



© A. HAVINGA, 2021

CERTIFICAT COMPLÉMENTAIRE EN GÉOMATIQUE

CARTOGRAPHIE DES MILIEUX DU MASSIF FORESTIER MONTAGNEUX DE SORATA (NORD DE MADAGASCAR)

Exploration du logiciel ArcGIS Pro® et des outils de télédétection

Travail réalisé par Cécile Donati

Septembre 2023

Sous la direction de

Dr. Pascal Martin, Dr. Louis Nusbaumer et Prof. Anthony Lehmann

Remerciements

Pour leur disponibilité, leur assistance et leurs recommandations avisées pendant la réalisation de ce travail, je désire remercier tout particulièrement :

- Pascal Martin, adjoint scientifique aux Conservatoire et Jardin botaniques de Genève, pour son implication et son aide précieuse apportées tout au long du présent travail, ainsi que
- Louis Nusbaumer, conservateur aux Conservatoire et Jardin botaniques de Genève, pour les mêmes raisons.

Je souhaite également remercier toutes les personnes ayant contribué d'une façon ou d'une autre à l'accomplissement du présent travail :

- Anthony Lehmann et Gregory Giuliani, pour le temps qu'ils m'ont consacré lorsque je rencontrais des difficultés ;
- L'ensemble de l'équipe des Conservatoire et Jardin botaniques de Genève, pour leur accueil chaleureux, leur bonne humeur et leurs précieux conseils ;
- Ma famille et mes proches qui ont su, dans les moments de doute, de remise en question, de désespoir et de frustration, m'apporter tout leur soutien et leurs encouragements.

Résumé

Le travail présenté ici s'inscrit dans une démarche à la fois expérimentale et exploratoire, avec pour objectif la création d'une cartographie des milieux du massif forestier de Sorata, situé au nord de Madagascar. Pour atteindre cet objectif, différentes approches et méthodes basées sur la télédétection ont été explorées, principalement en utilisant le logiciel ArcGIS Pro® d'ESRI®. Des classifications supervisées, basées notamment sur une approche par pixel, ont constitué la pierre angulaire de ce travail. Deux cartographies ont été produites, l'une basée sur les images satellites majoritairement, l'autre incluant des données de relevés de terrain effectués entre fin 2021 et début 2023.

Dans l'ensemble, les résultats obtenus se sont avérés satisfaisants, bien que des erreurs de classification fréquentes aient été identifiées. Ces erreurs ont découlé de divers facteurs, tels que la présence de nuages ou des regroupements incorrects au sein des classes. Les évaluations de précision des classifications, mesurées par le coefficient de Kappa qui évalue la concordance entre les classes prédites et réelles, ont donné des valeurs de 0.99 ± 0.0054 pour la méthode où les points d'évaluation de la précision ont été mis à jour automatiquement et de 0.718 ± 0.078 pour la méthode utilisant une mise à jour manuelle.

Ce travail a constitué une étape cruciale vers l'élaboration d'une cartographie précise des milieux du massif forestier de Sorata. Malgré les défis, cette cartographie représente un outil précieux pour identifier et protéger les milieux fragiles ou rares, contribuant ainsi à la gestion et à la préservation de ce site et de ses ressources. L'expérience développée lors de ce travail s'est révélée inestimable, offrant une opportunité de mettre en pratique les connaissances et les compétences acquises.

Table des matières

1.	Introduction	1
2.	Contexte et objectif du projet.....	1
3.	Concepts théoriques.....	2
3.1.	Systèmes d'Information Géographique (SIG).....	2
3.2.	Imagerie spatiale.....	3
3.2.1.	Satellites et télédétection	3
3.2.2.	Mission Sentinel-2.....	3
3.3.	Signatures spectrales	4
3.4.	Indices spectraux	5
3.5.	Classification	5
3.5.1.	Classification par pixel et par objet.....	6
3.5.2.	Classification non-supervisée et supervisée	6
3.5.3.	Méthodes de classification : les classificateurs.....	7
3.5.4.	Échantillons d'entraînement.....	8
3.5.5.	Évaluation de la précision	8
4.	Méthodologie	10
4.1.	Site d'étude.....	10
4.2.	Données utilisées et acquisition	11
4.3.	Prétraitements des images	13
4.3.1.	Découpage de l'image en fonction du site d'étude	13
4.3.2.	Création de mosaïques	15
4.3.3.	Création d'un masque de nuages	15
4.3.4.	Création d'indices spectraux.....	16
4.3.5.	Création de composites multibandes	17
4.4.	Classifications des images satellites sur ArcGIS Pro®	17
4.4.1.	Classifications non-supervisées.....	17
4.4.2.	Classifications supervisées	17
4.4.2.1.	Données d'entraînement.....	17
4.4.2.2.	Classification	18
4.5.	Évaluation de la précision des classifications.....	19
5.	Résultats	20

5.1.	Comparaison du paysage entre les saisons humide et sèche	20
5.2.	Classifications supervisées	20
5.2.1.	Classification « visuelle ».....	20
5.2.2.	Classification « réalité du terrain »	24
6.	Discussion	24
7.	Conclusion	25
8.	Bibliographie.....	27
9.	Annexes	32

Liste des abréviations

AOT	Épaisseur optique des aérosols (<i>Aerosol Optical Thickness</i>)
CJBG	Conservatoire et Jardin botaniques de Genève
EPSG	<i>European Petroleum Survey Group</i>
ESA	Agence spatiale européenne (<i>European Space Agency</i>)
FDHS	Forêts Dense Humide Sempervirente
GCS	<i>Geographic Coordinate System</i>
ISO	<i>Iterative Self Organising</i>
ISODATA	Analyse itérative des données auto-organisées (<i>Iterative Self Organising Data</i>)
MBG	<i>Missouri Botanical Garden</i>
MGRS	Système de référence de grille militaire (<i>Military Grid Reference System</i>)
ML	Maximum de vraisemblance (<i>Maximum Likelihood</i>)
NDBI	<i>Normalized Difference Built-up Index</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NDWI	<i>Normalized Difference Water Index</i>
OOC	Classification orientée objet (<i>Object-Oriented Classification</i>)
PBC	Classification basée sur les pixels (<i>Pixel-Based Classification</i>)
PIR	Proche infrarouge
RF	Forêts d'arbres décisionnels (<i>Random Forest</i>)
SCL	Classification des scènes (<i>Scene Classification</i>)
SIG	Systèmes d'Information Géographique
SVM	Machines vectorielles de support (<i>Support Vector Machine</i>)
SWIR	Infrarouge Court
UE	Union Européenne
UTC	<i>Universal Time Coordinated</i>
UTM	<i>Universal Transverse Mercator</i>
VRE	<i>Vegetation Red Edge</i>
WGS	<i>World Geodetic System</i>
WVP	<i>Water Vapour Pressure</i>

Liste des figures

Figure 1 Localisation du massif forestier de Sorata (encadré jaune). Le gradient de couleur représente des différences d'altitudes, allant du bleu (altitude basse) au rouge (altitude élevée) (modifié d'après Rakotovo et al., 2006).....	2
Figure 2 Comparaison des caractéristiques spécifiques des satellites Sentinel-1 et Sentinel-2 (modifié d'après Kpienbaareh et al., 2021).....	4
Figure 3 Signatures spectrales du sol, de la végétation et de l'eau (Source : https://seos-project.eu/)..	5
Figure 4 Végétation du massif forestier de Sorata cartographiée dans le cadre de l'Atlas de la végétation de Madagascar (UTM projection) (Moat & Smith, 2007).	11
Figure 5 Massif forestier de Sorata et d'Andravory, fond de carte World Imagery.	14
Figure 6 Indice NDVI calculé avec Raster Calculator pour le 14 février 2023 (saison humide) (image de gauche) et le 8 août 2021 (saison sèche) (image de droite).	20
Figure 7 Cartographie des milieux du massif forestier de Sorata et alentours, basée sur des images satellites ainsi que des données d'altitude et de courbure, réalisée par classification supervisée avec ArcGIS Pro®. Les classifications ont été effectuées avec un ensemble limité (10) de données d'entraînement (image de gauche) ainsi qu'avec le double de données (image de droite). La zone tampon définie pour les données d'entraînement est de 20 mètres. (FDHS = Forêts Dense Humide Sempervirente).....	21
Figure 8 Cartographie des milieux du massif forestier de Sorata et alentours réalisée par classification supervisée avec ArcGIS Pro® et incluant les données des relevés de terrain réalisés entre fin 2021 et début 2023. Les classifications ont été effectuées avec un ensemble limité de données d'entraînement (10, excepté pour les forêts de bambous (5)) (image de gauche) ainsi qu'avec le double de données (image de droite). La zone tampon définie pour les données d'entraînement est de 20 mètres. (FDHS = Forêts Dense Humide Sempervirente).....	24

Liste des tableaux

Tableau 1 Exemple de matrice de confusion présentant des valeurs fictives (Source : https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/compute-confusion-matrix.htm).....	9
Tableau 2 Informations relatives aux images satellites téléchargées (EO Browser, 2023).	12
Tableau 3 Informations relatives aux bandes spectrales disponibles avec les satellites Sentinel-2 (modifié d'après Taconet, 2019). La bande de cirrus 10 n'est pas fournie, car elle ne contient aucune information de surface (User Guides - Sentinel-2 MSI Level-2 Processing, 2022).	12
Tableau 4 Coordonnées des angles (vertex) relevés pour l'entité de découpage.	14
Tableau 5 Méthodes de priorisation d'affichage des images disponibles avec l'outil Mosaic Rasters (Source : https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/mosaic-rasters.htm).....	15
Tableau 6 Matrice de confusion générée automatiquement avec 100 points de précision de l'évaluation distribués de manière aléatoire stratifiée par classe (IC = intervalle de confiance).....	22
Tableau 7 Matrice de confusion générée manuellement avec 10 points de précision de l'évaluation distribués de manière aléatoire stratifiée par classe (IC = intervalle de confiance).....	23

1. Introduction

Selon Augustin-Pyramus de Candolle, précurseur de la botanique moderne à Genève, un jardin botanique doit remplir trois rôles distincts : l'enseignement, la recherche et l'acclimatation. À sa demande, la création d'un jardin botanique en 1816 a été rendue possible grâce au soutien de l'État de Genève. En 1824, un conservatoire botanique a été ajouté au jardin, devenant essentiel pour abriter les collections de plantes sèches, notamment les importantes collections d'herbiers de de Candolle et d'Edmond Boissier (Sigrist & Bungener, 2008).

Les Conservatoire et Jardin botaniques de Genève (CJBG) occupent une place centrale en tant qu'institution muséale. Avec environ 6 millions d'échantillons et une acquisition annuelle d'environ 12 000 nouveaux échantillons, provenant de collectes internes, d'échanges avec d'autres institutions, de dons et d'achats, les CJBG jouent aujourd'hui un rôle majeur dans l'étude, l'archivage et la conservation d'espèces (CJBG, 2023).

Depuis plusieurs dizaines d'années, les CJBG sont engagés dans des projets de recherche à différentes échelles à des fins de protection et conservation des espèces et milieux. En collaboration avec d'autres institutions, ils ont notamment mené des projets de recherche à Madagascar, en particulier dans le nord de l'île (Ammann, 2011; Callmander et al., 2018; Havinga, 2022; Nusbaumer et al., 2010; Ranirison et al., 2010; Trigui, 2010). L'un des derniers projets issus de cette collaboration se concentre, depuis 2021, sur le massif forestier montagneux de Sorata avec pour but de conserver au mieux ses milieux naturels les plus rares et fragiles.

2. Contexte et objectif du projet

Le massif montagneux forestier de Sorata se situe dans le nord de Madagascar (Figure 1). Malgré son étendue de forêts tropicales encore intactes et sa situation géographique au sein du Domaine du Centre, ce massif a été peu exploré botaniquement avant 2006 et n'a obtenu de statut de protection qu'en 2015 (*Étude botanique et conservation du massif forestier montagneux méconnu de Sorata*, 2021; Favre, 2013; Glaw et al., 2021; Goodman et al., 2018). Il est confronté à des pressions importantes, notamment la déforestation due à la culture de vanille, l'élevage de zébus, l'exploitation de produits forestiers ligneux, etc. (Favre, 2013; Glaw et al., 2021; Goodman et al., 2018; Scherz et al., 2015, 2017; Vieilledent et al., 2018).

Le projet « Étude botanique et conservation du massif forestier montagneux méconnu de Sorata (nord de Madagascar) » vise à améliorer les connaissances sur cette région et à renforcer sa protection. Les deux missions de terrain réalisées entre fin 2021 et début 2023 ont permis de faire d'importantes découvertes. Le travail actuel s'inscrit dans la lignée des précédents efforts déployés et est directement lié au travail de Master qui sera entrepris par la suite. À ce stade de la connaissance botanique du massif forestier de Sorata, nous estimons qu'une cartographie détaillée de la végétation réalisée en laboratoire puis vérifiée sur le terrain dans le cadre du mémoire est indispensable pour les objectifs de recherche et de conservation du projet. Ainsi, une cartographie satellitaire « théorique » du massif, et alentours, doit être réalisée.

Le travail en laboratoire s'est déroulé aux CJBG du 1^{er} mai au 30 juin sous la forme d'un stage à plein temps. L'objectif principal a été de créer une cartographie des différentes unités de végétation qui composent le massif de Sorata par la détermination et délimitation des typologies et unités de

végétation, en se basant sur des données satellitaires ainsi que des relevés de terrain. La cartographie finale, après vérification sur le terrain entre octobre 2023 et février 2024, permettra de prioriser les habitats les plus rares et menacés à protéger en priorité et leur espèces caractéristiques, en collaboration avec les associations de conservation malgaches partenaires.

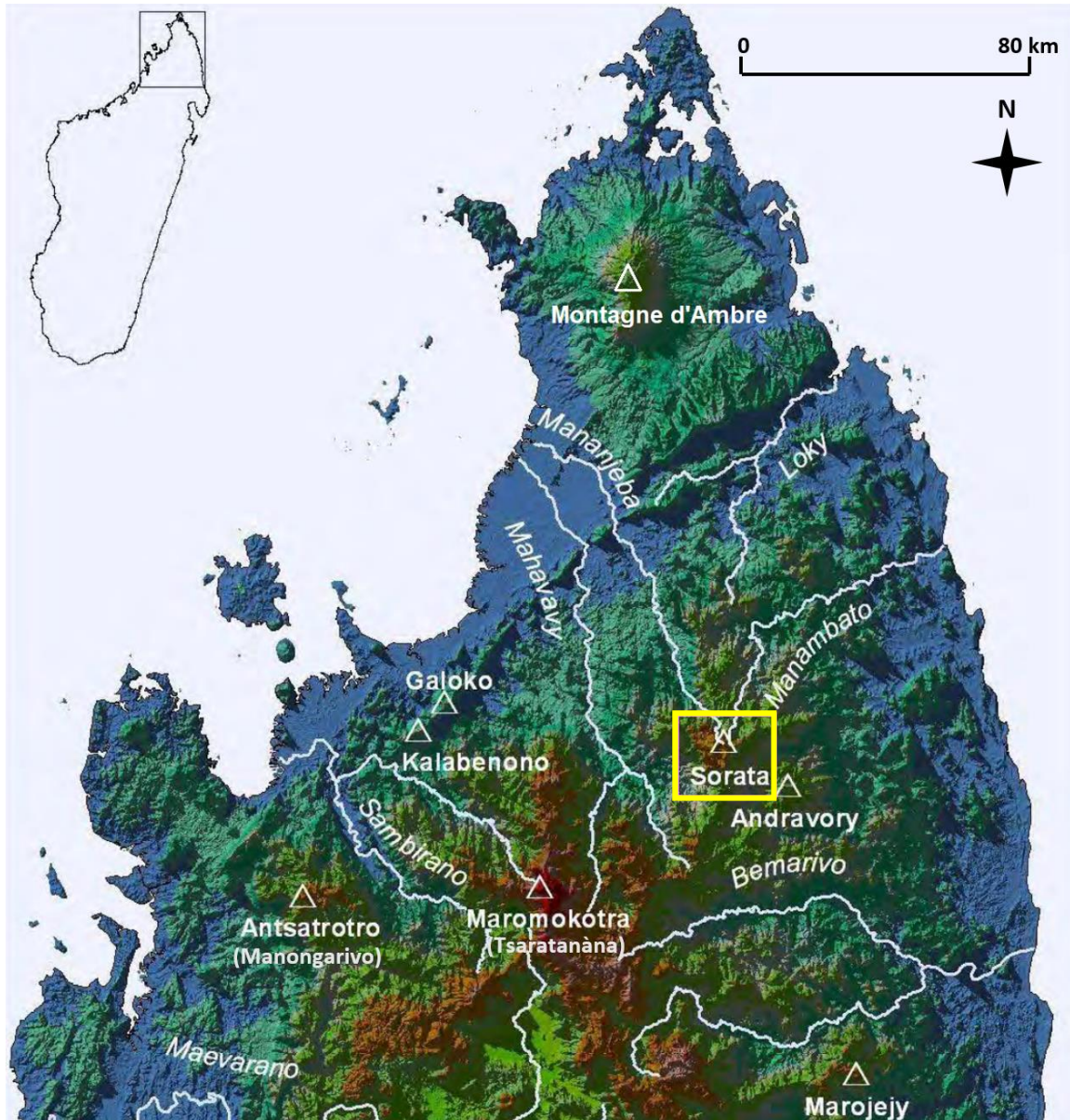


Figure 1 Localisation du massif forestier de Sorata (encadré jaune). Le gradient de couleur représente des différences d'altitudes, allant du bleu (altitude basse) au rouge (altitude élevée) (modifié d'après Rakotovo et al., 2006).

3. Concepts théoriques

3.1. Systèmes d'Information Géographique (SIG)

Les Systèmes d'Information Géographique (SIG) sont des outils technologiques puissants jouant un rôle crucial dans une multitude de domaines (géographie, météorologie, biologie, santé, etc.) grâce à leur capacité à visualiser, collecter, stocker, analyser et diffuser des informations et données géographiques représentées sous différentes formes (entités, rasters, attributs, topologies, réseaux, etc.). Dans un SIG, les données sont organisées en couches thématiques. Les SIG facilitent l'intégration de données

provenant de diverses sources, favorisant la collaboration multisectorielle (*Logiciel SIG, 2023; Qu'est-ce qu'un SIG ?, 2023*).

Un SIG comprend trois volets essentiels pour manipuler et afficher des données géographiques (géodonnées, géovisualisation et géotraitement) représentés dans ArcGIS® d'ESRI® par le catalogue et la géodatabase, la carte ainsi que la boîte à outils (*Logiciel SIG, 2023*).

Les données sont stockées principalement de deux manières : en mode « image » (raster) où le site étudié est composé d'une matrice de pixels, chacun portant une valeur unique, ou en mode « objet » (vecteur) où les objets sont représentés par des entités géométriques ponctuelles, linéaires ou polygonales auxquelles sont associés des attributs. Ces deux modes sont utilisés ensemble dans un même projet, les données pouvant être converties d'un format à l'autre selon les besoins (*Qu'est-ce qu'un SIG ?, 2023*).

3.2. Imagerie spatiale

3.2.1. Satellites et télédétection

L'observation de la Terre par satellite représente un outil précieux de surveillance, permettant d'évaluer précisément et régulièrement l'état et l'évolution de la planète face aux changements rapides de la couverture terrestre (Chuvienco, 2020; Phiri et al., 2020). Depuis le lancement du premier satellite Landsat en 1972, de nombreux satellites¹ ont été déployés afin de fournir des données de télédétections diverses et variées (Phiri et al., 2020; Wulder et al., 2018). Le terme « télédétection » désigne toute méthode permettant d'observer la Terre à distance, couvrant ainsi un large éventail d'aspects, de la collecte à l'analyse des données (Chuvienco, 2020). La télédétection, associée aux SIG notamment, permet de fournir une quantité importante d'informations sur la surface terrestre, contribuant ainsi à un meilleur suivi et une protection efficace des écosystèmes terrestres (Chuvienco, 2020; Santos et al., 2021).

Les progrès technologiques, contribuant à l'amélioration des outils de télédétection, combinés à une disponibilité croissante des données en ligne provenant de diverses sources et fournies gratuitement, les rendant plus accessibles au public, ont élevé la télédétection au statut de ressource primordiale pour les études à grande échelle (Chuvienco, 2020; Gislason et al., 2006; Melesse et al., 2007; Phiri et al., 2020; Simoes et al., 2021; Talukdar et al., 2020; Vali et al., 2020; Wulder et al., 2018; Xue & Su, 2017).

3.2.2. Mission Sentinel-2

La mission Copernicus, fruit de la collaboration entre l'Agence spatiale européenne (ESA) et le programme d'observation de la Terre de l'Union européenne (UE), a renforcé la surveillance de la Terre grâce à ses satellites Sentinel-2 (Phiri et al., 2020). Lancés dès 2015, ces satellites équipés d'instruments multispectraux capturent 13 bandes à large spectre, allant d'une résolution spatiale de 10 à 60 mètres et offrant un aperçu détaillé de la surface terrestre. Ils fournissent des données spatiales à haute résolution, contribuant ainsi à la bonne gestion des ressources naturelles (*Découvrez nos satellites, 2023; Pelletier et al., 2016; Phiri et al., 2020; Somvanshi & Kumari, 2020*).

Les images Sentinel-2 connaissent une utilisation croissante en raison de leur résolution spatiale et temporelle élevées ainsi que de la disponibilité de bandes spectrales rouges essentielles (Figure 2) (Phiri

¹ Les capteurs embarqués sur les satellites collectent des données à partir de différentes longueurs d'onde du spectre électromagnétique (Chuvienco, 2020; Vali et al., 2020).

et al., 2020). En outre, Copernicus Sentinel Hub² offre un accès gratuit à ces données (Chuvieco, 2020). Malgré certaines limitations liées aux conditions atmosphériques ou à la topographie par exemple, Sentinel-2 demeure un outil inestimable pour la surveillance environnementale à l'échelle mondiale.

	Sentinel-1	Sentinel-2
Resolution	5 × 20 m	10, 20, and 60 m
Band type	C-band	Coastal aerosol, Blue, Green, Red, NIR, and SWIR
Revisit time	6 days	5 days
Orbit Height	693 km	786 km
Orbit inclination	98.18°	98.62°
Spectral range	3.75–7.5 cm	0.44–2.19 μm

Figure 2 Comparaison des caractéristiques spécifiques des satellites Sentinel-1 et Sentinel-2 (modifié d'après Kpienbaareh et al., 2021).

3.3. Signatures spectrales

Les surfaces terrestres, qu'il s'agisse d'eau, de sol nu ou de végétation, possèdent des caractéristiques radiantes uniques, variant en fonction de la longueur d'onde, définies par leurs signatures spectrales (Figure 3). Ces signatures spectrales résultent de l'interaction de la lumière avec les propriétés optiques des objets, telles que leur structure et leur composition (Arkebauer, 2015; *Signatures spectrales*, 2009; Walter-Shea & Biehl, 1990; Xue & Su, 2017).

Dans le cas de la végétation, le rayonnement solaire est absorbé principalement dans les parties ultraviolettes, visibles et thermiques du spectre, tandis qu'il est réfléchi dans le rayonnement infrarouge proche (PIR) (Arkebauer, 2015; Xue & Su, 2017). La réflectance des surfaces végétales et de leur couvert est influencée par de multiples facteurs, notamment l'énergie solaire, l'angle d'illumination ou encore les propriétés du substrat sous-jacent par exemple (Walter-Shea & Biehl, 1990; Xue & Su, 2017).

En ce qui concerne les zones bâties, leur émissivité élevée est principalement due aux matériaux de construction dominants. En revanche, l'émissivité des surfaces de sol nu dépendra plutôt de sa composition en matière minérale et de son taux d'humidité. L'analyse spectrale des zones urbaines montre souvent une réflectance importante dans la plage des longueurs d'onde d'infrarouge court (SWIR) par rapport au PIR (As-syakur et al., 2012).

Enfin, les approches utilisées pour identifier l'eau se basent sur l'absorption prononcée de la lumière du PIR par l'eau qui se distingue de la réflectance élevée observée pour la végétation et les sols nus dans cette même gamme spectrale. Ainsi, les données PIR permettent de mettre en évidence les surfaces végétales en blanc, tandis que les étendues d'eau apparaissent en noir (McFeeters, 1996; *Signatures spectrales*, 2009).

² Lien d'accès : <https://scihub.copernicus.eu/>.

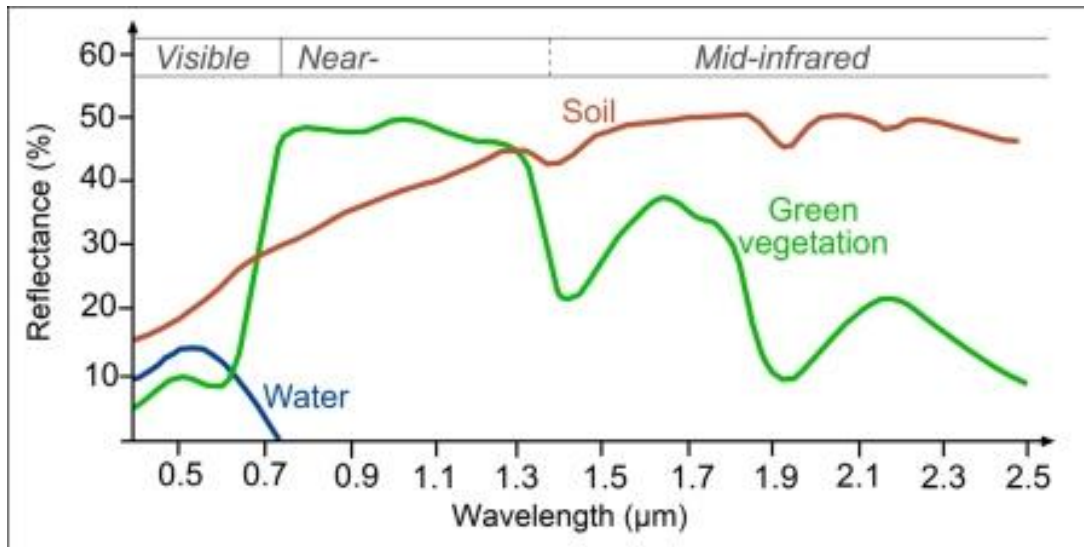


Figure 3 Signatures spectrales du sol, de la végétation et de l'eau (Source : <https://seos-project.eu/>).

3.4. Indices spectraux

Les indices spectraux jouent un rôle crucial en télédétection en mettant en évidence des caractéristiques spécifiques de l'image tout en minimisant les influences indésirables telles que les ombres notamment (*Indices gallery*, 2023). Ils reposent sur la réponse spectrale des caractéristiques de surface, mais leur performance est influencée par divers facteurs, notamment la résolution des données, le moment de l'acquisition, la localisation géographique et les conditions climatiques (Kaur & Pandey, 2022). Le principe fondamental derrière les indices spectraux consiste à manipuler mathématiquement différentes bandes spectrales pour isoler des caractéristiques particulières de l'image (*Indices gallery*, 2023; Kaur & Pandey, 2022).

Le choix de l'indice approprié se fait en fonction de l'objectif de l'analyse et du phénomène que l'on souhaite analyser. Chaque indice est spécialement conçu pour mettre en évidence des caractéristiques spécifiques telles que la végétation, l'eau, la chaleur, etc. Il est donc essentiel de choisir celui qui correspond le mieux à l'objectif de l'étude. De plus, il faut tenir compte des variations dues aux ombres causées par la topographie ou les nuages. Pour pallier cela, les indices sont conçus pour se concentrer davantage sur la couleur des objets plutôt que sur leur intensité ou leur luminosité (*Indices gallery*, 2023).

Parmi les indices les plus couramment utilisés, on peut citer le *Normalized Difference Vegetation Index* (NDVI) pour la végétation, le *Normalized Difference Water Index* (NDWI) pour l'eau, et le *Normalized Difference Built-up Index* (NDBI) pour les zones bâties. Les indices étant liés à des bandes spécifiques du spectre électromagnétique, ils ne sont valables que pour certains capteurs ou certaines catégories de capteurs (*Indices gallery*, 2023).

Malgré leur utilité, les indices spectraux présentent certaines limites, notamment leur sensibilité aux effets atmosphériques, aux variations de la réflectance spectrale des éléments présents à la surface terrestre et aux conditions de luminosité (Somvanshi & Kumari, 2020; Xue & Su, 2017).

3.5. Classification

La classification des images issues de données de télédétection revêt une importance capitale dans divers domaines d'application. La classification consiste à attribuer des pixels à des classes ou catégories

similaires, regroupant ainsi les pixels en classes homogènes en fonction de leurs caractéristiques spectrales (Jawak et al., 2015). De nombreux facteurs entrent en jeu lors d'une classification. La sélection adéquate des données de télédétection, le prétraitement des images, les stratégies de classification et la complexité du paysage sont autant de paramètres pouvant considérablement influencer les résultats (Jawak et al., 2015; Kulkarni & Lowe, 2016).

Pour garantir une classification précise, plusieurs étapes sont nécessaires, dont le choix d'un système et de méthodes de classification appropriés, la sélection des échantillons d'entraînement et l'évaluation de la précision notamment (Jawak et al., 2015).

3.5.1. Classification par pixel et par objet

Deux approches prédominent pour la classification : basée sur les pixels (PBC) et orientée objet (OOC). Le choix de l'une ou de l'autre approche dépend du contexte d'application et des jeux de données utilisés (Jawak et al., 2015; Phiri et al., 2020).

La PBC est une approche traditionnelle qui repose sur des méthodes statistiques. Elle permet d'assigner directement les informations spectrales des pixels à des classes thématiques. Cette méthode présente cependant certaines limites, notamment l'effet poivre-sel sur l'image classifiée ou la difficulté à distinguer des objets différents ayant des caractéristiques spectrales analogue (Jawak et al., 2015; Phiri et al., 2020).

L'approche OOC considère à la fois les caractéristiques spectrales et spatiales des données satellitaires. Elle se base sur la classification floue et la segmentation pour minimiser la variation intra-objet et ainsi améliorer la précision de la classification. L'utilisation des informations basées sur les objets permet de surmonter les limitations de la PBC, notamment l'effet poivre-sel mentionné précédemment. Toutefois, l'OOC peut être sensible à des erreurs de segmentation (sur- ou sous-segmentation), ce qui peut réduire la précision de la classification (Jawak et al., 2015).

3.5.2. Classification non-supervisée et supervisée

La classification des données multispectrales par apprentissage automatique³ est cruciale en télédétection et se décline en deux méthodes principales d'apprentissage : la classification supervisée et la classification non-supervisée (Talukdar et al., 2020).

La classification non-supervisée identifie des groupes ou structures au sein des données sans utiliser de données d'apprentissage préalables. Les valeurs des pixels de réflectance sont regroupées en classes distinctes, le nombre de classes souhaitées ayant été spécifié préalablement. Les classes obtenues en sortie de la classification sont étiquetées par l'utilisateur sur la base de données de référence (Jawak et al., 2015).

La classification supervisée repose sur l'utilisation « d'échantillons d'entraînement dont l'identité est connue pour classer les pixels de l'image dont l'identité est inconnue » (Jawak et al., 2015). Les échantillons d'entraînement servent à définir les règles de classification pour chaque classe, permettant ainsi d'affecter chaque pixel à la classe qui lui correspond le mieux. Cette méthode de classification comporte trois étapes : la formation manuelle des données d'entraînement, la classification à proprement parler et l'évaluation de la précision (Jawak et al., 2015). En général, les classifications

³ L'apprentissage automatique se révèle efficace dans la classification de données de télédétection, modélisant des signatures complexes avec une approche non paramétrique (Maxwell et al., 2018).

supervisées sont considérées comme étant plus précises que les non-supervisées, car elles offrent un meilleur contrôle du processus de classification et sont donc souvent privilégiées (Allenbach et al., 2023).

3.5.3. Méthodes de classification : les classificateurs

Les méthodes de classification dans le domaine de la télédétection offrent diverses options pour l'identification de classes. Ces méthodes comprennent des méthodes non-supervisées, telles que l'analyse itérative des données auto-organisées (ISODATA), ainsi que des méthodes supervisées paramétriques et non-paramétriques, comme le maximum de vraisemblance (ML), les forêts d'arbres décisionnels (RF) ou les machines vectorielles de support (SVM). Parmi ces méthodes, les SVM et les RF se démarquent généralement par une meilleure précision et sont alors souvent considérées comme les plus performants pour la classification (Maxwell et al., 2018; Phiri et al., 2020; Talukdar et al., 2020; Vali et al., 2020). Le choix du classificateur approprié pour une tâche spécifique peut s'avérer complexe en raison de la diversité des méthodes disponibles et des informations parfois divergentes mentionnées dans la littérature. Il dépend de nombreux facteurs, spécifiques à chaque cas, tels que les classes cartographiées, les données d'apprentissage et les variables prédictives. Par conséquent, il est généralement recommandé d'expérimenter plusieurs classificateurs pour déterminer celui qui convient le mieux au contexte (Maxwell et al., 2018; Phiri et al., 2020). Il faut souligner que les performances des classificateurs peuvent être affectées en fonction de plusieurs facteurs, notamment le nombre d'échantillons d'apprentissage, le nombre de classes défini, la nature du terrain, les méthodes de prétraitement des images, les exigences en termes de calcul et les vitesses d'apprentissage (Maxwell et al., 2018; Phiri et al., 2020).

Les classificateurs paramétriques tels que le ML se fondent sur des calculs de probabilité et des hypothèses de distribution des données. Ces informations sont utilisées pour estimer les probabilités d'appartenance des pixels à chaque classe, puis les pixels sont assignés à la classe qui présente la probabilité la plus élevée. Ces méthodes exigent une quantité suffisante de données d'apprentissage pour estimer correctement les paramètres de distribution, ce qui peut les rendre moins appropriés pour traiter des ensembles de données volumineux (*Classify*, 2023; Kulkarni & Lowe, 2016; Wulder et al., 2018).

Les méthodes SVM déterminent un hyperplan optimal pour séparer les données d'entraînement en un nombre prédéfini de classes tout en minimisant les taux de classification incorrecte. Ils se démarquent par leur aptitude à traiter efficacement des ensembles de données d'apprentissage de petite taille tout en maintenant des performances de classification élevées (Kulkarni & Lowe, 2016; Maxwell et al., 2018; Talukdar et al., 2020; Vali et al., 2020). Les SVM peuvent traiter en entrée des rasters segmentés ou standards et présentent une moindre sensibilité au bruit, aux bandes corrélées ainsi qu'à un nombre ou une taille inégale des échantillons d'entraînement au sein de chaque classe (*Classify*, 2023). L'utilisation des SVM implique un temps de traitement et calcul élevé, surtout lorsque le nombre de classes définies augmente (Cortes & Vapnik, 1995; Maxwell et al., 2018).

Les RF sont une méthode efficace pour gérer des images multibandes de diverses profondeurs de bits (*Classify*, 2023). Ils se basent sur des classificateurs arborescents où chaque arbre effectue une classification en émettant un vote individuel. La classe finale attribuée à chaque inconnue est déterminée par le vote majoritaire des arbres (*Classify*, 2023; Gislason et al., 2006; Kulkarni & Lowe, 2016; Maxwell et al., 2018; Talukdar et al., 2020; Wulder et al., 2018). Les RF sont capables de traiter une grande diversité de données et se révèlent être résistantes au bruit et au surapprentissage, tout en

restant relativement légères en termes de calcul (Gislason et al., 2006; Kulkarni & Lowe, 2016; Pelletier et al., 2016; Talukdar et al., 2020; Vali et al., 2020). Les deux paramètres à définir sont le nombre d'arbres et le nombre de caractéristiques dans chaque division (Maxwell et al., 2018; Talukdar et al., 2020). Il est important de noter que le nombre d'arbres dans la forêt a généralement peu d'impact sur la précision de la classification, à condition qu'il soit suffisamment élevé (Maxwell et al., 2018; Vali et al., 2020). Il est alors probable que le nombre d'arbres nécessaire pour garantir une classification précise soit spécifique à chaque cas, ce nombre augmentant avec la complexité des variables et la taille de l'ensemble de données (*Classify*, 2023; *How Forest-based Classification and Regression works*, 2023; Maxwell et al., 2018). En comparant les performances de différents classificateurs couramment employés, il est fréquent de constater que les RF surpassent généralement les autres méthodes en matière de précision et d'efficacité. Ceci s'explique par le fait qu'elles sont moins influencées par la configuration de leurs paramètres, offrant ainsi un bon compromis entre les performances de classification et le temps de calcul (Gislason et al., 2006; Maxwell et al., 2018; Pelletier et al., 2016; Talukdar et al., 2020). Toutefois, cette méthode peut présenter des limites lorsqu'elle est confrontée à des ensembles de données d'apprentissage déséquilibrés, en particulier pour les classes minoritaires (Gislason et al., 2006; Pelletier et al., 2016).

3.5.4. Échantillons d'entraînement

Les méthodes de classification supervisée reposent sur des échantillons d'entraînement pour développer des modèles de classification (Maxwell et al., 2018; Wulder et al., 2018). L'impact significatif de la taille de l'échantillon d'entraînement sur la précision de la classification, dépassant même l'influence de l'algorithme utilisé, a été mis en évidence dans une étude menée par Huang, Davis et Townshend (2002). Ainsi, la taille et la qualité de l'échantillon d'entraînement revêtent une importance majeure dans l'obtention de résultats satisfaisants dans le processus de classification (Maxwell et al., 2018; Pelletier et al., 2016; Santos et al., 2021; Simoes et al., 2021).

Il n'existe cependant pas de règles universelles concernant le nombre minimum d'échantillons requis pour les classifications par apprentissage automatique. Ce nombre peut considérablement varier en fonction de plusieurs facteurs, notamment l'algorithme de classification, le nombre de variables d'entrée et les caractéristiques spatiales de la région étudiée. Néanmoins, il est généralement reconnu qu'un grand nombre d'échantillons de haute qualité, couvrant au mieux la variabilité de chaque classe, améliore la précision de la classification (Huang et al., 2002; Kulkarni & Lowe, 2016; Maxwell et al., 2018; Santos et al., 2021; Vali et al., 2020). La collecte d'un grand nombre d'échantillons de formation répondant à ces critères peut présenter des défis en termes de temps, d'accès ou de disponibilité (Maxwell et al., 2018; Vali et al., 2020; Wulder et al., 2018). Un autre aspect important à prendre en considération est le déséquilibre des classes dans les données, susceptible de conduire à une sous-estimation des classes les moins fréquentes. Les algorithmes de classification, visant généralement à minimiser le taux d'erreur globale, ont tendance à privilégier les classes majoritaires (Maxwell et al., 2018). Puisque la qualité des échantillons d'entraînement revêt une importance capitale, il est recommandé d'envisager la mise en place de méthodes de contrôle de la qualité (Santos et al., 2021).

3.5.5. Évaluation de la précision

En sortie de la classification, deux principales sources d'erreurs peuvent être identifiées : le bruit de caractéristique, causé par des perturbations telles que les nuages par exemple, et le bruit de classe, résultant d'erreurs d'étiquetage et de regroupements incorrects dans les classes (Santos et al., 2021). Afin de garantir la fiabilité des résultats obtenus grâce à la classification, il est essentiel de procéder à

des évaluations de la précision, celle-ci représentant le degré de concordance entre la carte produite et de référence (Olofsson et al., 2013).

Pour évaluer la précision, plusieurs indicateurs de la qualité des classifications peuvent être calculés à l'aide d'une matrice de confusion : la précision globale, la précision de l'utilisateur, la précision du producteur et le coefficient d'accord Kappa. La matrice de confusion dans le cadre de la classification permet alors d'évaluer l'ampleur des erreurs de classification (Tableau 1) (Olofsson et al., 2013).

La précision globale représente la proportion de la zone cartographiée correctement et se calcule en divisant le nombre total de pixels correctement classés, selon les résultats obtenus dans la matrice de confusion (diagonale), par le nombre total de pixel (Allenbach et al., 2023; Olofsson et al., 2013). La précision de l'utilisateur évalue la probabilité que la zone cartographiée dans une certaine classe appartienne effectivement à celle-ci (Olofsson et al., 2013). Elle est calculée en divisant, pour une ligne donnée, le nombre de pixels correctement classés par le nombre total de pixels. Elle s'apparente à l'erreur de commission (inclusion) qui indique la proportion de pixels affectés à une classe autre que celle à laquelle ils appartiennent (Allenbach et al., 2023). La précision du producteur mesure la probabilité que les pixels correspondant à une certaine classe aient bel et bien été classés dans cette classe (Olofsson et al., 2013). Elle se calcule de la même manière que la précision de l'utilisateur, mais en tenant compte des colonnes de la matrice de confusion. Elle s'apparente à l'erreur d'omission (exclusion) qui indique les pixels qui, pour une certaine classe, n'ont pas été convenablement classés (Allenbach et al., 2023). Le coefficient d'accord Kappa constitue une évaluation globale de la précision de la classification. Il quantifie le degré de concordance entre les classes prédites et les classes réelles, en éliminant la concordance due au hasard. En d'autres termes, il permet de déterminer si la classification obtenue est plus précise que si les classes avaient été attribuées de manière aléatoire (Allenbach et al., 2023; Olofsson et al., 2013).

Ces indicateurs sont calculés à partir d'échantillons, ce qui peut introduire des incertitudes. Elles peuvent être représentées par un intervalle de confiance (Olofsson et al., 2013). Il est essentiel que les évaluations de précision reposent sur des données de validation indépendantes qui n'ont pas été utilisées lors de la création du modèle. En fonction des résultats de précision obtenus, il est possible d'ajuster les données d'entraînement ou l'approche de classification pour obtenir de meilleurs résultats (Wulder et al., 2018).

Tableau 1 Exemple de matrice de confusion présentant des valeurs fictives (Source : <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/compute-confusion-matrix.htm>).

		Données de référence				Précision de l'utilisateur	Kappa
		c_1	c_2	c_3	Total		
Données classifiées	c_1	49	4	4	57	0.8594	0
	c_2	2	40	2	44	0.9091	0
	c_3	3	3	59	65	0.9077	0
	Total	54	47	65	166	0	0
Précision du producteur		0.9074	0.8511	0.9077	0	0.8916	0
Kappa		0	0	0	0	0	0.8357

4. Méthodologie

Certaines étapes de la méthodologie ont été exécutées en dehors du délai initialement prévu pour le stage. Cette situation découle principalement de la durée relativement courte du stage qui s'est révélée insuffisante pour atteindre un niveau d'achèvement optimal tout en explorant, assimilant et exploitant l'éventail d'outils et de fonctionnalités offerts par ArcGIS Pro®. Plus précisément, l'intégration des données provenant des relevés de terrain dans le processus de classification s'est déroulée après la période officielle du stage. Cette distinction n'est pas clairement spécifiée dans la suite de la méthodologie mais l'est dans les résultats et la discussion.

La décision de poursuivre la cartographie au-delà de la fin officielle du stage s'est imposée comme une nécessité, car cette carte joue un rôle essentiel pour les étapes à venir du mémoire de Master, en particulier le travail sur le terrain programmé pour la mi-octobre. Ce travail sur le terrain revêt une importance cruciale dans la validation de la cartographie réalisée dans le présent travail.

4.1. Site d'étude

Le massif montagneux forestier de Sorata, situé dans le nord-est du Domaine du Centre, s'étend sur 19 000 ha et atteint une altitude d'environ 1850 m (Favre, 2013; Scherz et al., 2015). La prise en compte des facteurs pédoclimatiques, variant en fonction de l'altitude et de la topographie est essentielle pour la compréhension de la répartition des communautés végétales dans cette région (Gautier & Goodman, 2002; Nusbaumer, 2011; Ramandimbimanana, 2009; Randimbiarison et al., 2014; Ranirison, 2010; Ranirison et al., 2010; Razanajatovo, 2009). La topographie particulière de Madagascar, en particulier dans la partie nord du Domaine du Centre, joue un rôle majeur dans la répartition et mise en place d'une diversité importante de microclimats en raison notamment de son impact significatif sur les vents dominants (Alizés et Mousson), les précipitations et la température (Favre, 2013; Rakotoarison et al., 2020; Rakotovao et al., 2006; Randimbiarison et al., 2014; Ravet, 1952; Razanajatovo, 2009; Trigui, 2010). Les Alizés en particulier apportent des précipitations abondantes sur les régions montagneuses du Domaine du Centre, mais se dessèchent en descendant vers la côte ouest après avoir traversé les hauts plateaux centraux (Brenon, 1972; Cornet, 1974; Ramandimbimanana, 2009; Randimbiarison et al., 2014; Ranirison, 2010; Ravet, 1952; Razanajatovo, 2009; Trigui, 2010).

Soumis à un climat subhumide, le massif de Sorata reçoit en moyenne 1863 mm de précipitations annuelles (entre 1981 et 2017), principalement pendant la saison des pluies (novembre à avril). Les températures journalières varient de 13°C à 21.9°C en moyenne, avec des extrêmes allant de 6.7°C en saison froide (juin-août) à plus de 28.6°C en saison chaude (décembre-février) (Goodman et al., 2018). Les propriétés géologiques et pédologiques de l'île présentent une grande diversité, spécifique à chaque région (Rakotovao et al., 2006; Ramandimbimanana, 2009; Trigui, 2010). Le massif de Sorata repose sur des formations magmatiques et des sols acides de type acrisols et ferralsols (Besairie, 1964; Brenon, 1972; Favre, 2013; Goodman et al., 2018). Cette combinaison unique de conditions climatiques, géologiques et pédologiques confère à ce massif un intérêt écologique et floristique particulier, justifiant une exploration plus approfondie de cette région (Favre, 2013).

Au niveau de la végétation, le massif forestier de Sorata est décrit comme étant recouvert de végétation primaire dense et humide, avec une forêt humide dégradée en périphérie (Figure 4) (Moat & Smith, 2007). Selon un rapport du *Missouri Botanical Garden* (MBG), il est plus précisément composé de trois types de forêts : de moyenne altitude, de haute altitude et de transition pluristrates (2008 in Favre, 2013).

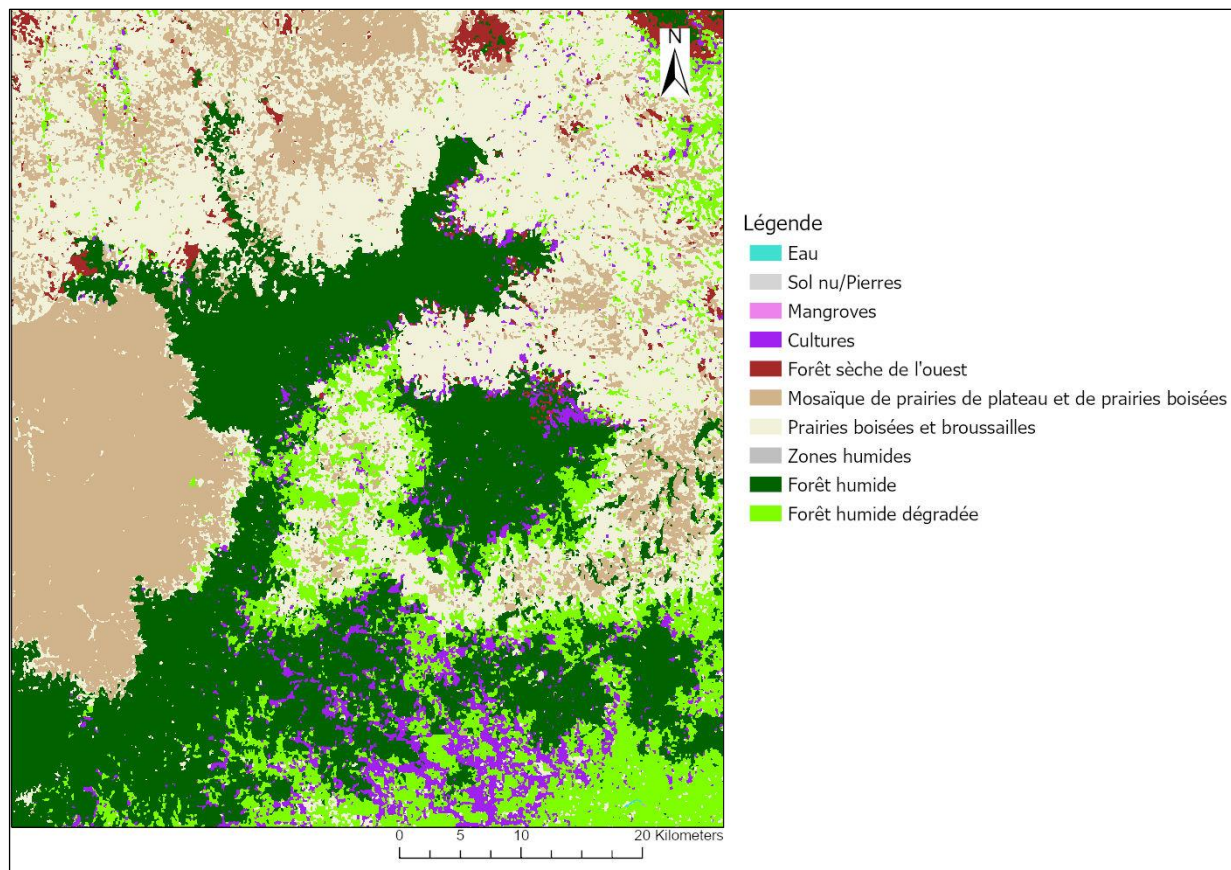


Figure 4 Végétation du massif forestier de Sorata cartographiée dans le cadre de l'Atlas de la végétation de Madagascar (UTM projection) (Moat & Smith, 2007).

4.2. Données utilisées et acquisition

Le présent travail a été effectué avec le logiciel ArcGIS Pro® et s'est basé sur des images satellites Sentinel-2. Elles ont été téléchargées via EO Browser depuis le site de Sentinel Hub⁴. EO Browser offre des données diverses provenant de différents satellites ainsi que de nombreuses fonctionnalités pour des images satellites dans la région, la période et la couverture nuageuse voulues. Les données permettent des analyses à long terme et sont disponibles gratuitement (EO Browser, 2023). Les images de Sentinel-2 L2A (corrigé au niveau de l'atmosphère) ont été retenues avec comme critère une couverture nuageuse maximum sur l'ensemble de la tuile de 20%. Le filtre temporel a été fixé de janvier 2020 à avril 2023. La carte finale devant représenter l'état actuel du massif forestier du Sorata, il était nécessaire de considérer les dates les plus récentes possibles tout en ayant suffisamment d'images satellites de bonne qualité pour représenter différentes périodes de l'année. Pour couvrir entièrement la zone d'étude, les tuiles correspondantes aux emplacements 39LUE et 39LUF du système de référence de grille militaire (MGRS) ont dues être considérées. Le téléchargement des images retenues (Tableau 2) se fait via un lien SciHub accessible gratuitement, le second lien étant récemment devenu payant. Les dossiers téléchargés au format ZIP, en particulier celui intitulé GRANULE, contient l'ensemble des bandes des images satellites (Tableau 3). Les bandes satellites sont également fournies avec une carte de l'épaisseur optique des aérosols (*Aerosol Optical Thickness (AOT)*), une carte de la pression de la vapeur d'eau (*Water Vapour Pressure (WVP)*) ainsi qu'une carte de la classification des scènes (*Scene Classification (SCL)*). Ces différentes données sont rassemblées dans trois dossiers distincts, en fonction

⁴ Lien d'accès : <https://www.sentinel-hub.com/explore/eobrowser/>.

de la résolution spatiale des bandes (10, 20 ou 60 mètres) (*User Guides - Sentinel-2 MSI Level-2 Processing*, 2022).

Tableau 2 Informations relatives aux images satellites téléchargées (EO Browser, 2023).

Nom de la source de données	Date de détection (jj.mm.aaaa)	Emplacement du MGRS	Heure de détection (UTC)	Couverture nuageuse (%)
Sentinel-2 L2A	17.10.2020	39LUE/39LUF	07:03:35/07:03:20	1.9/5.8
Sentinel-2 L2A	25.05.2021	39LUE/39LUF	07:03:31/07:03:17	0.3/0
Sentinel-2 L2A	03.08.2021	39LUE/39LUF	07:03:33/07:03:18	2.1/0.1
Sentinel-2 L2A	08.08.2021	39LUE/39LUF	07:03:34/07:03:19	0.5/10.7
Sentinel-2 L2A	15.05.2022	39LUE/39LUF	07:03:36/07:03:21	2.4/0
Sentinel-2 L2A	31.12.2022	39LUE/39LUF	07:03:30/07:03:15	2/1.2
Sentinel-2 L2A	14.02.2023	39LUE/39LUF	07:03:30/07:03:15	9.4/0.2
Sentinel-2 L2A	06.03.2023	39LUE/39LUF	07:03:34/07:03:19	1.3/19.5

La quasi-totalité des prétraitements et classifications ont été réalisés sur les images du 17 octobre 2020 en raison de sa qualité et de son absence de couverture nuageuse sur le massif d'intérêt par rapport aux autres dates retenues.

Tableau 3 Informations relatives aux bandes spectrales disponibles avec les satellites Sentinel-2 (modifié d'après Taconet, 2019). La bande de cirrus 10 n'est pas fournie, car elle ne contient aucune information de surface (*User Guides - Sentinel-2 MSI Level-2 Processing*, 2022).

Description	Bande	Longueur d'onde centrale (µm)	Résolution maximale (m)
Aérosol côtier	1	0.443	60
Bleu	2	0.490	10
Vert	3	0.560	10
Rouge	4	0.665	10
Vegetation Red Edge (VRE)*	5	0.705	20
VRE	6	0.740	20
VRE	7	0.783	20
Proche Infrarouge (PIR)	8	0.842	10
Proche Infrarouge (PIR)	8A	0.865	20
Vapeur d'eau	9	0.945	60
Infrarouge Court (SWIR) 1	11	1.610	20
Infrarouge Court (SWIR) 2	12	2.190	20

*VRE = pic de réflectance de la chlorophylle

En plus des images satellites, d'autres données ont été intégrées dans le processus de classification, principalement pour faciliter la création des données d'entraînement. Les données utilisées pour les différentes classes sont précisées dans le tableau en annexe 1. Il s'agit notamment de données d'altitude (courbes de niveau à intervalles de 100 mètres) et de courbure⁵ (résolution de 90 mètres).

⁵ La courbure est une mesure de l'ensoleillement des pixels. Une valeur élevée indique que le pixel est fortement exposé au soleil, tandis qu'une valeur faible signifie qu'il reçoit moins de lumière. Cette information est particulièrement utile pour détecter les crêtes et dépressions.

Ces informations étaient déjà disponibles dans un projet et une base de données globaux pour Madagascar réalisés au sein des CJBG. Le fond de carte *World Imagery* fourni par ArcGIS Pro® a également été utilisé en raison de sa résolution exceptionnellement élevée (1 mètre ou plus) et de sa qualité d'image. L'utilisation de ce fond de carte a permis d'apporter davantage de précision, en particulier en cas d'incertitude concernant une classe particulière. Des informations supplémentaires ont été obtenues à partir des données de récoltes⁶ et de relevés linéaires⁷ effectués lors des missions de terrain évoquées au chapitre 2. L'ensemble de ces données a été exploité pour générer deux cartes : l'une purement « visuelle », basée sur les images satellites et données d'altitude, et l'autre intégrant également les données collectées sur le terrain.

4.3. Prétraitements des images

Les images satellites brutes doivent subir certains traitements avant d'être utilisables pour la classification. Pour la plupart de ces prétraitements, il est indispensable de vérifier que les données initiales n'ont pas été modifiées. Cela peut se faire en comparant manuellement la valeur des pixels entre la couche initiale et celle obtenue ou en allant observer les informations raster dans les propriétés des couches.

Les outils utilisés sur ArcGIS Pro®, leurs informations relatives et les liens menant aux aides correspondantes sont regroupés en annexe 2. Les versions d'ArcGIS Pro® utilisées dans le présent travail sont les versions 2.4.0 et 2.5.2⁸. Les outils disponibles peuvent varier en fonction des versions. Il est donc recommandé de se référer aux aides correspondantes aux versions utilisées.

L'ensemble du projet a été réalisé dans les systèmes de coordonnées géographiques et projetées GCS WGS 1984 et WGS 1984 UTM Zone 39S (EPSG 32739). Les systèmes de coordonnées retenus correspondent à ceux couramment utilisés dans les projets en cours ou achevés à Madagascar. Si nécessaire, ces systèmes de coordonnées ont été définis ou modifiés à l'aide des outils *Define Projection* ou *Project/Project Raster*.

4.3.1. Découpage de l'image en fonction du site d'étude

Afin de ne considérer qu'une partie de la zone couverte par les images satellites, une zone d'étude a été délimitée. Il a été décidé de considérer une zone relativement large qui inclut l'ensemble des massifs forestiers de Sorata et d'Andravory (Figure 5). Une entité délimitant la zone retenue a été créée et a servi de base de référence pour les différentes couches à découper. La création d'une nouvelle classe d'entités (*Feature Class*) dans la géodatabase constitue la première étape de la démarche. L'onglet *Edit* d'ArcGIS Pro® permet ensuite de créer et modifier des entités. Une entité rectangulaire a été dessinée tout d'abord manuellement de sorte que les angles, appelés vertex, soient calés sur l'intersection des pixels aux différentes résolutions disponibles (10, 20 et 60 mètres). Les coordonnées des angles à l'intersection de ces pixels ont ensuite été relevées (Tableau 4) et insérées dans l'outil *Edit Vertices* afin d'assurer un calage parfait de l'entité délimitée sur les pixels de l'image. De cette manière, le redécoupage et recalcul des pixels de l'image sont évités. Pour évaluer la justesse de l'entité créée, il est nécessaire de vérifier que les coordonnées en x et y vont de pair, c'est-à-dire que l'on retrouve de chaque côté du rectangle délimité deux fois la même valeur pour x et pour y. En outre, la différence

⁶ Ces données comprennent les relevés réalisés selon la méthode Braun-Blanquet (coefficient d'abondance-dominance) ainsi que des informations essentielles sur l'aspect du milieu.

⁷ Les emplacements des relevés ont été choisis en raison du caractère homogène et représentatif du milieu étudié.

⁸ Les versions d'ArcGIS Pro® utilisées étaient paramétrées en anglais. Par conséquent, les noms des fonctions, outils ou autres éléments sont en anglais dans ce travail.

entre les coordonnées de mêmes points cardinaux divisée par la résolution maximale des pixels, ici 60 mètres, doit avoir pour résultat un chiffre rond.

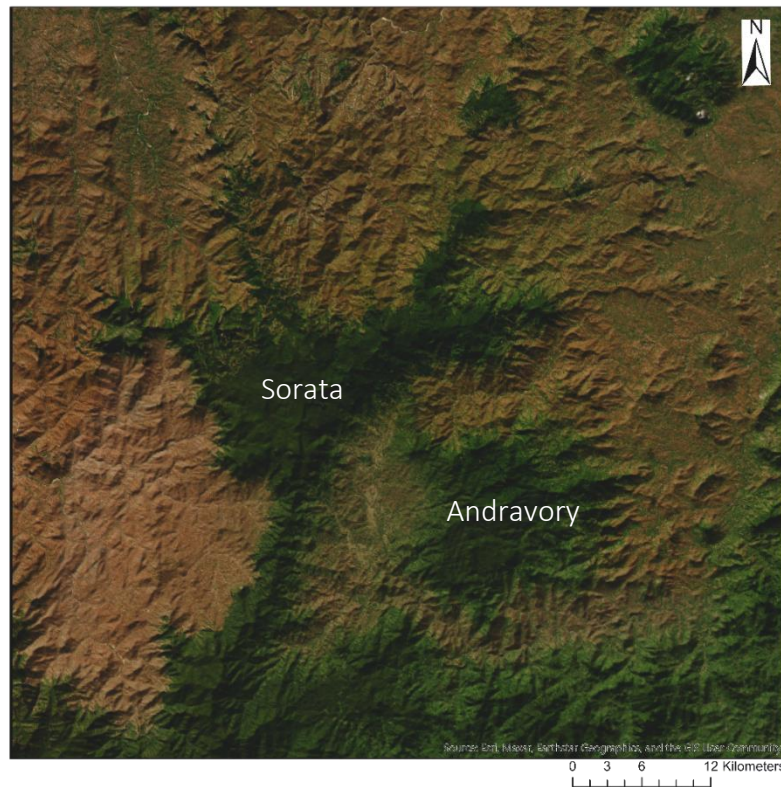


Figure 5 Massif forestier de Sorata et d'Andravory, fond de carte World Imagery.

Les bandes ont été découpées sans traitements préalables. Le nombre élevé de bandes à découper a motivé une automatisation du processus. La création d'un modèle à l'aide de *ModelBuilder*⁹ sur ArcGIS Pro® s'est faite en deux parties. Pour chacune, des dossiers d'entrée, contenant l'ensemble des images à traiter, et de sortie, permettant de récupérer les données traitées, ont été créés préalablement. Cette démarche vise à simplifier l'utilisation du *ModelBuilder* et à maintenir une structure organisée compte tenu des nombreuses sorties générées. Ces dossiers ont été définis comme *Workspace* dans les modèles. La notation *%name%* comme nom de fichier en sortie des outils est fortement conseillée, car elle permet de conserver les noms des fichiers d'entrée.

Tableau 4 Coordonnées des angles (vertex) relevés pour l'entité de découpage.

Coordonnées Est (X) (m)	Coordonnées Nord (Y) (m)
302'520	8'526'400
369'780	8'526'400
369'780	8'452'300
302'520	8'452'300

La conversion des images satellites du format JP2 au format TIF a constitué une première étape indispensable pour l'obtention de données utilisables avec l'outil de découpage. Un premier *ModelBuilder* a ainsi été réalisé avec l'outil *Raster To Other Format* et l'itérateur *Iterate Rasters*, qui

⁹ *ModelBuilder* est un langage de programmation qui facilite la création de flux de travail pour le géotraitement (*What is ModelBuilder ?*, 2023).

permet d'itérer l'outil sur chaque bande individuellement (annexe 3). La deuxième étape a été le découpage des images satellites avec l'outil *Extract By Mask*. Cet outil a été privilégié par rapport à l'outil *Clip Rasters* car, après plusieurs tentatives, il s'est avéré que celui-ci produisait des images en sortie dont l'étendue dépassait largement de la zone définie pour ce travail. La création d'un second *ModelBuilder* a été nécessaire étant donné qu'un modèle ne supporte qu'un seul itérateur (annexe 4). Le *Workspace* d'entrée de l'itérateur *Iterate Rasters* pour cette partie a été le dossier contenant les images générées à la première étape.

Une autre entité de découpe, nettement plus restreinte, a également été créée dans le but d'optimiser l'efficacité des procédures et outils utilisés. Après avoir validé le succès des démarches effectuées sur la zone restreinte, elles peuvent être appliquées sur la zone d'étude complète.

4.3.2. Création de mosaïques

Comme mentionné au chapitre 4.2, des tuiles correspondantes à deux emplacements MGRS différents ont été nécessaires afin de couvrir la totalité de la zone d'étude. Par conséquent, pour avoir une image unique, homogène et complète de la zone pour chaque bande, les images découpées ont subi un traitement de type mosaïque. Dans l'onglet *Analysis* d'ArcGIS Pro®, on peut retrouver l'ensemble des fonctions applicables aux rasters (*Raster Functions*), dont l'outil *Mosaic Rasters* utilisé ici. Les deux images d'une même bande aux emplacements MGRS différents doivent être sélectionnées. Lorsque les images se chevauchent, différentes méthodes de priorisation d'affichage sont proposées (Tableau 5) (*Mosaic Rasters function*, 2023). La méthode *Blend* a été choisie pour attribuer un poids relatif aux pixels qui se chevauchent, évitant ainsi d'en privilégier certains en fonction de l'ordre de sélection des images.

Tableau 5 Méthodes de priorisation d'affichage des images disponibles avec l'outil *Mosaic Rasters* (Source : <https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/mosaic-rasters.htm>).

Opération	Description
<i>First</i>	Affiche les pixels de la première image sélectionnée
<i>Last</i>	Affiche les pixels de la dernière image sélectionnée
<i>Min</i>	Affiche le pixel de valeur la plus faible de toutes les couches superposées
<i>Max</i>	Affiche le pixel de valeur la plus élevée de toutes les couches superposées
<i>Mean</i>	Calcule et affiche une moyenne des pixels
<i>Blend (Default value)</i>	Calcule et affiche une moyenne des pixels en donnant un poids supérieur aux pixels qui sont plus proches des images voisines

La création de mosaïques devant se faire manuellement, seules les bandes utilisées dans ce travail ont subi ce traitement. Les mosaïques ont finalement été exportées avec l'option *Export Raster* disponible dans les données (*Data*) de la couche créée. Pour préserver le résultat obtenu tel quel, il est crucial d'exporter la mosaïque directement dans la géodatabase.

4.3.3. Création d'un masque de nuages

Étant donné la complexité à obtenir des images satellites complètement exemptées de couverture nuageuse, la création d'un masque de nuages spécifique aux images satellites retenues pour les analyses s'est révélée indispensable.

Le raster SCL à 20 mètres pour les deux emplacements MGRS a été utilisé. Il est inclus dans le fichier MTD¹⁰ des images satellites téléchargées aux différentes dates et comprend une classification dans

¹⁰ Le fichier MTD contient des images multibandes.

laquelle neuf classes sont concernées par les nuages, les ombres, la neige, les pixels saturés, etc., et trois par la végétation, les sols nus et l'eau (*User Guides - Sentinel-2 MSI Level-2 Processing*, 2022). Sur la base de cette classification, un code d'extraction des nuages a été mis en place avec l'outil *Raster Calculator* :

$$\text{Con}(\sim("SCL-20m_MTD_MSIL2A" == 4) \& \sim("SCL-20m_MTD_MSIL2A" == 5) \& \sim("SCL-20m_MTD_MSIL2A" == 6), 1, 0)$$

Ce code permet l'obtention d'un raster binaire attribuant à chaque pixel considéré comme étant de la végétation, des sols nus ou de l'eau une valeur de 0, et à tous les autres une valeur de 1. Certains éléments peuvent être classés dans la mauvaise catégorie et il est alors nécessaire de pouvoir manipuler le raster obtenu en le transformant en couche d'entités vectorielles avec l'outil *Raster To Polygon*. Après cette étape seulement, la couche d'entités polygonales peut être découpée selon la délimitation définie au chapitre 4.3.1. Effectivement, si le découpage est effectué avant, la classification initiale du fichier SCL est perdue.

Pour modifier les valeurs des entités si petites qu'elles peuvent être considérées comme du bruit dans l'image, une sélection des entités ayant une aire inférieure ou égale à 35 000 m² a d'abord été effectuée avec l'outil *Select By Attributes*. Cette surface a été définie arbitrairement, en fonction des objets à conserver ou à supprimer. La valeur des entités sélectionnées peut ainsi être modifiée avec l'option *Calculate Field* accessible via la colonne gridcode de la table attributaire. Dans le cas présent, l'expression a simplement été « gridcode = 0 ». Les entités ayant une surface supérieure à 35 000 m² et classées dans la mauvaise catégorie ont été sélectionnées manuellement et leur valeur a été modifiée avec la même démarche que précédemment. Dans certains cas particuliers, lorsque deux objets ne devant pas être affiliés à la même catégorie ne formaient qu'une seule entité, l'outil *Split* a été utilisé.

Pour obtenir une couche ne contenant que les éléments considérés comme « indésirables », les polygones correspondant à ces entités (gridcode = 1) ont été sélectionnés avec *Select By Attributes*, puis exportés avec l'option *Export Features* disponible dans les données (*Data*) de la couche. L'outil *Merge-1* a ensuite été utilisé pour regrouper les différentes entités polygonales exportées en une seule et même entité. La démarche présentée jusqu'ici a été effectuée séparément pour les images des deux emplacements MGRS. L'outil *Merge-2* a permis de combiner les deux couches vectorielles polygonales en une nouvelle couche d'entités. L'outil *Merge-1* a encore été utilisé pour n'avoir finalement qu'une seule entité polygonale servant de masque.

Afin d'appliquer le masque, l'outil *Pixel Editor* a été utilisé avec l'opération *Set No Data*. Il est indispensable de sélectionner au préalable la couche raster que l'on souhaite modifier et, ensuite, la région (*Features To Region*) que l'on souhaite supprimer de l'image. Il est crucial d'appliquer le masque à ce stade car il ne fonctionne pas sur des images ayant subi une transformation, indice spectral ou composite notamment. Les démarches présentées dans les deux prochains chapitres ont été réalisées à partir de ces bandes sans nuages.

4.3.4. Création d'indices spectraux

Afin d'accentuer les distinctions entre les différentes surfaces composant la zone d'étude et d'ainsi discriminer l'information pour la classification, des indices spectraux ont été réalisés avec l'outil *Raster Calculator*. Le tableau en annexe 5 regroupe des informations relatives aux indices utilisés. Les différents indices n'ont été calculés que pour les images du 17 octobre 2020, excepté l'indice NDVI qui lui a été appliqué à toutes les dates afin de comparer la végétation entre les saisons sèche et humide.

4.3.5. Création de composites multibandes

La création de composites multibandes, réalisée avec l'outil *Composite Bands*, consiste à empiler plusieurs couches raster, simplifiant ainsi la manipulation simultanée de toutes les bandes qui les composent. Des composites vraies et fausses couleurs ainsi que d'analyse de la végétation ont été créés pour les différentes dates, cela dans le but d'apprécier visuellement les différences dans le paysage et de pouvoir les utiliser pour guider la classification. Pour réaliser une classification précise, une composite comprenant un maximum de bandes a également été créée permettant ainsi de faire ressortir au mieux la diversité des éléments composant le paysage. Toutes les bandes des images satellites de la date retenue, dans leur résolution la plus fine, ainsi que les indices spectraux créés au chapitre 4.3.4 ont été inclus dans cette composite multibandes (annexe 6). Il convient de souligner que pour obtenir une composite avec une résolution élevée, il est impératif de sélectionner en premier lieu dans l'outil *Composite Bands* un raster ayant la résolution la plus fine.

Dans certains cas, il peut être intéressant de tester la corrélation entre les différentes bandes à disposition afin de sélectionner les plus pertinentes. En outre, la création et manipulation d'une composite multibandes peuvent être limitées en termes de nombre de bandes supportées par la machine utilisée. Pour pallier ce genre de problème, un code en langage Python a été développé pour fournir, à partir des bandes en entrée, une matrice de corrélation au format CSV (annexe 7).

4.4. Classifications des images satellites sur ArcGIS Pro®

Les outils de classification d'images sont disponibles sous l'onglet *Imagery*. Des classifications non-supervisées et supervisées ont été réalisées sur la composite multibandes créée au chapitre 4.3.5 avec l'outil *Classify*. Les approches PBC et OOC ont toutes deux été testées avec les deux méthodes de classification. Cependant, il convient de noter que l'approche OOC a été explorée de manière expérimentale, mais n'a pas été concrètement intégrée dans la suite de ce travail. Pour pouvoir effectuer les classifications basées sur l'approche OOC, la composite multibandes a été préalablement segmentée avec l'outil *Segmentation* paramétré avec les valeurs par défaut.

4.4.1. Classifications non-supervisées

Dans ArcGIS Pro®, les classifications non-supervisées sont effectuées en utilisant le classificateur « *Iterative Self Organising (ISO) Cluster* ». Il est nécessaire de définir le nombre maximum de classes que l'on désire obtenir en sortie de la classification. Dans le cas présent, des classifications de 4, 5, 10, 15 et 20 classes ont été testées. Le nombre d'itérations maximum a été fixé à 20 afin d'éviter un processus trop long résultant d'un cycle sans fin dans lequel le seuil de convergence n'est jamais atteint (Allenbach et al., 2023). En spécifiant ces facteurs, la probabilité de sur- ou sous-estimation est réduite.

Les valeurs attribuées aux classes en sortie n'ont pas été étiquetées. La réalisation de classifications non-supervisées avait deux objectifs. Elle visait, d'une part, à évaluer les rendus possibles ainsi que les limites éventuelles de cette approche en termes de nombre de classes et, d'autre part, à évaluer visuellement ce que le classificateur identifiait comme appartenant à des classes distinctes.

4.4.2. Classifications supervisées

4.4.2.1. Données d'entraînement

Pour une classification supervisée, des zones d'entraînement sont délimitées manuellement avec l'outil *Training Samples Manager* et utilisées pour définir les signatures spectrales de la couverture du sol. Il

est indispensable que la couche à classifier, ici la composite multibandes créée au chapitre 4.3.5, soit sélectionnée pour accéder à l'outil.

La classification a été conduite en suivant une approche par étapes, avec une considération de différents niveaux. Cette démarche a permis d'analyser progressivement les résultats en sortie ainsi que d'évaluer leur cohérence, la possibilité de regrouper ou séparer des classes et la nécessité d'ajuster les données d'entraînement. Un récapitulatif des différents niveaux et classes définis est présenté en annexe 1. Pour chaque niveau, un schéma de classification a été créé en reprenant systématiquement comme base le schéma du niveau précédent.

Des données d'entraînement, contenant des informations de localisation et une classe de couverture de sol associée définie au préalable, ont été créées pour chaque classe des schémas de classification de chaque niveau (annexe 1). La création des données d'entraînement est une tâche délicate. Si les échantillons d'entraînement fournis ne représentent pas convenablement les classes de couverture du sol, la marge d'erreur peut être très importante. En outre, ils doivent inclure un nombre significatif de pixels pour procurer assez d'informations au classificateur (Allenbach et al., 2023).

La démarche adoptée pour réaliser les données d'entraînement a été conçue de manière à assurer une représentation équitable des échantillons pour chaque classe, évitant ainsi la sur- ou sous-estimation des classes. La création de données d'entraînement ponctuelles n'est pas proposée dans les versions d'ArcGIS Pro® utilisées. Elles ont donc été représentées par des cercles. Pour chaque nouveau schéma et campagne d'échantillonnage, il est conseillé d'utiliser l'option *Save As* pour ne pas écraser les données précédentes.

Après la création des données d'entraînement, il est nécessaire de les convertir en entités ponctuelles avec l'outil *Feature To Point*. L'outil *Buffer* a ensuite été utilisé pour appliquer une zone tampon autour de ces entités. Différentes largeurs de zones tampons ont été testées¹¹. Compte tenu de la résolution de la composite utilisée pour la classification, la zone tampon minimum retenue a été de 20 mètres, 10 mètres étant trop faible pour considérer plus d'un pixel. Afin d'évaluer l'impact potentiel du nombre de données d'entraînement sur la performance de la classification, deux approches ont été explorées pour chaque niveau : une classification basée sur un nombre relativement restreint de données d'entraînement et une autre avec le double de données par classe.

Par défaut, 10% des données d'entraînement sont écartés aléatoirement de la classification, créant ainsi deux sous-ensembles distincts : l'un pour l'apprentissage et l'autre pour la validation des données d'entraînement en entrée (*How Forest-based Classification and Regression works*, 2023).

4.4.2.2. Classification

La classification s'est basée sur les données d'entraînement créées au chapitre 4.4.2.1. Parmi les méthodes de classification supervisée proposées par ArcGIS Pro®, le classificateur *Random Trees* a été privilégié. Cependant, des essais ont également été faits avec le classificateur SVM.

Le nombre maximum d'arbres a été fixé à 20 afin de garantir un temps de traitement raisonnable, tout en assurant un nombre suffisant d'arbres de décision pour la classification. Les paramètres de profondeur des arbres et du nombre d'échantillons par classe¹² ont été laissés par défaut.

¹¹ Des largeurs de zones tampons de 20, 30 et 40 mètres ont été évaluées.

¹² La profondeur maximale des arbres représente le nombre de critères que chaque arbre peut utiliser pour prendre une décision. Pour le nombre maximal d'échantillons par classe, il est conseillé de garder la valeur par défaut lorsque le raster d'entrée n'est pas segmenté (*Classify*, 2023).

La classification génère deux fichiers en sortie : la composite multibandes classifiée et un fichier de définition de classificateur ESRI® (.ecd). Ce fichier contient des informations indispensables pour les modèles de classification, y compris le résultat de la validation réalisée avec les 10% d'échantillons exclus de la classification (Cross Validation Rate). Le fichier de définition stocke les informations relatives à la classification, facilitant une réutilisation ultérieure du modèle de classification.

La contribution de chaque bande constituant la composite à la classification a été évaluée. Le fichier de définition ESRI® inclut la valeur d'importance de chaque bande dans le cadre de la classification effectuée. La contribution relative de chacune des bandes, c'est-à-dire la valeur d'importance pondérée par l'inverse du nombre de bandes, a été calculé.

4.5.Évaluation de la précision des classifications

Afin d'évaluer la précision des classifications, plusieurs étapes ont été suivies. Tout d'abord, l'outil *Create Accuracy Assessment Points* a été utilisé avec les couches de données d'entraînement utilisées pour chaque classification en entrée. Pour cette étape, le champ cible a été *Ground Truth* et le nombre de points aléatoires par classe a été fixé arbitrairement à 100. La stratégie d'échantillonnage *Equalized stratified random* a été choisie pour garantir un nombre équivalent de points par classe. Ces points d'évaluation de la précision sont alors compris dans l'emprise des données d'entraînement. Les points d'évaluation ont ensuite été mis à jour grâce à l'outil *Update Accuracy Assessment Points* avec en entrée la couche classifiée. Le champ cible à cette étape a été *Classified*. Sur la base de ces points mis à jour, une matrice de confusion a été réalisée avec l'outil *Compute Confusion Matrix*.

Suivant une démarche similaire, une évaluation de la précision a également été réalisée manuellement. L'une des distinctions majeures réside dans le choix du champ cible lors de la création des points d'évaluation. Pour la première étape, le champ *Classified* a été sélectionné avec en entrée la couche classifiée et un nombre fixe de 10 échantillons aléatoires par classe. Les valeurs de classe réelle ont été ensuite mises à jour manuellement dans la colonne *Ground Truth* de la table attributaire des points d'évaluation de la précision.

La matrice de confusion présente les résultats obtenus en comparant les échantillons de classe prédits par le classificateur aux échantillons de classe réelle. Les erreurs sont évaluées en termes d'omission (résultats des colonnes) et de commission (résultats des lignes). Un intervalle de confiance a été calculé à partir de l'erreur moyenne. Celle-ci se calcule en soustrayant la valeur de précision globale à 1. La formule pour l'intervalle de confiance est (Allenbach & Dao, 2023) :

$$IC = \pm z \sqrt{\frac{p * q}{n}}$$

IC = intervalle de confiance

z = score de l'intervalle de confiance (1.96 pour 95%)

p = erreur moyenne

q = 1-p

n = nombre total de pixel

5. Résultats

5.1. Comparaison du paysage entre les saisons humide et sèche

Pour comparer l'évolution du paysage à différentes saisons, l'indice NDVI a été calculé pour le 14 février 2023 (saison humide) et le 8 août 2021 (saison sèche) (Figure 6). Des composites ont également été créés dans ce but (cf. chapitre 4.3.5) et sont présentées en annexe (8, 9 et 10).

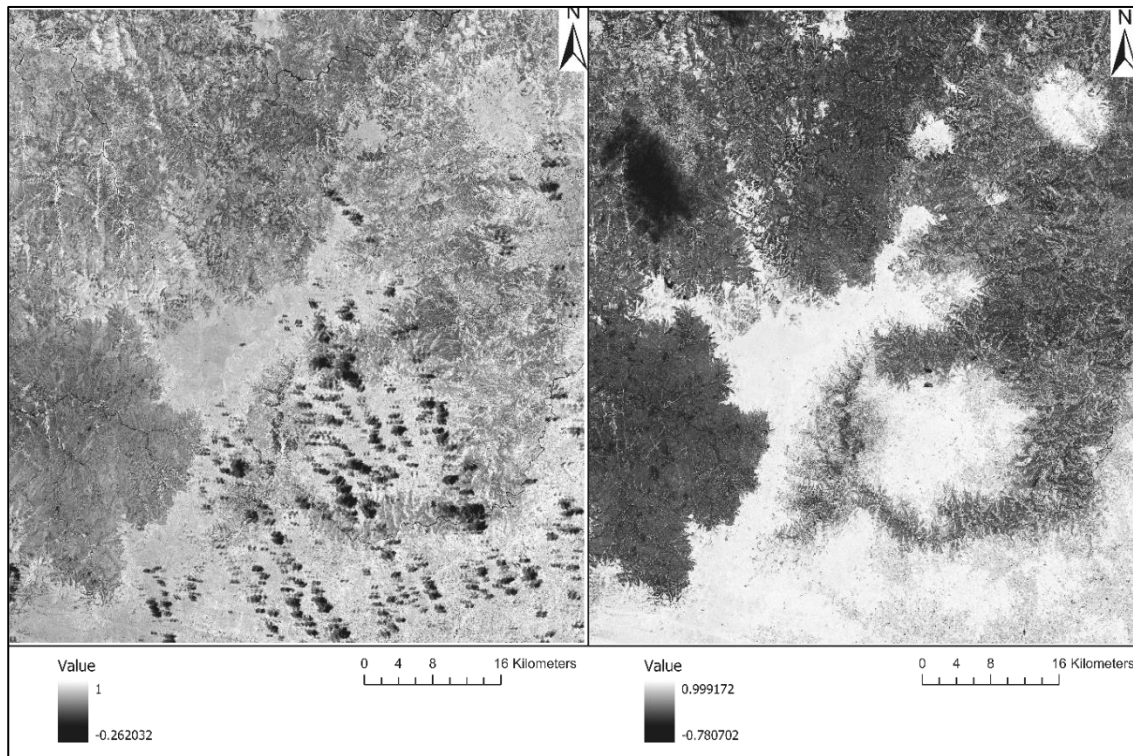


Figure 6 Indice NDVI calculé avec Raster Calculator pour le 14 février 2023 (saison humide) (image de gauche) et le 8 août 2021 (saison sèche) (image de droite).

Durant la saison humide, les valeurs de l'indice NDVI varient de 1 (activité photosynthétique élevée) à -0.26 (activité photosynthétique faible/sol nu). Les valeurs observées pour la saison sèche s'étendent quant à elles de 0.99 à -0.78. Ce déplacement de la gamme de valeurs vers des valeurs plus basses traduit de manière évidente les caractéristiques propres à cette période de l'année, soit une activité photosynthétique moindre liée à un manque d'eau.

L'analyse visuelle révèle une plus grande uniformité tant au niveau des teintes que des valeurs de NDVI pendant la saison humide, tandis que la saison sèche se caractérise par une variation paysagère plus marquée. Les zones avoisinant le massif de Sorata affichent des valeurs de NDVI nettement inférieures à celles observées dans le massif, témoignant d'une diminution significative de l'activité photosynthétique dans ces zones. Par ailleurs, le massif ne semble pas subir de perte de superficie de végétation photosynthétiquement active entre la saison sèche et humide.

5.2. Classifications supervisées

5.2.1. Classification « visuelle »

La composite utilisée pour effectuer les classifications est présentée en annexe 11. La classification brute finale, basée sur les images satellites ainsi que les données d'altitude et de courbure, est illustrée en figure 7. Les résultats des classifications des autres niveaux sont consignés en annexe (12 à 15).

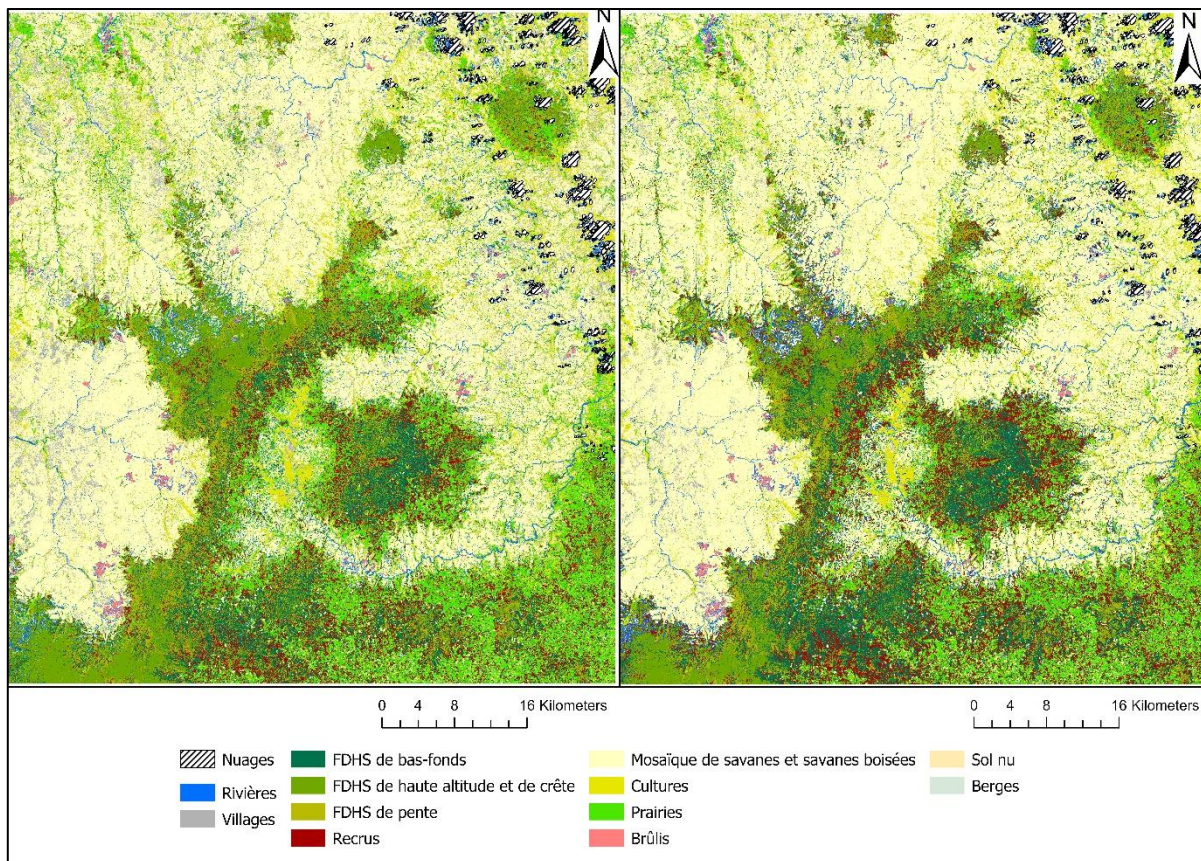


Figure 7 Cartographie des milieux du massif forestier de Sorata et alentours, basée sur des images satellites ainsi que des données d'altitude et de courbure, réalisée par classification supervisée avec ArcGIS Pro®. Les classifications ont été effectuées avec un ensemble limité (10) de données d'entraînement (image de gauche) ainsi qu'avec le double de données (image de droite). La zone tampon définie pour les données d'entraînement est de 20 mètres. (FDHS = Forêts Dense Humide Sempervirente)

Les deux cartographies présentent des différences, notamment en termes de sur- ou sous-estimation de certaines classes. Par exemple, dans le cas de la classification utilisant un ensemble limité de données d'entraînement, il semble que les zones de recrus soit sous-estimées, tandis que les zones urbaines paraissent être surestimées. Dans l'ensemble, les deux classifications demeurent toutefois relativement similaires dans leur représentation des différentes classes.

Les résultats obtenus pour les matrices de confusion de la classification réalisée avec un nombre plus important d'échantillons d'entraînement (20), générés automatiquement et manuellement, sont présentés ci-dessous (Tableaux 6 et 7). Ces résultats révèlent des différences significatives. En effet, que ce soit pour la précision de l'utilisateur ou du producteur ou bien encore pour le coefficient de Kappa, les valeurs obtenues automatiquement sont nettement supérieures, atteignant un coefficient de Kappa presque parfait de 0.99 ± 0.0054 . En revanche, les résultats obtenus manuellement sont plus modestes, avec une valeur de Kappa de 0.718 ± 0.078 . Cette approche se traduit par des performances globalement bonnes, bien que parfois médiocres, en particulier en ce qui concerne la précision de l'utilisateur de la classe « Villages » (valeur de 0.1).

Tableau 6 Matrice de confusion générée automatiquement avec 100 points de précision de l'évaluation distribués de manière aléatoire stratifiée par classe (IC = intervalle de confiance)

	Rivières	Villages	FDHS de bas-fonds	FDHS de haute altitude et de crête	FDHS de pente et de moyenne altitude	Recrus	Mosaïque de savanes et savanes boisées	Cultures	Prairies	Brûlis	Sol nu	Berges	Total	User's Accuracy	Kappa
Rivières	100	0	1	0	0	0	0	0	0	0	0	0	101	0.990099	0
Villages	0	100	0	0	0	0	0	0	0	0	0	0	100	1	0
FDHS de bas-fonds	0	0	98	2	1	2	0	0	0	0	0	0	103	0.951456	0
FDHS de haute altitude et de crête	0	0	0	98	1	1	0	0	0	0	0	0	100	0.98	0
FDHS de pente et de moyenne altitude	0	0	0	0	97	0	0	0	0	0	0	0	97	1	0
Recrus	0	0	1	0	1	96	0	0	0	0	0	0	98	0.979592	0
Mosaïque de savanes et savanes boisées	0	0	0	0	0	0	100	0	0	0	0	0	100	1	0
Cultures	0	0	0	0	0	0	0	100	0	0	0	0	100	1	0
Prairies	0	0	0	0	0	1	0	0	100	0	0	0	101	0.990099	0
Brûlis	0	0	0	0	0	0	0	0	0	100	0	0	100	1	0
Sol nu	0	0	0	0	0	0	0	0	0	0	100	0	100	1	0
Berges	0	0	0	0	0	0	0	0	0	0	0	100	100	1	0
Total	100	100	100	100	100	100	100	100	100	100	100	100	1200	0	0
Producer's Accuracy	1	1	0.98	0.98	0.97	0.96	1	1	1	1	1	1	0	0.990833	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9900

IC = 0.0054

Tableau 7 Matrice de confusion générée manuellement avec 10 points de précision de l'évaluation distribués de manière aléatoire stratifiée par classe (IC = intervalle de confiance)

	Rivières	Villages	FDHS de bas-fonds	FDHS de haute altitude et de crête	FDHS de pente et de moyenne altitude	Recrus	Mosaïque de savanes et savanes boisées	Cultures	Prairies	Brûlis	Sol nu	Berges	Total	User's Accuracy	Kappa
Rivières	3	0	0	1	0	3	3	0	0	0	0	0	10	0.3	0
Villages	0	1	0	0	0	1	6	0	0	0	2	0	10	0.1	0
FDHS de bas-fonds	0	0	8	0	1	1	0	0	0	0	0	0	10	0.8	0
FDHS de haute altitude et de crête	0	0	0	9	1	0	0	0	0	0	0	0	10	0.9	0
FDHS de pente et de moyenne altitude	0	0	0	1	9	0	0	0	0	0	0	0	10	0.9	0
Recrus	0	0	2	0	0	8	0	0	0	0	0	0	10	0.8	0
Mosaïque de savanes et savanes boisées	0	0	0	0	0	0	10	0	0	0	0	0	10	1	0
Cultures	0	0	0	0	0	4	2	4	0	0	0	0	10	0.4	0
Prairies	0	0	0	0	0	1	0	0	9	0	0	0	10	0.9	0
Brûlis	0	0	0	0	0	0	0	0	0	10	0	0	10	1	0
Sol nu	0	0	0	0	0	0	1	0	0	0	9	0	10	0.9	0
Berges	0	0	0	0	0	0	0	0	0	0	1	9	10	0.9	0
Total	3	1	10	11	11	18	22	4	9	10	12	9	120	0	0
Producer's Accuracy	1	1	0.8	0.818182	0.818182	0.444444	0.454545	1	1	1	0.75	1	0	0.741667	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.718182
															IC=0.078

5.2.2. Classification « réalité du terrain »

Les classifications obtenues en intégrant les données des relevés de terrain sont relativement similaires à celles obtenues au chapitre précédent. Néanmoins, certaines différences peuvent être relevées (Figure 8). Les forêts denses humides sempervirentes (FDHS) de bas-fonds en particulier semblent être moins sur-représentées que dans la classification « visuelle », notamment pour la classification avec peu d'échantillons (image de gauche). Par ailleurs, les forêts de bambous, indétectables sans les données de terrain¹³, semblent être réparties un peu partout dans le massif de Sorata. De manière générale, et à l'instar de la classification précédente, des différences peuvent être relevées entre les classifications avec un nombre restreint et plus élevé d'échantillons d'entraînement (sur- ou sous-estimation de certaines classes).

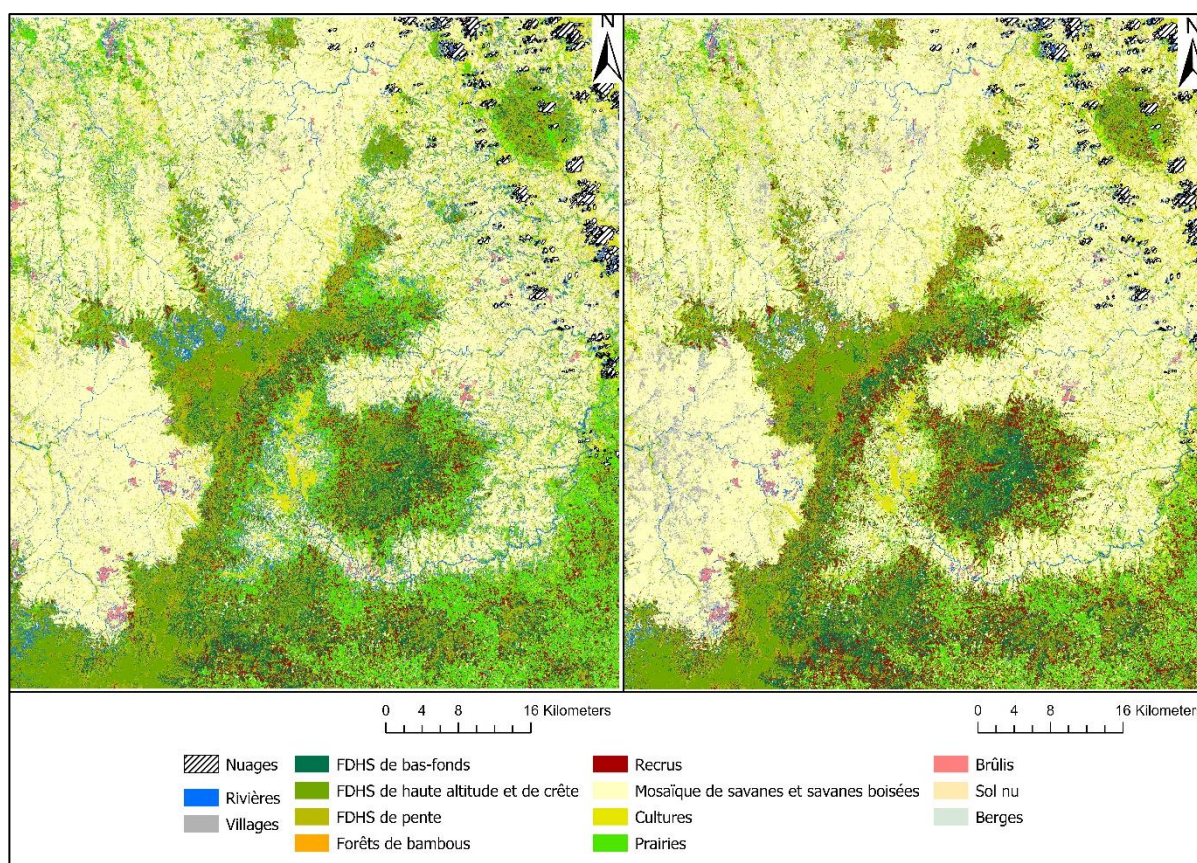


Figure 8 Cartographie des milieux du massif forestier de Sorata et alentours réalisée par classification supervisée avec ArcGIS Pro® et incluant les données des relevés de terrain réalisés entre fin 2021 et début 2023. Les classifications ont été effectuées avec un ensemble limité de données d'entraînement (10, excepté pour les forêts de bambous (5)) (image de gauche) ainsi qu'avec le double de données (image de droite). La zone tampon définie pour les données d'entraînement est de 20 mètres. (FDHS = Forêts Dense Humide Sempervirente)

6. Discussion

Les classifications obtenues semblent donner des résultats relativement cohérents, bien que certains problèmes persistent. Plus spécifiquement, des difficultés récurrentes ont été observées lors de la classification de certaines classes, notamment entre les sols nus, les formations végétales de type savane et les zones de villages. Ces surfaces sont en effet souvent difficiles à distinguer (As-syakur et al.,

¹³ Ces bambous d'environ 2-3 m maximum étaient sous le couvert de la canopée et donc jamais visibles sur les images satellites (Havinga & Nusbaumer, comm. pers.).

2012). Les autres erreurs de classification fréquentes concernaient la classe « Eau » qui rentrait alors souvent en conflit avec celle des zones brûlées ou avec des zones ombragées, liées aux nuages et falaises notamment. De manière générale, il a été relevé que les classifications réalisées avec un nombre supérieur de données d'entraînement présentaient de meilleurs résultats, notamment en ce qui concerne la résolution des conflits entre les classes difficiles à distinguer. Cependant, il est essentiel de noter que si certaines sources de bruit ont été atténuées et certains pixels mieux classés, d'autres zones ont continué de présenter des problèmes de classification similaires à ceux de l'approche avec un nombre limité d'échantillons.

Pour ce qui est des classifications des différents types de forêts définis, les résultats semblent relativement cohérents, à l'exception des FDHS de bas-fonds qui semblent être largement sur-représentées pour les deux classifications obtenues, mais plus particulièrement pour la classification « visuelle ». Toutefois, en l'absence de connaissances personnelles du terrain, il est difficile de tirer des conclusions définitives, bien que cela semble plausible. De plus, le schéma de classification final pourrait ne pas être suffisamment exhaustif pour représenter de manière adéquate la diversité des milieux présents dans la zone d'étude. Les FDHS de haute altitude et de crête semblent particulièrement bien représentées, à l'exception du massif forestier situé au nord-est du Sorata qui n'atteint pas d'altitude supérieure à 1400 m et ne devrait alors présenter que des formations de moyenne altitude (1000-1400 m).

Concernant les résultats des matrices de confusion, il n'est pas surprenant d'observer des valeurs extrêmement élevées, voire presque parfaites, pour l'évaluation « automatique » de la précision. En effet, étant donné que les points d'évaluation de la précision sont basés sur les données d'entraînement, il est naturel que les valeurs prédites soient similaires à celles de référence. L'évaluation de la précision faite manuellement, même si elle présente certains biais, a donné des résultats plus cohérents. Certaines valeurs de précision étaient très faibles avec cette démarche, ce qui peut être attribué à divers facteurs, tels qu'un échantillonnage inadéquat ou une représentation limitée de la diversité des milieux dans la zone d'étude par exemple.

Pour les résultats de classification obtenus en incluant les données des relevés de terrain, ils semblent relativement similaires à ceux des classifications purement visuelles, mais probablement plus précis. Il est cependant difficile de tirer des conclusions pour les mêmes raisons qu'évoquées plus haut. De plus, la quantité et l'exhaustivité des données de terrain étaient relativement limitées, en particulier pour les forêts de bambous.

7. Conclusion

Le présent travail s'est caractérisé par une démarche à la fois expérimentale et exploratoire, ayant pour objectif la création d'une cartographie de la végétation du massif forestier de Sorata, situé au nord de Madagascar. Pour atteindre cet objectif, diverses approches et méthodes reposant sur des principes de télédétection ont été explorées, utilisant principalement le logiciel ArcGIS Pro® d'ESRI®. Les classifications supervisées, principalement basées sur une approche par pixel, ont constitué la base de ce travail. Deux cartographies ont été élaborées, se distinguant par l'inclusion ou l'exclusion des données provenant de relevés de terrain effectués entre fin 2021 et début 2023. La carte excluant ces données a été réalisée durant la période de stage, tandis que, en raison de contraintes de temps, la seconde a été élaborée en dehors de ce cadre.

Dans l'ensemble, les résultats obtenus ont été satisfaisants, bien que des erreurs de classification récurrentes aient été identifiées. Ces erreurs étaient dues à divers facteurs tels que la présence de nuages, d'ombres, ou encore à des regroupements incorrects dans les classes. L'effet poivre-sel était également notable sur les cartes finales, un phénomène courant dans une approche basée sur la classification par pixel.

Pour améliorer les résultats, plusieurs approches peuvent être envisagées. Le traitement post-classification, notamment l'application de filtres pour éliminer le bruit (Wulder et al., 2018), devrait être considéré. Faute de temps, cette étape n'a malheureusement pas été réalisée dans le cadre de ce travail. Des outils tels que *Majority Filter*, *Boundary Clean* et *Region Group*, disponibles dans ArcGIS Pro®, peuvent être utiles à cet égard. L'utilisation d'indices améliorés, comme celui proposé par As-syakur et al. (2012), ainsi que l'intégration de données auxiliaires précises pourraient également contribuer à réduire les erreurs de classification. De plus, l'usage d'images radar pour atténuer les biais dus aux nuages et à leurs ombres est également une option à envisager, le masque créé dans ce travail n'ayant pas permis de parfaitement résoudre ce problème. L'utilisation de séries temporelles d'images satellites pourrait également offrir une solution à cette problématique, tout en permettant de suivre et d'analyser les variations au fil des années.

Il est essentiel de noter que l'obtention d'une classification représentant parfaitement la réalité est presque impossible, notamment en raison de la variabilité des données au sein de chaque classe, de la résolution employée, de la précision et de l'exhaustivité des données de terrain, etc. Obligatoirement, certains pixels seront mal classés (Simoes et al., 2021). Néanmoins, malgré ces défis, les classifications peuvent produire des résultats globaux satisfaisants.

Certaines tentatives exploratoires n'ont malheureusement pas abouti, notamment celles visant à exporter le fond de carte *World Imagery* dans un format manipulable. Malgré les efforts déployés pour extraire les données, le résultat obtenu ne reflétait pas la résolution exceptionnelle du fond de carte original. Il serait donc pertinent d'approfondir cette démarche afin de pleinement exploiter la qualité de cette donnée. En outre, la comparaison par soustraction des NDVI entre la saison sèche et humide n'a pas non plus donné de résultats probants. Si les résultats de NDVI obtenus individuellement pour les deux dates semblaient parfaitement cohérents, le résultat de la soustraction présentait des valeurs déconcertantes et complexes à interpréter, nécessitant alors une étude plus approfondie. En effet, les zones caractérisées par des variations significatives entre la saison sèche et la saison humide, telles que les zones de type savane, ont montré des résultats qui semblaient indiquer l'absence de changement.

La végétation du massif forestier de Sorata telle que décrite par Moat & Smith (2007) ne rend probablement pas pleinement compte de la réelle complexité de cette zone. De plus, la quantité, la précision et l'exhaustivité des données de terrain utilisées étaient relativement limitées. La mission de terrain prévue dès mi-octobre devrait permettre d'améliorer la qualité des résultats obtenus dans ce travail.

En conclusion, ce travail constitue une étape cruciale vers la création d'une cartographie précise des milieux du massif forestier de Sorata. Bien que des défis subsistent, cette cartographie demeure un outil précieux pour identifier et protéger les milieux fragiles ou rares, contribuant ainsi à la gestion et à la préservation de ce site et de ses ressources. L'expérience obtenue lors de ce stage aux CJBG s'avère inestimable, offrant une opportunité de développer et d'appliquer des connaissances et des compétences acquises.

8. Bibliographie

- Allenbach, K., & Dao, H. (2023). *GEOTOOLS-RS Classifications*. Document de cours.
- Allenbach, K., Giuliani, G., Dao, H., Planque, M., & Chambers, F. (2023). *TP5 : Classifications, accuracy assessment, statistiques de surface et comportement spectral*. Document de cours.
- Ammann, M. (2011). *La presque-île d'Ampasindava (Nord-Ouest de Madagascar) : Une région mal connue et menacée, capitale pour la compréhension de la mise en place de la flore du Domaine du Sambirano* [Travail de maîtrise universitaire (Master) en Biologie]. Laboratoire de Botanique Systématique et Biodiversité.
- Arkebauer, T. J. (2015). Leaf Radiative Properties and the Leaf Energy Budget. In J. L. Hatfield & J. M. Baker (Éds.), *Agronomy Monographs* (p. 93-103). American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. <https://doi.org/10.2134/agronmonogr47.c5>
- As-syakur, Abd. R., Adnyana, I. W. S., Arthana, I. W., & Nuarsa, I. W. (2012). Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. *Remote Sensing*, 4(10), 2957-2970. <https://doi.org/10.3390/rs4102957>
- Besairie, H. (1964). *Carte géologique de Madagascar, au 1:1'000'000*. Service Géologique, Antananarivo.
- Brenon, P. (1972). The Geology of Madagascar. In R. Battistini & G. Richard-Vindard (Éds.), *Biogeography and Ecology in Madagascar* (p. 27-86). Springer Netherlands. https://doi.org/10.1007/978-94-015-7159-3_2
- Callmander, M. W., Phillipson, P. B., Andriamahefarivo, L. D., & Lowry, P. P. (2018). Site 10. Galoko-Kalobinono. In *The terrestrial protected areas of Madagascar : Their history, description, and biota* (eds S. M. Goodman, M.J. Raherilalao&S. Wohlhauser, p. 589-593). Association Vahatra.
- Chuvieco, E. (2020). *Fundamentals of Satellite Remote Sensing—An Environmental Approach* (3rd Edition). CRC Press. <https://doi.org/10.1201/9780429506482>
- CJBG. (2023). *Présentation des herbiers*. Conservatoire et Jardin botaniques de Genève. <https://www.cjbg.ch/collections/herbiers/presentation-herbiers>
- Classify*. (2023). [ArcGIS Pro | Documentation]. Esri. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/classify.htm>
- Cornet, A. (1974). Essai de cartographie bioclimatique à Madagascar. *Notice Explicative de l'ORSTOM*, 55, 28.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Découvrez nos satellites*. (2023). Copernicus. <https://www.copernicus.eu/fr/propos-de-copernicus/infrastructure/decouvrez-nos-satellites>
- EO Browser*. (2023). Sentinel Hub. <https://www.sentinel-hub.com/explore/eobrowser/>

- Étude botanique et conservation du massif forestier montagneux méconnu de Sorata.* (2021). Conservatoire et Jardin botaniques de Genève. <https://www.cjbg.ch/fr/science/nos-activites-scientifiques/flores-catalogues/flore-vegetation-massifs-forestiers-0>
- Favre, S. (2013). *Recherche sur les récoltes botaniques du massif forestier de Sorata, nord-ouest de Madagascar* [Travail de baccalauréat universitaire en Biologie]. Université de Genève.
- Gautier, L., & Goodman, S. M. (2002). Description de la Réserve Spéciale de Manongarivo, Madagascar. In *Gautier, L. & Goodman, S.M. (eds.) Inventaire floristique et faunistique de la Réserve Spéciale de Manongarivo (NW Madagascar)* (p. 21-40).
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Glaw, F., Köhler, J., Hawlitschek, O., Ratsoavina, F., Rakotoarison, A., Scherz, M. D., & Vences, M. (2021). Extreme miniaturization of a new amniote vertebrate and insights into the evolution of genital size in chameleons. *Scientific Reports*, 11(1), 1-14. <https://doi.org/10.1038/s41598-020-80955-1>
- Goodman, S. M., Raherilalao, M. J., & Wohlauser, S. (Éds.). (2018). *Les aires protégées terrestres de Madagascar : Leur histoire, description et biote /The terrestrial protected areas of Madagascar : their history, description, and biota*. Association Vahatra.
- Havinga, A. (2022). *Rapport de mission—Sorata Madagascar 2021-2022* (p. 19) [Non publié].
- How Forest-based Classification and Regression works.* (2023). [ArcGIS Pro | Documentation]. Esri. <http://link.springer.com/10.1023/A:1010933404324>
- Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725-749. <https://doi.org/10.1080/01431160110040323>
- Indices gallery.* (2023). [ArcGIS Pro | Documentation]. Esri. <https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm>
- Jawak, S., Devliyil, P., & Luis, A. (2015). A Comprehensive Review on Pixel Oriented and Object Oriented Methods for Information Extraction from Remotely Sensed Satellite Images with a Special Emphasis on Cryospheric Applications. *Advances in Remote Sensing*, 4(3), 177-195. <https://doi.org/10.4236/ars.2015.43015>
- Kaur, R., & Pandey, P. (2022). A review on spectral indices for built-up area extraction using remote sensing technology. *Arabian Journal of Geosciences*, 15(5), 391. <https://doi.org/10.1007/s12517-022-09688-x>
- Kulkarni, A. D., & Lowe, B. (2016). Random Forest Algorithm for Land Cover Classification. *International Journal on Recent and Innovation Trends in Computing and Communication*, 4(3), 58-63.
- Logiciel SIG.* (2023). Esri France. <https://www.esrifrance.fr/produits/Logiciel-SIG.aspx>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing : An applied review. *International Journal of Remote Sensing*, 39(9), 2784-2817. <https://doi.org/10.1080/01431161.2018.1433343>

- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425-1432. <https://doi.org/10.1080/01431169608948714>
- Melesse, A., Weng, Q., Thenkabail, P., & Senay, G. (2007). Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. *Sensors*, 7(12), 3209-3241. <https://doi.org/10.3390/s7123209>
- Moat, J., & Smith, P. (2007). Atlas of the vegetation of madagascar. *Royal Botanic Gardens, Kew*, 126.
- Mosaic Rasters function*. (2023). [ArcGIS Pro | Documentation]. Esri. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/mosaic-rasters.htm>
- Nusbaumer, L. (2011). Species distribution patterns in steep environmental gradients : Downscaling of a biogeographical framework (Loky-Manambato Region, NE Madagascar). - *Ph.D. Thesis, Université de Genève*. <https://doi.org/10.13097/archive-ouverte/unige:17744>
- Nusbaumer, L., Ranirison, P., Gautier, L., Chatelain, C., Loizeau, P.-A., & Spichiger, R. (2010). Loky-Manambato : Point de rencontre des principales unités phytogéographiques de Madagascar. In *Systématique et Conservation des Plantes Africaines* (X. van der Burgt, J. van Maesen&J.-M. Onana, p. 253-264). Royal Botanical Garden.
- Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies : Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131. <https://doi.org/10.1016/j.rse.2012.10.031>
- Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156-168. <https://doi.org/10.1016/j.rse.2016.10.010>
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping : A Review. *Remote Sensing*, 12(14), 1-35. <https://doi.org/10.3390/rs12142291>
- Qu'est-ce qu'un SIG ?* (2023). UNIL. <https://www.unil.ch/gis/fr/home/menuinst/introduction/quest-ce-quun-sig.html>
- Rakotoarison, A., Scherz, M., Köhler, J., Rasoavina, F., Hawlitschek, O., Megson, S., Vences, M., & Glaw, F. (2020). Frogs of the genus *Platypelis* from the Sorata massif in northern Madagascar : Description of a new species and reports of range extensions. *Zoosystematics and Evolution*, 96(1), 263-274. <https://doi.org/10.3897/zse.96.47088>
- Rakotovoao, C., Razakamalala, R., Randriambololomamonjy, O., Ravelonarivo, D., Rakotonirina, N., Randrianarivelo, C., Antilahimena, P., & Callmander, M. W. (2006). Biogeography and endemism of the Northern Humid Moutains of Madagascar. *Missouri Botanical Garden, St-Louis*, 145.
- Ramandimbimananana, S. D. (2009). *Analyse floristique de quelques groupements végétaux des forêts des altitudes supérieures de la Montagne d'Ambre (Nord de Madagascar, Région DIANA)*. Université d'Antananarivo.

- Randimbarison, C. H., Ramavovololona, Roger, E., Rabarison, H., & Gautier, L. (2014). *Analyse structurale et floristique de la végétation sur coulées de lave de la montagne d'Ambre : Région Diana* [Mémoire]. Université d'Antananarivo.
- Ranirison, P. (2010). *LES MASSIFS FORESTIERS DE LA REGION DE LA LOKY-MANAMBATO (DARAINA), écorégion de transition Nord : Caractéristiques floristiques et structurales. Essai de modélisation des groupements végétaux*. Université d'Antananarivo.
- Ranirison, P., Nusbaumer, L., Roger, E., Rajeriarison, C., & Gautier, L. (2010). Les principaux facteurs environnementaux déterminant le répartition des groupements végétaux de la région Loky-Manambato (NE Madagascar). X. vander Burgt, J. van der Maesen & J.-M. Onana (eds), *Systématique et Conservation des Plantes Africaines*, 671-678.
- Ravet, J. (1952). *Notice sur la climatologie de Madagascar et des Comores*. 4(1), 1-36.
- Razanajatovo, M. H. (2009). *Caractérisation écologique et analyse structurale des forêts humides de la MONTagne d'Ambre (région DIANA)*. Université d'Antananarivo.
- Santos, L. A., Ferreira, K. R., Camara, G., Picoli, M. C. A., & Simoes, R. E. (2021). Quality control and class noise reduction of satellite image time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 177, 75-88. <https://doi.org/10.1016/j.isprsjprs.2021.04.014>
- Scherz, M. D., Hawlitschek, O., Andreone, F., Rakotoarison, A., Vences, M., & Glaw, F. (2017). A review of the taxonomy and osteology of the Rhombophryne serratopalpebrosa species group (Anura : Microhylidae) from Madagascar, with comments on the value of volume rendering of micro-CT data to taxonomists. *Zootaxa*, 4273(3), 301-340. <https://doi.org/10.11646/zootaxa.4273.3.1>
- Scherz, M. D., Rakotoarison, A., Hawlitschek, O., Vences, M., & Glaw, F. (2015). Leaping towards a saltatorial lifestyle? An unusually long-legged new species of Rhombophryne (Anura, Microhylidae) from the Sorata massif in northern Madagascar. *Zoosystematics and Evolution*, 91(2), Art. 2. <https://doi.org/10.3897/zse.91.4979>
- Signatures spectrales*. (2009). ESA - Eduspace FR; European Space Agency. https://www.esa.int/SPECIALS/Eduspace_FR/SEM9Q0S7NWF_0.html
- Sigrist, R., & Bungener, P. (2008). The first botanical gardens in Geneva (c. 1750–1830) : Private initiative leading science. *Studies in the History of Gardens & Designed Landscapes*, 28(3-4), 333-350. <https://doi.org/10.1080/14601176.2008.10404723>
- Simoes, R., Camara, G., Queiroz, G., Souza, F., Andrade, P. R., Santos, L., Carvalho, A., & Ferreira, K. (2021). Satellite Image Time Series Analysis for Big Earth Observation Data. *Remote Sensing*, 13(13), 2428. <https://doi.org/10.3390/rs13132428>
- Somvanshi, S. S., & Kumari, M. (2020). Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data. *Applied Computing and Geosciences*, 7, 1-10. <https://doi.org/10.1016/j.acags.2020.100032>
- Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., & Rahman, A. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*, 12(7), 1135. <https://doi.org/10.3390/rs12071135>

- Trigui, S. M. (2010). *Etude floristique et biogéographique des altitudes supérieures de la montagne d'Ambre (nord de Madagascar)* [Travail de maîtrise universitaire (Master) en Biologie]. Université de Genève.
- User Guides—Sentinel-2 MSI Level-2 Processing*. (2022). Sentinel Online. <https://copernicus.eu/user-guides/sentinel-2-msi/processing-levels/level-2>
- Vali, A., Comai, S., & Matteucci, M. (2020). Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data : A Review. *Remote Sensing*, 12(15), 2495. <https://doi.org/10.3390/rs12152495>
- Vieilledent, G., Grinand, C., Rakotomalala, F. A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T. F., & Achard, F. (2018). Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation*, 222, 189-197. <https://doi.org/10.1016/j.biocon.2018.04.008>
- Walter-Shea, E. A., & Biehl, L. L. (1990). Measuring vegetation spectral properties. *Remote Sensing Reviews*, 5(1), 179-205. <https://doi.org/10.1080/02757259009532128>
- What is ModelBuilder?* (2023). [ArcGIS Pro | Documentation]. Esri. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geoprocessing/modelbuilder/what-is-modelbuilder-.htm>
- Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, 39(12), 4254-4284. <https://doi.org/10.1080/01431161.2018.1452075>
- Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices : A Review of Developments and Applications. *Journal of Sensors*, 2017, 1-17. <https://doi.org/10.1155/2017/1353691>

9. Annexes

Annexe 1 Récapitulatif des niveaux et classes définis pour les classifications supervisées et données utilisées correspondantes.

Annexe 2 Récapitulatif des outils utilisés sur ArcGIS Pro® par chapitre.

Annexe 3 *ModelBuilder* réalisé avec l’outil *Raster To Other Format* et l’itérateur *Iterate Rasters*.

Annexe 4 *ModelBuilder* réalisé avec l’outil *Extract By Mask* et l’itérateur *Iterate Rasters*.

Annexe 5 Indices spectraux ayant servi à la création de la composite utilisée pour les classifications.

Annexe 6 Composites créées et bandes associées.

Annexe 7 Code en langage Python développé pour créer une matrice de corrélation des bandes en entrée en format CSV.

Annexe 8 Composites vraies couleurs du 14 février 2023 (saison humide) à gauche et du 8 août 2021 (saison sèche) à droite.

Annexe 9 Composites fausses couleurs du 14 février 2023 (saison humide) à gauche et du 8 août 2021 (saison sèche) à droite.

Annexe 10 Composites pour l’analyse de la végétation du 14 février 2023 (saison humide) à gauche et du 8 août 2021 (saison sèche) à droite.

Annexe 11 Composite multibandes utilisée pour les classifications.

Annexe 12 Classification supervisée de niveau 1 réalisée avec un nombre restreint de données d’entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d’entraînement est de 20 mètres).

Annexe 13 Classification supervisée de niveau 2 réalisée avec un nombre restreint de données d’entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d’entraînement est de 20 mètres).

Annexe 14 Classification supervisée de niveau 3 réalisée avec un nombre restreint de données d’entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d’entraînement est de 20 mètres).

Annexe 15 Classification supervisée de niveau 4 réalisée avec un nombre restreint de données d’entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d’entraînement est de 20 mètres).

Annexe 1 Récapitulatif des niveaux et classes définis pour les classifications supervisées et données utilisées correspondantes.

		Niveaux					Données utilisées (hors données de terrain)	Données de terrain utilisées
		1	2	3	4	5		
Classes	Végétation	Arborée	Forêts de haute altitude (> 1400 m)*	F. de crête	FDHS de haute altitude et de crête	FDHS de haute altitude et de crête	Composites vraies et fausses couleurs Fond de carte <i>World Imagery</i> Indice NDVI Courbes de niveau Curvature	Relevés Braun-Blanquet de la mission fin 2021-début 2022 Relevés linéaires de la mission fin 2022-début 2023
				F. de pente				
				F. de bas-fond				
			Forêts de moyenne altitude (800 - 1400 m)*	F. de crête	FDHS de pente	FDHS de pente		
				F. de pente				
				F. de bas-fond				
			Forêts de basse altitude (< 800 m)*	F. de crête	FDHS de bas-fonds	FDHS de bas-fonds		
				F. de pente				
				F. de bas-fond				
	Arbustive	Recrus	Recrus	Recrus	Recrus	Composites vraies et fausses couleurs Fond de carte <i>World Imagery</i> Indice NDVI		
	Herbacée	Savane	Savane/Savane boisée	Mosaïque de savanes et savanes boisées	Mosaïque de savanes et savanes boisées			
		Cultures	Cultures	Cultures	Cultures			
		Zones herbeuses	Zones herbeuses	Prairies	Prairies			
	Sols	Sols nus	Routes	Routes	Sol nu	Sol nu		Composite vraies couleurs Fond de carte <i>World Imagery</i>
Erosion			Erosion					
Berges			Berges					
Zones brûlées		Brûlis	Brûlis	Brûlis	Brûlis			
Eau	Rivières	Rivières	Rivières	Rivières	Rivières	Composite finale		
Urbain	Villages	Villages	Villages	Villages	Villages	Composite vraies couleurs Fond de carte <i>World Imagery</i>		

*Les seuils altitudinaux ont été fixés arbitrairement

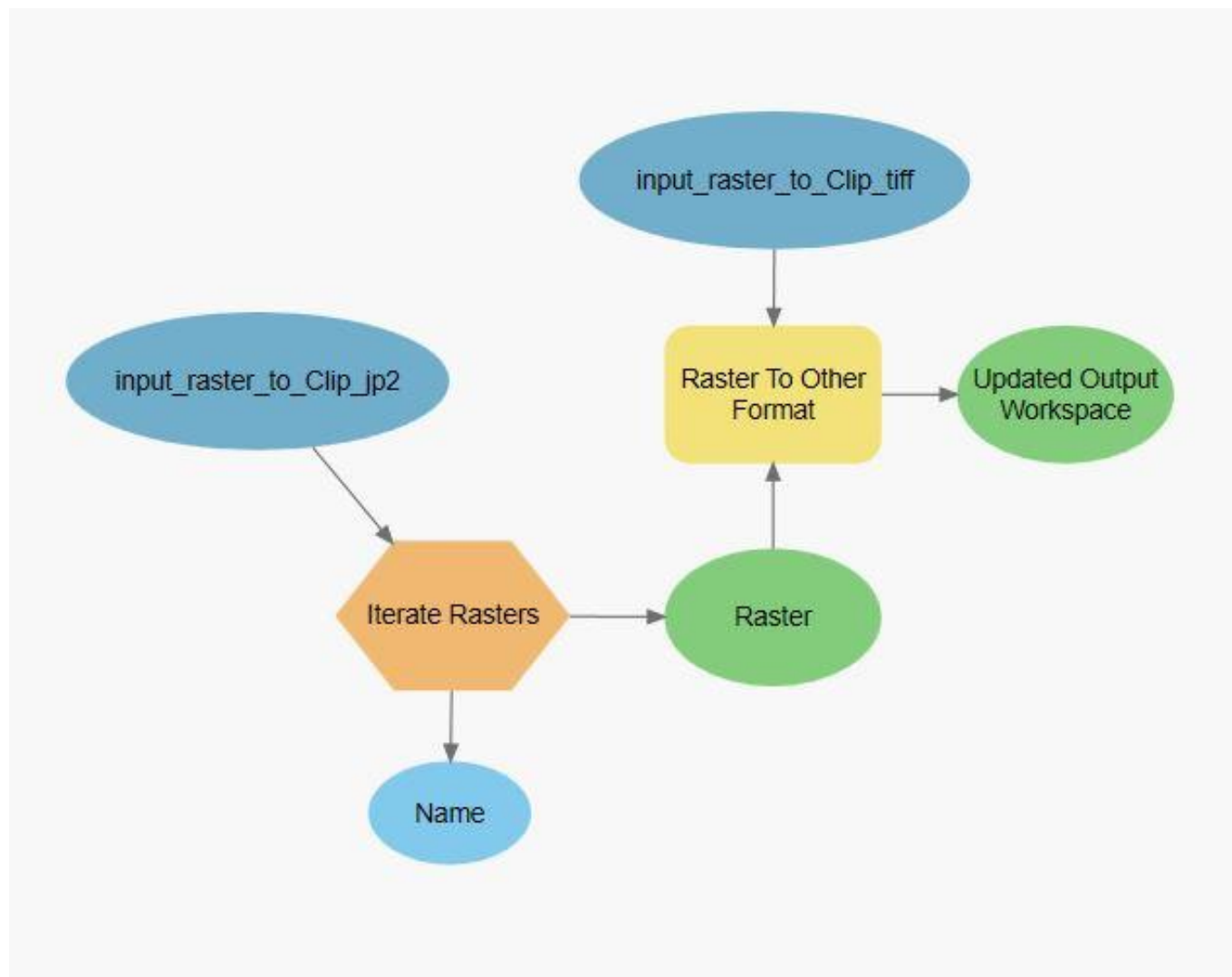
FDHS = Forêt Dense Humide Sempervirente

Annexe 2 Récapitulatif des outils utilisés sur ArcGIS Pro® par chapitre.

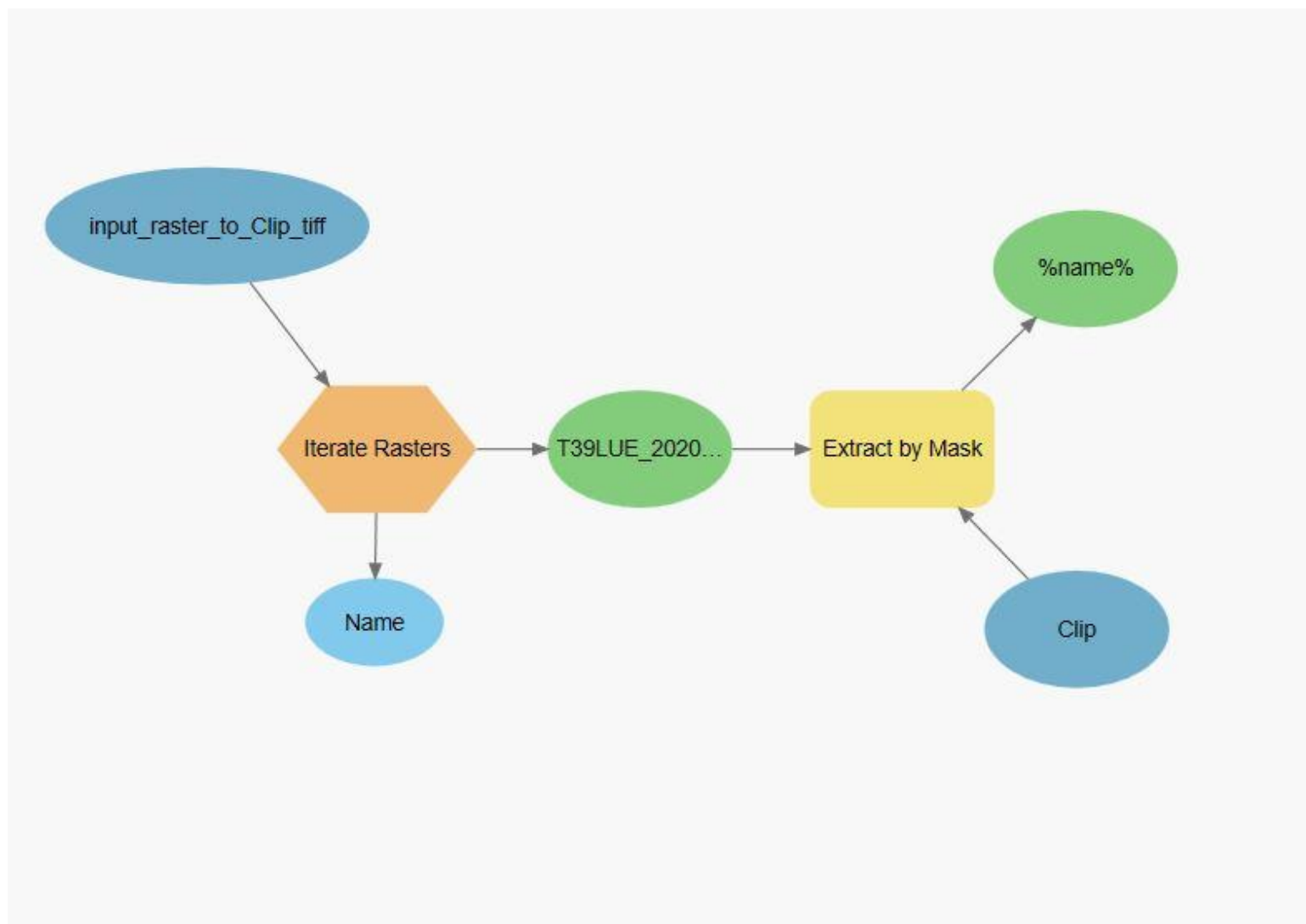
Chapitre	Outils	Onglet	Option	Référence de l'outil
4.3.	Define Projection	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/define-projection.htm
	Project	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/project.htm
	Project Raster	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/project-raster.htm
4.3.1.	Edit Vertices	Edit	Modify	https://pro.arcgis.com/en/pro-app/latest/help/editing/modify-feature-vertices.htm
	Raster To Other Format	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/conversion/raster-to-other-format-multiple.htm
	Extract By Mask	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/extract-by-mask.htm
	Clip Rasters	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/clip.htm
4.3.2.	Mosaic Rasters	Analysis	Raster Functions	https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/mosaic-rasters.htm
4.3.3.	Raster Calculator	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/raster-calculator.htm
	Raster To Polygon	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/conversion/raster-to-polygon.htm
	Select By Attributes	Map/Analysis	Select By Attributes/Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/select-layer-by-attribute.htm
	Split	Edit	Modify	https://pro.arcgis.com/en/pro-app/latest/help/editing/split-a-feature.htm
	Merge-1	Edit	Modify	https://pro.arcgis.com/en/pro-app/latest/help/editing/merge-features-into-one-feature.htm

	Merge-2	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/merge.htm
	Pixel Editor	Imagery	Pixel Editor	https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/pixel-editor-in-arcgis-pro.htm#:~:text=Pixel%20Editor%20contains%20a%20set,type%20of%20your%20raster%20dataset.
4.3.4.	Raster Calculator	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/raster-calculator.htm
4.3.5.	Composite Bands	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/composite-bands.htm
4.4.	Classify	Imagery	Classification Tools	https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/classify.htm
	Segmentation	Imagery	Classification Tools	https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/segmentation.htm
4.4.2.1.	Training Samples Manager	Imagery	Classification Tools	https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/training-samples-manager.htm
	Feature To Point	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/feature-to-point.htm
	Buffer	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/buffer.htm
4.5.	Create Accuracy Assessment Points	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/create-accuracy-assessment-points.htm
	Update Accuracy Assessment Points	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/update-accuracy-assessment-points.htm
	Compute Confusion Matrix	Analysis	Tools	https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/compute-confusion-matrix.htm

Annexe 3 *ModelBuilder* réalisé avec l'outil *Raster To Other Format* et l'itérateur *Iterate Rasters*.



Annexe 4 *ModelBuilder* réalisé avec l'outil *Extract By Mask* et l'itérateur *Iterate Rasters*.



Annexe 5 Indices spectraux ayant servi à la création de la composite utilisée pour les classifications.

Indice	Formule	Bandes utilisées	Domaine d'application	Référence de l'indice
Clay Minerals Ratio	SWIR1/SWIR2	11 et 12	Indice géologique	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm
MSAVI	$(1/2) * (2 * (NIR+1) - \sqrt{(2 * NIR+1)^2 - 8 * (NIR-R)})$	4 et 8	Indice de la végétation et des sols	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm ; https://custom-scripts.sentinel-hub.com/sentinel-2/savi/
Ferrous Minerals Ratio	SWIR1/NIR	8A et 11	Indice géologique	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm
PSRI	(R-B)/VRE	2, 4 et 6	Indice du stress de la végétation	https://custom-scripts.sentinel-hub.com/sentinel-2/psri/
NBR	$(NIR-SWIR)/(NIR+SWIR)$	8 et 12	Indice des zones brûlées	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm ; https://custom-scripts.sentinel-hub.com/sentinel-2/nbr/
NDBI	$(SWIR-NIR)/(SWIR+NIR)$	8 et 11	Indice des zones bâties	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm
Iron Oxide Ratio	R/B	2 et 4	Indice géologique	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm
MNDWI	$(G-SWIR)/(G+SWIR)$	3 et 11	Indice de l'eau	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm
VARI	$(G-R)/(G+R-B)$	2, 3 et 4	Indice de la végétation	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm
NDVI	$(NIR-R)/(NIR+R)$	4 et 8	Indice de la végétation	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm ; https://custom-scripts.sentinel-hub.com/sentinel-2/ndvi/
NDMI	$(NIR-SWIR1)/(NIR+SWIR1)$	8 et 11	Indice d'humidité	https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm ; https://custom-scripts.sentinel-hub.com/sentinel-2/ndmi/

Annexe 6 Composites créées et bandes associées (Source : <https://www.staridasgeography.gr/how-to-make-outstanding-maps-with-sentinel-2-and-arcgis-pro-part-1-band-combinations/>).

Composites	Bandes
Composite finale	B8 (10 m)
	NDVI (10 m)
	MNDWI (20 m)
	PSRI (20 m)
	MSAVI (10 m)
	Iron Oxide Ratio (10 m)
	Ferrous Minerals Ratio (20 m)
	Clay Minerals Ratio (20 m)
	NDMI (20 m)
	VARI (10 m)
	NBR (20 m)
	NDBI (20 m)
	B1 (60 m)
	B2 (10 m)
	B3 (10 m)
	B4 (10 m)
	B5 (20 m)
	B6 (20 m)
	B7 (20 m)
	B8A (20 m)
B9 (60 m)	
B11 (20 m)	
B12 (20 m)	
Vraies couleurs	B4 (10 m)
	B3 (10 m)
	B2 (10 m)
False color	B8 (10 m)
	B4 (10 m)
	B3 (10 m)
Analyse de la végétation/Agriculture	B11 (20 m)/B11 (20 m)
	B8 (10 m)/B8 (10 m)
	B4 (10 m)/B2 (10 m)

Annexe 7 Code en langage Python développé pour créer une matrice de corrélation des bandes en entrée en format CSV.

```
###-*- coding: cp1252 -*-

import arcpy
import os
import numpy as np
import pandas as pd

# Répertoire contenant les bandes satellites
band_directory = arcpy.GetParameterAsText(0)
arcpy.env.workspace = band_directory

# Récupérer tous les fichiers raster dans le répertoire
band_files = arcpy.ListRasters()

# Créer un dictionnaire pour stocker les données des bandes satellites
bands_data = {}

# Variables pour la vérification de la longueur des bandes
band_length = None
band_length_valid = True

# Parcourir les fichiers raster
for i, band_file in enumerate(band_files):
    # Chemin d'accès complet du fichier raster
    band_path = os.path.join(band_directory, band_file)
    # Convertir la bande en tableau NumPy
    band = arcpy.RasterToNumPyArray(band_path)
    # Vérification de la longueur de la bande
    if band_length is None:
        band_length = len(band.flatten())
    elif len(band.flatten()) != band_length:
        band_length_valid = False
        break
    # Ajouter les données de la bande au dictionnaire avec un nom de colonne dynamique
    column_name = f"Band{i+1}"
    bands_data[column_name] = band.flatten()

# Vérification de la longueur des bandes
if not band_length_valid:
    print("Les bandes satellites n'ont pas la même longueur.")
else:
    # Créer le DataFrame Pandas à partir du dictionnaire
    df = pd.DataFrame(bands_data)

    # Calcul de la matrice de corrélation
    correlation_matrix = df.corr()
```

```
# Nom du dossier de rasters en entrée
```

```
folder_name = os.path.basename(band_directory)
```

```
# Chemin et nom du fichier de sortie basés sur le nom du dossier de rasters
```

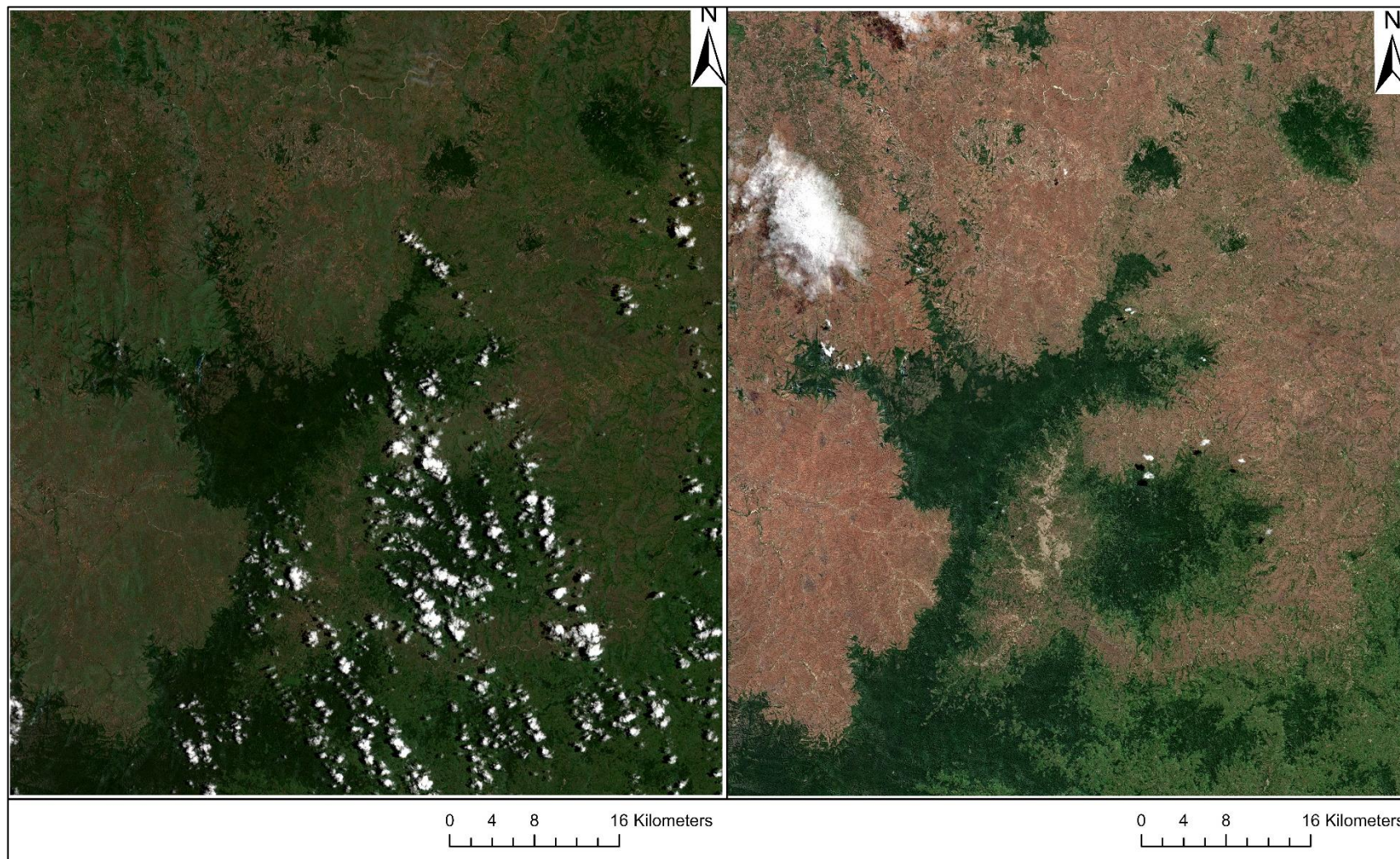
```
output_file =
```

```
os.path.join("C://Users//cecil//Desktop//MASTER//Certificat_Geomatique//Stage_Geomatique//Ma  
daCarto//Output_Correlation//", f"{folder_name}.csv")
```

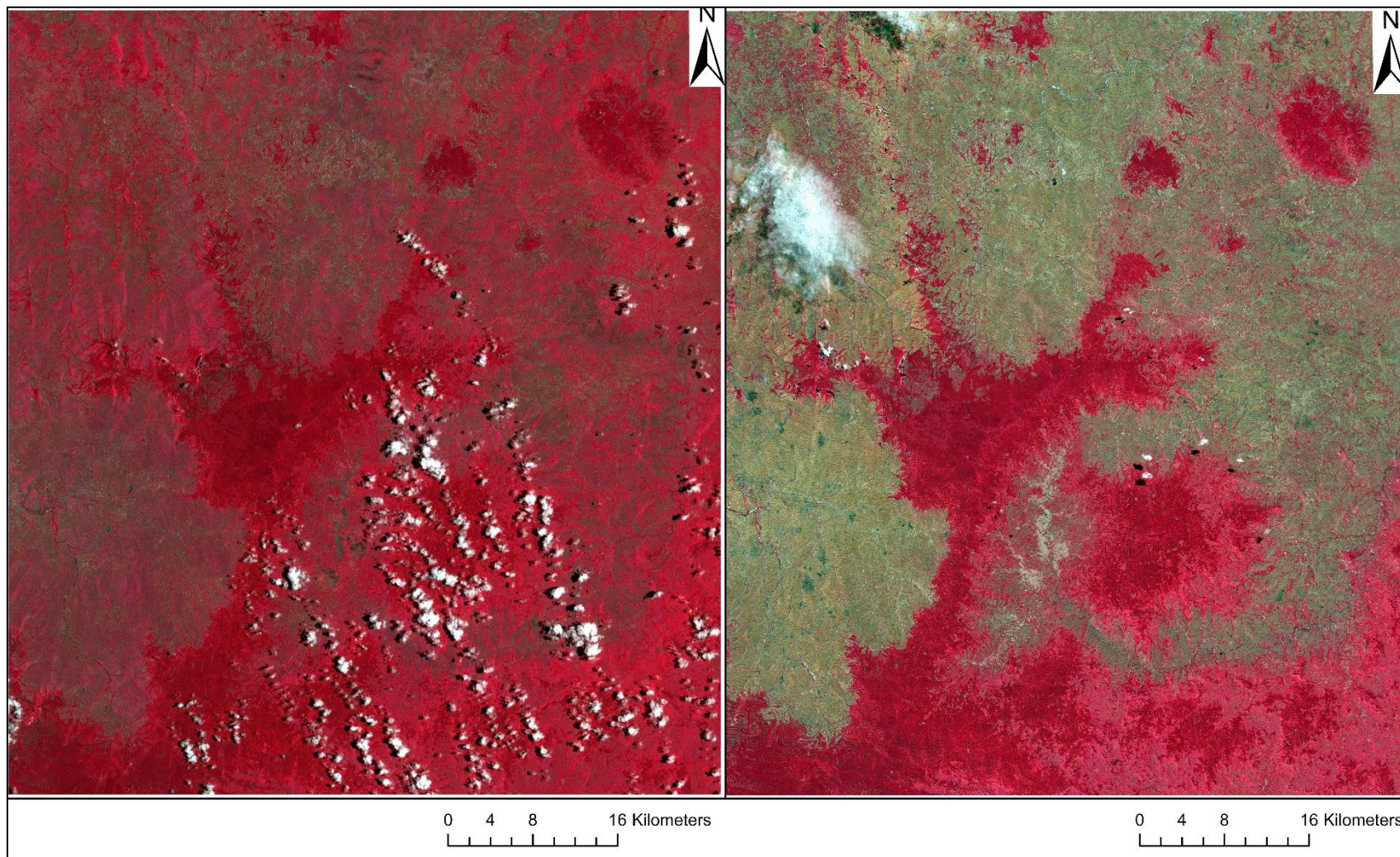
```
# Enregistrer la matrice de corrélation dans un fichier CSV
```

```
correlation_matrix.to_csv(output_file, index=False)
```

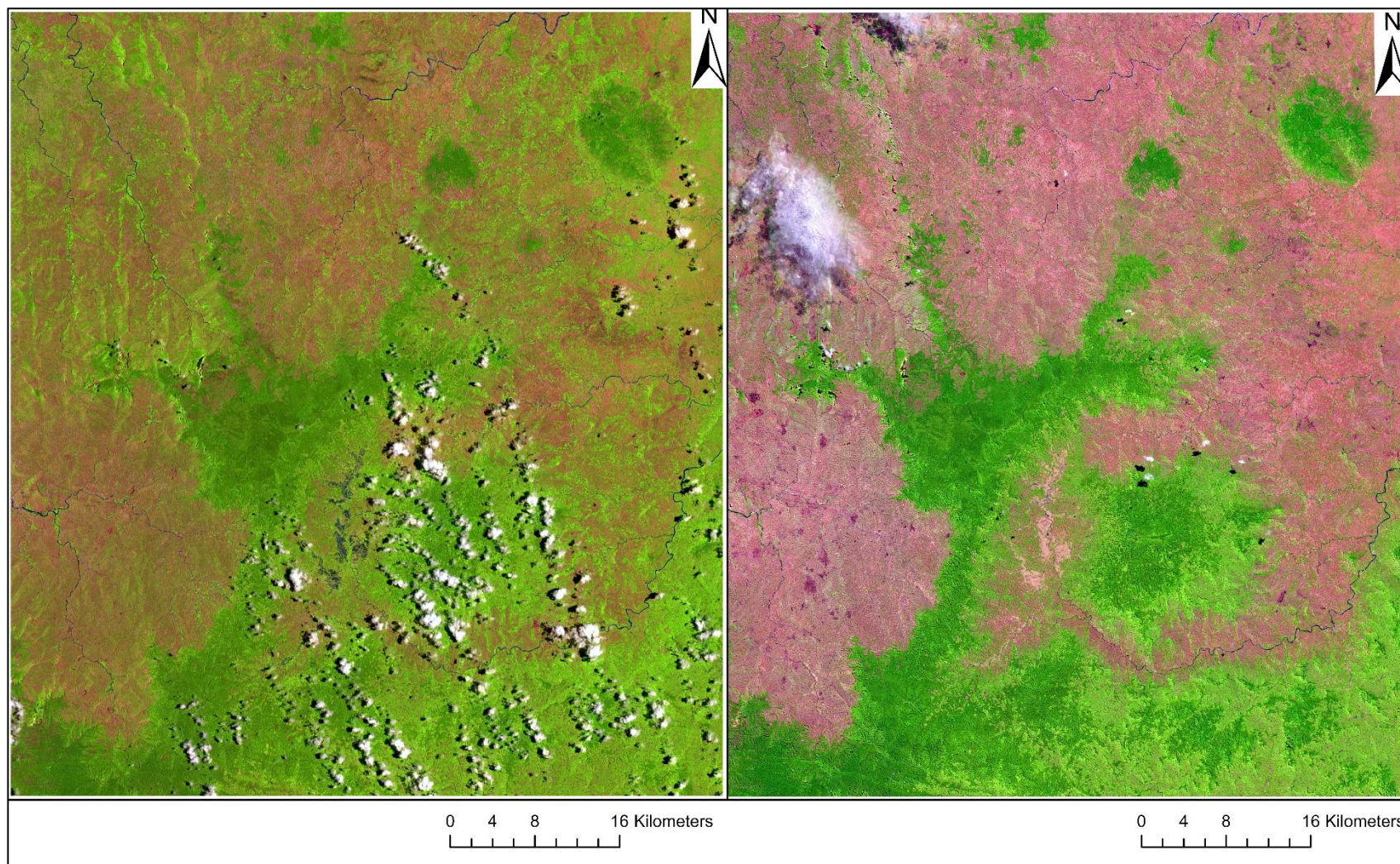

Annexe 8 Composites vraies couleurs du 14 février 2023 (saison humide) à gauche et du 8 août 2021 (saison sèche) à droite.



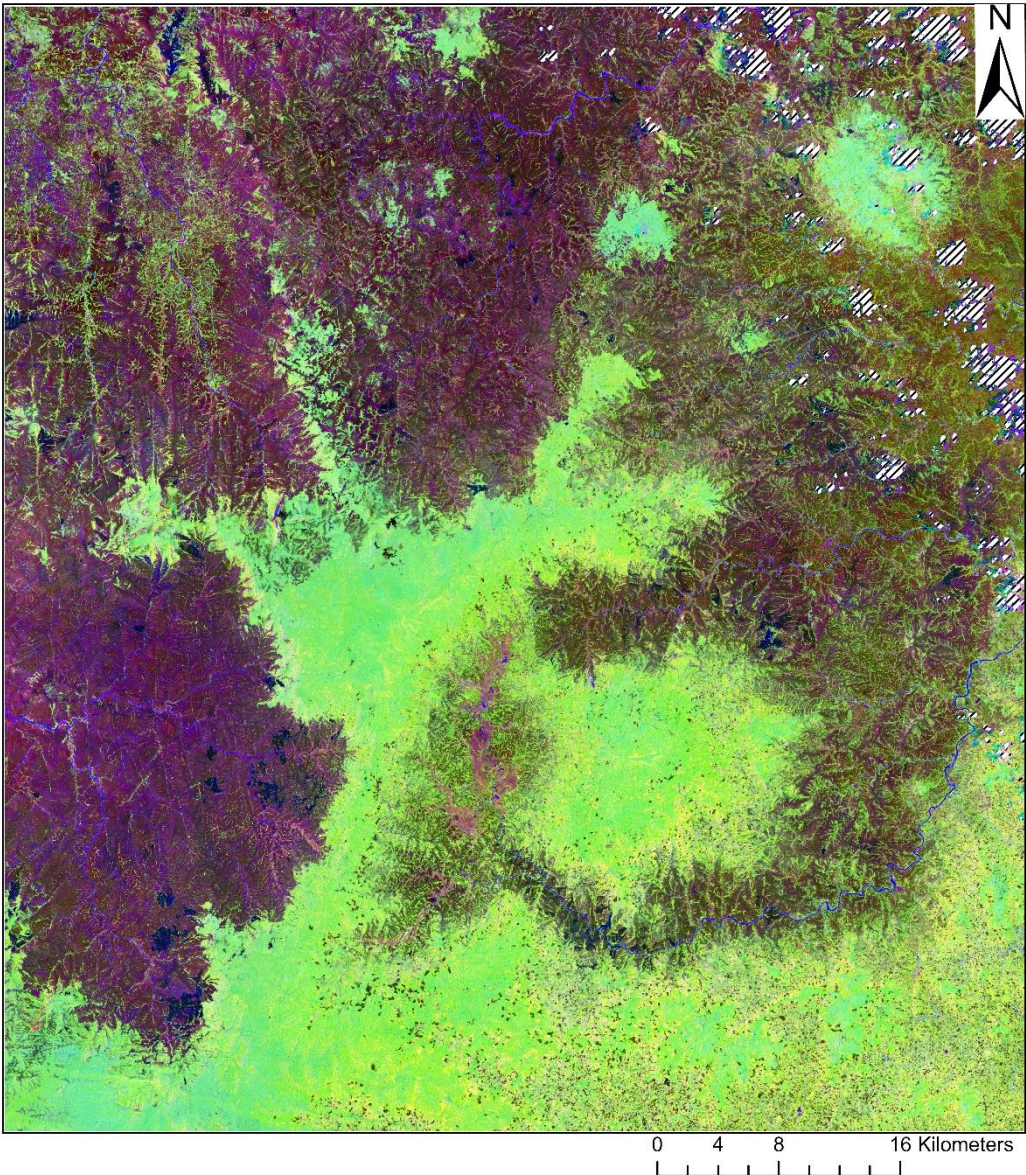
Annexe 9 Composites fausses couleurs du 14 février 2023 (saison humide) à gauche et du 8 août 2021 (saison sèche) à droite.



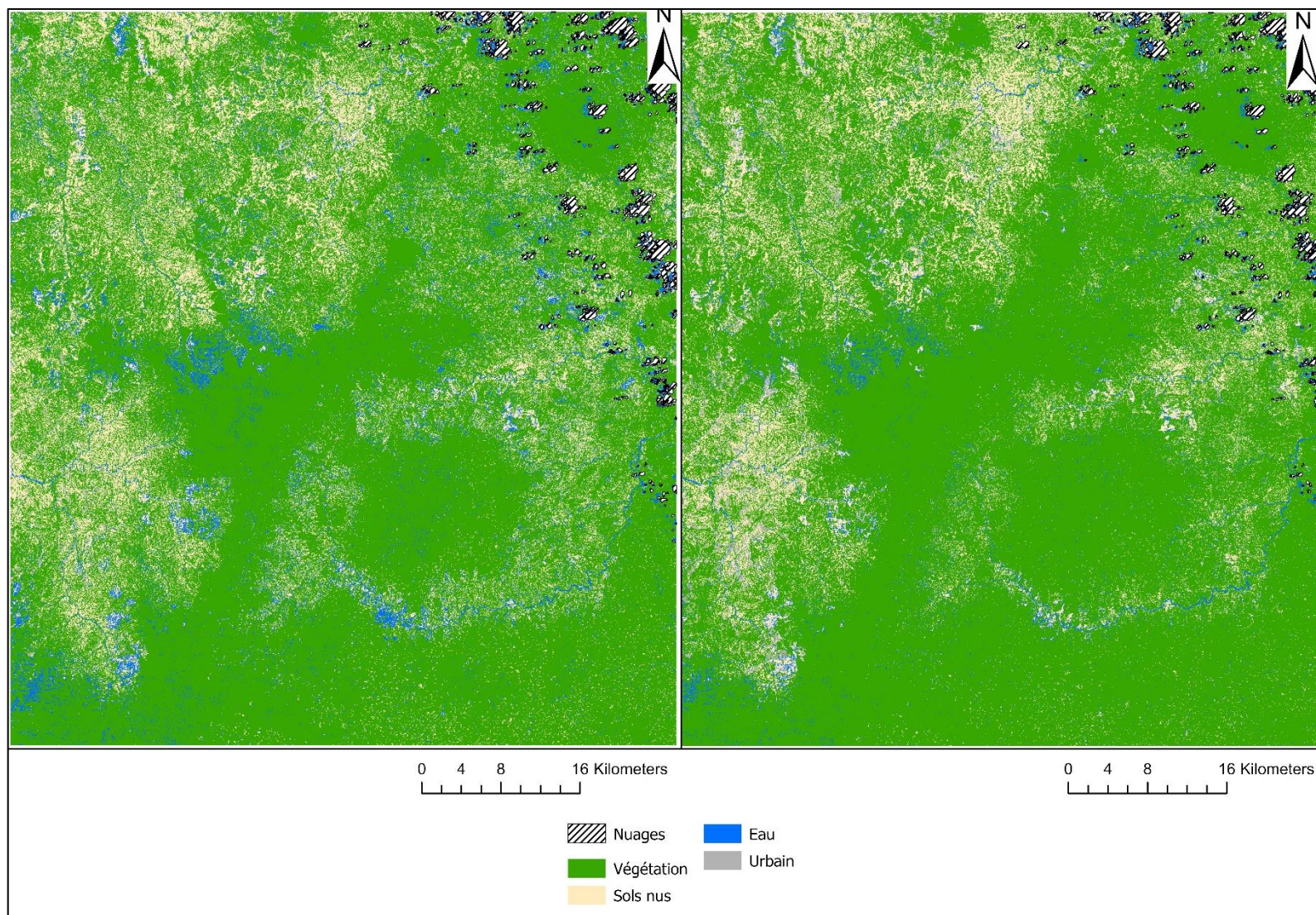
Annexe 10 Composites pour l'analyse de la végétation du 14 février 2023 (saison humide) à gauche et du 8 août 2021 (saison sèche) à droite.



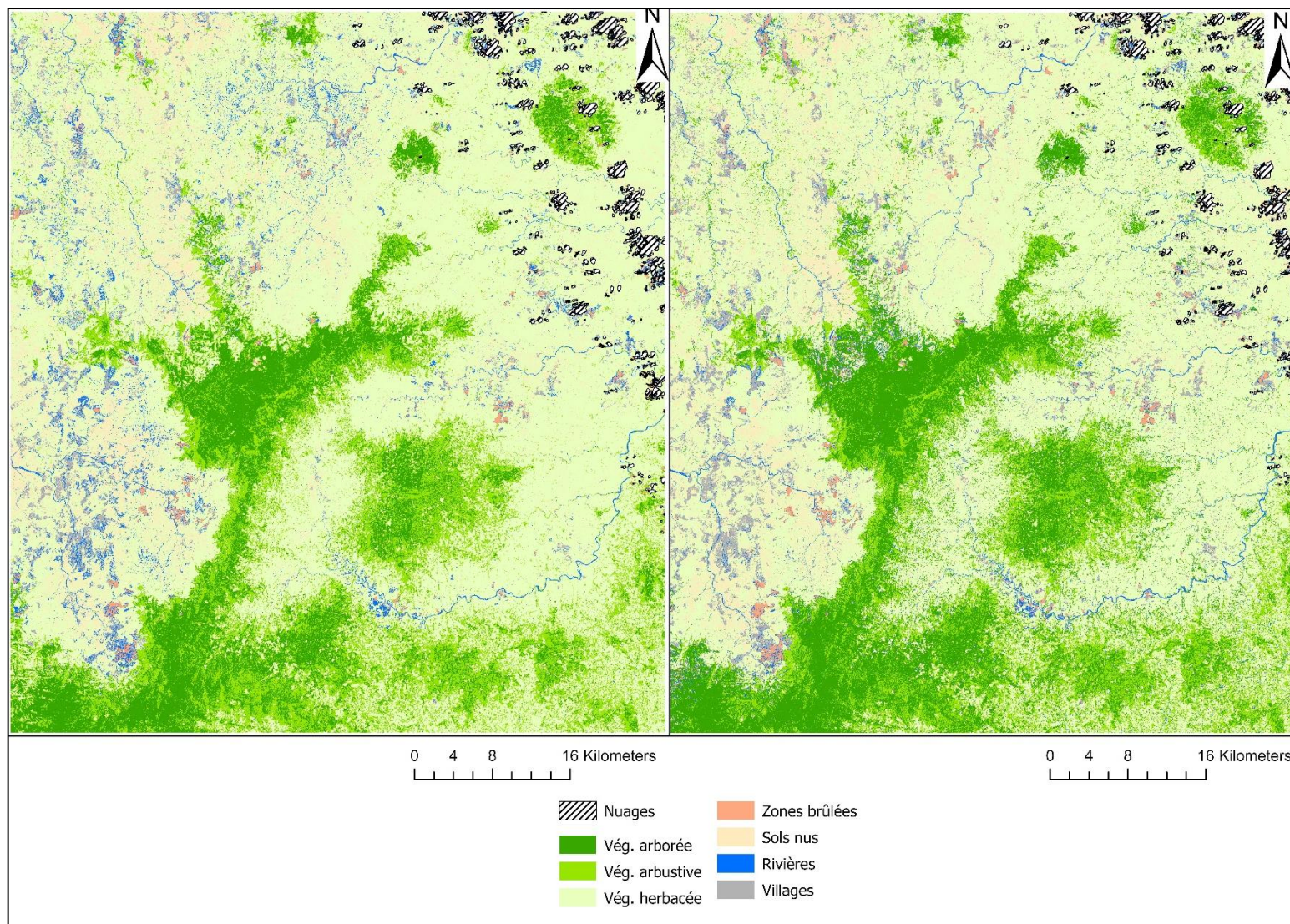
Annexe 11 Composite multibandes utilisée pour les classifications.



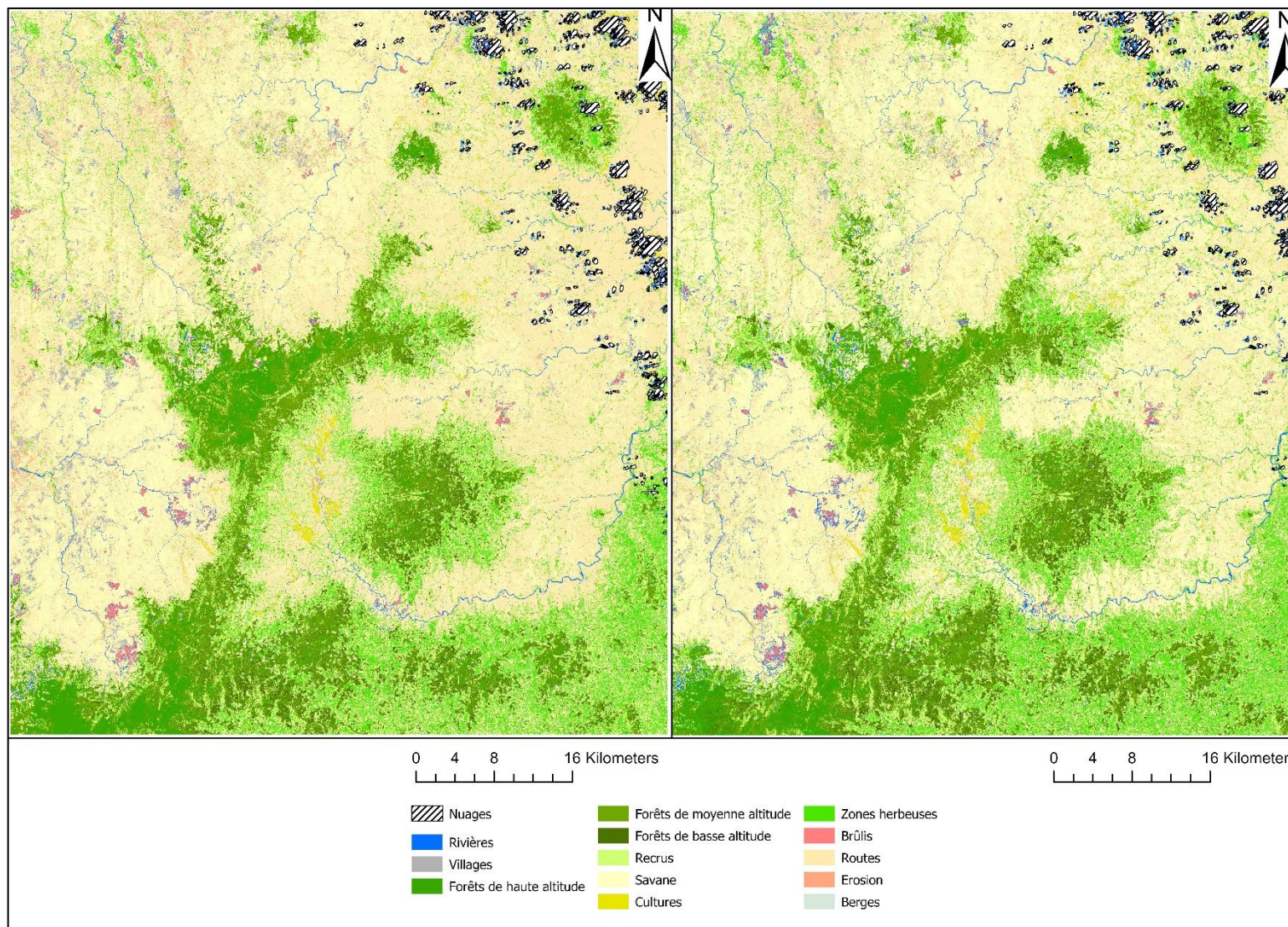
Annexe 12 Classification supervisée de niveau 1 réalisée avec un nombre restreint de données d'entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d'entraînement est de 20 mètres).



Annexe 13 Classification supervisée de niveau 2 réalisée avec un nombre restreint de données d'entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d'entraînement est de 20 mètres).



Annexe 14 Classification supervisée de niveau 3 réalisée avec un nombre restreint de données d’entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d’entraînement est de 20 mètres).



Annexe 15 Classification supervisée de niveau 4 réalisée avec un nombre restreint de données d’entraînement (image de gauche) et le double des données (image de droite) (la zone tampon autour des échantillons d’entraînement est de 20 mètres).

