles que le protocole développé ici adresse directement. Les auteurs soulignent également que 70 % des études publiées se limitent à une classification binaire forêt/non-forêt, ce qui place la discrimination en cinq classes de

couvert réalisée au Chapitre 6 au-delà de ce que la majorité des approches existantes produisent dans ces environnements.

Plusieurs écosystèmes présentent des conditions suffisamment similaires pour envisager une application directe de ce protocole, moyennant une recalibration locale des seuils d'échantillonnage. La Caatinga brésilienne, formation arbustive sèche à forte endémicité du nord-est du Brésil, partage avec les fourrés xérophiles du sud-ouest de Madagascar une structure ouverte et une phénologie caducifoliée prononcée (Aronson et al., 2018; Fernandes et al., 2020). Ces deux formations constituent des écosystèmes de terres sèches tropicales particulièrement sous-documentés, pour lesquels le déficit de données de référence terrain représente une contrainte majeure à la cartographie par télédétection (Aronson et al., 2018; Pennington et al., 2018). Les forêts claires de miombo du Mozambique et de la Zambie, dominées par des espèces des genres *Brachystegia* et *Julbernardia*, présentent également une hétérogénéité spatiale du couvert et des défis de cartographie par télédétection comparables (Deweese et al., 2010; Ryan et al., 2012). Les savanes arborées du Sahel ouest-africain, malgré leur différence floristique, partagent la même problématique de détection du couvert ligneux en milieu ouvert (Brandt et al., 2020). Dans tous ces contextes, la combinaison photo-interprétation CEO et modélisation RF sur séries Landsat-GEE constitue une approche techniquement accessible, puisqu'elle repose exclusivement sur des données et des outils gratuits disponibles en ligne.

Cependant, la transférabilité des seuils quantitatifs dérivés du Chapitre 6 — 15,5 à 41,4 placettes dispersées par 1 000 km² — doit être conditionnée à des caractéristiques écologiques comparables à celles du District de Toliara II. Ces seuils ont été établis dans un écosystème présentant une hétérogénéité spatiale du couvert modérée à forte, avec cinq classes de couvert bien représentées sur le gradient 0–100 %. Un écosystème plus homogène — par exemple une formation dense avec peu de variation spatiale du couvert — atteindrait probablement la saturation à une densité plus faible. À l'inverse, un écosystème très fragmenté, avec une forte variabilité à courte distance, pourrait nécessiter une densité plus élevée pour atteindre la même précision (Belgiu and Drăguț, 2016; Fassnacht et al., 2014). La recommandation pratique qui découle de ces travaux est donc de procéder à une analyse de saturation locale — en testant plusieurs densités d'échantillonnage sur un sous-ensemble de la zone — avant de définir la densité cible pour un nouveau contexte.

8.6.2 Incertitudes résiduelles et précautions d'interprétation

Il est important de délimiter clairement le périmètre des conclusions de cette thèse pour éviter toute extrapolation non fondée. Premièrement, les dynamiques temporelles reconstituées au Chapitre 7 sont basées sur un modèle calibré en 2022 et appliqué rétrospectivement à 1989. Cette approche suppose que les relations spectrales entre les variables Landsat et le couvert ligneux sont restées stables sur l'ensemble de la période après harmonisation inter-capteurs. Cette hypothèse est raisonnable mais non vérifiée formellement sur des données de terrain historiques indépendantes, qui n'existent pas pour cette région avant 2022. Deuxièmement, les facteurs explicatifs analysés — feux et distance aux habitations — sont deux indicateurs mesurables par

satellite parmi un ensemble plus large de pressions qui incluent le pâturage caprin, la production de charbon, l'agriculture sur brûlis et les sécheresses. Le fait que ces deux variables expliquent une part significative de la variance de la perte de couvert ne signifie pas qu'elles en sont les causes exclusives. Troisièmement, les conclusions sur la récupération partielle depuis 2019 restent incertaines, comme discuté en §9.3.3, et ne peuvent pas être interprétées comme une preuve de régénération écologique sans confirmation par des données terrains récents ou des capteurs à plus haute résolution spatiale.

Une dernière piste de réflexion, qui mériterait d'être explorée plus systématiquement dans des travaux futurs, concerne la probabilité même de détecter des changements statistiquement significatifs dans ce type d'écosystème. La variabilité inter-annuelle naturelle du couvert ligneux xérophile, liée notamment à la phénologie et aux cycles de sécheresse, constitue un bruit de fond qui peut masquer des tendances réelles ou, à l'inverse, produire de fausses significativités sur de courtes fenêtres temporelles. Cette probabilité dépend conjointement de l'ampleur du changement, de la durée de la série temporelle disponible, et du nombre de facteurs explicatifs considérés simultanément : un facteur isolé (les feux, par exemple) peut ne pas atteindre individuellement le seuil de signification statistique, alors que sa combinaison avec d'autres pressions identifiées plus haut (pâturage, agriculture sur brûlis, sécheresses, proximité aux habitations) explique conjointement une part substantielle de la variance observée. Cette interdépendance entre puissance statistique, durée d'observation et nombre de facteurs explicatifs considérés n'a pas été quantifiée dans le présent travail et constitue une piste pour des modèles multivariés futurs intégrant simultanément l'ensemble des pressions identifiées.

8.7. Limites méthodologiques transversales

Plusieurs limites sont communes aux trois chapitres empiriques de cette thèse. Les aborder ici de façon groupée permet d'éviter les répétitions dans chaque chapitre et d'en apprécier la portée réelle sur l'ensemble des conclusions.

La première limite concerne la résolution spatiale des images Landsat (30 m). Dans les formations très ouvertes, où le couvert ligneux est inférieur à 20 % et où les individus ligneux sont isolés ou de petite taille, un pixel de 30 m intègre à la fois la contribution spectrale de la végétation et celle du sol nu environnant. Cette confusion intra-pixel limite la précision des estimations de couvert dans les classes les plus basses du gradient, précisément celles qui correspondent aux stades les plus avancés de dégradation. L'intégration de Sentinel-2 (10 m) ou de données radar Sentinel-1 permettrait de partiellement résoudre ce problème, comme le montre d'ailleurs la supériorité du modèle Sentinel-1+2 sur le modèle Landsat dans le Chapitre 6. Le choix de Landsat a néanmoins été justifié par la nécessité de couvrir une période de 36 ans : Sentinel-2 n'est disponible qu'à partir de 2015, ce qui aurait réduit de moitié la fenêtre temporelle de l'analyse.

La deuxième limite porte sur l'absence de données de feux antérieures à 2001. Le produit MODIS MCD64A1, utilisé pour l'analyse des régimes de feux au Chapitre 7, n'est disponible qu'à partir de novembre 2000. La première phase de la trajectoire

temporelle — la légère expansion du couvert entre 1989 et 2001 (+0,055 %/an) — reste donc aveugle aux effets des feux sur cette période. Or, cette phase coïncide avec une période de relative stabilité institutionnelle à Madagascar avant les crises politiques des années 2000, durant laquelle la pression des feux pastoraux était probablement différente de celle documentée après 2001. L'interprétation de cette phase d'expansion doit donc rester prudente : elle pourrait refléter autant une réduction des feux qu'une réponse climatique ou une pression d'exploitation plus faible.

La troisième limite concerne le biais de sélection des placettes de terrain. Les 41 placettes collectées en 2022 et utilisées comme validation externe dans les Chapitres 7 et 8 ont été établies dans des zones accessibles par piste ou à pied depuis un point d'accès motorisé. Ce biais géographique est inhérent aux conditions logistiques du terrain dans le District de Toliara II, mais il implique que les formations les plus denses et les plus éloignées — précisément celles qui correspondent aux fourrés intacts à couvert élevé — sont sous-représentées dans l'échantillon de validation. La conséquence directe est que les métriques de performance du modèle ($R^2 = 0,75$, RMSE = 13,74 % pour le modèle Landsat) sont probablement légèrement optimistes pour les classes de couvert élevé, où la validation repose sur un nombre réduit de placettes. Cette limite a été partiellement compensée par la validation croisée par blocs spatiaux (blockCV), qui évalue la capacité de généralisation du modèle sur des zones géographiquement distinctes des données d'entraînement.

La quatrième limite est liée à la stationnarité temporelle du modèle Random Forest. Le modèle est calibré sur les relations entre les variables spectrales Landsat et le couvert ligneux mesuré en 2022, puis appliqué à chaque année de la série temporelle couvrant 1989 à 2025. Cette approche suppose que ces relations sont restées stables sur l'ensemble de la période après harmonisation inter-capteurs — une hypothèse raisonnable pour les propriétés physiques de la réflectance, mais qui ne peut pas être vérifiée formellement faute de données de terrain historiques indépendantes antérieures à 2022. Des changements dans la composition floristique des formations au fil du temps — par exemple le remplacement progressif d'espèces ligneuses par des espèces herbacées sous l'effet de la dégradation — pourraient modifier légèrement les relations spectrales et introduire un biais systématique dans les estimations historiques. Ce type de dérive temporelle est difficile à quantifier sans données de référence historiques, et constitue une limite commune à toutes les études de cartographie rétrospective par séries temporelles satellitaires (Hermosilla et al., 2015; Wulder et al., 2022).

La cinquième limite concerne la part de subjectivité inhérente à la photo-interprétation sur Collect Earth Online. Contrairement à la méthode point-intercept sur le terrain, qui repose sur des critères de mesure physiques objectifs, la photo-interprétation fait appel au jugement visuel de l'opérateur pour estimer le couvert ligneux à partir d'images à très haute résolution. Cette dimension subjective peut introduire une variabilité entre opérateurs qui affecte la cohérence des données de calibration. Dans les Chapitres 7 et 8, deux opérateurs ont suivi un protocole standardisé avec des critères d'interprétation définis et des sessions de calibration préalables, mais aucune évaluation formelle de la concordance inter-opérateurs n'a été

conduite. Cette lacune est particulièrement problématique dans les formations ouvertes et clairsemées, où la discrimination entre végétation ligneuse basse et végétation herbacée est visuellement ambiguë sur les images disponibles, même à très haute résolution. La comparaison directe entre les estimations photo-interprétées et les mesures terrain sur les 41 co-placettes ($R^2 = 0,775$; $RMSE = 0,125$; biais = $+0,042$) fournit une mesure partielle de cette incertitude, mais ne remplace pas une évaluation formelle de la variabilité inter-opérateurs. Les études futures devraient systématiquement inclure une étape de validation croisée sur un sous-échantillon commun interprété par plusieurs opérateurs indépendants, afin de quantifier et de minimiser ce biais (Meyer et al., 2024; Ploton et al., 2020).

8.8. Aspects innovants de la thèse

Les résultats de cette thèse mettent en évidence trois contributions majeures à l'étude de la végétation ligneuse dans les régions arides.

Premièrement, ils démontrent l'intérêt d'une approche fondée sur des variables continues de couverture végétale plutôt que sur une classification binaire forêt/non-forêt. Cette approche permet de mieux représenter les gradients de végétation et de détecter les processus progressifs de dégradation souvent masqués par les classifications catégorielles.

Deuxièmement, ils montrent qu'une calibration locale basée sur 505 placettes de référence améliore significativement la précision des estimations par rapport aux produits globaux existants, en particulier dans les fourrés xérophiles du sud-ouest de Madagascar, caractérisés par des signatures spectrales atypiques.

Troisièmement, ils permettent d'identifier des seuils d'échantillonnage opérationnels pour la production de cartes de végétation, dont la robustesse a été évaluée dans un second contexte africain.

Ces contributions méthodologiques et thématiques, ainsi que leur lien avec les hypothèses H1.1 à H2.2, sont synthétisés dans la conclusion générale (Chapitre 9). Elles ouvrent également plusieurs perspectives de recherche, notamment la validation des seuils d'échantillonnage dans d'autres environnements arides, l'automatisation de la détection des arbres et arbustes à partir d'images à très haute résolution, l'intégration de données radar et climatiques, le développement de dispositifs de suivi participatif, le couplage avec des modèles d'estimation de la biomasse et du carbone, ainsi que l'adaptation de cette approche méthodologique à d'autres types d'écosystèmes forestiers.

8.9. Perspectives

Les résultats de cette thèse ouvrent six axes de recherche dont la poursuite permettrait de consolider, d'étendre et d'opérationnaliser les contributions développées dans les trois chapitres empiriques.

La priorité concerne la validation des seuils d'échantillonnage établis dans le Chapitre 5 dans d'autres contextes arides tropicaux. Les seuils opérationnels dérivés — seuil minimal de 15,5 placettes dispersées par 1 000 km², seuil optimal de 20,7 à

41,4 placettes par 1 000 km² — ont été établis dans un écosystème spécifique à signatures spectrales atypiques. Le protocole a été repris par Mpanda Mukenza et al. (2025b, 2026) dans les forêts semi-arides de Lualaba, ce qui constitue une première application externe dans un contexte écologique distinct et suggère que les principes sous-jacents — si pas les seuils numériques exacts — sont transférables. Une validation croisée systématique sur d'autres écosystèmes arides permettrait de déterminer dans quelle mesure ces seuils varient en fonction de paramètres structuraux mesurables.

La deuxième perspective concerne le développement de méthodes de détection automatique des arbres et arbustes à partir d'images à très haute résolution. Les modèles de vision basés sur l'architecture Transformer — et en particulier les modèles à auto-attention comme DINO (Self-Distillation with NO labels) — ont récemment démontré leur capacité à segmenter les couronnes individuelles dans des paysages hétérogènes sans supervision explicite (Caron et al., 2021). Appliqués aux images Pléiades ou aux orthophotos disponibles sur Google Earth dans le District de Toliara II, ces modèles pourraient permettre de distinguer automatiquement les arbres des arbustes sur la base du diamètre de la couronne projetée au sol, résolvant ainsi la confusion arbre/arbuste qui constitue la source principale d'erreur identifiée dans le Chapitre 5 pour les produits globaux. Cette cartographie du couvert arboré basée sur le diamètre de la canopée — plutôt que sur un seuil de hauteur ou de classe spectrale — fournirait des données de calibration à la fois plus précises et plus cohérentes avec les définitions écologiques employées dans les inventaires terrain, réduisant le biais systématique documenté au Chapitre 2 (§2.4) entre couvert arboré et couvert ligneux agrégé.

La troisième perspective concerne l'intégration de données radar et de variables climatiques pour affiner l'interprétation du signal de récupération identifié sur 2019–2025. La résolution de 30 m du capteur Landsat ne permet pas de résoudre les stades précoces de régénération post-feu dans les formations succulentes aphyllées, et la série temporelle mobilisée ne distingue pas la contribution respective de la variabilité pluviométrique et de la réduction de la pression anthropique dans l'explication de ce rebond. L'intégration de données Sentinel-1, dont la sensibilité à la structure du couvert est indépendante des conditions d'éclairage (Reiche et al., 2018), et de séries climatiques mensuelles à haute résolution spatiale issues de CHIRPS (Funk et al., 2015) permettrait de décomposer ce signal et d'évaluer la résilience intrinsèque des fourrés xérophiles face aux perturbations.

Dans la même logique d'enrichissement des facteurs explicatifs, les analyses des régimes de feux et de la proximité aux zones bâties menées au Chapitre 7 pourraient être étendues à d'autres variables explicatives non testées dans le présent travail : la distance aux routes et aux cours d'eau, des variables topographiques (pente, altitude, exposition), ainsi que les types de sols. Ces variables, disponibles à l'échelle régionale sans nécessiter d'acquisition de données supplémentaires, pourraient être intégrées dans un modèle explicatif multivarié de la dégradation et ainsi préciser la part de variance respectivement attribuable à l'accessibilité, au relief et aux propriétés

pédologiques, en complément des facteurs feux et proximité aux habitations déjà documentés.

La quatrième perspective porte sur le développement d'un dispositif de suivi écologique intégré, associant la surveillance satellitaire à l'implication directe des communautés locales. Si la télédétection permet de quantifier le couvert ligneux à l'échelle du district, la composition floristique, la structure verticale et les usages associateurs aux formations ne peuvent être appréhendés que par des relevés participatifs conduits avec les communautés d'éleveurs et d'agriculteurs du district. Un protocole de monitoring communautaire basé sur des transects annuels simples, couplé à la mise à jour des cartes CEO, permettrait aux gestionnaires VOI de détecter localement les événements de dégradation (feux non autorisés, coupes illégales) avant qu'ils ne soient détectables par le satellite, et de valider les signaux de récupération identifiés dans la série temporelle. Ce couplage télédétection-relevés participatifs constitue une voie directe pour renforcer l'appropriation locale des outils de surveillance et leur intégration dans les systèmes de gestion durable des ressources naturelles.

La cinquième perspective concerne le couplage des cartes de couvert avec des équations allométriques pour l'estimation de la biomasse aérienne et des stocks de carbone. Les équations allométriques spécifiques aux arbustes xérophiiles du sud-ouest malgache récemment développées par Randriamalala et al., (2022) pour les principales espèces des familles Didiereaceae, Euphorbiaceae et Fabaceae constituent une ressource directement mobilisable. Ce couplage permettrait de reconstruire les trajectoires de stock de carbone sur la période 1989–2025 et de quantifier les flux associés aux phases de dégradation et de récupération identifiées dans le Chapitre 7 — fournissant les éléments nécessaires à l'intégration de ces formations dans les mécanismes REDD+, dont leur exclusion actuelle des bilans forestiers conventionnels constitue précisément l'une des injustices documentées dans le Chapitre 2 (§2.4).

Une sixième perspective, à plus long terme, concerne la généralisation du cadre méthodologique aux forêts sèches et aux forêts humides. Le protocole développé dans cette thèse a été conçu spécifiquement pour des formations ligneuses ouvertes à signatures spectrales atypiques ; son application aux forêts denses tropicales nécessiterait des adaptations. Dans les forêts sèches à phénologie fortement caducifoliée (Caatinga, forêts de transition à l'est de Madagascar), la fenêtre temporelle d'acquisition optimale devrait être recalée sur les périodes de feuillaison maximale pour maximiser le contraste spectral entre végétation et sol — un principe que Mpanda Mukenza et al., (2025) ont appliqué, à partir du cadre développé ici, dans les forêts semi-arides de Lualaba. Dans les forêts humides à canopée fermée, la saturation spectrale du signal Landsat au-delà de 70 % de couvert limiterait fortement la capacité discriminante du modèle ; l'intégration de données LiDAR GEDI ou de données radar SAR serait alors indispensable pour capturer la structure verticale de la canopée. Ces adaptations représentent une voie d'extension naturelle vers des biomes forestiers plus diversifiés.

Chapitre 9

Conclusion générale

Cette thèse avait pour objectif de développer et de valider un cadre méthodologique adapté à la cartographie et au suivi du couvert ligneux des fourrés xérophiles du District de Toliara II, dans le sud-ouest de Madagascar, et d'en analyser les dynamiques sur 36 ans à partir d'une série temporelle Landsat harmonisée. Elle s'inscrit dans un contexte de déficit documentaire reconnu pour les milieux arides tropicaux, où les outils de surveillance disponibles à l'échelle mondiale peinent à restituer fidèlement la structure et la dynamique des végétations ligneuses ouvertes (Bastin et al., 2017; Brandt et al., 2020). Deux hypothèses principales ont guidé la démarche, déclinées en quatre sous-hypothèses testées empiriquement à travers trois chapitres empiriques.

La première hypothèse (H1.1), qui postulait la prédominance des méthodes classiques de télédétection et la sous-représentation des végétations ouvertes dans la littérature cartographique sur le sud-ouest de Madagascar, est confirmée. Le Chapitre 2 (§2.4) documente que les études disponibles reposent majoritairement sur des classifications non supervisées à faible résolution temporelle, et que les formations xérophiles — pourtant dominantes dans le domaine subaride — sont systématiquement marginalisées au profit des forêts denses dans les bilans nationaux.

La deuxième hypothèse (H1.2), relative à l'hétérogénéité substantielle des produits globaux et à son lien avec la composition floristique, est également confirmée. Le Chapitre 5 établit un coefficient de variation inter-produits de 78,94 % pour le couvert arboré, structuré par la nature des formations : les Euphorbiaceae et Didiereaceae succulentes génèrent les plus forts désaccords entre produits, en raison de leur signature spectrale atypique. Cette hétérogénéité se réduit à 20,75 % lorsque l'on agrège arbres et arbustes, ouvrant une voie d'utilisation conditionnelle des produits globaux sous réserve de validation locale. Par ailleurs, la comparaison directe entre le meilleur produit global testé — Dynamic World ($R^2=0,70$ sur les 41 placettes terrain indépendantes) — et les modèles Random Forest locaux calibrés sur données CEO ($R^2=0,75-0,88$ selon le capteur) quantifie la valeur ajoutée de la calibration locale dans ce type d'écosystème.

La troisième hypothèse (H2.1), qui affirmait la primauté du plan d'échantillonnage de calibration sur la sophistication des capteurs, est confirmée et enrichie. Le Chapitre 6 démontre que 75 placettes photo-interprétées via Collect Earth Online, bien distribuées sur le gradient de couvert, surpassent systématiquement les 41 placettes terrain plus précises mais spatialement biaisées ($R^2=0,88$ contre $0,46-0,66$). Un seuil minimal de 15,5 placettes dispersées par 1 000 km² assure des performances acceptables ($R^2 \geq 0,82$), tandis qu'un seuil optimal de 20,7 à 41,4 placettes par 1 000 km² maximise le rapport précision/effort — constituant le premier repère empirique quantifié pour la calibration des modèles de couvert ligneux en milieu aride tropical. L'échantillonnage dispersé atteint en outre des performances équivalentes à l'échantillonnage groupé avec moitié moins de placettes, ce qui représente un avantage logistique et économique considérable pour les programmes de surveillance à ressources limitées. Enfin, la validation croisée par blocs spatiaux (blockCV) s'impose comme le seul protocole de validation garantissant la généralisation géographique des

modèles, une exigence méthodologique sous-utilisée dans les études de cartographie par télédétection en milieu aride.

La quatrième hypothèse (H2.2), relative à la non-linéarité des dynamiques temporelles et aux effets de seuil dans la réponse aux perturbations, est confirmée. Le Chapitre 7 identifie quatre phases distinctes dans la trajectoire du couvert entre 1989 et 2025 et démontre qu'une perte nette limitée ($-3,2\%$; $-0,090\%/an$) masque une transformation structurelle profonde : les fourrés intacts ($>80\%$ de couvert) ont reculé de $62,8\%$ à $29,0\%$ du paysage, au bénéfice des fourrés dégradés ($60-80\%$) ; une régression robuste (Theil-Sen) indique cependant qu'aucune de ces quatre tendances de phase ne diffère significativement de zéro au seuil de 95% , la transformation structurelle (intacts vers dégradés) demeurant le résultat le plus robuste de cette analyse. La dégradation s'accélère non linéairement au-delà de $8,9$ feux cumulés depuis 2000 — soit approximativement un feu tous les trois ans sur la période 2000–2025 —, et le taux de perte décroît de façon monotone avec la distance aux habitations, de $-0,166\%/an$ à 1 km à $-0,090\%/an$ à 10 km ($\rho=+0,985$). Une phase de récupération partielle est identifiée sur 2019–2025 ($+0,304\%/an$), cohérente avec une réduction de la pression des feux dans les secteurs où la fréquence cumulée reste inférieure au seuil critique — résultat non anticipé qui ouvre des perspectives sur la résilience passive de ces formations. Par ailleurs, dans les zones pastorales où la fréquence des feux demeure sous ce seuil, le couvert ligneux se révèle stable ou légèrement croissant, suggérant que le pâturage extensif de petits ruminants n'induit pas, en lui-même, de perte directionnelle du couvert — un résultat contre-intuitif aux implications directes pour la gestion.

Sur le plan méthodologique, cette thèse apporte quatre contributions originales. Elle fournit le premier cadre empirique quantifié pour l'optimisation du plan de calibration des modèles de couvert ligneux en zone aride tropicale, en isolant simultanément la source des données de référence, la configuration spatiale de l'échantillon et la densité de placettes. Elle démontre, pour la première fois dans ce type d'écosystème, que la photo-interprétation sur Collect Earth Online constitue une alternative robuste et plus efficiente que les inventaires terrain pour la calibration spatiale, à condition que l'échantillon soit suffisamment dispersé. Elle établit que l'échantillonnage dispersé divise par deux l'effort nécessaire par rapport à l'échantillonnage groupé pour une précision équivalente, et que la validation croisée par blocs spatiaux est indispensable pour évaluer la généralisation géographique réelle des modèles. Elle établit enfin que la variable continue de couvert ($0-100\%$) est indispensable pour détecter la dégradation qualitative des fourrés xérophiles, que les classifications binaires sont structurellement incapables de révéler. La portée de ces contributions méthodologiques dépasse le seul contexte malgache. Le protocole CEO + Random Forest + blockCV élaboré dans cette thèse a été repris et appliqué dans des écosystèmes forestiers comparables par Mpanda Mukenza et al., (2026, 2025), qui l'ont adapté à la cartographie des types forestiers de la Province de Lualaba (RD Congo), attestant que les principes méthodologiques développés ici sont opérationnels au-delà du contexte malgache.

Sur le plan thématique, cette thèse fournit la première reconstruction annuelle du couvert ligneux de Toliara II sur 36 ans, identifie des seuils critiques de fréquence de feux opérationnellement utilisables pour la gestion spatiale de la conservation, caractérise avec précision le gradient d'influence anthropique et ses effets non linéaires sur la structure du paysage, et démontre quantitativement que les modèles de cartographie localement calibrés surpassent les produits globaux dans les écosystèmes xérophiiles tropicaux.

Références

Abduallah, M., 2018. Satellite vs. UAVs Remote Sensing of Arid Ecosystems: A Review with in an Ecological Perspective. EAES 2. <https://doi.org/10.31031/EAES.2018.02.000540>

Abdullah, M.M., Al-Ali, Z.M., Abdullah, M.T., Al-Anzi, B., 2021. The Use of Very-High-Resolution Aerial Imagery to Estimate the Structure and Distribution of the *Rhanterium epapposum* Community for Long-Term Monitoring in Desert Ecosystems. *Plants* 10, 977. <https://doi.org/10.3390/plants10050977>

Abel, C., Abdi, A.M., Tagesson, T., Horion, S., Fensholt, R., 2023. Contrasting ecosystem vegetation response in global drylands under drying and wetting conditions. *Global Change Biology* 29, 3954–3969. <https://doi.org/10.1111/gcb.16745>

Abel, C., Horion, S., Tagesson, T., Brandt, M., Fensholt, R., 2019. Towards improved remote sensing based monitoring of dryland ecosystem functioning using sequential linear regression slopes (SeRGS). *Remote Sensing of Environment* 224. <https://doi.org/10.1016/j.rse.2019.02.010>

Acharki, S., Frison, P.L., Amharref, M., Khoj, H., Bernoussi, S., 2021. Complémentarité des images optiques SENTINEL-2 avec les images radar SENTINEL-1 et ALOS-PALSAR-2 pour la cartographie de la couverture végétale : application à une aire protégée et ses environs au Nord-Ouest du Maroc via trois algorithmes d'apprentissage automatique. *RFPT* 223, 143–158. <https://doi.org/10.52638/rfpt.2021.599>

Adjognon, G.S., Rivera-Ballesteros, A., van Soest, D., 2019. Satellite-based tree cover mapping for forest conservation in the drylands of Sub Saharan Africa (SSA): Application to Burkina Faso gazetted forests. *Development Engineering* 4, 100039.

Adzhar, R., Kelley, D.I., Dong, N., George, C., Torello Raventos, M., Veenendaal, E., Feldpausch, T.R., Phillips, O.L., Lewis, S.L., Sonké, B., Taedoumg, H., Schwantes Marimon, B., Domingues, T., Arroyo, L., Djagbletey, G., Saiz, G., Gerard, F., 2022. MODIS Vegetation Continuous Fields tree cover needs calibrating in tropical savannas. *Biogeosciences* 19, 1377–1394. <https://doi.org/10.5194/bg-19-1377-2022>

Adzhar, R., Kelley, D.I., Dong, N., Torello Raventos, M., Veenendaal, E., Feldpausch, T.R., Philips, O.L., Lewis, S., Sonké, B., Taedoumg, H., Schwantes Marimon, B., Domingues, T., Arroyo, L., Djagbletey, G., Saiz, G., Gerard, F., 2021. Assessing MODIS Vegetation Continuous Fields tree cover product(collection 6): performance and applicability in tropical forests and savannas. <https://doi.org/10.5194/bg-2020-460>

Agassounon, B.M., Assede, E.S.P., Bastin, J.-F., Biaou, S.S.H., 2025. Remote sensing applied in land use and land cover change (LUCC) in arid and semi-arid ecosystems: Current status, challenges and prospects – A systematic review. *Environ Monit Assess* 197, 1266. <https://doi.org/10.1007/s10661-025-14755-3>

Aleman, J.C., Blarquez, O., Staver, C.A., 2016. Land-use change outweighs projected effects of changing rainfall on tree cover in sub-Saharan Africa. *Global Change Biology* 22, 3013–3025. <https://doi.org/10.1111/gcb.13299>

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Moghaddam, S.H.A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., Brisco, B., 2020. Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, 5326–5350. <https://doi.org/10.1109/JSTARS.2020.3021052>

Antonelli, A., Smith, R.J., Perrigo, A.L., Crottini, A., Hackel, J., Testo, W., Farooq, H., Torres Jiménez, M.F., Andela, N., Andermann, T., Andriamanohera, A.M., Andriambololonera, S., Bachman, S.P., Bacon, C.D., Baker, W.J., Belluardo, F., Birkinshaw, C., Borrell, J.S., Cable, S., Canales, N.A., Carrillo, J.D., Clegg, R., Clubbe, C., Cooke, R.S.C., Damasco, G., Dhanda, S., Edler, D., Faurby, S., De Lima Ferreira, P., Fisher, B.L., Forest, F., Gardiner, L.M., Goodman, S.M., Grace, O.M., Guedes, T.B., Henniges, M.C., Hill, R., Lehmann, C.E.R., Lowry, P.P., Marline, L., Matos-Maraví, P., Moat, J., Neves, B., Nogueira, M.G.C., Onstein, R.E., Papadopulos, A.S.T., Perez-Escobar, O.A., Phelps, L.N., Phillipson, P.B., Pironon, S., Przelomska, N.A.S., Rabarimanarivo, M., Rabehevitra, D., Raharimampionona, J., Rajaonah, M.T., Rajaonary, F., Rajaovelona, L.R., Rakotoarinivo, M., Rakotoarisoa, A.A., Rakotoarisoa, S.E., Rakotomalala, H.N., Rakotonasolo, F., Ralaiveloarisoa, B.A., Ramirez-Herranz, M., Randriamamonjy, J.E.N., Randriamboavonjy, T., Randrianasolo, V., Rasolohery, A., Ratsifandrihamanana, A.N., Ravololomanana, N., Razafiniary, V., Razanajatovo, H., Razanatosoa, E., Rivers, M., Sayol, F., Silvestro, D., Vorontsova, M.S., Walker, K., Walker, B.E., Wilkin, P., Williams, J., Ziegler, T., Zizka, A., Ralimanana, H., 2022. Madagascar's extraordinary biodiversity: Evolution, distribution, and use. *Science* 378, eabf0869. <https://doi.org/10.1126/science.abf0869>

Archibald, S., 2016. Managing the human component of fire regimes: lessons from Africa. *Phil. Trans. R. Soc. B* 371, 20150346. <https://doi.org/10.1098/rstb.2015.0346>

Archibald, S., Lehmann, C.E.R., Gómez-Dans, J.L., Bradstock, R.A., 2013. Defining pyromes and global syndromes of fire regimes. *Proceedings of the National Academy of Sciences* 110, 6442–6447. <https://doi.org/10.1073/pnas.1211466110>

Aronson, J.C., Phillipson, P.B., Le Floch, E., Raminosoa, T., 2018. Dryland tree data for the Southwest region of Madagascar: alpha-level data can support policy decisions for conserving and restoring ecosystems of arid and semiarid regions. *MCD* 13, 60. <https://doi.org/10.4314/mcd.v13i1.7>

Aubert, S., Rahajason, F., Ganomanana, T., 2012. La modélisation d'accompagnement pour le Suivi de l'Impact des Transferts de Gestion à Madagascar. *vertigo* 11–3. <https://doi.org/10.4000/vertigo.11888>

Bachantourian, M., Chaleplis, K., Gemitzi, A., Kalabokidis, K., Palaiologou, P., Vasilakos, C., 2022. Evaluation of MODIS, Climate Change Initiative, and CORINE Land Cover Products Based on a Ground Truth Dataset in a Mediterranean Landscape. *Land* 11, 1453. <https://doi.org/10.3390/land11091453>

Bajirao, T.S., Kumar, P., Kumar, A., 2018. Spatio-Temporal Variability of Land use/Land Cover within Koyna River Basin. *Int. J. Curr. Microbiol. App. Sci* 7, 944–953.

Bastin, J.-F., Berrahmouni, N., Grainger, A., Maniatis, D., Mollicone, D., Moore, R., Patriarca, C., Picard, N., Sparrow, B., Abraham, E.M., 2017. The extent of forest in dryland biomes. *Science* 356, 635–638.

Bastin, J.-F., Finegold, Y., Garcia, C., Mollicone, D., Rezende, M., Routh, D., Zohner, C.M., Crowther, T.W., 2019. The global tree restoration potential. *Science* 365. <https://doi.org/10.1126/science.aax0848>

Bastin, J.-F., Latte, N., Bogaert, J., Garcia, C.A., Berzaghi, F., Maestre, F.T., Svenning, J.-C., Assede, E., Barima, S., Besisa, T., Bouchoms, S., Crowther, T.W., de Haulleville, T., de Lame, H., Depoortere, P., Dufrêne, M., Hoek Van Dijke, A.J., Lhoest, S., Mahy, G., Messier, C., Mollicone, D., Ramalason, F.N., Peaucelle, M., Plumacker, A., Quétier, F., Rakotondrasoa, O., Sambieni, K.R., Sparrow, B., Tarelkin, Y., Sikuzani, Y.U., Vander Linden, A., Lejeune, P., 2025. Global alternatives of natural vegetation cover. *Nat Commun* 16, 6484. <https://doi.org/10.1038/s41467-025-61520-8>

Baylis, J., Smith, S., Owens, P. (Eds.), 2017. *The Globalization of World Politics*. Oxford University Press. <https://doi.org/10.1093/hepl/9780198739852.001.0001>

Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J.-F., Moore, R., Federici, S., Rezende, M., Patriarca, C., Turia, R., Gamoga, G., Abe, H., Kaidong, E., Miceli, G., 2016. Collect Earth: Land Use and Land Cover Assessment through Augmented Visual Interpretation. *Remote Sensing* 8. <https://doi.org/10.3390/rs8100807>

Bond, W.J., Keeley, J.E., 2005. Fire as a global ‘herbivore’: the ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution* 20, 387–394. <https://doi.org/10.1016/j.tree.2005.04.025>

Bond, W.J., Van Wilgen, B.W., 1996. *Fire and plants*, 1. ed. ed, Population and community biology series. Chapman & Hall, London.

Brandt, J., Stolle, F., 2021. A global method to identify trees outside of closed-canopy forests with medium-resolution satellite imagery. *International Journal of Remote Sensing* 42, 1713–1737.

Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A.A., Mbow, C., Mougin, E., Fensholt, R., 2016. Woody plant cover estimation in drylands from Earth Observation based seasonal metrics.

Brandt, M., Rasmussen, K., Hiernaux, P., Herrmann, S., Tucker, C.J., Tong, X., Tian, F., Mertz, O., Kergoat, L., Mbow, C., David, J.L., Melocik, K.A., Dendoncker, M., Vincke, C., Fensholt, R., 2018. Reduction of tree cover in West African woodlands and promotion in semi-arid farmlands. *Nature Geoscience* 11, 328–333. <https://doi.org/10.1038/s41561-018-0092-x>

Brandt, M., Rasmussen, K., Peñuelas, J., Tian, F., Schurgers, G., Verger, A., Mertz, O., Palmer, J.R.B., Fensholt, R., 2017. Human population growth offsets climate-

driven increase in woody vegetation in sub-Saharan Africa. *Nat Ecol Evol* 1, 0081. <https://doi.org/10.1038/s41559-017-0081>

Brandt, M., Tucker, C.J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L.V., Hiernaux, P., Diouf, A.A., Kergoat, L., Mertz, O., Igel, C., Gieseke, F., Schöning, J., Li, S., Melocik, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Montagu, A., Dendoncker, M., Fensholt, R., 2020. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* 587, 78–82. <https://doi.org/10.1038/s41586-020-2824-5>

Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>

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? *Agriculture, Ecosystems & Environment* 195. <https://doi.org/10.1016/j.agee.2014.06.008>

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. *Sci Data* 9, 251. <https://doi.org/10.1038/s41597-022-01307-4>

Brown, C.F., Kazmierski, M.R., Pasquarella, V.J., Rucklidge, W.J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L.L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P., 2025. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. <https://doi.org/10.48550/ARXIV.2507.22291>

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020. Copernicus Global Land Cover Layers—Collection 2. *Remote Sensing* 12, 1044. <https://doi.org/10.3390/rs12061044>

Cao, L., Pan, J., Li, R., Li, J., Li, Z., 2018. Integrating Airborne LiDAR and Optical Data to Estimate Forest Aboveground Biomass in Arid and Semi-Arid Regions of China. *Remote Sensing* 10. <https://doi.org/10.3390/rs10040532>

Cardille, J.A., Crowley, M.A., Saah, D., Clinton, N.E. (Eds.), 2024. *Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-031-26588-4>

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging Properties in Self-Supervised Vision Transformers. <https://doi.org/10.48550/ARXIV.2104.14294>

Carrière, S., Randriambanona, H., Labeyrie, V., Hervé, D., Mariel, J., Razanaka, S., Randriamalala, J., 2022. Biocultural landscapes diversity shaped by agricultural systems in Madagascar.

Casse, T., Milhøj, A., Ranaivoson, S., Randriamanarivo, J.R., 2004. Causes of deforestation in southwestern Madagascar: what do we know? *Forest Policy and Economics* 6, 33–48. [https://doi.org/10.1016/S1389-9341\(02\)00084-9](https://doi.org/10.1016/S1389-9341(02)00084-9)

- Chanteau, S. (Ed.), 2006. Les pluies et la végétation, in: Atlas de la peste à Madagascar. IRD Éditions, pp. 19–22. <https://doi.org/10.4000/books.irdeditions.6600>
- Chazdon, R.L., Guariguata, M.R., 2016. Natural regeneration as a tool for large-scale forest restoration in the tropics: prospects and challenges. *Forests, Trees and Agroforestry*. URL <https://www.foreststreesagroforestry.org/natural-regeneration-as-a-tool-for-large-scale-forest-restoration-in-the-tropics-prospects-and-challenges/> (accessed 6.24.26).
- Cherif, E.K., Lucas, R., Ait Tchakoucht, T., Gama, I., Ribeiro, I., Domingos, T., Proença, V., 2024. Predicting Fractional Shrub Cover in Heterogeneous Mediterranean Landscapes Using Machine Learning and Sentinel-2 Imagery. *Forests* 15, 1739. <https://doi.org/10.3390/f15101739>
- Chomitz, K.M., Gray, D.A., 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. *World Bank Econ Rev* 10, 487–512. <https://doi.org/10.1093/wber/10.3.487>
- Cochran, W.G., 1977. Sampling techniques, 3. ed. ed, Wiley series in probability and mathematical statistics. Wiley, New York, NY.
- Colditz, R.R., Schmidt, M., Conrad, C., Hansen, M.C., Dech, S., 2011. Land cover classification with coarse spatial resolution data to derive continuous and discrete maps for complex regions. *Remote Sensing of Environment* 115, 3264–3275. <https://doi.org/10.1016/j.rse.2011.07.010>
- Congalton, R.G., Green, K., 2019. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third Edition, 3rd ed. CRC Press. <https://doi.org/10.1201/9780429052729>
- Cornet, A., Guillaumet, J.L., Cornet, A., Guillaumet, J.L., 1976. Floristic divisions and vegetation units in Madagascar. *Cahiers ORSTOM, Biologie* 11, 35–42.
- Crawford, C.J., Roy, D.P., Arab, S., Barnes, C., Vermote, E., Hulley, G., Gerace, A., Choate, M., Engebretson, C., Micijevic, E., Schmidt, G., Anderson, C., Anderson, M., Bouchard, M., Cook, B., Dittmeier, R., Howard, D., Jenkerson, C., Kim, M., Kleyians, T., Maiersperger, T., Mueller, C., Neigh, C., Owen, L., Page, B., Pahlevan, N., Rengarajan, R., Roger, J.-C., Sayler, K., Scaramuzza, P., Skakun, S., Yan, L., Zhang, H.K., Zhu, Z., Zahn, S., 2023. The 50-year Landsat collection 2 archive. *Science of Remote Sensing* 8, 100103. <https://doi.org/10.1016/j.srs.2023.100103>
- Crist, E.P., Cicone, R.C., 1984. A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing* GE-22, 256–263. <https://doi.org/10.1109/TGRS.1984.350619>
- Croker, A.R., Woods, J., Kountouris, Y., 2023. Community-Based Fire Management in East and Southern African Savanna-Protected Areas: A Review of the Published Evidence. *Earth's Future* 11, e2023EF003552. <https://doi.org/10.1029/2023EF003552>
- Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M., Smith, J.R., Hintler, G., Duguid, M.C., Amatulli, G., Tuanmu, M.-N., Jetz, W., Salas, C., Stam, C., Piotto, D., Tavani, R., Green, S., Bruce, G., Williams, S.J., Wiser, S.K., Huber, M.O., Hengeveld, G.M., Nabuurs, G.-J., Tikhonova, E., Borchardt, P.,

Li, C.-F., Powrie, L.W., Fischer, M., Hemp, A., Homeier, J., Cho, P., Vibrans, A.C., Umunay, P.M., Piao, S.L., Rowe, C.W., Ashton, M.S., Crane, P.R., Bradford, M.A., 2015. Mapping tree density at a global scale. *Nature* 525, 201–205. <https://doi.org/10.1038/nature14967>

D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, T. L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE* 50, 885–900. <https://doi.org/10.13031/2013.23153>

Dardel, C., Kergoat, L., Hiernaux, P., Mougin, E., Grippa, M., Tucker, C.J., 2014. Re-greening Sahel: 30years of remote sensing data and field observations (Mali, Niger). *Remote Sensing of Environment* 140, 350–364. <https://doi.org/10.1016/j.rse.2013.09.011>

De Marzo, T., Pflugmacher, D., Baumann, M., Lambin, E.F., Gasparri, I., Kuemmerle, T., 2021. Characterizing forest disturbances across the Argentine Dry Chaco based on Landsat time series. *International Journal of Applied Earth Observation and Geoinformation* 98. <https://doi.org/10.1016/j.jag.2021.102310>

DeVries, B., Pratihast, A.K., Verbesselt, J., Kooistra, L., Herold, M., 2016. Characterizing Forest Change Using Community-Based Monitoring Data and Landsat Time Series. *PLoS ONE* 11, e0147121. <https://doi.org/10.1371/journal.pone.0147121>

DeVries, B., Verbesselt, J., Kooistra, L., Herold, M., 2015. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sensing of Environment* 161, 107–121. <https://doi.org/10.1016/j.rse.2015.02.012>

Deweese, P.A., Campbell, B.M., Katerere, Y., Siteo, A., Cunningham, A.B., Angelsen, A., Wunder, S., 2010. Managing the Miombo Woodlands of Southern Africa: Policies, Incentives and Options for the Rural Poor. *Journal of Natural Resources Policy Research* 2, 57–73. <https://doi.org/10.1080/19390450903350846>

Diédhiou, I., Mering, C., Sy, O., Sané, T., 2020. Cartographie par télédétection l'occupation du sol et ses changements: Application à l'analyse de la dynamique des paysages forestiers sénégalais entre 1972 et 2016. *echogeo* 54. <https://doi.org/10.4000/echogeo.20510>

D'Odorico, P., Bhattachan, A., Davis, K.F., Ravi, S., Runyan, C.W., 2013. Global desertification: drivers and feedbacks. *Advances in Water Resources* 51, 326–344. <https://doi.org/10.1016/j.advwatres.2012.01.013>

Dostálová, A., Doubková, M., Sabel, D., Bauer-Marschallinger, B., Wagner, W., 2014. Seven Years of Advanced Synthetic Aperture Radar (ASAR) Global Monitoring (GM) of Surface Soil Moisture over Africa. *Remote Sensing* 6, 7683–7707. <https://doi.org/10.3390/rs6087683>

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P.L., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing* 1, 100002. <https://doi.org/10.1016/j.srs.2020.100002>

Eldridge, D.J., Bowker, M.A., Maestre, F.T., Roger, E., Reynolds, J.F., Whitford, W.G., 2011. Impacts of shrub encroachment on ecosystem structure and functioning: towards a global synthesis. *Ecology letters* 14, 709–722. <https://doi.org/10.1111/j.1461-0248.2011.01630.x>

Ermakov, M.S., 2023. Chi-Squared Test for Hypothesis Testing of Homogeneity. *J Math Sci* 273, 763–777. <https://doi.org/10.1007/s10958-023-06539-2>

Fagan, M.E., 2020. A lesson unlearned? Underestimating tree cover in drylands biases global restoration maps. *Global Change Biology* 26, 4679–4690. <https://doi.org/10.1111/gcb.15187>

FAO, 2020a. Évaluation des ressources forestières mondiales 2020, Évaluation des ressources forestières mondiales 2020. FAO. <https://doi.org/10.4060/ca9825fr>

FAO, 2020b. Global Forest Resources Assessment 2020 -Keys findings. FAO, Rome. <https://doi.org/10.4060/ca8753en>

FAO, 2010. Global Forest Resources Assessment 2010 [WWW Document]. GlobalFRA. URL <https://www.fao.org/forest-resources-assessment/past-assessments/fra-2010/en> (accessed 4.20.26).

Faramalala, M.A., 1988. Etude de la vegetation de madagascar a l'aide des donnees spatiales (thesis). Toulouse 3.

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 Sensing of Environment* 154, 102–114. <https://doi.org/10.1016/j.rse.2014.07.028>

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment* 186, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>

Feldman, A.F., Reed, S., Amaral, C., Babst-Kostecka, A., Babst, F., Biederman, J., Devine, C., Fu, Z., Green, J.K., Guo, J., Hanan, N.P., Kokaly, R., Litvak, M., MacBean, N., Moore, D., Ojima, D., Poulter, B., Scott, R.L., Smith, W.K., Swap, R., Tucker, C.J., Wang, L., Watts, J., Wessels, K., Zhang, F., Zhang, W., 2024. Adaptation and Response in Drylands (ARID): Community Insights for Scoping a NASA Terrestrial Ecology Field Campaign in Drylands. *Earth's Future* 12. <https://doi.org/10.1029/2024ef004811>

Feng, S., Fu, Q., 2013. Expansion of global drylands under a warming climate. *Atmos. Chem. Phys.* 13, 10081–10094. <https://doi.org/10.5194/acp-13-10081-2013>

Fernandes, M.F., Cardoso, D., De Queiroz, L.P., 2020. An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species

richness and endemism. *Journal of Arid Environments* 174, 104079. <https://doi.org/10.1016/j.jaridenv.2019.104079>

Foody, G.M., 2009. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sensing of Environment* 113, 1658–1663. <https://doi.org/10.1016/j.rse.2009.03.014>

Francis, R.J., Kingsford, R.T., Moseby, K., Read, J., Pedler, R., Fisher, A., McCann, J., West, R., 2023. Tracking landscape scale vegetation change in the arid zone by integrating ground, drone and satellite data. *Remote Sensing in Ecology and Conservation* 10, 374–387. <https://doi.org/10.1002/rse2.375>

Frison, P.-L., Jarlan, L., Mougin, E., Hiernaux, P., Dezetter, A., Alhassane, A., Rudant, J.-P., 2018. Forest biomass estimation in Tanzanian miombo woodlands using Sentinel-1 and Sentinel-2 data. *Remote Sensing* 10, 1439. <https://doi.org/10.3390/rs10091439>

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data* 2, 150066. <https://doi.org/10.1038/sdata.2015.66>

Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58, 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)

Gaspard, D.T., Venegas-Gonzalez, A., Beeckman, H., Randriamalala, R.J., Tomazello-Filho, M., De Ridder, M., Ramanantoandro, T., 2018. Tree ring responses to climate variability of xerophytic thickets from South Soalara, Madagascar - ScienceDirect.

Geist, H.J., Lambin, E.F., 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *BioScience* 52, 143. [https://doi.org/10.1641/0006-3568\(2002\)052\[0143:PCAUDF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2)

Gessner, U., Conrad, C., Huttich, C., Keil, M., Schmidt, M., Schramm, M., Dech, S., 2008. A Multi-Scale Approach for Retrieving Proportional Cover of Life Forms, in: *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*. Presented at the IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Boston, MA, USA, p. III-700-III-703. <https://doi.org/10.1109/IGARSS.2008.4779444>

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L., Justice, C.O., 2018. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sensing of Environment* 217, 72–85. <https://doi.org/10.1016/j.rse.2018.08.005>

Goldblatt, R., Rivera Ballesteros, A., Burney, J., 2017. High Spatial Resolution Visual Band Imagery Outperforms Medium Resolution Spectral Imagery for Ecosystem Assessment in the Semi-Arid Brazilian Sertão. *Remote Sensing* 9, 1336. <https://doi.org/10.3390/rs9121336>

- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guida-Johnson, B., Villagra, P.E., Alvarez, L.M., Rojas, F., Alvarez, J.A., 2021. Finding woodlands in drylands: Bases for the monitoring of xeric open forests in a cloud computing platform. *Remote Sensing Applications: Society and Environment* 22. <https://doi.org/10.1016/j.rsase.2021.100528>
- Guirado, E., Alcaraz-Segura, D., Cabello, J., Puertas-Ruiz, S., Herrera, F., Tabik, S., 2020. Tree cover estimation in global drylands from space using deep learning. *Remote Sensing* 12. <https://doi.org/10.3390/rs12030343>
- Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., Herrera, F., 2017. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: *Ziziphus lotus* as Case Study. *Remote Sensing* 9. <https://doi.org/10.3390/rs9121220>
- Habiba, U., Kogachi, A., Huy, N., Shaw, R., 2012. Understanding Adaptation Practices in Arid Land ecosystem, in: Uy, N., Shaw, R. (Eds.), *Community, Environment and Disaster Risk Management*. Emerald Group Publishing Limited, pp. 123–144. [https://doi.org/10.1108/S2040-7262\(2012\)0000012013](https://doi.org/10.1108/S2040-7262(2012)0000012013)
- Hänke, H., Barkmann, J., Coral, C., Enfors Kaustky, E., Marggraf, R., 2017. Social-ecological traps hinder rural development in southwestern Madagascar. *E&S* 22, art42. <https://doi.org/10.5751/ES-09130-220142>
- 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, 850–853. <https://doi.org/10.1126/science.1244693>
- Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A., Potapov, P.V., Suarez, D.R., Roman-Cuesta, R.M., Saatchi, S.S., Slay, C.M., Turubanova, S.A., Tyukavina, A., 2021. Global maps of twenty-first century forest carbon fluxes. *Nat. Clim. Chang.* 11, 234–240. <https://doi.org/10.1038/s41558-020-00976-6>
- Hawkins, D.M., 2004. The problem of Overfitting. *Journal of Chemical Information and Computer Sciences*.
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sensing of Environment* 158, 220–234. <https://doi.org/10.1016/j.rse.2014.11.005>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global

reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146, 1999–2049. <https://doi.org/10.1002/qj.3803>

Hervé, D., Randriamalala, J.R., Randriambanona, H., Razanaka, S., Rakotoarimanana, V., Ranaivoson, R.E., Rafidison, V., Carrière, S.M., 2022. Valorisation durable des formations sèches de l’océan indien.

Hervé, D., Razanaka, S., Rakotondraompiana, S., Rafamantanantsoa, F., Carrière, S., 2015. Transitions agraires au sud de Madagascar : résilience et viabilité, deux facettes de la conservation : actes du séminaire de synthèse du projet FPPSM.

Hester, A.J., Scogings, P.F., Trollope, W.S.W., 2006. Long-term impacts of goat browsing on bush-clump dynamics in a semi-arid subtropical savanna. *Plant Ecol* 183, 277–290. <https://doi.org/10.1007/s11258-005-9039-6>

Hirota, M., Holmgren, M., Van Nes, E.H., Scheffer, M., 2011. Global Resilience of Tropical Forest and Savanna to Critical Transitions. *Science* 334, 232–235. <https://doi.org/10.1126/science.1210657>

Holling, C.S., 1973a. Resilience and Stability of Ecological Systems. *Annu. Rev. Ecol. Syst.* 4, 1–23. <https://doi.org/10.1146/annurev.es.04.110173.000245>

Holling, C.S., 1973b. Resilience and Stability of Ecological Systems. *Annu. Rev. Ecol. Syst.* 4, 1–23. <https://doi.org/10.1146/annurev.es.04.110173.000245>

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring* 83, 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

Huy, T.N., Adjognon, G.S., Van Soest, D., 2024. Combatting forest fires in the drylands of SUB-SAHARAN Africa: Quasi-experimental evidence from Burkina Faso. *American J Agri Economics* 106, 1684–1713. <https://doi.org/10.1111/ajae.12464>

IEFN, 1996. Inventaire écologique forestier national : recueil botanique de 200 espèces forestières.

INSTAT, 2020. TROISIEME RECENSEMENT GENERAL DE LA POPULATION ET DE L’HABITATION (RGPH-3) - RÉSULTATS GLOBAUX DU RECENSEMENT GÉNÉRAL DE LA POPULATION ET DE L’HABITATION DE 2018 DE MADAGASCAR (RGPH-3).

IPCC, 2023. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC. Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland.* <https://doi.org/10.59327/IPCC/AR6-9789291691647>

Jordan, C.F., 1969. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* 50, 663–666. <https://doi.org/10.2307/1936256>

Karakizi, C., Gounari, O., Sofikiti, E., Begkos, G., Karantzalos, K., Symeonakis, E., 2023. Assessing the Contribution of Optical and SAR Data for Fractional Savannah

Woody Vegetation Mapping, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Pasadena, CA, USA, pp. 3118–3121. <https://doi.org/10.1109/IGARSS52108.2023.10282969>

Karl, J.W., McCord, S.E., Hadley, B.C., 2017. A comparison of cover calculation techniques for relating point-intercept vegetation sampling to remote sensing imagery. *Ecological Indicators* 73, 156–165. <https://doi.org/10.1016/j.ecolind.2016.09.034>

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sensing of Environment* 114, 2897–2910. <https://doi.org/10.1016/j.rse.2010.07.008>

Key, C.H., Benson, N.C., 2006. Landscape Assessment (LA). In: Lutes, Duncan C.; Keane, Robert E.; Caratti, John F.; Key, Carl H.; Benson, Nathan C.; Sutherland, Steve; Gangi, Larry J. 2006. FIREMON: Fire effects monitoring and inventory system. Gen. Tech. Rep. RMRS-GTR-164-CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. LA-1-55 164.

Klemeš, V., 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* 31, 13–24. <https://doi.org/10.1080/02626668609491024>

Koechlin, J., Guillaumet, J.-L., Morat, P., 1974. Flore et végétation de Madagascar, https://horizon.documentation.ird.fr/exl-doc/pleins_textes/2025-01/010007501.pdf.

Lasry, F., Grouzis, M., Milleville, P., Razanaka, S., 2004. Dynamique de la déforestation et agriculture pionnière dans le sud-ouest de Madagascar : exploitation diachronique de l'imagerie satellitale haute résolution. *Photo-Interprétation* 26-35 + p. 45-47-26-35 + p. 45-47.

Laurance, W.F., Lovejoy, T.E., Vasconcelos, H.L., Bruna, E.M., Didham, R.K., Stouffer, P.C., Gascon, C., Bierregaard, R.O., Laurance, S.G., Sampaio, E., 2002. Ecosystem Decay of Amazonian Forest Fragments: a 22-Year Investigation. *Conservation Biology* 16, 605–618. <https://doi.org/10.1046/j.1523-1739.2002.01025.x>

Laxmi Goparaju, Ahmad, F., Sinha, D., 2019. Quantification and Conservation Status of Forests Fragments of Tropical Dry Deciduous Forests—A Geospatial Analysis Running Head: Tropical Dry Deciduous Forests. *Contemporary Problems of Ecology* 12. <https://doi.org/10.1134/S1995425519060131>

Lebzak, E.V., 2022. Modern problems and directions of development of forest cartography. *Interexpo GEO-Siberia 1*, 198–205. <https://doi.org/10.33764/2618-981X-2022-1-198-205>

Leprieur, C., Kerr, Y.H., Mastorchio, S., Meunier, J.C., 2000. Monitoring vegetation cover across semi-arid regions: Comparison of remote observations from various scales. *International Journal of Remote Sensing* 21, 281–300. <https://doi.org/10.1080/014311600210830>

Leprun, J.-C., Grouzis, M., Randriambanona, H., 2009. Post-cropping change and dynamics in soil and vegetation properties after forest clearing: Example of the semi-arid Mikea Region (southwestern Madagascar) - Science Direct.

Lewin, A., Murali, G., Rachmilevitch, S., Roll, U., 2024. Global evaluation of current and future threats to drylands and their vertebrate biodiversity. *Nat Ecol Evol* 8, 1448–1458. <https://doi.org/10.1038/s41559-024-02450-4>

Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2015. Remote sensing and image interpretation, Seventh edition. ed. John Wiley & Sons, Inc, Hoboken, N.J.

Los, S.O., Rosette, J.A.B., Kljun, N., North, P.R.J., Chasmer, L., Suárez, J.C., Hopkinson, C., Hill, R.A., Van Gorsel, E., Mahoney, C., Berni, J.A.J., 2012. Vegetation height and cover fraction between 60° S and 60° N from ICESat GLAS data. *Geosci. Model Dev.* 5, 413–432. <https://doi.org/10.5194/gmd-5-413-2012>

Ma, X., Ding, J., Wang, T., Lu, L., Sun, H., Zhang, F., Cheng, X., Nurmemet, I., 2023. A Pixel Dichotomy Coupled Linear Kernel-Driven Model for Estimating Fractional Vegetation Cover in Arid Areas From High-Spatial-Resolution Images. *IEEE Trans. Geosci. Remote Sensing* 61, 1–15. <https://doi.org/10.1109/TGRS.2023.3289093>

Maestre, F.T., Eldridge, D.J., Soliveres, S., Kéfi, S., Delgado-Baquerizo, M., Bowker, M.A., García-Palacios, P., Gaitán, J., Gallardo, A., Lázaro, R., Berdugo, M., 2016. Structure and Functioning of Dryland Ecosystems in a Changing World. *Annu. Rev. Ecol. Evol. Syst.* 47, 215–237. <https://doi.org/10.1146/annurev-ecolsys-121415-032311>

Markham, B.L., Storey, J.C., Williams, D.L., Irons, J.R., 2004. Landsat sensor performance: history and current status. *IEEE Transactions on Geoscience and Remote Sensing* 42, 2691–2694. <https://doi.org/10.1109/TGRS.2004.840720>

Martin-Ortega, P., Garcia-Montero, L., Pascual, C., Garcia-Robredo, F., Picard, N., Bastin, J.F., Sibelet, N., 2017. Global Drylands Assessment using Collect Earth tools and opportunities. *Forêt méditerranéenne*.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat Surface Reflectance Dataset for North America, 1990–2000. *IEEE Geosci. Remote Sensing Lett.* 3, 68–72. <https://doi.org/10.1109/LGRS.2005.857030>

Masezamana, H.N., Andrianarisoa, J.H., Raoliarivelo, L.I., Randriamalala, R.J., 2013. Identification et Analyse d'Activités Alternatives à la Fabrication de Charbon de Bois dans le District de Toliara II. Q096 Project Final Report, DERAD, Eastern and Southern Africa Partnership Program, Antananarivo.

McRoberts, R., Tomppo, E., 2007. Remote sensing support for national forest inventories. *Remote Sensing of Environment* 110, 412–419. <https://doi.org/10.1016/j.rse.2006.09.034>

McRoberts, R.E., Chen, Q., Gormanson, D.D., Walters, B.F., 2018. Probability-and model-based approaches to inference for proportion forest using satellite imagery as auxiliary data. *Remote sensing of environment* 204, 642–649. <https://doi.org/10.1016/j.rse.2017.09.032>

MEFT, USAID, CI, 2009. Evolution de la couverture de forêts naturelles a Madagascar, 1990- 2000-2005.

Mellor, A., Boukir, S., Haywood, A., Jones, S., 2015. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, 155–168. <https://doi.org/10.1016/j.isprsjprs.2015.03.014>

Mertens, B., Lambin, E.F., 2000. Land-Cover-Change Trajectories in Southern Cameroon. *Annals of the Association of American Geographers* 90, 467–494. <https://doi.org/10.1111/0004-5608.00205>

Meyer, H., Ludwig, M., Milà, C., Linnenbrink, J., Schumacher, F., 2024. The CAST package for training and assessment of spatial prediction models in R [WWW Document]. *arXiv.org*. URL <https://doi.org/10.48550/arXiv.2404.06978> (accessed 10.28.25).

Middleton, N. J., Thomas, D. S. G., 1997. *World Atlas of Desertification*. Arnold. - References - Scientific Research Publishing [WWW Document]. URL <https://www.scirp.org/reference/referencespapers?referenceid=4203453> (accessed 6.24.26).

Morat, P.P., 1969. Note sur l’application à Madagascar du quotient pluviothermique d’Emberger.

Mpanda Mukenza, M., Kikuni Tchowa, J., Delame, H., Besisa Nguba, T., Bogaert, J., Useni, Y.S., Bastin, J.-F., 2025. Mapping forest types of semi-arid regions of DR Congo: accounting for leaf phenology to inform forest management. <https://doi.org/10.2139/ssrn.5561496>

Mpanda Mukenza, M., Kikuni Tchowa, J., Nantenaina Ramalason, F., Khoji Muteya, H., Bogaert, J., Useni Sikuzani, Y., Bastin, J.-F., 2026. Quantifying Categorical Information Loss in Forest Compositional Mapping: Implications for the Accuracy of Forest Assessment in Lualaba Province (DR Congo). <https://doi.org/10.20944/preprints202604.1653.v1>

Noromiarilanto, F., Brinkmann, K., Faramalala, M.H., Buerkert, A., 2016. Assessment of food self-sufficiency in smallholder farming systems of south-western Madagascar using survey and remote sensing data. *Agricultural Systems* 149, 139–149. <https://doi.org/10.1016/j.agsy.2016.09.005>

Nuzzo, R.L., 2019. Histograms: A Useful Data Analysis Visualization. *PM&R* 11, 309–312. <https://doi.org/10.1002/pmrj.12145>

Nzabarinda, V., Bao, A., Xu, W., Uwamahoro, S., Xiaoran, H., Habiyakare, T., Sindikubwabo, C., Habumugisha, J.M., Itangishaka, A.C., 2022. A simple model to predict the spatiotemporally vegetation dynamics in terms of precipitation and temperature. *Environmental Development* 44, 100769. <https://doi.org/10.1016/j.envdev.2022.100769>

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>

ONE, DGF, FTM, MNP, CI, 2013. Evolution de la couverture de forêts naturelles à Madagascar 2005–2010.

ONE, DGF, MNP, WCS, ETCTerra, 2015. Changement de la couverture de forêts naturelles à Madagascar, 2005-2010- 2013. Antananarivo.

Pande, C.B., Moharir, K.N., Kumar Singh, S., Varade, A.M., Elbeltagi, A., Khadri, S.F.R., Choudhari, P., 2021. Estimation of crop and forest biomass resources in a semi-arid region using satellite data and GIS. *Journal of the Saudi Society of Agricultural Sciences* 20. <https://doi.org/10.1016/j.jssas.2021.03.002>

Pausas, J.G., Ribeiro, E., 2017. Fire and plant diversity at the global scale. *Global Ecol Biogeogr* 26, 889–897. <https://doi.org/10.1111/geb.12596>

Pearson, T.R.H., Brown, S., Murray, L., Sidman, G., 2017. Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon Balance Manage* 12, 3. <https://doi.org/10.1186/s13021-017-0072-2>

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing* 11, 523. <https://doi.org/10.3390/rs11050523>

Pengra, B.W., Stehman, S.V., Horton, J.A., Dockter, D.J., Schroeder, T.A., Yang, Z., Cohen, W.B., Healey, S.P., Loveland, T.R., 2007. Quality control and assessment of interpreter consistency of annual land cover reference data in an operational national monitoring program. *ISPRS journal of photogrammetry and remote sensing* 128, 299–306. <https://doi.org/10.1016/j.isprsjprs.2017.04.010>

Pennington, R.T., Lehmann, C.E.R., Rowland, L.M., 2018. Tropical savannas and dry forests. *Current Biology* 28, R541–R545. <https://doi.org/10.1016/j.cub.2018.03.014>

Pérez-Luque, A.J., Ramos-Font, M.E., Tognetti Barbieri, M.J., Tarragona Pérez, C., Calvo Renta, G., Robles Cruz, A.B., 2022. Vegetation Cover Estimation in Semi-Arid Shrublands after Prescribed Burning: Field-Ground and Drone Image Comparison. *Drones* 6, 370. <https://doi.org/10.3390/drones6110370>

Phelps, L.N., Andela, N., Gravey, M., Davis, D.S., Kull, C.A., Douglass, K., Lehmann, C.E.R., 2022. Madagascar’s fire regimes challenge global assumptions about landscape degradation. *Global Change Biology* 28, 6944–6960. <https://doi.org/10.1111/gcb.16206>

Picard, N., Saint André, L., Henry, M., 2012. Manuel de construction d’équations allométriques pour l’estimation du volume et la biomasse des arbres : de la mesure de terrain à la prédiction. CIRAD.

Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling* 216, 316–322. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., others, 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communications* 11, 4540. <https://doi.org/10.1038/s41467-020-18321-y>

Pouliot, D., Latifovic, R., Pasher, J., Duffe, J., 2018. Landsat Super-Resolution Enhancement Using Convolution Neural Networks and Sentinel-2 for Training. *Remote Sensing* 10. <https://doi.org/10.3390/rs10030394>

Prävălie, R., Bandoc, G., Patriche, C., Sternberg, T., 2019. Recent changes in global drylands: Evidences from two major aridity databases. *CATENA* 178, 209–231. <https://doi.org/10.1016/j.catena.2019.03.016>

Probst, P., Wright, M.N., Boulesteix, A., 2019. Hyperparameters and tuning strategies for random forest. *WIREs Data Min & Knowl* 9. <https://doi.org/10.1002/widm.1301>

Provencher-Nolet, L., Bernier, M., Lévesque, E., 2014. Short term change detection in tundra vegetation near Umiujaq, subarctic Quebec, Canada, in: 2014 IEEE Geoscience and Remote Sensing Symposium. Presented at the 2014 IEEE Geoscience and Remote Sensing Symposium, pp. 4668–4670. <https://doi.org/10.1109/IGARSS.2014.6947534>

Puyravaud, J.-P., 2003. Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management* 177, 593–596. [https://doi.org/10.1016/S0378-1127\(02\)00335-3](https://doi.org/10.1016/S0378-1127(02)00335-3)

Qu, L., Lu, H., Tian, Z., School, J.M., Huang, B., Liang, Yonghong, Qiu, D., Liang, Yin, 2024. Spatial prediction of soil sand content at various sampling density based on geostatistical and machine learning algorithms in plain areas. *CATENA* 234, 107572. <https://doi.org/10.1016/j.catena.2023.107572>

Rakotoarimanana, V., Grouzis, M., 2006. Influence of fire and of grazing on the richness and the floristic diversity of a savanna dominated by *Heteropogon contortus* in the south-west of Madagascar (region of Sakaraha).

Ramalason, F.N., 2015. Etude de la résilience des forêts sèches de la région du Menabe face à la variabilité des précipitations. Antananarivo.

Ramalason, F.N., Rakotondrasoa, O.L., Bastin, J.-F., 2025a. Recension des travaux antérieurs sur les méthodes utilisées dans la cartographie des écosystèmes forestiers des zones arides : cas du sud-ouest de Madagascar, in: LES CAHIERS DE L'ED GRND. Presented at the Colloque de l'ED GRND 2023, ED GRND, Antananarivo, Madagascar.

Ramalason, F.N., Rakotondrasoa, O.L., Linden, A.V., Renard, G., Ratsimba, H.R., Bogaert, J., Bastin, J.-F., 2025b. Potential of global vegetation maps in capturing xerophytic vegetation cover: insights from Madagascar's arid ecosystems. *Journal of Arid Environments* 229, 105406. <https://doi.org/10.1016/j.jaridenv.2025.105406>

Ramalason, F.N., Rakotondrasoa, O.L., Vander Linden, A., Renard, G., Randriamalala, J.R., Vereecken, N.J., Razakamiaramanana, A., Ranaivoharivelo, M.F., Raholinarivo, S., Rakoto Ratsimba, H., Bogaert, J., Bastin, J.-F., 2026. The importance of the sampling design in mapping woody cover in arid ecosystems. *GIScience & Remote Sensing* 63, 2658305. <https://doi.org/10.1080/15481603.2026.2658305>

Ramond, J.-B., Jordaan, K., Díez, B., Heinzelmann, S.M., Cowan, D.A., 2022. Microbial Biogeochemical Cycling of Nitrogen in Arid Ecosystems. *Microbiol Mol Biol Rev* 86, e00109-21. <https://doi.org/10.1128/membr.00109-21>

Ranaivoharivelo, M.F., Hantanirina, I.H., Randriamalala, J.R., Randrianariveloseheno, A.J.M., Jungblut, N., Leclercq, S., Rakotondrasoa, O.L., Bindelle, J., 2026. Diet selection and forage nutritional quality of goats in the xerophytic thickets of Southwestern Madagascar. *Trop Anim Health Prod* 58, 200. <https://doi.org/10.1007/s11250-026-05010-z>

Randriamalala, J., Hosnah, H., Ranaivoarimanana, S., Ravonjimalala, H., 2015. Dynamique spatiale des fourrés xérophiles de Betioky-Sud et Soalara-Sud (sud-ouest, Madagascar) Spatial dynamics of xerophytic thickets of Betioky Sud and Soalara Sud (southwestern. *Transitions agraires* 113.

Randriamalala, J.R., Andrianarisoa, J.H., Raoliarivelo, L.I., Masezamana, H.N., Radobarimanjaka, R., Ehrensperger, A., 2016. Gestion de pâturage des petits ruminants en zone semi-aride de Madagascar, cas de la commune rurale de Soalara-Sud, Toliara II. *Akon'ny ala* 33, 31–38.

Randriamalala, J.R., Hervé, D., Letourmy, P., Carrière, S.M., 2015. Effects of slash-and-burn practices on soil seed banks in secondary forest successions in Madagascar. *Agriculture, Ecosystems & Environment* 199, 312–319. <https://doi.org/10.1016/J.AGEE.2014.09.010>

Randriamalala, J.R., Radosy, H.O., Ramanakoto, M., Razafindrahanta, H., Ravoninjatovo, J.-M., Haingomanantsoa, R.S., Ramananantoandro, T., 2022. Allometric models to predict the individual aboveground biomass of shrubs of Malagasy xerophytic thickets. *Journal of Arid Environments* 202, 104751. <https://doi.org/10.1016/j.jaridenv.2022.104751>

Randriamalala, J.R., Randrianomana, M., Ranaivoson, R.E., Rabemananjara, Z.H., Hervé, D., 2021. Estimating wood charcoal supply to Toliara town in southwestern Madagascar, a comparison of methods. *Scientific African*. <https://doi.org/10.1016/j.sciaf.2021.e01011>

Randriamalala, J.R., Randriarimalala, J., Hervé, D., Carrière, S.M., 2019. Slow recovery of endangered xerophytic thickets vegetation after slash-and-burn cultivation in Madagascar. *Biological Conservation* 233, 260–267. <https://doi.org/10.1016/J.BIOCON.2019.03.006>

Randriamiharisoa, L., Klinges, D.H., Razafindranaivo, S., Scheffers, B.R., 2024. Community-sourced knowledge improves biodiversity monitoring in Madagascar's National Parks. *Discov Conserv* 1, 14. <https://doi.org/10.1007/s44353-024-00015-x>

Randrianasolo, Z.H., Razafimahatratra, A.R., Razafinarivo, R.N.G., Randrianary, T., Rakotovololonalimanana, H., Rajemison, A.H., Mamitiana, A., Andriamanalina, R.L., Rakotosoa, A., Ramananantoandro, T., 2019. Which allometric models are the most appropriate for estimating aboveground biomass in secondary forests of Madagascar with *Ravenala madagascariensis*? *Scientific African* 6, e00147. <https://doi.org/10.1016/j.sciaf.2019.e00147>

Raoliarivelo, L.I.B., Rabeniala, R., Masezamana, H.N., Andrianarisoa, J.H., Randriamalala, R.J., 2010. Impact de la fabrication de charbon de bois sur la production et la disponibilité fourragère de pâturage en zone subaride, cas de la commune de Soalara-Sud, Toliara II: Rapport final. University of Bern. <https://doi.org/10.7892/BORIS.69311>

Raselimanana, A.P., Raherilalao, M.J., Soarimalala, V., Gardner, C.J., Jasper, L.D., Schoeman, C.M., Goodman, S.M., 2012. Un premier aperçu de la faune de vertébrés du bush épineux de Salary-Bekodoy, à l'ouest du Parc National de Mikea, Madagascar. *Malagasy Nature* 6.

Ravonjimalala, H.R., Bogaert, J., Hervé, D., Razanaka, S., Ranaivo, J., Randriambanona, H., Rakotondraompiana, S., 2017. Spatio-Temporal Evolution of the Fragmentation Classes of the Mikea Dry Deciduous Forest (Southwestern Madagascar), in: *Water and Land Security in Drylands*. Springer International Publishing, pp. 209–219. https://doi.org/10.1007/978-3-319-54021-4_19

Razanaka, S., 2004. La forêt des Mikea : un espace et des ressources assiégés – diversité des types de végétation et dynamique post-culturelle d'une forêt dense sèche du sud-ouest de Madagascar (Thèse de doctorat d'Etat es-Science). UNIVERSITE D'ANTANANARIVO, Antananarivo.

Reiche, J., Verhoeven, R., Verbesselt, J., Hamunyela, E., Wielaard, N., Herold, M., 2018. Characterizing tropical forest cover loss using dense Sentinel-1 data and active fire alerts. *Remote Sensing* 10, 777.

Reynolds, J.F., Stafford Smith, D.M., Lambin, E.F., Turner, B.L., Mortimore, M., Batterbury, S.P., Downing, T.E., Dowlatabadi, H., Fernández, R.J., Herrick, J.E., others, 2007. Global desertification: building a science for dryland development. *Science* 316, 847–851. <https://doi.org/10.1126/science.1131634>

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Aroita, 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, 913–929. <https://doi.org/10.1111/ecog.02881>

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS.

Roy, D.P., Kovalsky, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment, Landsat 8 Science Results* 185, 57–70. <https://doi.org/10.1016/j.rse.2015.12.024>

Ryan, C.M., Hill, T., Woollen, E., Ghee, C., Mitchard, E., Cassells, G., Grace, J., Woodhouse, I.H., Williams, M., 2012. Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. *Global Change Biology* 18, 243–257. <https://doi.org/10.1111/j.1365-2486.2011.02551.x>

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. *Environmental Modelling & Software* 118, 166–171. <https://doi.org/10.1016/j.envsoft.2019.05.004>

Scales, I.R. (Ed.), 2014. *Conservation and Environmental Management in Madagascar*. Routledge, London. <https://doi.org/10.4324/9780203118313>

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. *Nature* 413, 591–596. <https://doi.org/10.1038/35098000>

Schneider, F.D., Morsdorf, F., Schmid, B., Petchey, O.L., Hueni, A., Schimel, D.S., Schaepman, M.E., 2017. Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nat Commun* 8, 1441. <https://doi.org/10.1038/s41467-017-01530-3>

Sedano, F., Silva, J.A., Machoco, R., Meque, C.H., Siteo, A., Ribeiro, N., Anderson, K., Ombe, Z.A., Baule, S.H., Tucker, C.J., 2016. The impact of charcoal production on forest degradation: a case study in Tete, Mozambique. *Environ. Res. Lett.* 11, 094020. <https://doi.org/10.1088/1748-9326/11/9/094020>

Senanayake, I.P., Yeo, I.-Y., Kuczera, G.A., 2023. A Random Forest-Based Multi-Index Classification (RaFMIC) Approach to Mapping Three-Decadal Inundation Dynamics in Dryland Wetlands Using Google Earth Engine. *Remote Sensing* 15, 1263. <https://doi.org/10.3390/rs15051263>

Sengani, D., Ramoelo, A., Archer, E., 2023. A review of fusion framework using optical sensors and Synthetic Aperture Radar imagery to detect and map land degradation and sustainable land management in the semi-arid regions. *Geocarto International* 38, 2278325. <https://doi.org/10.1080/10106049.2023.2278325>

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., 2015. Conservation policy and the measurement of forests. *Nature Climate Change* 6, 192–196.

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth* 6, 427–448.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S., 2020. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 13, 6308–6325. <https://doi.org/10.1109/JSTARS.2020.3026724>

Shirvani, Z., Abdi, O., Buchroithner, M.F., 2020. A new analysis approach for long-term variations of forest loss, fragmentation, and degradation resulting from road-network expansion using Landsat time-series and object-based image analysis. *Land Degradation & Development* 31. <https://doi.org/10.1002/ldr.3530>

Siemers, F.M., Bajorath, J., 2023. Differences in learning characteristics between support vector machine and random forest models for compound classification revealed by Shapley value analysis. *Sci Rep* 13, 5983. <https://doi.org/10.1038/s41598-023-33215-x>

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y., Keyzers, D., Neumann, M., Cisse, M., Quinn, J., 2021. Continental-Scale Building Detection from High Resolution Satellite Imagery. <https://doi.org/10.48550/arXiv.2107.12283>

Smit, I.P.J., Asner, G.P., Govender, N., Kennedy-Bowdoin, T., Knapp, D.E., Jacobson, J., 2010. Effects of fire on woody vegetation structure in African savanna. *Ecological Applications* 20, 1865–1875. <https://doi.org/10.1890/09-0929.1>

Staver, A.C., Archibald, S., Levin, S.A., 2011. The Global Extent and Determinants of Savanna and Forest as Alternative Biome States. *Science* 334, 230–232. <https://doi.org/10.1126/science.1210465>

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment* 231, 111199. <https://doi.org/10.1016/j.rse.2019.05.018>

Stock, A., 2025. Choosing blocks for spatial cross-validation: lessons from a marine remote sensing case study. *Front. Remote Sens.* 6, 1531097. <https://doi.org/10.3389/frsen.2025.1531097>

Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 36, 111–133. <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>

Strassburg, B.B.N., Iribarrem, A., Beyer, H.L., Cordeiro, C.L., Crouzeilles, R., Jakovac, C.C., Braga Junqueira, A., Lacerda, E., Latawiec, A.E., Balmford, A., Brooks, T.M., Butchart, S.H.M., Chazdon, R.L., Erb, K.-H., Brancalion, P., Buchanan, G., Cooper, D., Díaz, S., Donald, P.F., Kapos, V., Leclère, D., Miles, L., Obersteiner, M., Plutzer, C., De M. Scaramuzza, C.A., Scarano, F.R., Visconti, P., 2020. Global priority areas for ecosystem restoration. *Nature* 586, 724–729. <https://doi.org/10.1038/s41586-020-2784-9>

Streiner, D.L., 2018. *Statistics Commentary Series: Commentary No. 24: Box Plots.* *J Clin Psychopharmacol* 38, 5–6. <https://doi.org/10.1097/JCP.0000000000000828>

Symeonakis, E., Korkofigkas, A., Higginbottom, T., Boyd, J., Arnau-Rosalen, E., Stamou, G., Karantzalos, K., 2022. Towards a Deep Learning Fractional Woody Vegetation Cover Monitoring Framework, in: *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Kuala Lumpur, Malaysia, pp. 5905–5908.* <https://doi.org/10.1109/IGARSS46834.2022.9883347>

Teferi, H.H., 2018. Land use land cover changes along topographic gradients in Hugumburda national forest priority area, Northern Ethiopia. *Land* 7.

Tian, F., Brandt, M., Liu, Y.Y., Rasmussen, K., Fensholt, R., 2017. Mapping gains and losses in woody vegetation across global tropical drylands. *Global change biology* 23, 1748–1760.

Torbick, N., Qi, J., Ge, J., Olsen, J., Lusch, D., 2005. An assessment of africover and GLC2000 using general agreement and videography techniques, in: *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05*. Presented at the 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., IEEE, Seoul, Korea, pp. 5005–5008. <https://doi.org/10.1109/IGARSS.2005.1526801>

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. *Remote Sensing of Environment* 120, 9–24. <https://doi.org/10.1016/j.rse.2011.05.028>

Tovondrazane, C.A., Holède, B.I., Andriatsiaronandroy, R., Razanakoto, T., Rabemananjara, Z.H., 2020. Effet de l'exploitation minière sur les facteurs productifs comme l'agriculture et la production de charbon de bois et la dynamique de déforestation dans le Sud-ouest de Madagascar. *Malagasy Nature* 6–23.

Trabucco, A., Zomer, R., 2019. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. <https://doi.org/10.6084/M9.FIGSHARE.7504448.V3>

Tsendbazar, N., Herold, M., Li, L., Tarko, A., de Bruin, S., Masiliunas, D., Lesiv, M., Fritz, S., Buchhorn, M., Smets, B., Van De Kerchove, R., Duerauer, M., 2021. Towards operational validation of annual global land cover maps. *Remote Sensing of Environment* 266, 112686. <https://doi.org/10.1016/j.rse.2021.112686>

UNCCD, 2024. The Global Threat of Drying Lands: Regional and global aridity trends and future projections. A Report of the Science-Policy Interface.

UNEP, F., 2021. *Becoming #GenerationRestoration: Ecosystem Restoration for People, Nature and Climate* | UNEP - UN Environment Programme [WWW Document]. URL <https://www.unep.org/resources/ecosystem-restoration-people-nature-climate> (accessed 5.4.26).

Urban, M., Schellenberg, K., Morgenthal, T., Dubois, C., Hirner, A., Gessner, U., Magonong, B., Zhang, Z., Baade, J., Collett, A., Schullius, C., 2021. Using Sentinel-1 and Sentinel-2 Time Series for Slangbos Mapping in the Free State Province, South Africa. *Remote Sensing* 13, 3342. <https://doi.org/10.3390/rs13173342>

Valavi, R., Elith, J., Lahoz-Monfort, J., Guillera-Arroita, G., 2018. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution* 10. <https://doi.org/10.1111/2041-210X.13107>

Valencia, B.N.M., 2014. La gestion durable de l'eau et la conservation du sol dans la nouvelle aire protégée de Tsinjoriake, sud-ouest de Madagascar (Diplôme D'études Approfondies (DEA)). Université de Toliara.

Valjarević, A., Djekić, T., Stevanović, V., Ivanović, R., Jandziković, B., 2018. GIS numerical and remote sensing analyses of forest changes in the Toplica region for the period of 1953–2013. *Applied Geography* 92, 131–139. <https://doi.org/10.1016/j.apgeog.2018.01.016>

Valjarević, A., Šiljeg, A., Šiljeg, S., Vujović, F., Sahay, A., 2025. GIS-based water stress analysis in North African drylands. *Journal of Arid Environments* 230, 105427. <https://doi.org/10.1016/j.jaridenv.2025.105427>

Van Passel, J., De Keersmaecker, W., Somers, B., 2020. Monitoring woody cover dynamics in tropical dry forest ecosystems using sentinel-2 satellite imagery. *Remote Sensing* 12, 1276.

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L.E.O.C., Nasi, R., 2021. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Sci. Adv.* 7, eabe1603. <https://doi.org/10.1126/sciadv.abe1603>

Veldman, J.W., Overbeck, G.E., Negreiros, D., Mahy, G., Le Stradic, S., Fernandes, G.W., Durigan, G., Buisson, E., Putz, F.E., Bond, W.J., 2015. Where Tree Planting and Forest Expansion are Bad for Biodiversity and Ecosystem Services. *BioScience* 65, 1011–1018. <https://doi.org/10.1093/biosci/biv118>

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment* 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>

Vetter, S., 2005. Rangelands at equilibrium and non-equilibrium: recent developments in the debate. *Journal of Arid Environments* 62, 321–341. <https://doi.org/10.1016/j.jaridenv.2004.11.015>

Vieilledent, G., Grinand, C., Rakotomalala, F.A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnut, 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. *Biological Conservation* 222, 189–197. <https://doi.org/10.1016/j.biocon.2018.04.008>

Vieilledent, G., Nourtier, M., Grinand, C., Pedrono, M., Clausen, A., Rabetrano, T., Rakotoarijaona, J.-R., Rakotoarivelo, B., Rakotomalala, F.A., Rakotomalala, L., Razafimpahanana, A., Ralison, J.M., Achard, F., 2020. It's not just poverty: unregulated global market and bad governance explain unceasing deforestation in Western Madagascar (preprint). *Ecology*. <https://doi.org/10.1101/2020.07.30.229104>

Walker, B., Holling, C.S., Carpenter, S.R., Kinzig, A.P., 2004. Resilience, Adaptability and Transformability in Social-ecological Systems. *E&S* 9, art5. <https://doi.org/10.5751/ES-00650-090205>

Wang, G., Gertner, G., Anderson, A.B., 2005. Sampling design and uncertainty based on spatial variability of spectral variables for mapping vegetation cover. *International Journal of Remote Sensing* 26, 3255–3274. <https://doi.org/10.1080/01431160500114748>

Wang, J., Yang, X., Wang, Z., Cheng, H., Kang, J., Tang, H., Li, Y., Bian, Z., Bai, Z., 2022. Consistency Analysis and Accuracy Assessment of Three Global Ten-Meter

Land Cover Products in Rocky Desertification Region—A Case Study of Southwest China. *IJGI* 11, 202. <https://doi.org/10.3390/ijgi11030202>

Watson, J.E.M., Venter, O., 2019. Mapping the Continuum of Humanity's Footprint on Land. *One Earth* 1, 175–180. <https://doi.org/10.1016/j.oneear.2019.09.004>

White, A., Sparrow, B., Leitch, E., Foulkes, J., Flitton, R., Lowe, A.J., Caddy-Retalic, S., 2012. *AusPlots Rangelands survey protocols manual*. University of Adelaide Press.

White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., Guindon, L., 2014. Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Canadian Journal of Remote Sensing* 40, 192–212. <https://doi.org/10.1080/07038992.2014.945827>

Wigley, B.J., Bond, W.J., Hoffman, M.T., 2009. Bush encroachment under three contrasting land-use practices in a mesic South African savanna. *African Journal of Ecology* 47, 62–70. <https://doi.org/10.1111/j.1365-2028.2008.01051.x>

Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Healey, S., Zhu, Z., Scambos, T.A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C.J., Masek, J.G., Hermosilla, T., White, J.C., Belward, A.S., Schaaf, C., Woodcock, C.E., Huntington, J.L., Lymburner, L., Hostert, P., Gao, F., Lyapustin, A., Pekel, J.-F., Strobl, P., Cook, B.D., 2022. Fifty years of Landsat science and impacts. *Remote Sensing of Environment* 280, 113195. <https://doi.org/10.1016/j.rse.2022.113195>

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P.D., Allen, G.H., Pavelsky, T.M., 2019. MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research* 55, 5053–5073. <https://doi.org/10.1029/2019WR024873>

Zeileis, A., Grothendieck, G., 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. *J. Stat. Soft.* 14. <https://doi.org/10.18637/jss.v014.i06>

Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* 24, 583–594. <https://doi.org/10.1080/01431160304987>

Zhang, X., Pan, Q., Zhao, Y., Huang, Q., 2005. Research on spatial deviation analysis model of land-use change, in: Li, D., Ma, H. (Eds.), . Presented at the MIPPR 2005 Image Analysis Techniques, Wuhan, China, p. 60440C. <https://doi.org/10.1117/12.652320>

Zhang-Zheng, H., Adu-Bredu, S., Duah-Gyamfi, A., Moore, S., Addo-Danso, S.D., Amissah, L., Valentini, R., Djagbletey, G., Anim-Adjei, K., Quansah, J., Sarpong, B., Owusu-Afriyie, K., Gvozdevaite, A., Tang, M., Ruiz-Jaen, M.C., Ibrahim, F., Girardin, C.A.J., Rifai, S., Dahlsjö, C.A.L., Riutta, T., Deng, X., Sun, Y., Prentice, I.C., Oliveras Menor, I., Malhi, Y., 2024. Contrasting carbon cycle along tropical forest aridity gradients in West Africa and Amazonia. *Nat Commun* 15, 3158. <https://doi.org/10.1038/s41467-024-47202-x>

Annexes

Annexe 1: Chapitre 2.4

Nombre de résultats obtenus par questions et moteurs de recherche (30/09/2022)

Questions de recherche	Nombre de résultats	
	Google Scholar	BASE
("remote sensing" OR "geographic mapping") AND (deforestation OR "open forest" OR "tree cover" OR "forest cover") AND ("arid zones" OR dryland)	10 100	73
("remote sensing" OR "geographic mapping") AND (deforestation OR "open forest" OR tree cover" OR "forest cover") AND ("arid zones" OR dryland) AND Madagascar	1	120

Annexe 2: Chapitre 5 - Appendix

Appendix A: Photographs of vegetation types

Figure A1: Photographs of vegetation types taken in August 2023: (A) xerophytic shrubs and trees over 5 meters in height, and (B) xerophytic shrubs under 5 meters in height. These images provide visual examples of the plant species present in the study area, illustrating the characteristics of xerophytic vegetation in relation to height classes.



Xerophytic Shrubs in Tsimanapetsotsa National Park

Appendix B: Descriptive statistics of tree cover, shrub cover, and wood cover data from different authors

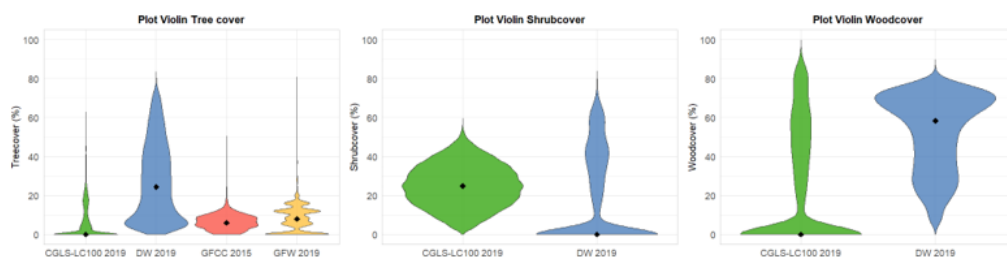


Figure B1: Comparison of tree cover, shrub cover, and wood cover data from different authors. The data from GFCC are from 2015, while the data from GFW, CGLS-LC100, and DW are based on 2019. This figure presents a comparative analysis of vegetation cover estimates across various global datasets, highlighting differences in tree, shrub, and wood cover across the study area.

Tableau B1: The average trend and standard deviation of tree cover, shrub cover, and wood cover data from different authors. The data from GFCC are from 2015, while the data from GFW, CGLS-LC100, and DW are based on 2019. The mentioned years correspond to the reference years used in the study. This figure illustrates the variability and trends in vegetation cover estimates across different global datasets.

Products	Year	Tree cover		Shrub cover		Wood cover	
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
GFCC	2015	6,56	4,01	-	-	-	-
GFW	2019	7,67	6,65	-	-	-	-
CGLS-LC100	2019	5,51	7,91	19,74	23,02	25,24	29,71
DW	2019	27,28	19,56	25,03	10,65	52,30	19,66

Tableau B2: Statistical test results for comparing different vegetation cover products. This table presents the results of normality and homogeneity of variance tests (Shapiro-Wilk and Levene's tests) for each cover type dataset, along with the subsequent Kruskal-Wallis non-parametric test results. All p-values indicate highly significant differences between products ($p < 0.001$) for all cover types. The high Kruskal-Wallis chi-squared values for Treecover (45674) and Shrubcover (23714) suggest substantial disparities between products, while the relatively lower value for Woodcover (87.149) indicates more subtle but still statistically significant differences.

Cover type	Shapiro-Wilk Test (p-value)	Levene's test (p-value)	Test used	Kruskal-Wallis χ^2	p-value
Tree cover	8.38e-42	0	Kruskal-Wallis	45674	< 2.2e-16
Shrub cover	8.15e-15	0	Kruskal-Wallis	23714	< 2.2e-16
Wood cover	3.25e-32	3.98e-36	Kruskal-Wallis	87.149	< 2.2e-16

Appendix C: Linear regression parameter values for tree cover, shrub cover, and wood cover data from different authors

Tableau C1: Slopes and intercepts of regression analysis for tree cover, shrub cover, and wood cover data. This table presents the results of the linear regression analysis comparing predicted data from different authors to field data. The slopes and intercepts represent the relationship between the predicted vegetation cover values and the field data for each cover type (tree, shrub, and wood cover).

Products	Year	Tree cover		Wood cover		Shrub cover	
		Slope	Intercept	Slope	Intercept	Slope	Intercept
GFCC	2015	1,64	-1,15	-	-	-	-
GFW	2019	1,27	0,22	-	-	-	-
CGLS-LC100	2019	0,39	-3,19	1,02	4,28	0,45	44,45
DW	2019	0,80	-4,58	1,24	-9,28	0,66	29,9

Appendix D: Model biases of different products - Influence of vegetation typology on model errors

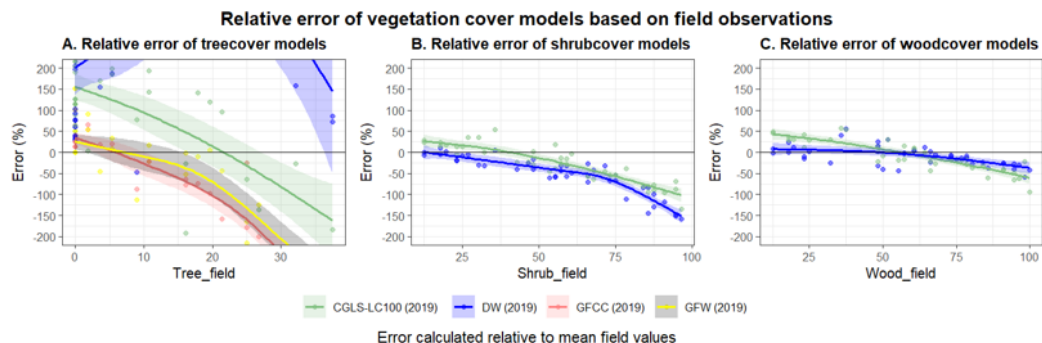


Figure D1: Relative error (%) of vegetation cover models based on 2022 field observations in southwestern Madagascar. (A) Tree cover, (B) Shrub cover, and (C) Wood cover estimates from four global products (GFCC, GFW, CGLS-LC100, and DW) compared to field data. Relative error is calculated as the difference between product estimates and mean field values, expressed as a percentage. Colored curves represent smoothed error trends per product, with shaded areas indicating 95% confidence intervals. The black horizontal line marks zero error. The plots highlight underestimation or overestimation patterns across vegetation density gradients

Tableau D-2: Results of Wilcoxon tests assessing the sensitivity of vegetation cover estimates from four global products (GFCC, GFW, CGLS-LC100, DW) to ecological

gradients derived from field data: floristic composition (top) and vegetation structure (bottom). Tests were performed on tree, shrub, and combined woody cover values across typological groups. The use of a non-parametric test was justified by the lack of normality and homogeneity of variances. Significance levels are as follows:

NS = not significant, ($p < 0.1$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

Products	Year	Tree cover		Shrub cover		Wood cover	
		p-value	Significance	p-value	Significance	p-value	Significance
Floristic composition							
GFCC	2015	0,02	* ($p < 0.05$)	-	-	-	-
GFW	2019	0,16	NS	-	-	-	-
CGLS-LC100	2019	0,00	*** ($p < 0.001$)	0,00	** ($p < 0.01$)	1,20E-05	*** ($p < 0.001$)
DW	2019	0,02	* ($p < 0.05$)	0,12	NS	0,80	NS
Structure							
GFCC	2015	3,97E-05	*** ($p < 0.001$)	-	-	-	-
GFW	2019	9,99E-02	. ($p < 0.1$)	-	-	-	-
CGLS-LC100	2019	1,58E-07	*** ($p < 0.001$)	0,001	*** ($p < 0.001$)	6,51E-06	*** ($p < 0.001$)
DW	2019	1,73E-04	*** ($p < 0.001$)	0,063	. ($p < 0.1$)	3,12E-01	NS

Annexe 3 : Chapitre 6 - Supporting Information

Contents

- S1. Satellite Data Processing
- S2. Spatial Cross-Validation Design
- S3. Random Forest Hyperparameters
- S4. Field Sampling Protocol
- S5. Photo-interpretation Protocol
- S6. Complete R Code
- S7. Performance Metrics Formulas

S1. Satellite Data Processing

S1.1 Google Earth Engine Scripts

Sentinel-1 Acquisition (April 2022)

```
// Load study area boundary
```

```
var zone = ee.FeatureCollection('users/Limite_District_STII');
```

```
var geometry = zone.geometry();
```

```
// Sentinel-1 data acquisition
```

```
var s1_2022 = ee.ImageCollection('COPERNICUS/S1_GRD')
```

```
.filterDate('2022-04-01', '2022-04-30')
.filterBounds(zone)
.filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV'))
.filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VH'))
.filter(ee.Filter.eq('instrumentMode', 'IW'))
.select(['VV', 'VH'])
.median()
.clip(geometry);

// Visualization
Map.centerObject(zone, 10);
var s1Vis = {bands: ['VV'], min: -25, max: 5};
Map.addLayer(s1_2022, s1Vis, 'Sentinel-1 April 2022');

// Export
Export.image.toDrive({
  image: s1_2022,
  description: 'S1_April_2022',
  folder: 'GEE_Exports',
  fileNamePrefix: 'S1_2022_04',
  region: geometry,
  scale: 10,
  crs: 'EPSG:4326',
  maxPixels: 1e13
});
Sentinel-2 Acquisition (April 2022)
// Sentinel-2 data acquisition
var s2_2022 = ee.ImageCollection('COPERNICUS/S2_SR')
  .filterDate('2022-04-01', '2022-04-30')
  .filterBounds(zone)
  .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 20))
  .select(['B2', 'B3', 'B4', 'B8', 'B11', 'B12'])
  .median()
  .clip(geometry);

// Visualization
var s2Vis = {bands: ['B4', 'B3', 'B2'], min: 0, max: 3000, gamma: 1.4};
Map.addLayer(s2_2022, s2Vis, 'Sentinel-2 April 2022');

// Export
Export.image.toDrive({
```

```

image: s2_2022,
description: 'S2_April_2022',
folder: 'GEE_Exports',
fileNamePrefix: 'S2_2022_04',
region: geometry,
scale: 10,
crs: 'EPSG:4326',
maxPixels: 1e13
});
AlphaEarth Embeddings Acquisition (2022)
// Configuration
var YEAR = 2022;
var TABLE_ASSET = 'users/Limite_District_STII';
var OUTPUT_NAME = 'alphaearth_embeddings_MDG_District_T2_' + YEAR;

// Load geometry
var districts = ee.FeatureCollection(TABLE_ASSET);
var region = districts.geometry();

// Load AlphaEarth dataset
var
                                coll
                                =
ee.ImageCollection('GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL')
  .filterDate(ee.Date.fromYMD(YEAR, 1, 1), ee.Date.fromYMD(YEAR+1, 1, 1))
  .filterBounds(region);

// Mosaic and clip
var img = coll.mosaic().clip(region);

// Visualization (RGB-like representation)
Map.centerObject(region, 7);
Map.addLayer(img, {bands: ['A01','A02','A03'], min: -0.3, max: 0.3}, 'AlphaEarth
RGB-ish');

// Export
Export.image.toDrive({
  image: img,
  description: OUTPUT_NAME,
  folder: 'earthengine',
  fileNamePrefix: OUTPUT_NAME,
  region: region,
  scale: 10,

```

```

maxPixels: 1e13,
fileFormat: 'GeoTIFF'
});

```

S1.2 Satellite imagery used for analysis

Tableau S1. Sentinel-1 images used (April 2022)

Source: Google Earth Engine — COPERNICUS/S1_GRD | DOI: 10.5270/S1_-wvqbeuw |
Polarisations: VV, VH | Instrument mode: IW | Level: GRD | Resolution: 10 m | Satellite: Sentinel-1A

#	Image ID	Date	Direction
1	S1A_IW_GRDH_1SDV_20220401T153350_20220401T153419_042581_051458_08C1	2022-04-01	ASCENDING
2	S1A_IW_GRDH_1SDV_20220402T022855_20220402T022920_042588_051495_78C0	2022-04-02	DESCENDING
3	S1A_IW_GRDH_1SDV_20220402T022920_20220402T022952_042588_051495_88B8	2022-04-02	DESCENDING
4	S1A_IW_GRDH_1SDV_20220406T154154_20220406T154223_042654_0516D4_E515	2022-04-06	ASCENDING
5	S1A_IW_GRDH_1SDV_20220407T023710_20220407T023740_042661_051708_2DD2	2022-04-07	DESCENDING
6	S1A_IW_GRDH_1SDV_20220413T153350_20220413T153419_042756_051A45_81C2	2022-04-13	ASCENDING
7	S1A_IW_GRDH_1SDV_20220418T154155_20220418T154224_042829_051CB2_8030	2022-04-18	ASCENDING
8	S1A_IW_GRDH_1SDV_20220419T023710_20220419T023741_042836_051CE5_B5D9	2022-04-19	DESCENDING
9	S1A_IW_GRDH_1SDV_20220426T022855_20220426T022920_042938_05203D_3210	2022-04-26	DESCENDING
10	S1A_IW_GRDH_1SDV_20220426T022920_20220426T022953_042938_05203D_16D5	2022-04-26	DESCENDING

Notes: Images composited using median reducer in Google Earth Engine (Gorelick et al., 2017; doi: 10.1016/j.rse.2017.06.031). Orbit directions: ASCENDING (day passes) and DESCENDING (night passes). Full image IDs provided above for reproducibility.

Tableau S2. Sentinel-2 images used (April 2022)

Source: Google Earth Engine — COPERNICUS/S2_SR | DOI: 10.5270/S2_-742ikth | Bands: B2, B3, B4, B8, B11, B12 | Resolution: 10 m | Level: Surface Reflectance (SR)

#	Image ID	Satellite	Date	MGRS Tile	Cloud (%)
1	20220401T072211_20220401T072255_T38KLU	Sentinel-2A	2022-04-01	38KLU	0.002
2	20220401T072211_20220401T072255_T38KLV	Sentinel-2A	2022-04-01	38KLV	0.004
3	20220403T071159_20220403T071158_T38KLU	Sentinel-2B	2022-04-03	38KLU	1.534
4	20220403T071159_20220403T071158_T38KLV	Sentinel-2B	2022-04-03	38KLV	2.575
5	20220403T071159_20220403T071158_T38KMU	Sentinel-2B	2022-04-03	38KMU	0.332
6	20220403T071159_20220403T071158_T38KMV	Sentinel-2B	2022-04-03	38KMV	0.926
7	20220406T072159_20220406T072159_T38KLU	Sentinel-2B	2022-04-06	38KLU	0.002
8	20220406T072159_20220406T072159_T38KLV	Sentinel-2B	2022-04-06	38KLV	0.047
9	20220408T071201_20220408T071329_T38KLU	Sentinel-2A	2022-04-08	38KLU	1.487

10	20220408T071201_20220408T071329_T38KLV	Sentinel-2A	2022-04-08	38KLV	0.020
11	20220408T071201_20220408T071329_T38KMU	Sentinel-2A	2022-04-08	38KMU	3.301
12	20220408T071201_20220408T071329_T38KMV	Sentinel-2A	2022-04-08	38KMV	1.229
13	20220413T071159_20220413T071541_T38KLU	Sentinel-2B	2022-04-13	38KLU	0.015
14	20220413T071159_20220413T071541_T38KLV	Sentinel-2B	2022-04-13	38KLV	0.013
15	20220413T071159_20220413T071541_T38KMU	Sentinel-2B	2022-04-13	38KMU	0.000
16	20220413T071159_20220413T071541_T38KMV	Sentinel-2B	2022-04-13	38KMV	0.001
17	20220416T072159_20220416T072535_T38KLU	Sentinel-2B	2022-04-16	38KLU	2.757
18	20220416T072159_20220416T072535_T38KLV	Sentinel-2B	2022-04-16	38KLV	0.510
19	20220418T071211_20220418T071205_T38KLU	Sentinel-2A	2022-04-18	38KLU	18.242
20	20220418T071211_20220418T071205_T38KLV	Sentinel-2A	2022-04-18	38KLV	3.990
21	20220418T071211_20220418T071205_T38KMU	Sentinel-2A	2022-04-18	38KMU	10.488
22	20220418T071211_20220418T071205_T38KMV	Sentinel-2A	2022-04-18	38KMV	8.374
23	20220428T071211_20220428T071336_T38KLU	Sentinel-2A	2022-04-28	38KLU	11.599
24	20220428T071211_20220428T071336_T38KLV	Sentinel-2A	2022-04-28	38KLV	0.196
25	20220428T071211_20220428T071336_T38KMU	Sentinel-2A	2022-04-28	38KMU	1.363

Notes: Cloud cover filter applied: < 20%. Images composited using median reducer in Google Earth Engine (Gorelick et al., 2017; doi: 10.1016/j.rse.2017.06.031). Full image IDs provided above for reproducibility.

SI.3 Preprocessing Summary

Sentinel-1:

- Temporal aggregation: Median composite of all April 2022 acquisitions
- Radiometric calibration: Ground Range Detected (GRD) processing
- Speckle filtering: Applied during GRD processing
- Terrain correction: Applied automatically within the COPERNICUS/S1_GRD GEE pipeline (SRTM 1 Arc-Second Global, DOI: 10.5066/F7PR7TFT) — not a user-introduced predictor variable
- Output bands: VV, VH polarizations at 10m resolution

Sentinel-2:

- Cloud filtering: CLOUDY_PIXEL_PERCENTAGE < 20%
- Temporal aggregation: Median composite of cloud-free pixels
- Atmospheric correction: Level 2A (surface reflectance)
- Output bands: B2, B3, B4, B8, B11, B12 at 10m resolution

- Derived indices: $NDVI = (B8 - B4) / (B8 + B4)$

AlphaEarth:

- Pre-processed annual embeddings for 2022 (GEE collection: GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL)
- 64 bands (A00-A63) representing multidimensional ecosystem features
- Integration of spectral, textural, topographic (Copernicus DEM GLO-30), and climatic information (ERA5-Land)
- Native resolution: 10m

S2. Spatial Cross-Validation Design

S2.1 Fold Creation Algorithm

Spatial folds were created using k-means clustering of plot coordinates to ensure geographic separation between training and test sets (Valavi et al., 2018):

```
# Create 5 spatial folds using k-means
set.seed(123)
coords <- terrain_df[, c("x", "y")]
folds <- kmeans(coords, centers = 5, nstart = 25)$cluster
terrain_df$fold <- folds
```

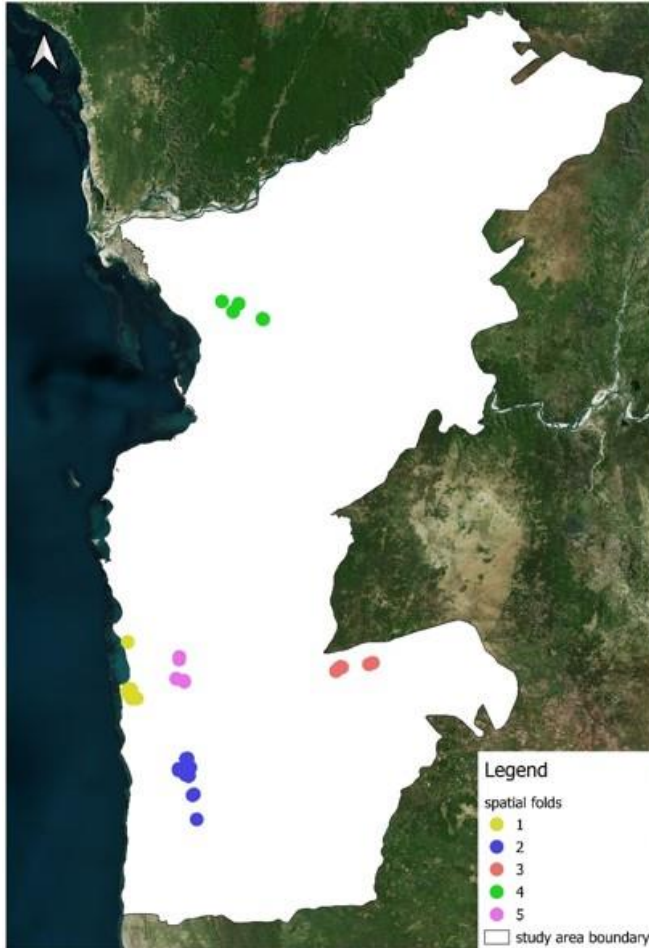


Figure S2.1: Distribution of spatial folds

The spatial distribution ensures:

- Geographic separation between training and validation data
- Balanced sample sizes across folds (n=7-9 plots per fold)
- Representative coverage of environmental gradients

S2.2 Spatial Autocorrelation Assessment

Moran's I test confirmed absence of spatial autocorrelation in model residuals:

- Moran's I = -0.025
- p-value = 0.5
- Interpretation: No significant spatial autocorrelation, validating independence assumption

S2.3 Cross-Validation Procedure

For each of the 5 folds:

1. Training: 4 folds (~33 plots)
2. Testing: 1 fold (~8 plots)
3. Model fitting with square-root transformed response
4. Prediction on test fold
5. Back-transformation for evaluation

This procedure was applied consistently for both field-calibrated and photo-calibrated models.

S3. Random Forest Hyperparameters

S3.1 Configuration-Specific Parameters

Configuration	Variables	mtry Space	Search Optimal mtry	ntree	nodesize
Sentinel-1	2 (VV, VH)	$\sqrt{p} \approx 1-2$	1	500	5
Sentinel-2	7 (B2-B12, NDVI)	$\sqrt{p} \approx 2-3$	3	500	5
S1+S2	9 (combined)	$\sqrt{p} \approx 3$	3	500	5
AlphaEarth bands)	(64 64 (A00-A63)	[8, 16, 32, 48]	16	500	5
AlphaEarth (Top 20)	20 (top bands)	[4, 6, 8, 12, 16]	8	500	5

S3.2 Hyperparameter Optimization

Grid Search Protocol:

Example for AlphaEarth 64 bands

```
mtry_values <- c(8, 16, 32, 48)
```

```
oob_errors <- numeric(length(mtry_values))
```

```
for(i in 1:length(mtry_values)) {
  rf_temp <- randomForest(
    x = train_data,
    y = train_resp,
    ntree = 100,
    mtry = mtry_values[i]
  )
}
```

```
oob_errors[i] <- tail(rf_temp$mse, 1)
}
```

```
optimal_mtry <- mtry_values[which.min(oob_errors)]
```

Rationale:

- `ntree = 500`: Ensures stability of predictions (Probst et al., 2019)
- `mtry` optimization: Balances between bias and variance
- `nodesize = 5`: Standard for regression problems (Breiman, 2001)

S3.4 Variable Transformation

Response variable (woody cover) underwent square-root transformation for variance stabilization:

```
# Transformation
```

```
terrain_df$response_sqrt <- sqrt(terrain_df$response)
```

```
# Model fitting
```

```
rf_model <- randomForest(
  x = train[, predictor_candidates],
  y = train$response_sqrt,
  ntree = 500,
  mtry = optimal_mtry
)
```

```
# Back-transformation
```

```
preds_sqrt <- predict(rf_model, test)
```

```
preds <- preds_sqrt^2
```

This transformation addresses heteroscedasticity common in proportion data

S3.3 Variable importance

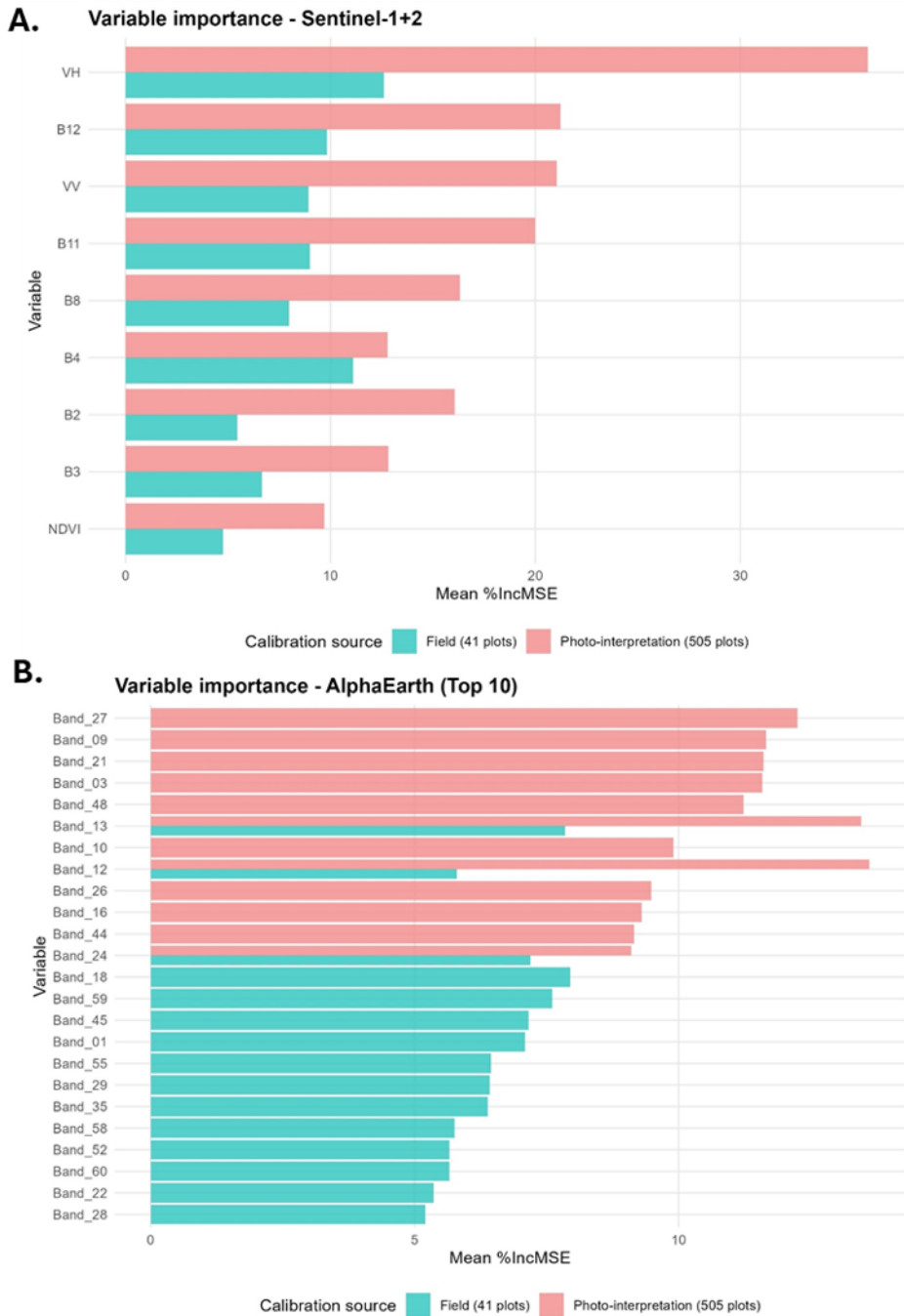
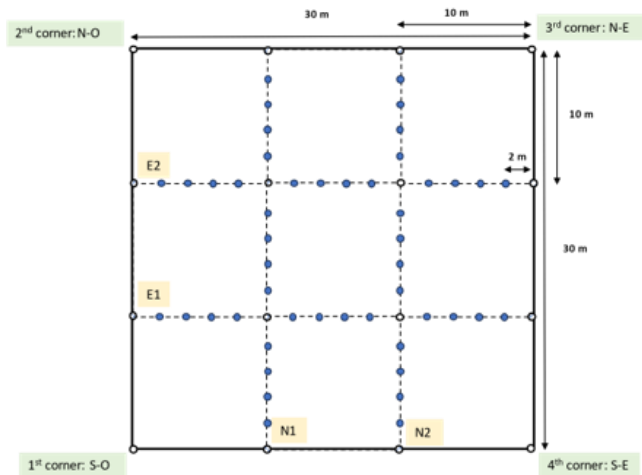


Figure S3.3: Variable importance (Mean %IncMSE) for woody cover prediction models under field calibration (41 plots) and photointerpretation (505 plots). (A) Sentinel-1+2: VH cross-polarised backscatter ranked highest under both calibration

sources, followed by SWIR (B12, B11) and NIR (B8) bands. (B) AlphaEarth (Top 10): importance was concentrated in few latent dimensions under field calibration, whereas photointerpretation produced a more diffuse distribution across a broader set of embedding dimensions.

S4. Field Sampling Protocol

S4.1 Point-Intercept Method



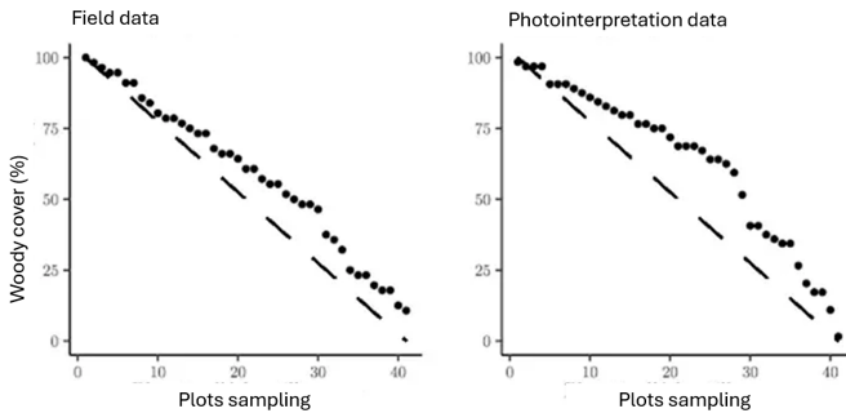


Figure S4.2: Distribution of woody cover values in field data and photointerpretation data

S5. Photo-interpretation Protocol

S5.1 Sampling Design

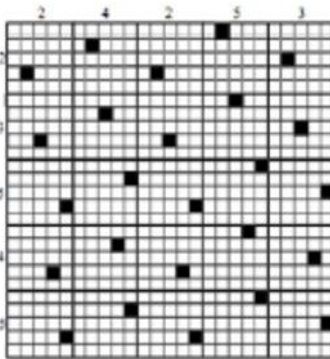


Figure S5.1: Sampling design of the distribution of photo-interpretation plots (Stratified Systematic Unaligned Sampling- SSUS)

Stratified Systematic Unaligned Sampling:

1. Created systematic grid overlay of study area
2. Randomly selected one point per grid cell
3. Environmental stratification based on:
 - Topography (elevation from NASADEM — displayed as contextual layer in Collect Earth Online, not used as model predictor)
 - Climate (mean annual precipitation)
 - Vegetation indices (NDVI, EVI, NDWI)

This design ensures:

- Unbiased spatial distribution
- Coverage of full environmental gradient
- Operational efficiency in photo-interpretation

S5.2 Plot Configuration

Protocol details:

- Plot size: 50 × 50 m (2,500 m²)
- Grid configuration: 7 × 7
- Total points per plot: 49
- Point spacing: ~7 m
- Data source: VHR orthophotographs (2021-2022)
- Platform: Collect Earth Online (Bey et al., 2016)
- Measurement criterion: Same as field (woody ≥1 m height)

Calculation:

Woody Cover = (Number of points with woody presence / 49) × 100%

S5.3 Quality Control

- All photo-interpretation performed by two trained operators following a standardized protocol
- Independent cross-validation of random subset (10%, n=50) performed by second operator for inter-rater reliability assessment
- Inter-rater reliability: $\kappa = 0.87$ (substantial agreement); no formal inter-rater reliability assessment was conducted across the full dataset — constitutes a recognized limitation (see Section 4.4)
- Ambiguous cases reviewed with field reference data

S6. Complete R Code

S6.1 Core Analysis: Sentinel Data

```
# =====
# WOODY COVER ANALYSIS WITH RANDOM FOREST
# COMPARISON: FIELD vs PHOTO-INTERPRETATION CALIBRATION
# DATA: SENTINEL-1 & SENTINEL-2
# =====

# 0. LIBRARIES =====
```

```

library(sf)
library(raster)
library(randomForest)
library(dplyr)
library(ggplot2)
library(gridExtra)

# 1. DATA LOADING =====
terrain_data <- st_read("Final_Data42.shp")
photo_data <- st_read("Final_Data505.shp")
sentinel1 <- stack("Sentinel1_042022.tif")
sentinel2 <- stack("Sentinel2_042022.tif")
predictors <- stack(sentinel1, sentinel2)

# 2. DATA PREPARATION =====
# Transform CRS
terrain_data <- st_transform(terrain_data, crs(predictors))
photo_data <- st_transform(photo_data, crs(predictors))

# Extract coordinates and values
terrain_coords <- st_coordinates(st_centroid(terrain_data))
photo_coords <- st_coordinates(st_centroid(photo_data))
terrain_values <- raster::extract(predictors, terrain_data,
                                  fun = mean, df = TRUE, na.rm = TRUE)
photo_values <- raster::extract(predictors, photo_data,
                                 fun = mean, df = TRUE, na.rm = TRUE)

# Create dataframes
terrain_df <- cbind(terrain_values,
                  response = terrain_data$FieldWC,
                  x = terrain_coords[,1],
                  y = terrain_coords[,2])
photo_df <- cbind(photo_values,
                 response = photo_data$PhtInWC,
                 x = photo_coords[,1],
                 y = photo_coords[,2])

# Calculate NDVI and sqrt transformation
terrain_df$NDVI <- with(terrain_df, (B8 - B4) / (B8 + B4))
photo_df$NDVI <- with(photo_df, (B8 - B4) / (B8 + B4))
terrain_df$response_sqrt <- sqrt(terrain_df$response)

```

```

photo_df$response_sqrt <- sqrt(photo_df$response)

# Impute missing values
terrain_df <- terrain_df %>%
  mutate(across(where(is.numeric),
    ~ifelse(is.na(.), median(., na.rm = TRUE), .)))
photo_df <- photo_df %>%
  mutate(across(where(is.numeric),
    ~ifelse(is.na(.), median(., na.rm = TRUE), .)))

# Predictor candidates
predictor_candidates <- c("VV", "VH", "B2", "B3", "B4",
  "B8", "B11", "B12", "NDVI")

# 3. OPTIMIZE mtry =====
best_mtry <- tuneRF(
  x = terrain_df[, predictor_candidates],
  y = terrain_df$response_sqrt,
  ntreeTry = 500,
  stepFactor = 1.5,
  improve = 0.01,
  trace = TRUE,
  plot = TRUE
)
optimal_mtry <- best_mtry[which.min(best_mtry[, 2]), 1]

# 4. CREATE SPATIAL FOLDS =====
set.seed(123)
folds <- kmeans(terrain_df[, c("x", "y")],
  centers = 5, nstart = 25)$cluster
terrain_df$fold <- folds

# 5. CROSS-VALIDATION =====
results_terrain <- data.frame()
results_photo <- data.frame()
predictions_terrain <- data.frame()
predictions_photo <- data.frame()

for (f in 1:5) {
  cat("Processing fold", f, "/5...\n")

```

```
train <- terrain_df[terrain_df$fold != f, ]
test <- terrain_df[terrain_df$fold == f, ]

# --- TERRAIN MODEL ---
rf_terrain <- randomForest(
  x = train[, predictor_candidates],
  y = train$response_sqrt,
  ntree = 500,
  mtry = optimal_mtry,
  importance = TRUE
)

preds_terrain_sqrt <- predict(rf_terrain,
                             test[, predictor_candidates])
preds_terrain <- preds_terrain_sqrt^2

rmse_terrain <- sqrt(mean((test$response - preds_terrain)^2))
r2_terrain <- cor(test$response, preds_terrain)^2

results_terrain <- rbind(results_terrain,
                         data.frame(Fold = f,
                                    RMSE = rmse_terrain,
                                    R2 = r2_terrain))

predictions_terrain <- rbind(predictions_terrain,
                              data.frame(Fold = f,
                                         Observed = test$response,
                                         Predicted = preds_terrain))

# --- PHOTO MODEL ---
rf_photo <- randomForest(
  x = photo_df[, predictor_candidates],
  y = photo_df$response_sqrt,
  ntree = 500,
  mtry = optimal_mtry,
  importance = TRUE
)

preds_photo_sqrt <- predict(rf_photo,
                           test[, predictor_candidates])
preds_photo <- preds_photo_sqrt^2
```

```

rmse_photo <- sqrt(mean((test$response - preds_photo)^2))
r2_photo <- cor(test$response, preds_photo)^2

results_photo <- rbind(results_photo,
  data.frame(Fold = f,
    RMSE = rmse_photo,
    R2 = r2_photo))

predictions_photo <- rbind(predictions_photo,
  data.frame(Fold = f,
    Observed = test$response,
    Predicted = preds_photo))
}

# 6. RESULTS =====
# Global metrics
rmse_terrain_global <- sqrt(mean((predictions_terrain$Observed -
  predictions_terrain$Predicted)^2))
r2_terrain_global <- cor(predictions_terrain$Observed,
  predictions_terrain$Predicted)^2

rmse_photo_global <- sqrt(mean((predictions_photo$Observed -
  predictions_photo$Predicted)^2))
r2_photo_global <- cor(predictions_photo$Observed,
  predictions_photo$Predicted)^2

cat("FIELD - RMSE:", round(rmse_terrain_global, 4),
  "| R^2:", round(r2_terrain_global, 4), "\n")
cat("PHOTO - RMSE:", round(rmse_photo_global, 4),
  "| R^2:", round(r2_photo_global, 4), "\n")

```

S6.2 Sampling Density Analysis

```

# =====
# SENSITIVITY ANALYSIS: PHOTO-INTERPRETATION SAMPLE SIZE
# =====

# 1. PARAMETERS =====
sample_sizes <- c(10, 25, 50, 75, 100, 150, 200, 250,
  300, 350, 400, 450, 505)
n_repetitions <- 100

```

```

# 2. SAMPLING FUNCTIONS =====
# Dispersed sampling
get_dispersed_sample <- function(photo_data, centroids, n, seed = NULL) {
  if (!is.null(seed)) set.seed(seed)

  bbox <- st_bbox(photo_data)
  grid_size <- ceiling(sqrt(n * 2))
  grid_sf <- st_make_grid(photo_data, n = c(grid_size, grid_size))
  grid_sf <- st_sf(id = 1:length(grid_sf), geometry = grid_sf)

  intersects <- st_intersects(centroids, grid_sf)
  grid_ids <- sapply(intersects, function(x) ifelse(length(x) > 0, x[1], NA))

  temp_data <- cbind(as.data.frame(centroids), grid_id = grid_ids)
  random_indices <- c()

  for (grid_id in unique(temp_data$grid_id[!is.na(temp_data$grid_id)])) {
    points_in_grid <- which(temp_data$grid_id == grid_id)
    if (length(points_in_grid) > 0) {
      sampled_point <- sample(points_in_grid, 1)
      random_indices <- c(random_indices, sampled_point)
    }
  }

  if (length(random_indices) < n) {
    remaining_points <- setdiff(1:nrow(centroids), random_indices)
    additional_points <- sample(remaining_points,
                               min(n - length(random_indices),
                                   length(remaining_points)))
    random_indices <- c(random_indices, additional_points)
  }

  if (length(random_indices) > n) {
    random_indices <- sample(random_indices, n)
  }

  return(random_indices)
}

# Clustered sampling
get_clustered_sample <- function(photo_data, centroids, n, seed = NULL) {

```

```

if (!is.null(seed)) set.seed(seed)

coords <- st_coordinates(centroids)
center_idx <- sample(1:nrow(coords), 1)
center_point <- coords[center_idx, , drop = FALSE]

distances <- sqrt((coords[, 1] - center_point[1])^2 +
                  (coords[, 2] - center_point[2])^2)
sorted_idx <- order(distances)
clustered_indices <- sorted_idx[1:min(n, length(sorted_idx))]

return(clustered_indices)
}

# 3. MODEL EVALUATION =====
evaluate_model <- function(train_data, validation_data, raster_names) {
  if (nrow(train_data) < 10) return(NULL)

  rf_formula <- as.formula(paste("woodcover ~",
                                paste(raster_names, collapse = " + ")))

  n_trees <- min(500, max(100, nrow(train_data) * 2))
  mtry_param <- min(length(raster_names),
                    max(2, floor(length(raster_names)/3)))

  model <- tryCatch({
    randomForest(
      formula = rf_formula,
      data = train_data,
      ntree = n_trees,
      mtry = mtry_param,
      importance = TRUE
    )
  }, error = function(e) NULL)

  if (is.null(model)) return(NULL)

  preds <- predict(model, newdata = validation_data)
  rmse <- sqrt(mean((preds - validation_data$woodcover)^2, na.rm = TRUE))
  r2 <- cor(preds, validation_data$woodcover, use = "complete.obs")^2

```

```
mean_observed <- mean(validation_data$woodcover, na.rm = TRUE)
rel_error_mean <- (rmse / mean_observed) * 100
```

```
return(list(
  metrics = list(
    rmse = rmse,
    r2 = r2,
    rel_error_mean = rel_error_mean,
    n_train = nrow(train_data),
    n_validation = nrow(validation_data)
  ),
  model = model,
  predictions = preds,
  observations = validation_data$woodcover
))
}
```

```
# 4. RUN EXPERIMENTS =====
```

```
all_results <- list()
```

```
for (n in sample_sizes) {
  message("Processing sample size: ", n)

  if (n == 505) {
    # Full dataset evaluation
    full_eval <- evaluate_model(photo_df, terrain_df,
                               predictor_candidates)
    if (!is.null(full_eval)) {
      all_results[[paste0("dispersed_505")]] <- list(
        size = 505,
        rep = 1,
        strategy = "dispersed",
        evaluation = full_eval
      )
    }
  }
  next
}
```

```
for (rep in 1:n_repetitions) {
  rep_seed <- 123 + n + (rep * 1000)
```

```

# Dispersed strategy
dispersed_indices <- get_dispersed_sample(photo_data,
                                         centroids_photo,
                                         n, seed = rep_seed)
if (length(dispersed_indices) >= 10) {
  dispersed_df <- photo_df[dispersed_indices, ]
  dispersed_eval <- evaluate_model(dispersed_df, terrain_df,
                                  predictor_candidates)
  if (!is.null(dispersed_eval)) {
    all_results[[paste0("dispersed_", n, "_rep", rep)]] <- list(
      size = n,
      rep = rep,
      strategy = "dispersed",
      evaluation = dispersed_eval
    )
  }
}

# Clustered strategy
clustered_indices <- get_clustered_sample(photo_data,
                                          centroids_photo,
                                          n, seed = rep_seed)
if (length(clustered_indices) >= 10) {
  clustered_df <- photo_df[clustered_indices, ]
  clustered_eval <- evaluate_model(clustered_df, terrain_df,
                                  predictor_candidates)
  if (!is.null(clustered_eval)) {
    all_results[[paste0("clustered_", n, "_rep", rep)]] <- list(
      size = n,
      rep = rep,
      strategy = "clustered",
      evaluation = clustered_eval
    )
  }
}
}
}
}

```

S7. Performance Metrics Formulas

S7.1 Root Mean Square Error (RMSE)

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

- y_i : observed value
- \hat{y}_i : predicted value
- n : number of observations

Interpretation: RMSE quantifies the average magnitude of prediction errors in the same units as the response variable (proportion of woody cover). Lower values indicate better model performance.

S7.2 Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- y_i : observed value
- \hat{y}_i : predicted value
- \bar{y} : mean of observed values
- n : number of observations

Interpretation: R^2 represents the proportion of variance in the observed data explained by the model. Values range from 0 to 1, with 1 indicating perfect prediction.

S7.3 Mean Relative Error

$$\text{Relative Error}_i = 100 \times \frac{|\hat{y}_i - y_i|}{y_i}$$

Interpretation: Mean Relative Error = $\frac{1}{n} \sum_{i=1}^n \text{Relative Error}_i$ Expresses prediction errors as percentages of observed values, facilitating comparisons across different woody cover ranges. Lower values indicate better accuracy.

S7.4 Coefficient of Variation for Uncertainty

$$\text{CV} = 100 \times \frac{\sigma}{\mu}$$

- σ : standard deviation of predictions across the 5 spatial folds
- μ : mean of predictions across the 5 spatial folds

Interpretation: CV quantifies prediction uncertainty as a percentage. Values <10% indicate low uncertainty, 10-25% moderate uncertainty, and >25% high uncertainty.

Annexe 4: Chapter 7 - Supplementary materials

Tableau S1. Inventory of Landsat scenes used for annual woody cover mapping (1989–2025). A total of 1,527 scenes were selected from Landsat Collection 2 Level-2 products after cloud filtering. For each scene, the following information is provided: GEE scene identifier, acquisition date, sensor (TM/ETM+/OLI/OLI-2), path/row, and cloud cover percentage. Years 2002 and 2003 are absent due to insufficient scene availability following Scan Line Corrector failure (Markham et al., 2004). For years 1992, 2000, 2009, and 2010, scenes from adjacent years were pooled to constitute three-year median composites.

Year	N scenes	Sensor	Mean cloud cover (%)	Notes
1989	19	L5 TM	2.6	
1990	35	L5 TM	5.1	
1991	39	L5 TM	4.9	
1992	18	L5 TM	2.2	3-yr pool (1991–1993)
1993	47	L5 TM	3.2	
1994	24	L5 TM	3.7	
1995	47	L5 TM	5.3	
1996	35	L5 TM	5.8	
1997	22	L5 TM	5.5	
1998	56	L5 TM	4.5	
1999	24	L5 TM / L7 ETM+	3.6	
2000	40	L5 TM / L7 ETM+	4.5	3-yr pool (1999–2001)
2001	40	L5 TM / L7 ETM+	4.3	
2002	—	—	—	Excluded — SLC failure + insufficient scenes
2003	—	—	—	Excluded — SLC failure + insufficient scenes
2004	39	L5 TM / L7 ETM+	3.2	
2005	41	L5 TM / L7 ETM+	4.0	
2006	67	L5 TM / L7 ETM+	4.1	
2007	44	L5 TM / L7 ETM+	1.9	
2008	33	L5 TM / L7 ETM+	3.9	
2009	27	L5 TM / L7 ETM+	5.0	3-yr pool (2008–2010)
2010	19	L5 TM / L7 ETM+	5.2	3-yr pool (2009–2011)
2011	39	L5 TM / L7 ETM+	4.3	
2012	28	L7 ETM+	4.8	
2013	36	L8 OLI	2.8	
2014	47	L8 OLI	2.4	
2015	49	L8 OLI	3.3	

2016	52	L8 OLI	2.9	
2017	51	L8 OLI	2.8	
2018	61	L8 OLI	3.6	
2019	48	L8 OLI	1.4	
2020	58	L8 OLI	3.2	
2021	67	L8 OLI / L9 OLI-2	4.1	
2022	55	L9 OLI-2	3.9	
2023	45	L9 OLI-2	2.7	
2024	57	L9 OLI-2	2.0	
2025	49	L9 OLI-2	5.1	
Total	1 527			

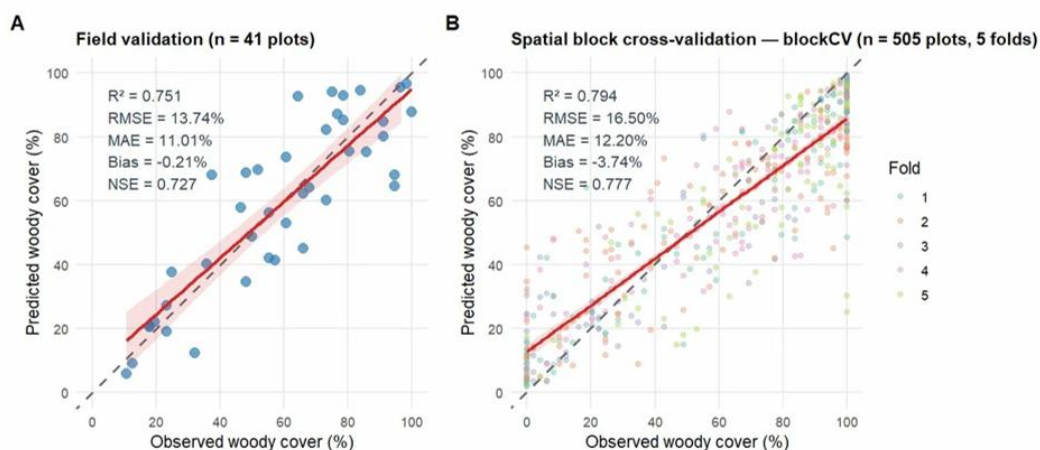


Figure S1. Predicted vs. observed woody cover for model validation (2022 reference year). (A) External validation on 41 independent field plots (April–May 2022); each point represents one 30×30 m plot. (B) Spatial block cross-validation (blockCV, 10-km blocks, 5 folds) on 505 photo-interpreted plots; colors indicate spatial folds. In both panels, the dashed line represents the 1:1 perfect agreement line and the red line the fitted linear regression with 95% confidence interval. Performance metrics: R^2 = coefficient of determination; RMSE = root mean square error; MAE = mean absolute error; Bias = mean prediction error; NSE = Nash-Sutcliffe efficiency coefficient.

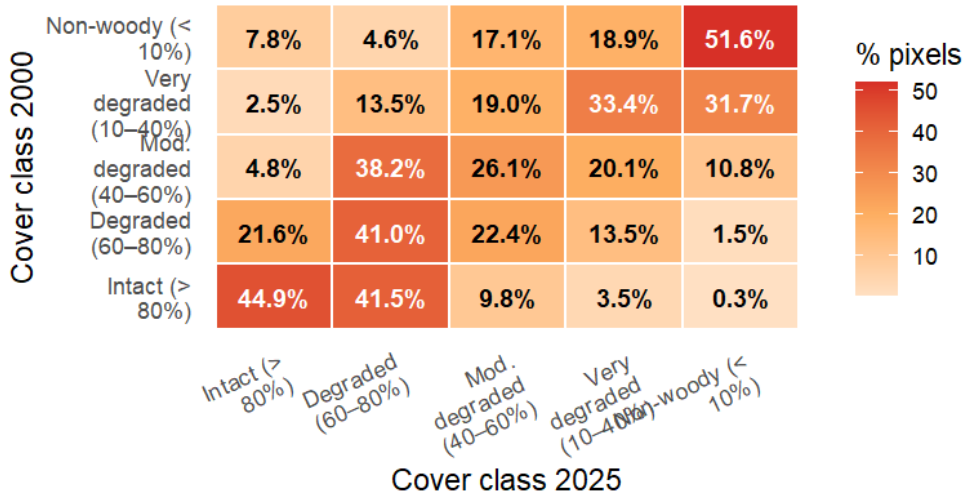


Figure S2. Cover class transition matrix (2000–2025). Values represent the percentage of pixels transitioning from each initial cover class (rows, 2000) to each final cover class (columns, 2025). Diagonal values indicate class persistence. $n = 20,000$ pixels sampled randomly across the study area.