

Mathieu Rahm
Consultant international
Expert REDD+/Télédétection

Termes de référence (Code : 311)



Secrétariat Général

Bureau National des Changements Climatiques, du Carbone et de la REDD



Cartographie de l'utilisation et occupation des terres
dans la zone du Programme de réduction des
émissions Atiala-Atsinanana (PREAA)
- Rapport final -

Mars 2020

Contact :
Mathieu Rahm

+32 (0) 493964503
mathieu.rahm973@gmail.com

Crédit photo: Sentinel-2 (Zone PREAA)

Rapport final - Livrable 1.3

Table des matières

I 	INTRODUCTION	7
II 	OBJECTIFS	9
III 	MATERIEL ET METHODE.....	12
III.1.	CONSIDERATIONS METHODOLOGIQUES.....	12
III.2.	VUE D'ENSEMBLE DES DONNEES UTILISEES.....	13
III.2.1	<i>Données satellitaires</i>	<i>13</i>
III.2.1.1	Les données Sentinel-2.....	13
III.2.1.2	Les données SRTM.....	15
III.2.2	<i>Données ancillaires.....</i>	<i>16</i>
III.3.	METHODE DE CARTOGRAPHIE DE L'UOT.....	17
III.3.1	<i>Etapas de Prétraitement : Préparation des données.....</i>	<i>19</i>
III.3.1.1	Sélection, prétraitement, téléchargement et stockage des données.....	19
III.3.1.2	Séquence de critères de sélection des images de niveau 2A.....	21
III.3.1.3	Variables d'entrée	22
III.3.1.3.1	Calcul de l'indice de végétation NDVI	22
III.3.1.3.2	Préparation des données SRTM.....	23
III.3.1.4	Construire une image multibande	23
III.3.1.5	Géo-référencement.....	25
III.3.1.6	Découper les images selon la grille d'analyse PREAA	25
III.3.1.7	Extraire et appliquer un masque de nuages	26
III.3.2	<i>Etapas de Traitement des données : Classification supervisée.....</i>	<i>28</i>
III.3.2.1	Créer des zones d'entraînement (ROI)	30
III.3.2.2	Calculer les statistiques de l'image	31
III.3.2.3	Entraîner le classificateur d'images	31
III.3.2.4	Créer une classification d'image	33
III.3.2.5	Créer une mosaïque de classification	33
III.3.3	<i>Etapas de Post-traitement des données : Vérification et amélioration de la classification</i>	<i>34</i>
III.3.3.1	Tamiser les résultats de classification.....	34
III.3.3.2	Ajustement des classifications.....	34
III.3.3.3	Remplissage des zones de « no data » résiduelles	35
III.4.	METHODE DE CONTROLE ET D'EVALUATION DE LA QUALITE	35
III.4.1	<i>Finalisation de la carte</i>	<i>35</i>
III.4.2	<i>Plan d'échantillonnage.....</i>	<i>36</i>
III.4.3	<i>Conception de la réponse</i>	<i>37</i>
III.4.4	<i>Protocole d'analyse</i>	<i>38</i>
IV 	RESULTATS.....	41
IV.1.	SPECIFICITES TECHNIQUES	41
IV.1.1	<i>Classes UOT retenues</i>	<i>41</i>
IV.1.2	<i>Données d'entrée</i>	<i>41</i>
IV.1.3	<i>Unité cartographique minimale</i>	<i>42</i>
IV.2.	CARTOGRAPHIE DES UOT A L'ECHELLE DU PREAA	42
IV.2.1	<i>Carte de la zone PREAA</i>	<i>42</i>
IV.2.2	<i>Statistiques et évaluation de l'exactitude de la carte</i>	<i>43</i>
IV.2.2.1	Plan d'échantillonnage et matrice de confusion	43
IV.2.2.2	Indicateurs statistiques de la qualité des résultats.....	44
IV.2.2.3	Mesure et estimation de la couverture des terres	45
IV.3.	CARTOGRAPHIE DES UOT A L'ECHELLE REGIONALE.....	46
IV.3.1	<i>Atlas cartographique régional.....</i>	<i>46</i>
IV.3.2	<i>Couverture des terres par région.....</i>	<i>46</i>
IV.4.	CARTOGRAPHIE DES UOT A L'ECHELLE DES DISTRICTS	48
IV.4.1	<i>Atlas cartographique des districts.....</i>	<i>48</i>
IV.5.	CARTOGRAPHIE DES UOT A L'ECHELLE COMMUNALE.....	49
IV.5.1	<i>Atlas cartographique communal.....</i>	<i>49</i>
V 	CONCLUSION	50

VI ANNEXES	52
ANNEXE 1 : RESULTATS DES TESTS REALISES SUR L'INTEGRATION DE LA BANDE SRTM COMME VARIABLE D'ENTREE	52
ANNEXE 2 : EXTRAIT DU CATALOGUE D'IMAGES SENTINEL-2 ET OUTIL DE SUIVI DE LA PRODUCTION	53
ANNEXE 3 : DEFINITION DU PLAN D'ECHANTILLONNAGE POUR L'ANALYSE DE L'EXACTITUDE DES RESULTATS	54
REFERENCES	55

Liste des figures

FIGURE 1: CARTES DE LOCALISATION DE LA ZONE DU PROGRAMME DE REDUCTION DES EMISSIONS ATIALA ATSIANANA (PREAA)	9
FIGURE 2: COMPARAISON DES BANDES SPECTRALES DE LANDSAT 7 ET 8 AVEC SENTINEL-2	15
FIGURE 3: COMPARAISON DE LA RESOLUTION SPATIALE ENTRE CLICHES ISSUS DE LA PLATEFORME GOOGLE EARTH, D'UNE IMAGE SENTINEL-2 ET D'UNE IMAGE LANDSAT 8 SUR LA MEME ZONE AU SEIN DU PREAA.....	15
FIGURE 4: MOSAÏQUE SRTM PRODUITE POUR L'ILE DE MADAGASCAR	16
FIGURE 5: RELEVES DE DONNEES D'ENTRAINEMENT AU SEIN DE LA ZONE PREAA (POINTS GPS, PHOTOS ET INFORMATION SUR L'UOT) POUR LA MISSION CARTOGRAPHIQUE A PARTIR DE L'APPLICATION MOBILE QFIELD	17
FIGURE 6 : ILLUSTRATION DE L'INTERFACE QGIS ET DE SA BOITE A OUTILS DE TRAITEMENT INTEGRANT LES FONCTIONNALITES DE PLUSIEURS LOGICIELS LIBRES	18
FIGURE 7 : DIAGRAMME DES ETAPES DE LA CHAINE DE TRAITEMENT SENTINEL-2 POUR LA PRODUCTION DE LA CARTE D'UOT DANS LA ZONE PREAA.....	19
FIGURE 8 : SUPERPOSITION DE LA ZONE PREAA (EN VERT) PAR LA GRILLE MONDIALE D'ACQUISITION DES TUILES SENTINEL-2 (EN JAUNE).....	20
FIGURE 9 : COMPOSITIONS COLOREES DIFFERENTES PRODUITES A PARTIR D'UNE IMAGE MULTIBANDE SENTINEL-2 DE LA ZONE PREAA. A : COMPOSITION COLOREE VRAIE COULEUR (R =R ;V =V ;B =B) ; B : COMPOSITION FAUSSE COULEUR INFRAROUGE 1 (R =IR, V =R ;B =V) ; C : COMPOSITION FAUSSE COULEUR MOYEN INFRAROUGE (R=SWIR, V=IR, B=R) ; D : COMPOSITION FAUSSE COULEUR INFRAROUGE 2 (R=IR, V=SWIR, B=SWIR)	24
FIGURE 10 : DECOUPAGE DES IMAGES SENTINEL-2 SUIVANT L'EMPRISE DE LA ZONE D'ETUDE PREAA, ELLE-MEME DECOUPEE EN 18 CELLULES CORRESPONDANT A LA GRILLE MONDIALE D'ACQUISITION DES DONNEES SENTINEL-2	26
FIGURE 11 : ILLUSTRATION DE LA CONSTRUCTION DU MASQUE DES NUAGES TEMPOREL A PARTIR D'UNE SERIE TEMPORELLE DE DEUX IMAGES SENTINEL-2	27
FIGURE 12 : RELATION ENTRE OBJETS CONSIDERES ET RESOLUTION SPATIALE: (A) MOYENNE RESOLUTION: PIXELS NETTEMENT PLUS GRANDS QUE LES OBJETS. (B) HAUTE RESOLUTION: LA TAILLE DES PIXELS ET DES OBJETS EST DU MEME ORDRE, LES TECHNIQUES ORIENTEES PIXEL SONT APPROPRIEES. (C) TRES HAUTE RESOLUTION: LES PIXELS SONT SIGNIFICATIVEMENT PLUS PETITS QUE LES OBJETS, L'AGREGATION DE PIXELS EN OBJETS EST RECOMMANDEE.	29
FIGURE 13 : EXEMPLE DE ZONES D'ENTRAINEMENT (ROI) NUMERISEES SUR UNE IMAGE SENTINEL-2 DE LA ZONE PREAA POUR LES BESOINS DE CLASSIFICATION DE L'UOT	31
FIGURE 14 : EXTRAIT D'UNE IMAGE SENTINEL-2 DE LA ZONE PREAA CLASSIFIEE A PARTIR DE L'ALGORITHME RANDOM FOREST	33
FIGURE 15 : ILLUSTRATION DE L'EFFET DE « TAMIS » SUR LES RESULTATS DE CLASSIFICATION	34
FIGURE 16 : MATRICE DE CONFUSION ISSUE DE L'EVALUATION DE L'EXACTITUDE DES RESULTATS CARTOGRAPHIQUES SUR LA ZONE PREAA.....	39
FIGURE 17 : CARTE DES UOT SUR LA ZONE PREAA.....	42
FIGURE 18 : REPARTITION DES ECHANTILLONS DE POINTS DESTINES A L'EVALUATION DE LA QUALITE DE LA CARTE PREAA	43
FIGURE 19 : DISTRIBUTION DES CLASSES UOT AU SEIN DE LA ZONE PREAA (SURFACES MESUREES A PARTIR DE LA CARTE ET SURFACES « REELLES », ESTIMEES STATISTIQUEMENT).....	45
FIGURE 20 : DISTRIBUTION DES CLASSES UOT AU SEIN DES 5 REGIONS DE LA ZONE PREAA (SURFACES MESUREES A PARTIR DE LA CARTE).....	47
FIGURE 21 : PHOTOS DE GROUPE DES ATELIERS DE VALIDATION DE LA CARTE UOT DU PREAA A L'ECHELLE REGIONALE (A GAUCHE) ET NATIONALE (A DROITE) REALISES RESPECTIVEMENT A TAMATAVE ET ANKERANA EN JANVIER 2020	51
FIGURE 22 : COMPARAISON DES RESULTATS DES TESTS REALISES SUR LA PERFORMANCES GLOBALE (INDEX KAPPA) DE L'ALGORITHME RANDOM FOREST, PRODUIT A PARTIR D'UNE TUILE MULTISPECTRALE SENTINEL-2 AVEC ET SANS L'AJOUT DES DONNEES SRTM 52	

Liste des tableaux

TABLEAU 1 : CLASSES UOT DEFINIES POUR MADAGASCAR (SOURCE : HTTPS://WWW.BNC-REDD.MG/IMAGES/DOCUMENTS/MNV/LIV7_UOT_30_05_2018.PDF)	10
TABLEAU 2 : CONTENU, PROPRIETES ET OBJECTIFS DES BANDES SPECTRALES SENTINEL-2 (SOURCE : EUROPEAN SPACE AGENCY, 2015)	14
TABLEAU 3 : COMPOSITION DE L'IMAGE MULTIBANDE SENTINEL-2 PRODUITE POUR LES BESOINS DE CARTOGRAPHIE DE L'UOT DE LA ZONE PREAA	24
TABLEAU 4 : CLASSES UOT RETENUES POUR LES TRAVAUX DE CARTOGRAPHIE DE LA ZONE PREAA	41
TABLEAU 5 : LISTE DES IMAGES SENTINEL-2 UTILISEES POUR LA CARTOGRAPHIE DE LA ZONE PREAA (NOMENCLATURE : CAPTEUR_TUILE_DATE D'ACQUISITION).....	42
TABLEAU 7 : MATRICE DE CONFUSION, COMPARANT, EN LIGNE, LES RESULTATS CARTOGRAPHIQUES ET, EN COLONNE, LES RESULTATS DE L'ANALYSE DES 1866 ECHANTILLONS PAR PHOTO-INTERPRETATION	43
TABLEAU 8 : PRECISION (« PRECISION » EN ANGLAIS), RAPPEL (« RECALL » EN ANGLAIS) ET F-SCORE MESURES POUR LES DIFFERENTES CLASSES UOT CARTOGRAPHIEES SUR LA ZONE PREAA	44
TABLEAU 9 : PRECISION (« PRECISION » EN ANGLAIS), RAPPEL (« RECALL » EN ANGLAIS) ET F-SCORE MESURES POUR LES DIFFERENTES CLASSES UOT CARTOGRAPHIEES SUR LA ZONE PREAA	44
TABLEAU 10 : MESURE CARTOGRAPHIQUE DES SURFACES ET ESTIMATION STATISTIQUE DES SURFACES « REELLES » DES CLASSES UOT DE LA ZONE PREAA.....	46
TABLEAU 11 : SURFACES MESUREES ET COUVERTURE DES TERRES POUR LES CINQ REGIONS COUVRANT LA ZONE PREAA	47
TABLEAU 12 : REPARTITION PAR REGION DES CLASSES UOT DE LA ZONE PREAA.....	48

Liste des équations

ÉQUATION 1 : CALCUL DU NDVI (ROUSE ET AL., 1973). OU NIR CORRESPOND A LA BANDE PROCHE INFRAROUGE ET R A LA BANDE ROUGE	23
ÉQUATION 2 : CALCUL DE LA TAILLE DE L'ECHANTILLON GLOBAL REPRESENTATIF POUR LA REALISATION DU PLAN D'ECHANTILLONNAGE STRATIFIE (COCHRAN, 1977). OU, N EST LE NOMBRE D'UNITES DANS LA ZONE D'INTERET (NOMBRE DE PIXELS GLOBAUX SI L'UNITE SPATIALE EST UN PIXEL), $S(\hat{O})$ EST L'ERREUR TYPE DE L'EXACTITUDE GLOBALE ESTIMEE QUE NOUS SOUHAITONS OBTENIR, W_i EST LA PROPORTION DE SURFACE CARTOGRAPHIEE POUR LA CLASSE i , ET S_i EST L'ECART TYPE DE LA CLASSE i	36
ÉQUATION 3 : CALCUL DE P.K POUR L'AJUSTEMENT DE L'ESTIMATION DE LA SURFACE DES CLASSES OBTENUE A PARTIR DES TRAVAUX CARTOGRAPHIQUES.	39

Acronymes

BN-CCCREDD+	Bureau National des Changements Climatiques, du Carbone et de la REDD+
CLM1	Masque des nuages de résolution 1 (<i>Cloud mask resolution 1</i> en anglais)
CNES	Centre national d'études spatiales français
DEM	Modèle numérique d'élévation (<i>Digital elevation model</i> en anglais)
EROS/ESPA	Earth Resources Observation and Science Center Science Processing Architecture
FCPF	Fonds de partenariat pour le carbone forestier
FRE	Réflectance de surface (<i>Flat REflectance</i> en anglais)
GES	Gaz à effet de serre
LOFM	Laboratoire d'Observation des Forêts de Madagascar
MEDD	Ministère de l'Environnement et du Développement Durable
NASA	Administration nationale de l'aéronautique et de l'espace américaine (National Aeronautics and Space Administration)
NDVI	Indice de végétation par différence normalisé (<i>Normalized Difference Vegetation Index</i>)
NGA	Agence nationale d'intelligence géospatiale américaine (National Geospatial-Intelligence Agency)
NIMA	Agence nationale d'imagerie et de cartographie américaine (National Imagery and Mapping Agency)
NIR	Proche infrarouge (<i>Near Infrared</i>)
OOB	Out-Of-the-Bag error
PA	Exactitude du producteur (<i>Producer Accuracy</i>)
PEPS	Plateforme d'Exploitation des Produits Sentinel
PRE AA	Programme de Réduction des Émissions dans les régions de Atiala – Atsinanana
REDD+	Réduction des émissions provenant du déboisement et de la dégradation des forêts, associées à la gestion durable des forêts, la conservation et l'amélioration des stocks de carbone forestier
RF	Random Forest
RGB	Rouge, vert et bleu (<i>Red, Green, Blue</i>)
RMSE	Erreur quadratique moyenne (<i>Root Mean Square Error</i>)
ROI	Zone d'entraînement (<i>Region of Interest</i>)
SEPAL	System for Earth observations, data access, Processing & Analysis for Land monitoring
SSTS	Système de Surveillance des Terres par Satellite

SVM	Support Vector Machine
UA	Exactitude de l'utilisateur (<i>User Accuracy</i>)
UOT	Utilisations et Occupations des Terres
UTM	Universal Transverse Mercator
WGS	World Geodetic System

I | Introduction

La réduction des émissions dues à la déforestation et la dégradation des forêts (REDD+) est un mécanisme qui vise simultanément l'atténuation des effets du changement climatique, la conservation des forêts et le développement durable. Le principe de la REDD+ est d'octroyer une compensation financière à des pays forestiers qui réduisent les émissions de gaz à effet de serre (GES) issues du secteur forestier. Ces émissions ou absorptions peuvent être issues de cinq activités recensées par les Accords de Cancún : a) la réduction des émissions dues à la déforestation ; b) la réduction des émissions dues à la dégradation des forêts ; c) la conservation des stocks de carbone forestiers ; d) la gestion durable des forêts ; et e) le renforcement des stocks de carbone forestiers.

Dans ce cadre, le Bureau National des Changements Climatiques, du Carbone et de la REDD+ (BN-CCCREDD+) a été créé au sein du Ministère de l'Environnement et du Développement Durable (MEDD) dans l'objectif de piloter et de coordonner toutes les initiatives sur la REDD+ ainsi que les projets carbonés forestiers en général. Dans le cadre de la mise en place de la REDD+ à Madagascar, le BN-CCCREDD+ est en train de développer un Programme de Réduction des Émissions dans les régions de Atiala – Atsinanana (PREAA).

Le BN-CCCREDD+ intègre trois composantes principales relatives à la REDD+ :

- Composante 1 : Dispositions institutionnelles pour la REDD+ ;
- Composante 2 : Développement de la stratégie nationale REDD+ intégrant les consultations et l'Evaluation Environnementale Stratégique et Sociale ;
- Composante 3 : Développement du scénario de référence national et du système de suivi de la REDD+.

En lien avec la composante 3 et dans le cadre de la mise en œuvre du PREAA, le BN-CCCREDD+ s'attèle au développement d'un Système de Surveillance des Terres par Satellite (SSTS). Ce système, hébergé au sein du Laboratoire d'Observation des Forêts de Madagascar (LOFM), a pour objectif de produire des données cartographiques et statistiques nécessaires à la surveillance des forêts et des changements des terres, incluant l'évolution des émissions de gaz à effet de serre du secteur forestier.

Dans ce cadre, une cartographie de référence des Utilisations et Occupations des Terres (UOT) a été produite dans la zone du PREAA par les techniciens du LOFM, avec l'assistance du consultant international recruté pour former et guider les équipes à travers les étapes de production.

Les cartes et informations générées serviront d’outil de base nécessaire à la planification des activités à entreprendre dans la mise en œuvre du PREAA. Ces données sont également cruciales pour la mise en œuvre de la REDD+ dans cette zone car elle permettra notamment la mise en place de Schémas d’aménagement, qui est une voie vers la REDD+.

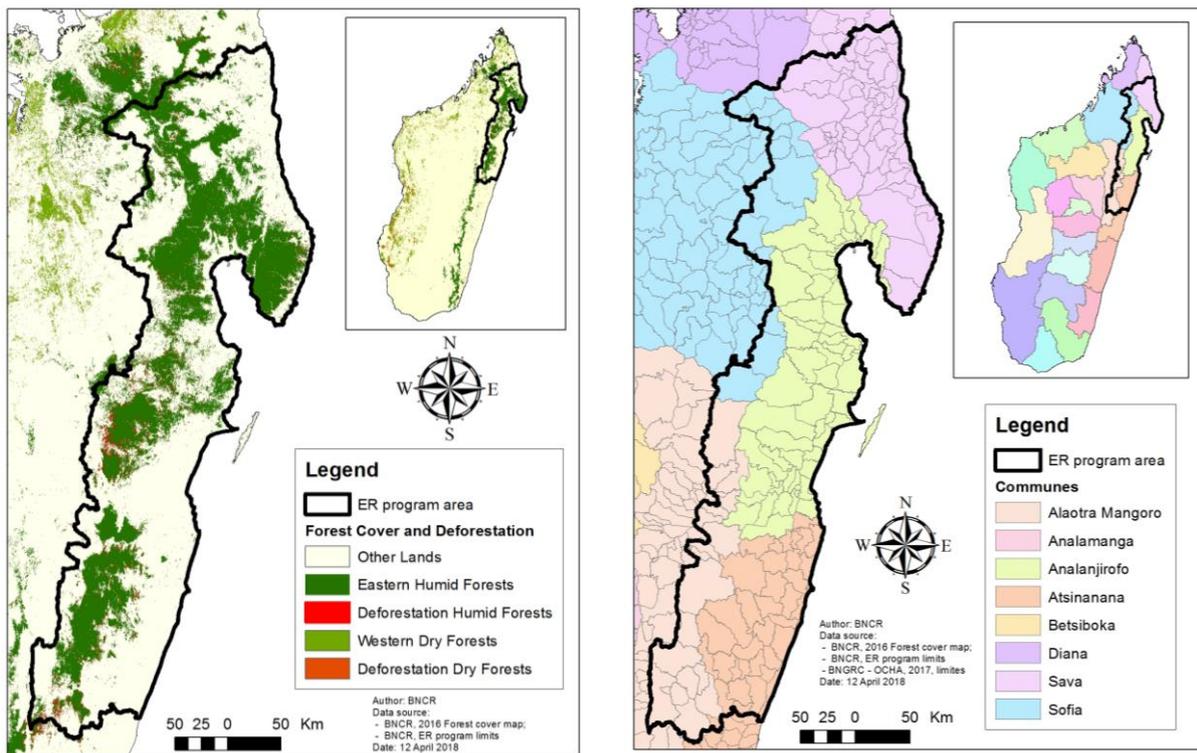
En effet, dans les communes qui se trouvent dans la zone du PREAA, l’occupation et l’utilisation des terres représente une information cruciale pour l’aménagement de l’espace, la gestion durable et la valorisation des ressources naturelles pour guider les processus de prise de décision. La couverture terrestre est l’expression des activités humaines. Par conséquent, l’occupation des terres est une information qui constitue une connaissance de base pour des applications allant du suivi des forêts, la production de statistiques, la planification, la biodiversité, le changement climatique, à la lutte contre la dégradation des terres et la désertification.

L’objectif de ce rapport final est de présenter la méthodologie et les résultats de cartographie par télédétection de l’utilisation et occupation des terres au sein de la zone du PREAA. Un livrable précédent intermédiaire, réalisé sous la forme d’un tutoriel technique, détaille les outils utilisés (logiciel, fonctionnalités, algorithmes, etc.) et permet de répliquer l’ensemble des étapes méthodologiques décrites dans ce document. Ce dernier est accompagné d’un atlas cartographique rassemblant l’ensemble des cartes produites et leurs statistiques à l’échelle globale de la zone du PREAA, ainsi qu’à l’échelle des régions, districts et communes. Il est également complété par un rapport annexe qui détaille les activités de renforcement et transfert de capacités à destination des agents du MEDD réalisée dans le cadre d’un atelier de trois jours à Antsirabe en janvier 2020.

II | Objectifs

L'objectif de l'étude est de réaliser la cartographie des classes d'utilisation et d'occupation des terres (UOT) présentes dans la zone du Programme de réduction des émissions Atiala Atsinanana (PREAA), approuvé par le Fonds de partenariat pour le carbone forestier (FCPF). Le PREAA se situe dans la région Est de Madagascar, caractérisée par un climat tropical et la présence de forêts denses humides. Elle couvre partiellement cinq régions du pays (Figure 1), à savoir :

- La région de Sava
- La région de Sofia
- La région d'Analanjirifo
- La région Atsinanana
- La région Alaotra Mangoro



Le niveau de détail fourni par la carte doit être le plus élevé possible, tant au niveau spatial que thématique, compte tenu des limitations liées au coût du système ainsi qu'à l'accès aux données et outils de télédétection optique par satellite. Au niveau thématique, au-delà des contraintes liées

aux données et outils, l'objectif est de respecter le système de classification national des UOT adopté en mai 2018 par Madagascar¹. Ce système, qui a pour but de standardiser les légendes et d'assurer la comparabilité entre les cartes produites sur le territoire, fournit une définition exhaustive et complète de l'ensemble des classes UOT présentes dans le pays. Les classes, définies par plusieurs critères (aspect spatial, stratification, altitude), sont catégorisées selon trois niveaux de détail agrégés, du moins détaillé au plus détaillé (Tableau 1) :

- Niveau 1 (6 Classes)
- Niveau 2 (30 classes)
- Niveau 3 (42 classes)

#	NIVEAU 1	NIVEAU 2	NIVEAU 3
1	Forêts	Forêt sempervirente	Forêt humide
2			Forêt humide dégradée
3			Forêt littorale
4			Forêt littorale dégradée
5			Forêt sclérophylle
6			Forêt sclérophylle dégradée
7			Forêt sub-humide
8			Forêt sub-humide dégradée
9			Forêt secondaire ou savoka
10		Forêt caducifoliée	Forêt dense sèche
11			Forêt sèche dégradée
12		Forêt ripicole	
13		Forêt galerie	
14		Raphière et bamboueraie	
15		Mangroves	Mangroves denses
16			Mangroves denses dégradées
17			Mangroves éparses
18		Fourrés xérophiles	
19		Plantation	Plantation monospécifique
20			Plantation mixte
21		Agroforesterie	
22	Savanes	Savane arborée	
23		Savane arbustive	
24		Savane arborée et arbustive	
25		Savane herbeuse	
26		Zones arbustives	
27		Zones arbustives arborées	
28		Formations rupicoles	
29	Zones cultivées	Culture irriguée	
30		Culture pluviale	
31		Tavy	
32		Jachère	
33	Surface bâtie	Zones habitées	
34		Infrastructure	
35	Zones nues	Zones rocheuse	
36		Zones sableuses	
37		Karst	
38		Autres zones nues	
39	Zones humides	Rivière	
40		Delta	
41		Lac	
42		Marais	

Tableau 1 : Classes UOT définies pour Madagascar (Source : https://www.bnc-redd.mg/images/documents/MNV/Liv7_UOT_30_05_2018.pdf)

¹ https://www.bnc-redd.mg/images/documents/MNV/Liv7_UOT_30_05_2018.pdf

Au-delà des objectifs de production cartographique, cette étude consiste également à former et renforcer les capacités des agents du MEDD afin qu'ils soient en mesure de mettre en œuvre et de répliquer dans le futur l'ensemble des étapes de la chaîne de traitement. L'appropriation des outils et méthodes par les agents locaux est donc particulièrement importante car elle conditionne le succès de la mise en place d'un système pérenne et robuste. Par conséquent, cette étude fait partie intégrante d'un processus d'apprentissage basé sur la méthode du « Learn by doing », où les agents du LOFM sont en charge de la production, orientés et assistés par le consultant international. Pour garantir le succès de l'étude et la viabilité économique du système, l'utilisation d'un nombre limité d'outils accessibles, intuitifs et à moindre coût est ainsi privilégiée.

III | Matériel et méthode

Ce chapitre décrit les données et les méthodes utilisées pour la production cartographique de l'UOT sur la zone du PREAA, ainsi que les étapes de contrôle qualité et d'évaluation de l'exactitude des résultats.

III.1. Considérations méthodologiques

Avant de décrire les étapes méthodologiques, il est important de rappeler certaines limites et contraintes auquel nous sommes confrontés dans le cadre de ces travaux cartographiques basés sur l'utilisation d'algorithmes de classification d'images. En effet, l'analyse d'images peut nous donner des informations précieuses sur l'occupation des terres à un instant T, sans pour autant nous renseigner sur l'utilisation des terres. A titre d'exemple, une terre cultivée peut passer par plusieurs types d'occupation des terres tout en conservant une utilisation agricole. En effet, dans le cas d'une rizière par exemple, l'occupation des terres peut varier selon le stade de culture et la période de l'année, en passant d'une phase de sol nu avant le semis, à une phase végétative, ou encore par la suite à une phase inondée semblable à une zone marécageuse ou de l'eau. En absence de données cartographiques ancillaires, de données terrain ou encore de photo-interprétation manuelle en post-traitement, fournir le détail sur l'utilisation des terres peut s'avérer délicat voire impossible. Typiquement, en prenant l'exemple d'une zone minière définie par la classification UOT et principalement caractérisée par une occupation du sol de type « sol nu », l'analyse automatisée d'images ne nous permettra pas de détecter une activité minière sur ces zones et ainsi de discriminer ces zones de sol nu par rapport aux autres. En effet, la réponse spectrale de l'élément « sol nu » analysé par l'algorithme de classification ne sera probablement pas différente qu'il s'agisse d'une zone minière, d'une zone rocheuse naturelle ou encore d'un terrain déboisé.

De la même façon nous ne confondrons pas l'utilisation foncière des terres et leur nature en termes d'occupation des terres. Il est important de préciser que la carte d'occupation des terres ne permettra pas de cartographier le foncier, mais que par ailleurs il sera possible de superposer ces deux informations à conditions que celles-ci soient disponibles aux dates prévues par l'étude avec toute l'exhaustivité nécessaire pour fiabiliser la prise de décision. En conséquence, en absence de données ancillaires fiables et suffisantes de quelque nature que ce soit fournissant de l'information sur l'utilisation des terres, on retiendra les classes "d'occupation des terres" et non "d'utilisation des terres". La définition de l'utilisation des terres doit donc se faire généralement de manière manuelle à partir de photo-interprétation visuelle lors des étapes de post-traitement.

Par ailleurs, la définition des classes de forêts considère différentes hauteurs d'arbres et différents niveaux de couvertures arborées. La variation de ces paramètres d'une classe à l'autre peut s'avérer compliqué à gérer d'un point de vue opérationnel pour les travaux de cartographie à partir d'images à résolution moyenne, et lorsque l'on ne dispose pas de modèle numérique de terrain et de surface fiables ; le signal capté par le satellite étant intégré verticalement, le paramètre de hauteur est soit ignoré, soit relégué à une carte de référence produite par un tiers. Par conséquent, les images satellites optiques de type Landsat ou Sentinel-2 ne nous permettent pas de discriminer une classe sur base d'un paramètre de hauteur, sauf si la variation de ce paramètre engendre une variation de la réponse spectrale pouvant être détectée par l'algorithme de classification. De plus, certaines classes peuvent présenter des caractéristiques spectrales et texturales similaires limitant leur discrimination avec un niveau de confiance suffisant.

En conclusion, il est donc important lors de la réalisation et de l'analyse des travaux cartographiques de conserver à l'esprit les limites liées aux données disponibles (données ancillaires fiables, résolution spatiale et temporelle...) et aux méthodes de télédétection automatisées (algorithmes de classification basés sur l'analyse des propriétés spectrales des images à un temps T).

III.2. Vue d'ensemble des données utilisées

III.2.1 Données satellitaires

Au regard des objectifs précités et dans le but de s'approcher au mieux du niveau le plus élevé du système de classification UOT défini pour Madagascar (niveau 3), il est recommandé de baser l'analyse cartographique sur l'utilisation de données satellites optiques récentes et gratuites, proposant la plus haute résolution spatiale et spectrale. L'utilisation des données Sentinel-2 possédant des bandes spectrales entre 10 et 60m de résolution spatiale est par conséquent priorisée.

III.2.1.1 Les données Sentinel-2

Héritière des programmes Landsat de l'Administration nationale de l'aéronautique et de l'espace américaine (NASA) et SPOT (Satellite pour l'Observation de la Terre) du CNES, la mission Sentinel-2 fait partie du programme d'observation et de surveillance de la Terre, Copernicus, conduit par l'Union Européenne. L'Agence spatiale européenne (ESA) est responsable du développement des deux satellites jumeaux Sentinel-2A et 2B, de leurs instruments et du segment sol.

Tous les 5 jours environ, la mission Sentinel-2 est capable de fournir des clichés de notre planète, avec l'objectif de suivre à haute résolution l'évolution de la végétation, de l'occupation des terres et l'impact du réchauffement climatique avec une haute fréquence d'observation. L'imageur multi-spectral MSI de Sentinel-2 effectue des prises d'images dans treize bandes spectrales allant du

visible au moyen infrarouge sur une largeur de 290 km. Le Tableau 2 détaille le contenu des bandes Sentinel-2. Les trois bandes visibles (bandes 2 à 4) et la bande 8 dans le proche infrarouge sont fournies avec une résolution de 10 m. Les bandes 1, 9 et 10 sont destinées aux corrections atmosphériques avec une résolution de 60 m tandis que dans les six bandes restantes, correspondantes au Red Edge (bandes 5 à 7) et au moyen infrarouge (bandes 11 et 12), la résolution est de 20 m.

Sentinel-2 bands	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)	Objective
Band 1 – Coastal aerosol	443	21	60	Aerosol correction
Band 2 – Blue	490	66	10	Aerosol correction, Land measurement
Band 3 – Green	560	36	10	Land measurement
Band 4 – Red	665	31	10	Land measurement
Band 5 – Vegetation red edge	705	16	20	Land measurement
Band 6 – Vegetation red edge	740	15	20	Land measurement
Band 7 – Vegetation red edge	783	20	20	Land measurement
Band 8 – NIR	842	106	10	Water vapor correction, Land measurement
Band 8A – Narrow NIR	865	22	20	Water vapor correction, Land measurement
Band 9 – Water vapour	940	21	60	Water vapor correction
Band 10 – SWIR – Cirrus	1375	30	60	Cirrus detection
Band 11 – SWIR	1610	94	20	Land measurement
Band 12 – SWIR	2190	185	20	Aerosol correction, Land measurement

Tableau 2 : Contenu, propriétés et objectifs des bandes spectrales Sentinel-2 (Source : European Space Agency, 2015)

En comparaison, les capteurs du satellite Landsat-8 produisent des données dans 11 bandes spectrales. Neuf bandes spectrales sont acquises par le radiomètre multispectral OLI (Operational Land Imager) allant du visible au moyen infra-rouge à 30 m de résolution, à l'exception de la bande panchromatique 8 à 15 m de résolution (Figure 2). À titre d'illustration concernant la résolution spatiale des données, la Figure 3 fournit une comparaison de données à très haute résolution extraites de la plateforme Google Earth et à haute et moyenne résolution fournies par les capteurs Sentinel-2 et Landsat-8 sur la même zone au sein du PREAA.

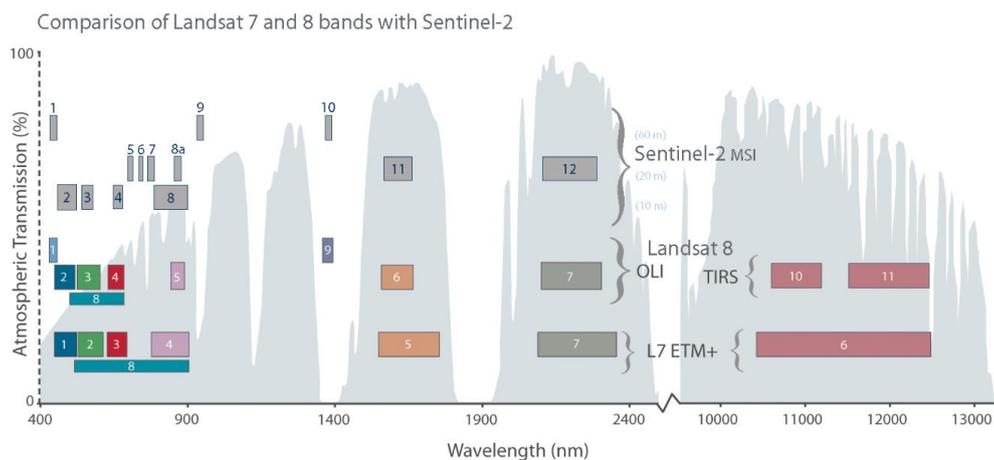


Figure 2: Comparaison des bandes spectrales de Landsat 7 et 8 avec Sentinel-2

(Source: <https://landsat.gsfc.nasa.gov/wp-content/uploads/2015/06/Landsat.v.Sentinel-2.png>)

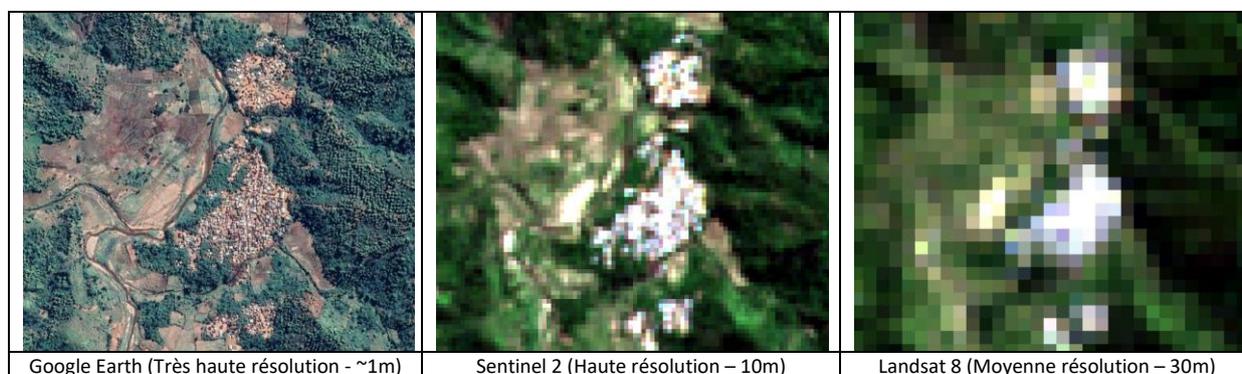


Figure 3: Comparaison de la résolution spatiale entre clichés issus de la plateforme Google Earth, d’une image Sentinel-2 et d’une image Landsat 8 sur la même zone au sein du PREAA

III.2.1.2 Les données SRTM

Shuttle Radar Topography Mission (SRTM) fait référence à des fichiers matriciels et vectoriels topographiques produits conjointement par deux agences spatiales américaines, à savoir la NASA et l’Agence nationale d’intelligence géospatiale (NGA), anciennement Agence nationale d’imagerie et de cartographie (NIMA). Ces données altimétriques ont été recueillies au cours d’une mission de onze jours en février 2000 par la navette spatiale Endeavour (STS-99) à une altitude de 233 km en utilisant l’interférométrie radar. Les données d’élévation SRTM 1 arc-seconde global² offrent une couverture mondiale des données remplies des vides à 1 seconde d’arc (30 mètres).

L’île de Madagascar étant caractérisée par un relief contrasté, la plus-value de l’intégration de ce Modèle numérique d’élévation (DEM) dans la chaîne de traitement en améliorant les performances de l’algorithme de classification a été testée et validée lors de la phase de production (voir Annexe 1).

² https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects

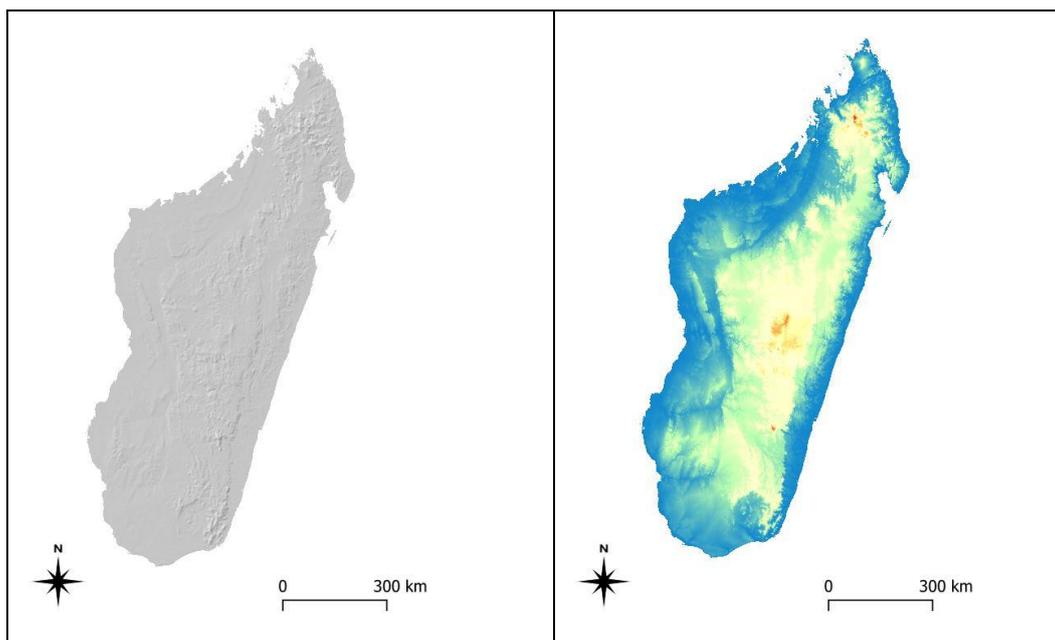


Figure 4: Mosaïque SRTM produite pour l'île de Madagascar

III.2.2 Données ancillaires

Les données ancillaires sont des données préexistantes à toute mission cartographique, nécessaires pour l'étalonnage, le traitement ou l'analyse des données de la mission. Il s'agit notamment de travaux cartographiques préexistants sur la zone, de données de terrain, de données issues de plateformes globales telles que Google Earth, Open street map, etc. L'ensemble de ces données présentent sur la zone PREAA et pouvant s'avérer utile au travail d'analyse ont été collectées lors de la phase de préparation du projet.

Trois visites de terrain ont notamment été réalisées au sein de la zone PREAA afin de confronter les images Sentinel-2 et des produits de classification intermédiaires à la vérité terrain. L'objectif de ces missions était de renforcer les connaissances des opérateurs sur les classes UOT de la zone du PREAA, de renforcer leurs capacités en relevés de données terrain à partir d'outils numériques mobiles et de calibrer la photo-interprétation de chacun afin d'assurer l'homogénéité des résultats de classification. La Figure 5 ci-dessous illustre les outils et types d'informations relevés permettant d'entraîner au mieux l'algorithme de classification. Ces relevés ont été effectués à l'aide de l'application mobile QFIELD compatible avec QGIS et tout appareil Android. A partir d'un projet préconfiguré, QFIELD permet l'affichage des couches d'informations désirées (images satellites, fonds google, etc.), la prise de points GPS liés à un formulaire de saisie d'information (classe UOT, commentaires sur la zone, etc.), ainsi que la prise de photos.



Figure 5: Relevés de données d'entraînement au sein de la zone PREAA (points GPS, photos et information sur l'UOT) pour la mission cartographique à partir de l'application mobile QFIELD

III.3. Méthode de cartographie de l'UOT

La chaîne de traitement et les éléments techniques proposés ci-dessous respectent les recommandations méthodologiques REDD+ et visent à proposer les outils les plus appropriés au regard des cadres méthodologiques du Fonds de partenariat pour le carbone forestier (FCPF) et du Voluntary Carbon Standard (VCS). Dans le but de faciliter et d'assurer au mieux le transfert des compétences et l'appropriation des méthodes par les agents locaux, la méthodologie privilégie l'utilisation d'un nombre limité d'outils, accessibles à tous et proposant des interfaces intuitives que l'on peut qualifier comme étant « user friendly ». De plus, afin d'assurer la pérennité du système et d'optimiser son coût, les données et outils proposés sont gratuits, open source et pérennes (l'accès aux outils étant ainsi assuré dans le temps). Sur cette base, il a été décidé de réaliser l'entièreté de la chaîne de traitement au sein du logiciel libre QGIS. Ce logiciel, basé sur la communauté open source, intègre toutes les fonctionnalités d'un système d'information géographique (SIG) nécessaire à la mise en œuvre de la chaîne de traitement et présente l'avantage d'intégrer un grand nombre d'outils d'analyse et de traitement d'images. Outre ses outils intégrés, il permet également à travers une interface harmonisée d'accéder aux fonctionnalités d'autres logiciels libres tels que GRASS, SAGA, R, Monteverdi Orfeo toolbox (OTB)... (Figure 6).

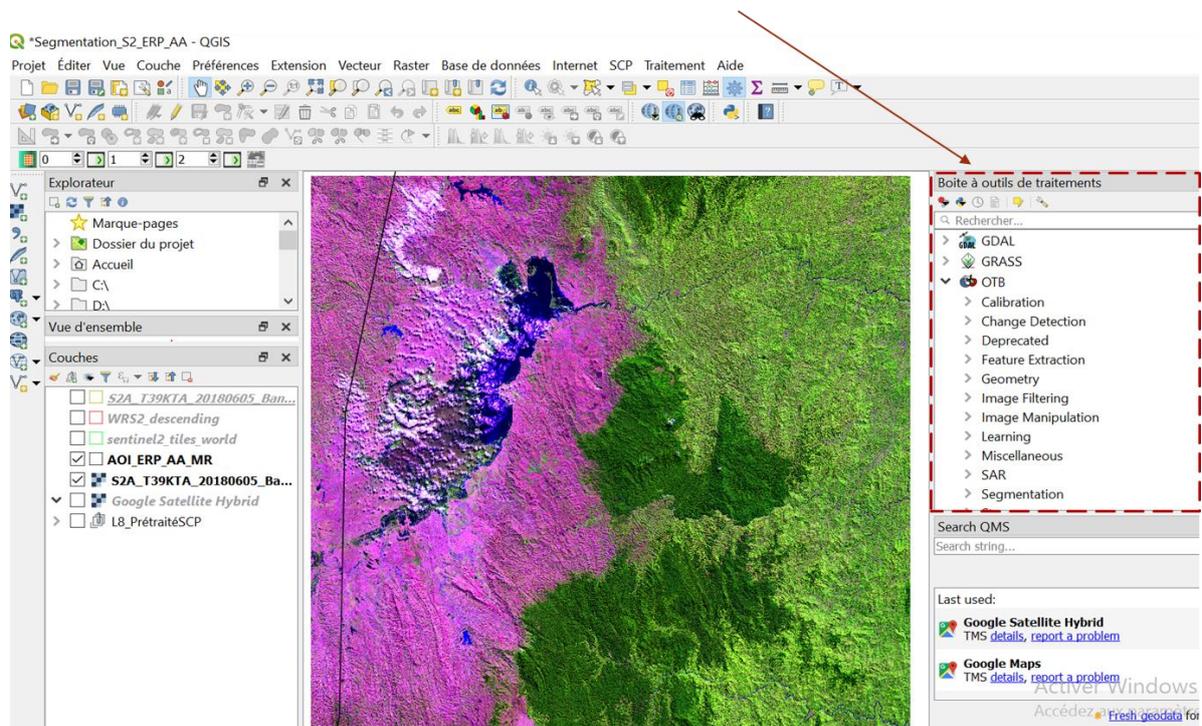


Figure 6 : Illustration de l'interface QGIS et de sa boîte à outils de traitement intégrant les fonctionnalités de plusieurs logiciels libres

Les principales étapes de la chaîne de traitement pour la réalisation de la cartographie de l'occupation des terres sur la zone PREAA sont illustrées à la Figure 7. On distingue trois grandes phases :

1. **Le Pré-traitement**, destiné à la préparation et mise en forme des données d'entrée ;
2. **Le Traitement**, à savoir la phase de classification semi-automatique ;
3. **Le Post-traitement**, dédié au « nettoyage » et à l'amélioration des résultats de classification préliminaires

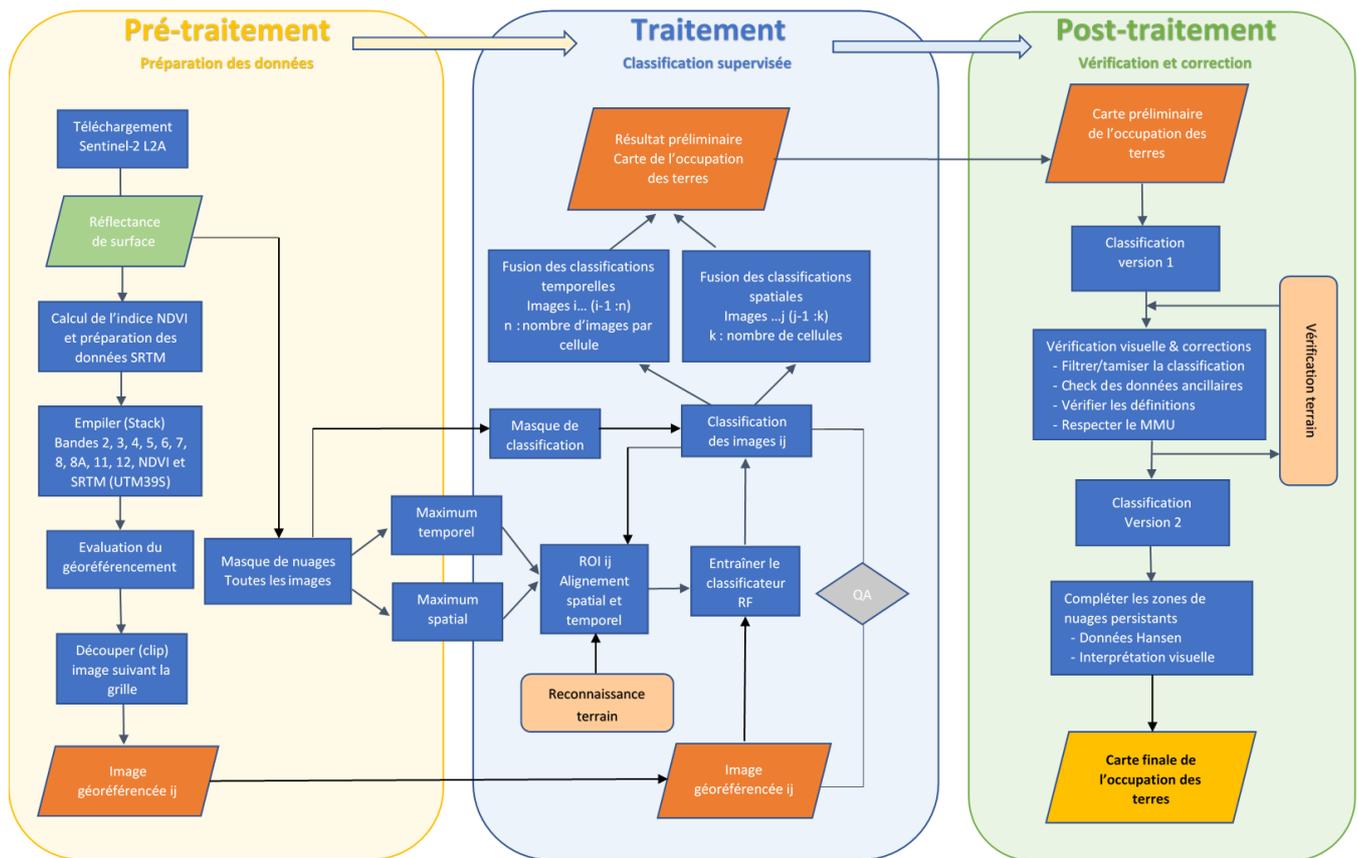


Figure 7 : Diagramme des étapes de la chaîne de traitement Sentinel-2 pour la production de la carte d'UOT dans la zone PREAA

III.3.1 Etapes de Prétraitement : Préparation des données

La phase de prétraitement comprend les étapes suivantes :

1. Sélectionner, télécharger et convertir les données en réflectance de surface
2. Identifier les images prioritaires à traiter
3. Définir et mettre en forme les variables d'entrée (Bandes spectrales, NDVI, SRTM)
4. Construire les images multibandes
5. Vérifier la qualité du géoréférencement des images et améliorer la précision des données si nécessaire
6. Découper les images multibandes selon la grille d'analyse de la zone PREAA
7. Extraire et appliquer un masque des nuages

III.3.1.1 Sélection, prétraitement, téléchargement et stockage des données

La Figure 8 illustre la zone PREAA recouverte par la grille d'acquisition mondiale des tuiles (ou granules) Sentinel-2. On voit que l'utilisation de 18 tuiles Sentinel-2 sont nécessaires pour couvrir l'ensemble de la zone. Chaque tuile correspond au découpage d'une image Sentinel-2 selon une grille

de 110*110 km, autorisant un chevauchement de 10 km entre les tuiles adjacentes. On distingue également deux grilles d'acquisition dont l'orientation est différente. Ceci provient du fait que Madagascar est traversé par deux systèmes de coordonnées de référence (SCR) dans lesquels sont référencés les images Sentinel-2, à savoir le WGS/UTM38S (EPSG : 32738) à l'ouest et le WGS/UTM39S (EPSG : 32739) à l'est. La zone d'étude PREAA étant presque entièrement située dans le SCR WGS/UTM39S, ce dernier a été adopté comme système de coordonnées de référence pour l'étude. Vu que certaines données projetées dans le SCR WGS/UTM38S seront utiles pour couvrir l'extrême sud-est de la zone, celles-ci devront donc être reprojeterées dans le SCR WGS/UTM39S de l'étude.

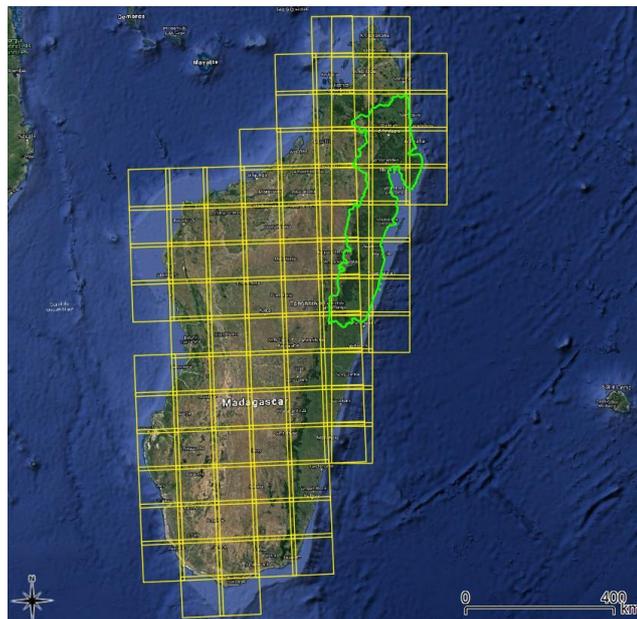


Figure 8 : Superposition de la zone PREAA (en vert) par la grille mondiale d'acquisition des tuiles Sentinel-2 (en jaune)

La première étape consiste à sélectionner et télécharger les images les plus pertinentes pour l'étude. En prenant la période la plus récente comme référence, à savoir 2018-2019, on sélectionne les images possédant le niveau de couverture nuageuse le plus faible. La sélection de tuiles acquises à la même date sont privilégiées pour assurer l'homogénéité de l'information. Dans certaines zones, afin de combler le manque de données dû à la couverture nuageuse sur une date, l'utilisation d'images multi-dates peut s'avérer nécessaire pour compléter l'analyse.

La sélection et le téléchargement des données Sentinel-2 peuvent se faire via plusieurs plateformes telles que le [Copernicus Open Access Hub de l'ESA](#), le [Sentinel hub](#) ou encore la [Plateforme d'Exploitation des Produits Sentinel \(PEPS\)](#) mis à disposition par le Centre national d'études spatiales français (CNES). Dans le cadre de cette étude, la plateforme PEPS du CNES a été privilégiée car elle offre la possibilité d'effectuer des traitements en ligne supplémentaires sur les

données pour obtenir des produits corrigés des effets atmosphériques par l'intermédiaire du [pôle thématique THEIA](#). Chaque tuile Sentinel-2 est ainsi sélectionnée au niveau 1C (donnée ortho-rectifiée, exprimée en réflectance au sommet de l'atmosphère) pour ensuite être traitée en ligne au niveau 2A³. Les images au niveau 2A, converties en réflectance de surface, ont été corrigées des [effets atmosphériques](#), des [effets d'environnement](#) et des [effets des pentes](#). De plus, ce niveau supérieur au niveau 1C fournit plusieurs couches de masque, dont un masque des nuages et des ombres des nuages, utile pour les étapes suivantes de la chaîne de traitement.

La sélection et le téléchargement des données Landsat et SRTM peuvent également se faire gratuitement sur plusieurs plateformes après enregistrement d'un compte utilisateur. Comme pour Sentinel-2, il est possible, via la plateforme de l'[USGS - Earth Resources Observation and Science Center Science Processing Architecture' \(EROS/ESPA\)](#), de traiter les images Landsat au niveau 2 pour obtenir des données converties en [réflectance de surface](#), corrigées des effets atmosphériques. Ce traitement fournit également plusieurs masques dont un masque des nuages et des ombres, et il est également possible d'obtenir des indices de végétation dérivés tel que le NDVI. Néanmoins, l'utilisation des images Landsat n'interviendra qu'en cas de besoin pour combler d'éventuels manques d'information sur les images Sentinel-2 (nuages ou ombres).

Lorsque les images Sentinel-2 sélectionnées ont été traitées au niveau 2A, celles-ci peuvent être directement téléchargées sous format .zip pour intégrer la base de données. Chaque image présente dans la base de données est intégrée dans un catalogue de données organisé sous forme de tableur qui renseigne pour chaque image son identifiant, le lien hypertexte vers les métadonnées, le capteur, l'identifiant géographique, la date d'acquisition, le pourcentage de couverture nuageuse, le niveau de priorité d'utilisation dans la chaîne de traitement. Ce catalogue sert ensuite aux opérateurs comme outil de suivi de production ; En reprenant chronologiquement chaque étape de la chaîne de traitement ainsi que la nomenclature des fichiers en sortie et des dossiers de destination, il représente un outil clé d'assurance et de contrôle qualité à chaque étape de la chaîne. À titre d'illustration, un extrait de ce fichier de suivi est repris en Annexe 2.

III.3.1.2 Séquence de critères de sélection des images de niveau 2A

Pour combiner les images sur une même zone, on sélectionne d'abord l'image de priorité 1, également appelée image « P1 », ayant la couverture nuageuse la plus faible sur la zone. En cas de couverture nuageuse similaire entre deux images, on sélectionne de préférence en priorité 1 les tuiles acquises à la même date que les tuiles voisines. Cette image P1 sert de couche de base pour

³ Attention, il s'agit du format des données de Niveau 2A Sentinel-2 produites par THEIA/MUSCATE. Le format des données produites par l'outil SEN2COR de l'ESA est différent, tout comme le format des données produites par Sen2AGRI. La description officielle du format des données THEIA est disponible [ici](#).

notre analyse et les images de priorité secondaire (P2), tertiaire (P3) ou davantage si nécessaire viennent compléter l'information couverte par les nuages. Par conséquent, la dispersion et la localisation des nuages est importante dans le processus de sélection des images de second ou troisième ordre. De manière générale, on essaie dans la mesure du possible à limiter l'analyse à 4 images maximum par zone.

III.3.1.3 Variables d'entrée

De nombreuses variables peuvent être utilisées comme données d'entrée dans un processus de classification d'images. Ces caractéristiques peuvent être: (1) des caractéristiques spectrales en combinant des bandes spectrales, (2) des caractéristiques spatiales en utilisant des informations spatiales ou en travaillant au niveau objet au lieu des pixels, et (3) des caractéristiques temporelles pour caractériser les catégories de couverture terrestre qui évoluent au fil du temps, tels que les terres cultivées.

Pelletier et al. (2016) ont testé, à partir des bandes spectrales de données SPOT 4 et Landsat 8, la plus-value de l'ajout de variables spectrales et temporelles sur les performances de classificateurs Random Forest (RF) et Support vector Machine (SVM). Les résultats de cette analyse ont montré que l'ajout de variables temporelles et spectrales offrait peu ou pas de plus-value en comparaison avec l'utilisation des bandes spectrales des données SPOT ou Landsat complétées de l'indice de végétation NDVI (Normalized Difference Vegetation Index). Par ailleurs, Rakotomala et al. (2015) ont montré lors d'une étude sur le suivi de la déforestation à Madagascar que, en complément du NDVI, l'ajout des variables d'altitude dérivées d'un Modèle Numérique de Terrain à 30 mètres de résolution permettait d'améliorer les performances du modèle RF. Sur base de ces résultats, la plus-value de l'ajout de ces deux variables aux diverses bandes spectrales des images Sentinel-2 a été testée et validée dans le cadre de notre étude. En effet, après avoir mené plusieurs tests de performance, il ressort que l'ajout des variables NDVI et SRTM améliorent les performances de l'algorithme de classification.

III.3.1.3.1 Calcul de l'indice de végétation NDVI

Le NDVI est l'un des indices de végétation les plus utilisés pour la cartographie de l'occupation des terres. Il est calculé par le rapport entre le facteur de réflectance de la bande rouge (R) et de la bande proche infrarouge (NIR). Le NDVI d'une zone à végétation dense aura tendance à afficher des valeurs positives, tandis que l'eau et les zones bâties seront représentées par des valeurs proches de zéro ou négatives. La formule de calcul du NDVI est fournie par Rouse et al. (1973) comme suit:

$$\text{NDVI} = (\text{NIR}-\text{R})/(\text{NIR}+\text{R})$$

Équation 1 : Calcul du NDVI (Rouse et al., 1973). Où NIR correspond à la bande proche infrarouge et R à la bande rouge

III.3.1.3.2 Préparation des données SRTM

À partir des tuiles SRTM à 30m de résolution téléchargées via le site de l'USGS, une mosaïque couvrant l'ensemble de la zone PREAA est produite. Celle-ci est ensuite découpée selon l'emprise de chaque tuile Sentinel-2 pour être empilée aux différentes variables d'entrée.

III.3.1.4 Construire une image multibande

Les images satellites optiques brutes téléchargées (Sentinel-2 ou autre) sont fournies selon un format où l'information acquise par le capteur dans une bande spectrale constitue une image à part entière. Le produit est donc composé d'autant d'images qu'il y a de bandes spectrales et on parle généralement d'images « monospectrales » (image en noir et blanc). Pour les besoins de classification d'images, l'objectif est de créer, à partir de ces différentes images, une image unique contenant l'information acquise dans les différentes bandes spectrales (opération de « Stacking » en anglais). Le produit résultant est une image multibande ou multi-spectrale. La représentation par combinaison de ces bandes d'information numérique en utilisant les trois couleurs primaires (rouge, vert, bleu - RVB) permet alors d'obtenir des images en couleurs. Les données de chaque bande sont représentées comme une couleur primaire et, selon la luminosité relative (c.-à-d. valeur numérique) de chaque pixel dans chaque bande, les couleurs primaires se combineront en proportions différentes pour produire des couleurs distinctes. Selon les bandes d'information placées dans les canaux RVB, il est possible de créer différentes compositions colorées. La Figure 9 illustre différentes compositions colorées réalisées à partir d'une image multispectrale Sentinel-2 de la zone PREAA produite pour les besoins de l'étude.

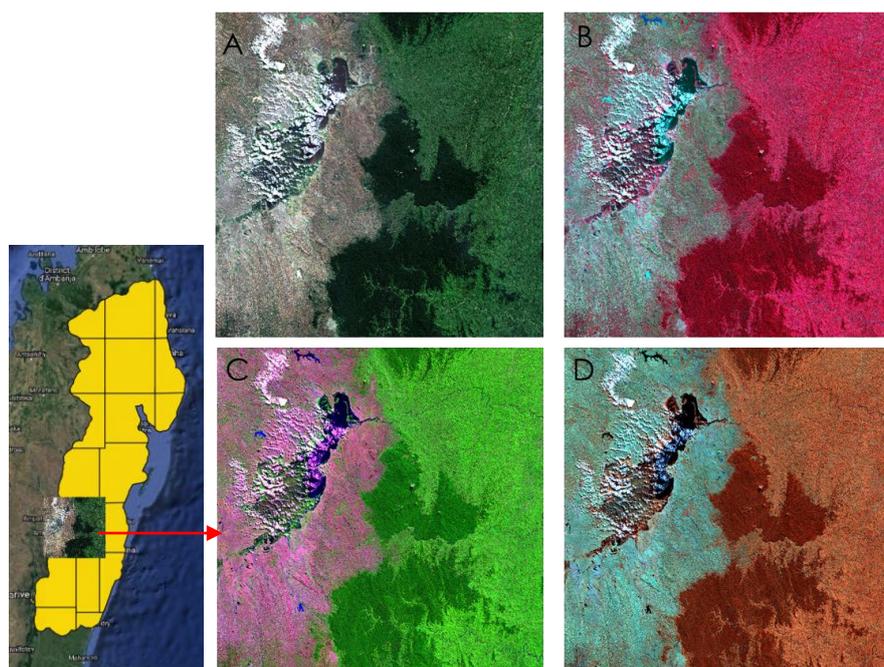


Figure 9 : Compositions colorées différentes produites à partir d'une image multibande Sentinel-2 de la zone PREAA. A : Composition colorée vraie couleur (R =R ;V =V ;B =B) ; B : Composition fausse couleur infrarouge 1 (R =IR, V =R ;B =V) ; C : Composition fausse couleur moyen infrarouge (R=SWIR, V=IR, B=R) ; D : Composition fausse couleur infrarouge 2 (R=IR, V=SWIR, B=SWIR)

Les images multibandes produites pour les besoins de cette étude rassemblent au total 12 bandes d'information et se composent comme suit :

# bandes de l'image multispectrale	Contenu des bandes	Résolution
1	Band 2 – Blue	10m
2	Band 3 – Green	10m
3	Band 4 – Red	10m
4	Band 5 – Vegetation red edge	rééchantillonné à 10m
5	Band 6 – Vegetation red edge	rééchantillonné à 10m
6	Band 7 – Vegetation red edge	rééchantillonné à 10m
7	Band 8 – NIR	10m
8	Band 8A – Narrow NIR	rééchantillonné à 10m
9	Band 11 – SWIR	rééchantillonné à 10m
10	Band 12 – SWIR	rééchantillonné à 10m
11	NDVI (NIR -RED)/(RED+NIR)	10m
12	SRTM 30m	rééchantillonné à 10m

Tableau 3 : Composition de l'image multibande Sentinel-2 produite pour les besoins de cartographie de l'UOT de la zone PREAA

Les bandes spectrales 2 à 8, 8a, 11 et 12 de Sentinel-2 sont donc empilées et complétées par les bandes d'information relatives au NDVI et SRTM. Les bandes spectrales 1, 9 et 10 de Sentinel-2 à

60m de résolution (voir Figure 2) sont ignorées car celles-ci n'offrent aucune plus-value pour les travaux de classification de l'UOT.

III.3.1.5 Géo-référencement

La co-registration est le processus d'alignement géométrique des pixels, qui représentent la même zone sur deux images ou plus. Un enregistrement précis d'image à image doit être garanti pour tous les jeux de données multi-temporels et multi-capteurs, car un ajustement spatial insuffisant peut conduire à la détection de « faux » changements dans le cas d'analyse de changement (Sundaresan et al ; 2007). Dans le cas de notre étude, bien que l'objectif n'est pas d'effectuer d'analyse de changement à l'heure actuelle, ceci pourrait s'avérer nécessaire dans le futur et il est donc important de corriger les éventuels décalages spatiaux, d'autant plus que nous combinons des images multi-dates.

Bien que les images téléchargées au niveau 2A soient ortho-rectifiées, il n'est pas rare d'observer de légers décalages spatiaux qu'il convient de corriger autant que possible. Par conséquent, l'évaluation systématique de la précision géométrique des données sera réalisée afin d'obtenir une liste des images devant faire l'objet d'une correction. Pour le processus de coregistration, un minimum de 6 points de calage seront uniformément répartis sur chaque image. Cette opération est réalisée dans le logiciel QGIS, à l'aide de l'outil de géoréférencement. Dans le cas de décalage nécessitant une simple translation d'une image vers une autre, l'utilisation de transformations linéaire ou polynomiale d'ordre 1 au maximum seront privilégiés pour éviter toute distorsion de l'image. Un ré-échantillonnage au plus proche voisin permettra également de conserver les valeurs des pixels en entrée. A l'issue de cette opération, l'objectif est d'atteindre dans la mesure du possible une erreur quadratique moyenne (RMSE) résiduelle inférieure à la taille du pixel de l'image.

III.3.1.6 Découper les images selon la grille d'analyse PREAA

Une mosaïque d'images de la zone pourrait être créée pour produire les résultats de classification sur une zone étendue. Toutefois, Pelletier et al. (2016) ont montré que les performances de l'algorithme de classification Random Forest diminuaient lorsque les paysages diffèrent de ceux pour lesquels des zones d'entraînement ont été définies. Par conséquent, afin d'améliorer les performances de classification, il est recommandé de stratifier les territoires étendus en plus petites zones d'analyse afin de garantir l'homogénéité des paysages. De plus, effectuer des mosaïques à partir d'images acquises à différentes périodes peut poser des problèmes d'homogénéité de l'information (phénologie végétale différente liée à la période de l'année ou

variation des valeurs de réflectance dues aux conditions d'acquisition), pouvant mener à des erreurs de classification

Au regard de ces recommandations et dans le but de répartir le travail de production au sein de l'équipe du LOFM, mais également de limiter le volume de données à traiter en un bloc, la zone d'étude PREAA est découpée en 18 cellules suivant la grille mondiale d'acquisition des données Sentinel-2. Chaque image multibande intégrant la chaîne de traitement est donc découpée selon la cellule correspondante à son emprise et les différentes cellules sont traitées simultanément par différents opérateurs (Figure 10). Les prétraitements et la classification des images s'opère donc cellule par cellule et les résultats sont ensuite compilés à l'échelle de la zone d'étude.

Afin d'assurer l'homogénéité des résultats de classification entre cellules, une validation croisée est réalisée à chaque étape et une attention particulière est portée aux zones de transition/recouvrement des données. Pour éviter l'apparition éventuelle d'effets de « couture » au niveau de ces zones, les classifications sont produites de manière séquentielle et l'utilisation de zones d'entraînement communes est préconisée.

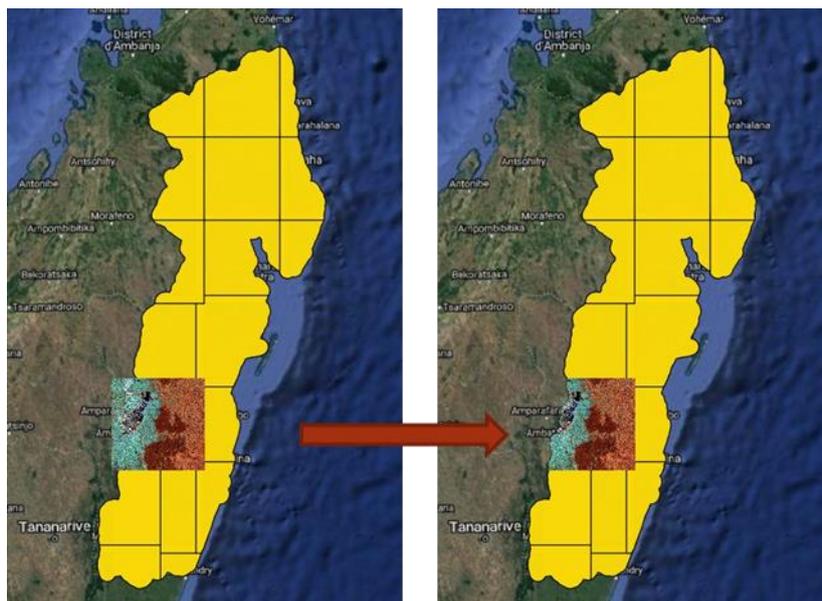


Figure 10 : Découpage des images Sentinel-2 suivant l'emprise de La zone d'étude PREAA, elle-même découpée en 18 cellules correspondant à la grille mondiale d'acquisition des données Sentinel-2

III.3.1.7 Extraire et appliquer un masque de nuages

En cas de présence de nuage sur une image, un masque des nuages et de leur ombre est réalisé afin d'ignorer ces zones lors des étapes de classification. Au cours du traitement des images Sentinel-2 au niveau 2A via la plateforme PEPS, un masque automatique des nuages et de leur ombre est produit et livré avec l'image corrigée. A partir de l'information contenue dans ce raster appelé

CLM1 (Cloud mask de résolution 1), un masque binaire (présence vs absence de nuage ou ombre) est réalisé à l'aide de la calculatrice raster dans QGIS, dont la formule sera dépendante des valeurs à conserver selon la qualité du masque. Une fois extrait, le masque binaire peut ensuite être amélioré si besoin par édition manuelle.

Lorsque les versions finales des masques de nuages sont obtenues pour les différentes images de la série temporelle d'une cellule, un **masque des nuages temporel** est créé en combinant les différents masques (Figure 11). Selon le même procédé, lorsque les masques de nuages temporels sont créés pour chaque cellule, il est possible de créer un **masque des nuages spatial** qui combinent les masques de nuages temporels de cellules adjacentes afin d'identifier les zones couvertes par les nuages sur l'ensemble des images dans la zone de recouvrement des deux cellules. Ces masques temporels et spatiaux permettent de prioriser la numérisation de zones d'entraînement (Region of Interest – ROI) dans les zones dépourvues de nuages et d'optimiser ainsi l'étape de production des ROI (étape limitante de la chaîne de traitement) en facilitant leur réutilisation d'une image à l'autre. De plus, l'utilisation des mêmes ROI d'une image à une autre pour entraîner l'algorithme renforce l'homogénéité des résultats de classification.

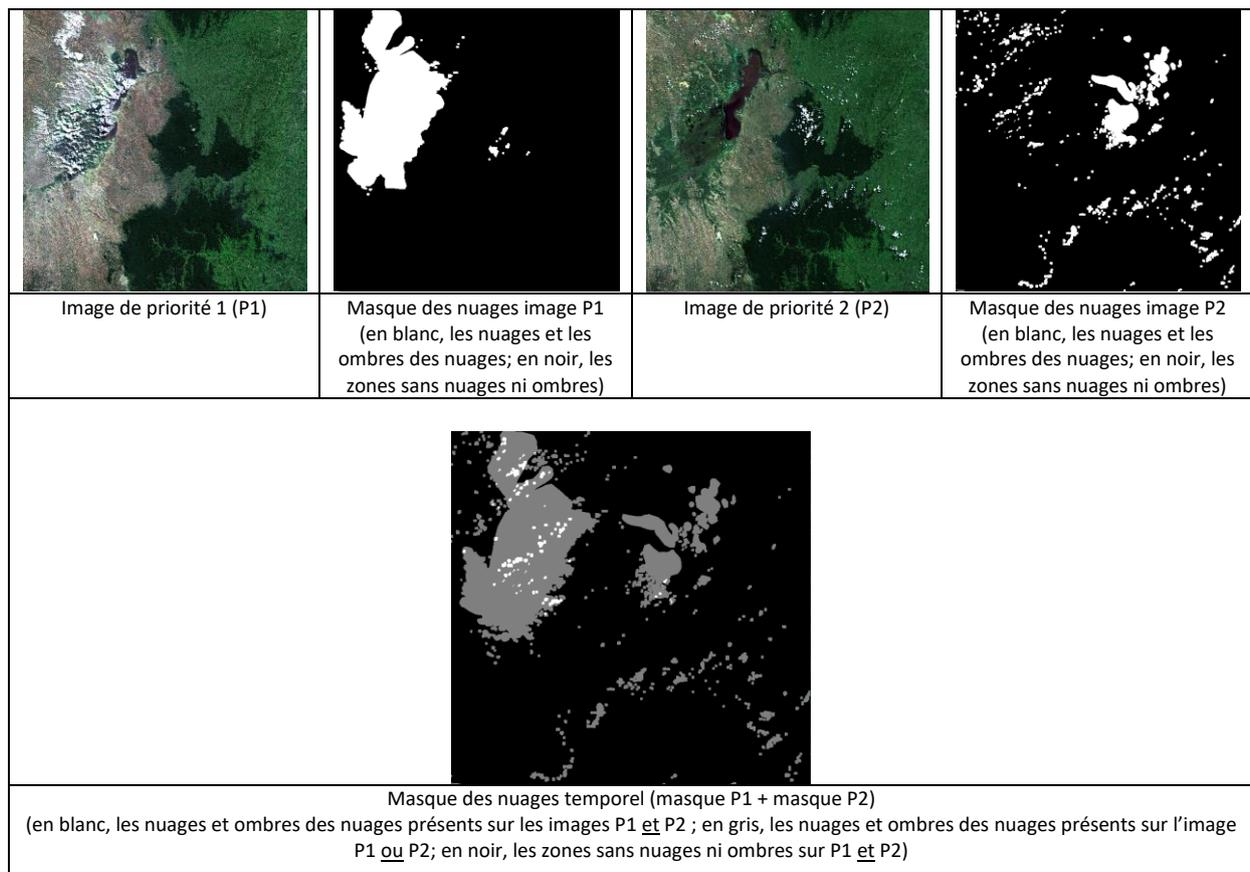


Figure 11 : Illustration de la construction du masque des nuages temporel à partir d'une série temporelle de deux images Sentinel-2

III.3.2 Etapes de Traitement des données : Classification supervisée

Classification supervisée vs classification non supervisée

On distingue deux grands types de méthode de classification : les méthodes de classification supervisée et non supervisée.

Lors de l'utilisation d'une méthode de classification supervisée, l'opérateur entraîne un algorithme de classification à partir d'échantillons numérisés par photo-interprétation, qui sont représentatifs des différentes classes UOT à discriminer. Ces échantillons forment un ensemble de données-tests, appelées zones d'entraînement ou ROI (Region of interest en anglais). L'opérateur supervise donc la classification d'un ensemble spécifique de classes. Les statistiques extraites pour chacune des bandes et pour chaque pixel de ces ensembles sont ensuite utilisées par l'algorithme de classification pour reconnaître des régions aux propriétés similaires à chaque classe sur l'ensemble de l'image.

La classification non supervisée procède de la façon contraire. Les classes sont automatiquement produites sur l'ensemble de l'image sur base de l'information numérique des données et celles-ci sont ensuite associées par un opérateur à des classes d'information utile (si possible). Dans ce cas, outre la possibilité de définir le nombre de classes à discriminer, l'opérateur n'a pas la possibilité de superviser la classification en entraînant l'algorithme sur base d'échantillons prédéfinis.

De nombreuses études ont été réalisées pour comparer les performances de ces deux types de méthodes à partir d'algorithme de classification variés et celles-ci ont démontré que les méthodes supervisées — Maximum Likelihood (ML), Neural Networks (NN), Support Vector Machines (SVM) (Vapnik, 1995, 1998), and Decision Trees (DT) (Breiman et al., 1984; Hansen et al., 1996; Friedl and Brodley, 1997) —performent mieux que les méthodes non supervisées (Szuster et al., 2011; Khatami et al., 2016).

Analyse orientée pixel vs orientée objet

On distingue également deux échelles d'analyse lors des travaux de classification d'images: les analyses orientées pixel ou les analyses orientées objet. Une méthode de classification orientée pixel analyse l'information contenue dans l'image à l'échelle du pixel alors que l'orientée objet crée des segments ou objets composés d'un agrégat de pixels généré par un ou plusieurs critères d'homogénéité dans une ou plusieurs dimensions. Ainsi, les segments ont des informations spectrales supplémentaires par rapport aux pixels individuels (valeurs moyennes par bande, mais aussi valeurs médianes, valeurs minimale et maximale, rapports moyens, variance, etc.), et

présentent l'avantage d'intégrer de l'information spatiale (distance, voisinage, forme, etc.) (Blaschke et Hay, 2001; Darwish et al., 2003; Flanders et al., 2003; Benz et al., 2004; van der Werff et van der Meer, 2008; Hay et Castilla, 2008).

Toutefois, lorsque la taille des pixels de l'image (résolution spatiale) est plus grossière ou, au mieux, de taille similaire à celle des objets d'intérêt, l'accent est généralement mis sur l'analyse orientée pixel. Bien qu'il est délicat de définir des seuils d'application génériques, trois situations nécessitant des techniques complètement différentes pour démêler les informations des jeux de données peuvent être différenciées sémantiquement (Figure 12). Comme l'illustre la figure, dans le cas de l'utilisation d'images à moyenne ou haute résolution (situations (a) et (b)), l'intérêt de l'analyse par objet est limité. Toutefois, lorsque la résolution augmente davantage (situation (c)), les avantages spécifiques de l'approche orientée objet peuvent être déployés, même si de telles approches de régionalisation ont également été appliquées à d'autres situations, pour des données à résolution moyenne ou grossière (Dorren et al., 2003; Duveiller et al., 2008; Myint et al., 2008; Jobin et al., 2008).

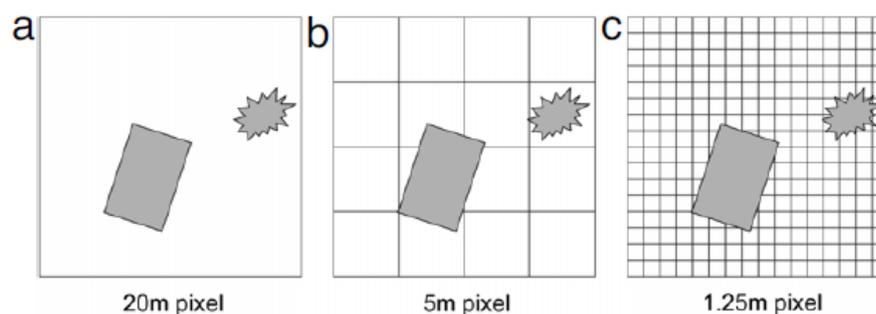


Figure 12 : relation entre objets considérés et résolution spatiale: (a) Moyenne résolution: pixels nettement plus grands que les objets. (b) Haute résolution: la taille des pixels et des objets est du même ordre, les techniques orientées pixel sont appropriées. (c) Très haute résolution: les pixels sont significativement plus petits que les objets, l'agrégation de pixels en objets est recommandée.

Choix de la méthode de classification à partir des données Sentinel-2 pour la zone PREAA

Sur base des éléments précités, une classification supervisée orientée pixel est recommandée pour l'analyse. Les classes UOT étant prédéfinies, il est préférable pour l'opérateur d'entraîner l'algorithme de classification sur base de ses connaissances du paysage et sur les classes à discriminer. Pour ce qui est de l'échelle d'analyse, tel qu'illustré précédemment, une analyse orientée pixel est appropriée compte tenu de la résolution des images Sentinel-2. De plus, la plupart des études concluantes réalisées à partir d'une méthode orientée objet à partir de données à moyenne résolution ont été réalisées à partir du logiciel commercial eCognition; et, à ce jour, tel que

démontré par Baisantry et al (2017), les logiciels gratuits ou libres proposant des techniques de segmentation (Monteverdi Orfeo Toolbox, Spring, MultiSpec, ILWIS) n'égalent pas ses performances.

Etapes de traitement des données par classification supervisée orientée pixel

Les étapes de traitement consistent en la classification supervisée des données prétraitées à partir des outils du logiciel open source Monteverdi Orfeo Toolbox (OTB)⁴ accessibles dans QGIS, dans le but de produire la carte d'occupation des terres de version 1. Le processus de classification comprend 5 étapes principales :

- 1) Créer des zones d'entraînement (ROI) ;
- 2) Calculer les statistiques de l'image ;
- 3) Entraîner le classificateur d'images (construire le modèle);
- 4) Créer une classification d'image ;
- 5) Créez une mosaïque de classification.

III.3.2.1 Créer des zones d'entraînement (ROI)

La méthode de classification sélectionnée pour cette étude étant une méthode supervisée, celle-ci implique la délimitation de zones d'entraînement (ROI) pour calibrer l'algorithme de classification. Ces ROI sont numérisés directement dans le logiciel QGIS par photo-interprétation des images Sentinel-2 (Figure 13). Pour numériser et labelliser les ROI, plusieurs compositions colorées et différents niveaux de contrastes sont affichés afin de considérer au mieux l'information spectrale contenue dans les images. L'échantillonnage des ROI se fait de façon raisonné et répartie sur l'ensemble de l'image à classifier en valorisant surtout la connaissance terrain des équipes. En complément, l'ensemble des données ancillaires disponibles sont utilisées pour orienter et confirmer l'interprétation (images satellites à très haute résolution disponibles dans Google Earth, données Open street map, données de terrain...), ainsi que les masques de nuages temporels et spatiaux afin d'optimiser la production et la réutilisation des ROI d'une image à une autre.

⁴ <https://www.orfeo-toolbox.org/>

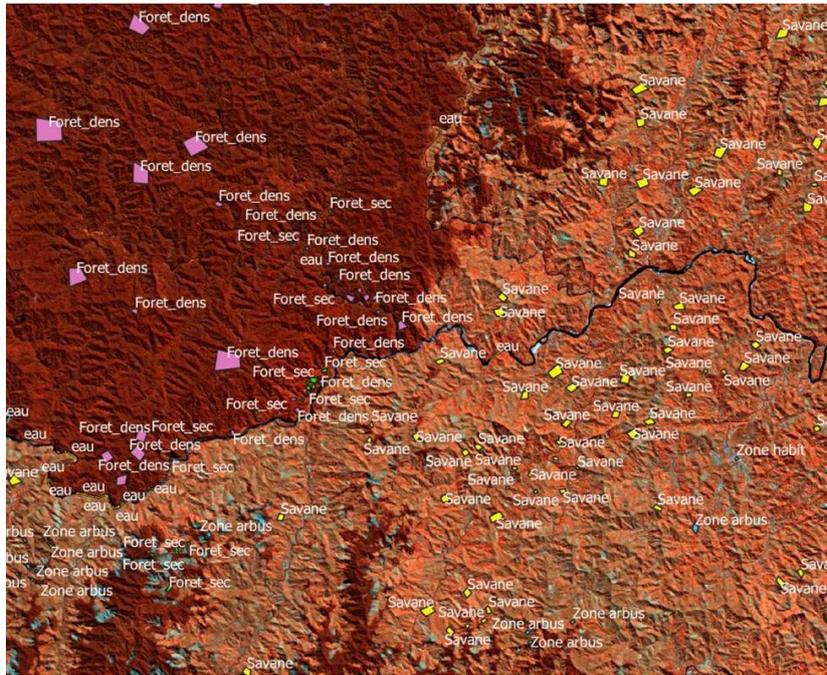


Figure 13 : Exemple de zones d’entraînement (ROI) numérisées sur une image Sentinel-2 de la zone PREAA pour les besoins de classification de l’UOT

Pour rappel, la zone PREAA étant couverte par 18 cellules, la carte UOT produite sur l’entièreté de la zone est une mosaïque des classifications produites par cellule. En moyenne, trois images par cellule sont nécessaires pour combler les zones nuageuses. Les classifications temporelles et spatiales doivent donc être impérativement rationalisées, en réutilisant autant que possible les ROI. Pour ce faire, l’utilisation des masques de nuages temporels et spatiaux est primordiale lors de cette phase de création des ROI afin d’optimiser la production mais aussi, et surtout, afin de garantir l’homogénéité des résultats de classification.

III.3.2.2 Calculer les statistiques de l’image

Cette application disponible dans l’OTB calcule une moyenne globale et un écart type pour chaque bande d’un ensemble d’images et enregistre les résultats dans un fichier XML. Le XML de sortie est destiné à être utilisé en entrée pour entraîner l’algorithme de classification afin de normaliser les échantillons avant l’apprentissage.

III.3.2.3 Entraîner le classificateur d’images

Il existe un grand nombre d’algorithme de classification et les méthodes d’apprentissage d’ensemble (bootstrap, boosting, etc.) ont récemment suscité un vif intérêt. Elles consistent à apprendre plusieurs classificateurs faibles pour générer un classificateur avec une règle de décision forte. Il est proposé dans le cadre de cette étude d’utiliser l’algorithme de classification Random Forest (RF). RF de Breiman (2001) est une méthode d’apprentissage d’ensemble bien connue pour

ses bonnes capacités prédictives de l'occupation des terres (Belgiu et Draǵut., 2016) et largement utilisée dans la littérature en télédétection au cours des dernières décennies. Il a obtenu des performances comparables aux classificateurs classiques tels que DT ou SVM, avec un temps de calcul inférieur (Inglada et al., 2015; Rodriguez-Galiano et al., 2012; Gislason et al., 2006).

RF est un algorithme de classification supervisée non paramétrique qui combine un algorithme d'arbres de décision et une technique d'agrégation (bagging). L'algorithme construit une multitude d'arbres de décision par tirage aléatoire d'observation (pixels) et de variables. L'affectation de la classe finale est réalisée par vote majoritaire à partir de l'ensemble des arbres. La mise en œuvre de RF nécessite le réglage de quatre paramètres : (1) K, le nombre maximum d'arbres; (2) m, le nombre de caractéristiques sélectionnées au hasard sur chaque nœud; (3) max_depth, la profondeur maximale de chaque arbre; et (4) min_samples, le nombre minimal d'échantillons par nœud. Dans le cadre de notre étude, les valeurs par défaut renseignées par l'OTB sont adaptées sur base des conclusions de l'étude de Pelletier et al. (2016) qui compare les performances de l'algorithme sur bases de différents réglages. Le réglage optimal suivant est adopté : (1) K, le nombre maximum d'arbres = 100 ; (2) m, le nombre de caractéristiques sélectionnées au hasard sur chaque nœud = 0; (3) max_depth, la profondeur maximale de chaque arbre = 25; et (4) min_samples, le nombre minimal d'échantillons par nœud = 25.

La fonction 'Entraîner le classificateur d'images Random Forest' ('Train Images Classifier (rf)') effectue un apprentissage du classificateur à partir de plusieurs paires d'images d'entrée et des ROI créées. Les échantillons sont composés de valeurs de pixels dans chaque bande, éventuellement normalisées à l'aide d'un fichier de statistiques XML généré par l'application 'Compute Images Statistics'. En sortie d'entraînement, l'algorithme crée un fichier de modèle contenant la « mémoire » de l'étape d'apprentissage. Dans notre cas, le modèle est construit en limitant le nombre d'échantillons d'entraînement et de validation par classe d'occupation des terres à 1000. Par conséquent, RF utilise la moitié du jeu de données d'apprentissage pour la calibration et l'autre moitié pour la validation. En comparant les résultats de classification générés avec les échantillons de validation, une matrice de confusion est produite et l'algorithme calcule les performances du modèle. Ces performances sont calculées au niveau global à partir d'un indice de précision de la classification appelé Out-Of-the-Bag (OBB) (Bylander et al., 2002), ainsi que l'indice de performance global appelé Kappa (Cohen et al., 1960). Plus l'indice OOB tend vers zéro et, à l'inverse, plus le coefficient Kappa tend vers 1, plus le modèle est performant. Les performances sont également calculées séparément pour chaque classe à partir des indicateurs suivants : la précision (Precision), le rappel (Recall) et le F-score ; La précision est le rapport entre les observations positives correctement prédites et le total des observations positives prédites ; Le rappel est le rapport entre les

observations positives correctement prédites et toutes les observations de la classe actuelle ; Et, le F-score est la moyenne pondérée de la précision et du rappel, selon la formule suivante : $2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$. Par conséquent, ce score prend en compte à la fois les faux positifs et les faux négatifs. Plus ces indicateurs tendent vers 1, plus le modèle est performant pour la classe considérée. Après généralement plusieurs itérations, le modèle correspondant aux meilleurs indicateurs de qualité est utilisé pour la classification de l'image correspondante.

III.3.2.4 Créer une classification d'image

Cette application effectue une classification d'image sur la base du fichier de modèle généré par l'application 'Train Images Classifier'. Chaque pixel de l'image en sortie se voit attribué une valeur correspondante à l'une des classes UOT définies au départ, à partir des ROI. Lors de cette étape, un masque de classification peut être introduit pour produire la classification uniquement sur une partie de l'image et plus particulièrement, dans notre cas, dans les zones dépourvues de nuages. Le masque des nuages de l'image créé lors des étapes de prétraitement est donc réutilisé ici et convertit en masque de classification. L'image en sortie contiendra donc des valeurs nulles dans les zones masquées. Ces valeurs seront ensuite converties lors de la fusion des classifications de la série temporelle, décrite ci-après.

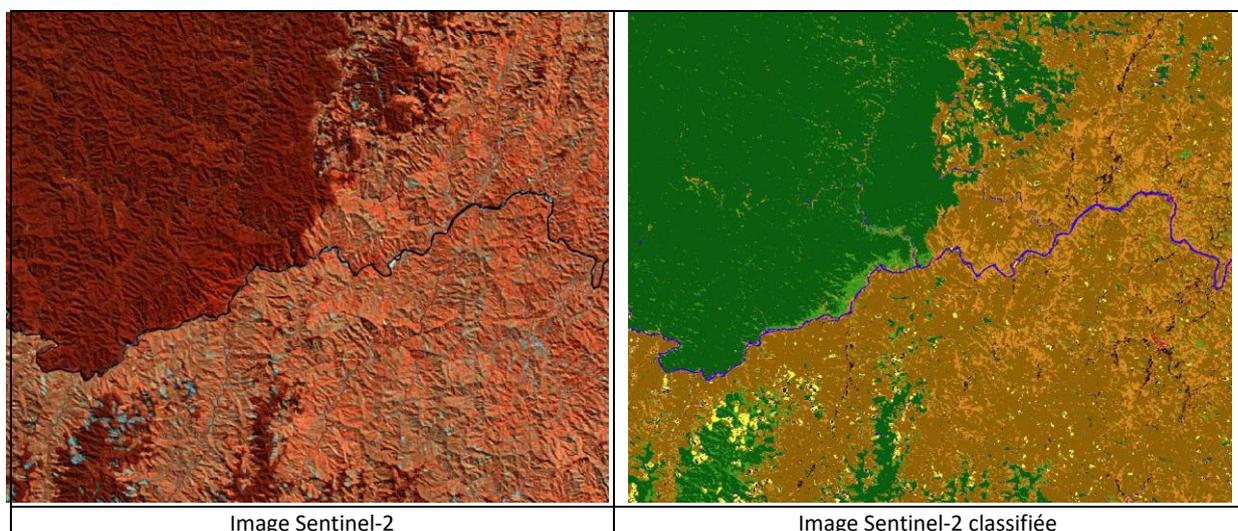


Figure 14 : Extrait d'une image Sentinel-2 de la zone PREAA classifiée à partir de l'algorithme Random Forest

III.3.2.5 Créer une mosaïque de classification

Lorsque l'ensemble des images de la série temporelle d'une cellule ont été classifiées, une mosaïque temporelle est réalisée afin de combler les zones couvertes par les nuages sur l'image de priorité 1 (P1). A la suite de cette étape, un contrôle de la qualité de la mosaïque temporelle est effectué pour s'assurer de l'alignement correct et de l'homogénéité des différentes classifications qui la composent.

Selon le même principe, lorsque les mosaïques temporelles ont été créées pour chaque cellule, une mosaïque spatiale est réalisée pour produire la carte d'occupation des terres version 1 et intégrer les étapes de post-traitement.

III.3.3 Etapes de Post-traitement des données : Vérification et amélioration de la classification

III.3.3.1 Tamiser les résultats de classification

Outre le contrôle qualité qui consiste à vérifier le rendu général de la mosaïque et s'assurer de l'alignement correct des résultats des différentes cellules, la première étape de prétraitement a pour objectif de tamiser les résultats de classification afin d'éliminer les pixels ou agrégats de pixels isolés tout en respectant l'unité de cartographie minimum sélectionnée de 0,5ha, correspondant à la définition de la forêt pour Madagascar. Cette opération, qui a pour effet de lisser et simplifier les résultats, s'effectue dans QGIS à l'aide de la fonction Tamiser (Sieve) où le seuillage du filtre (49 pixels=4,9ha) et le niveau de connectivité entre les pixels (Connectivité=4) est paramétré.

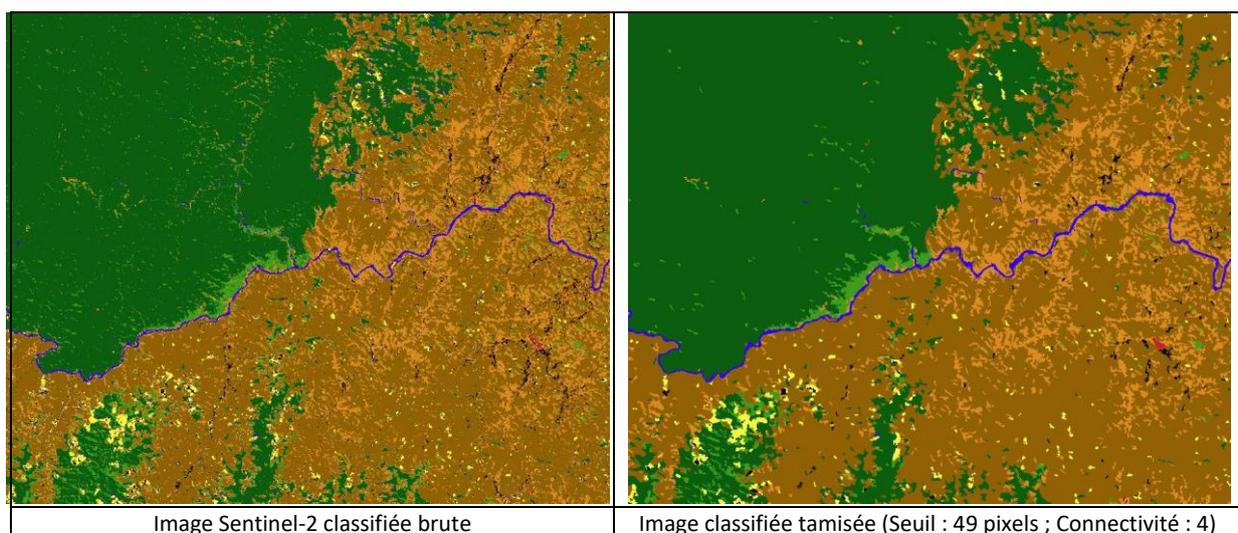


Figure 15 : Illustration de l'effet de « tamis » sur les résultats de classification

III.3.3.2 Ajustement des classifications

Les résultats tamisés de la classification semi-automatique de version 1 sont ensuite vérifiés et ajustés visuellement à l'aide des images Sentinel-2 correspondantes et de l'ensemble des données ancillaires disponibles. Cette opération peut s'effectuer de différentes manières selon les besoins. Les corrections peuvent se faire dans QGIS directement à partir du format raster ou après vectorisation des résultats. Elles peuvent également être réalisées par un opérateur à la fois sur la couche globale ou cette dernière peut également être découpée selon la grille d'analyse définie au départ pour permettre l'édition simultanée des différentes cellules.

Cette étape est particulièrement importante pour corriger les erreurs de classification mais également pour rajouter les classes d'utilisation des terres qui n'ont pas été intégrées au processus de classification, telles que les classes infrastructures et zone minière par exemple.

III.3.3.3 Remplissage des zones de « no data » résiduelles

A l'issue des travaux de classification, il se peut que quelques zones de « no data » persistent pour cause de couverture nuageuse persistante au sein de la série temporelle des images. Cette dernière étape consiste donc à combler ces zones à partir de photo-interprétation visuelle des images Sentinel-2 complémentaires sur la zone ou l'utilisation de données ancillaires pouvant fournir l'information.

III.4. Méthode de contrôle et d'évaluation de la qualité

Les valeurs réelles de la couverture terrestre varient des valeurs pouvant être extraites directement des travaux de classification. Toutefois, ces valeurs réelles peuvent être estimées après une évaluation de l'exactitude et la création d'une matrice d'erreur de classification (Olofsson, Foody, Stehman et Woodcock, 2013). Le principe sous-jacent de l'évaluation de l'exactitude est qu'elle compare la classification cartographiée à des données de référence de qualité supérieure, collectées selon une approche basée sur des échantillons. Les données de référence de meilleure qualité peuvent être obtenues à partir des données collectées au sol, mais comme elles sont coûteuses et fastidieuses, elles sont plus souvent obtenues par imagerie satellite ou photographie aérienne à résolution spatiale plus fine que les données utilisées pour créer la classification.

La conception et la mise en œuvre de l'évaluation de l'exactitude et de l'estimation des surfaces réelles de la carte UOT produite pour la zone PREAA est basée sur les recommandations de « bonnes pratiques » de la FAO, elles-mêmes basées sur les travaux de Olofsson et al. (2014). Lorsque la carte UOT est finalisée, la méthode d'évaluation se décline en trois étapes : le plan d'échantillonnage (« Sampling design »), la conception de la réponse (« Response design ») et l'analyse (« Analysis »).

III.4.1 Finalisation de la carte

La première étape est un contrôle général de la qualité des données cartographiques. Avant d'exécuter l'évaluation de l'exactitude des classes UOT, les données cartographiques doivent être considérées comme définitives.

III.4.2 Plan d'échantillonnage

Le plan d'échantillonnage est le protocole de sélection du sous-ensemble d'unités spatiales (pixels ou polygones, par exemple) qui constituera la base de l'évaluation de l'exactitude. Pour choisir un plan d'échantillonnage, il faut prendre en compte les objectifs spécifiques de l'évaluation de l'exactitude et dresser une liste priorisée des critères de conception souhaitables. Dans le cas de notre étude, pour s'assurer qu'aucune classe rare/plus petite ne soit manquée ou sous-représentée, une approche d'échantillonnage aléatoire stratifié est mise en œuvre.

L'échantillonnage aléatoire stratifié est un plan d'échantillonnage probabiliste, facile à mettre en œuvre et couramment utilisé par la communauté de la télédétection dans l'évaluation de l'exactitude (voir Cakir, Khorram et Nelson, 2006 ; Huang et al., 2010 ; Mayaux et al., 2006 ; Olofsson et al., 2011). La stratification a deux objectifs principaux. Tout d'abord, les strates peuvent présenter un intérêt pour la communication des résultats, c'est-à-dire l'exactitude par classe de couverture terrestre ou sous-région. Le deuxième objectif de la stratification est d'assurer une représentation suffisante des classes rares (par exemple, qui ne représentent qu'une faible proportion de la zone d'intérêt). C'est souvent le cas pour les classes de changement mais cela peut être le cas également pour une analyse à une date. La stratification permet dans ce cas de définir un nombre suffisant d'échantillons dans ces classes rares afin de diminuer l'erreur type pour l'estimation de l'exactitude de l'utilisateur. Pour cette raison, la stratification dans cette étude est basée sur les classes UOT et un échantillon indépendant est produit pour chaque classe.

Pour construire l'échantillon, la première étape consiste à définir sa taille. Celle-ci doit être représentative du nombre d'unités spatiales (pixels ou surface), le rendant suffisamment grand pour obtenir des estimations fiables, mais aussi petit que possible afin de réduire le temps d'analyse, et par conséquent les coûts. La détermination de la taille de cet échantillon est une science inexacte, car elle dépend d'informations précises et géographiques inconnues avant l'évaluation. Bien qu'il existe des formules permettant de calculer la taille globale de l'échantillon et sa distribution, il appartient à l'utilisateur de décider du meilleur moyen de déterminer la taille de l'échantillon. Pour notre étude, la taille d'échantillon globale adéquate pour l'échantillonnage aléatoire stratifié sera estimée à partir de l'équation de Cochran (1977) suivante :

$$n = \frac{(\sum W_i S_i)^2}{[S(\bar{O})]^2 + (1/N)\sum W_i S_i^2} \approx \left(\frac{\sum W_i S_i}{S(\bar{O})} \right)^2$$

Équation 2 : Calcul de la taille de l'échantillon global représentatif pour la réalisation du plan d'échantillonnage stratifié (Cochran, 1977). Où, N est le nombre d'unités dans la zone d'intérêt (nombre de pixels globaux si l'unité spatiale est un pixel), $S(\bar{O})$ est l'erreur type de l'exactitude globale estimée que nous souhaitons obtenir, W_i est la proportion de surface cartographiée pour la classe i , et S_i est l'écart type de la classe i .

La taille globale de l'échantillon résultant de ce calcul peut ensuite être répartie entre les classes de plusieurs manières, soit par répartition égale, soit par répartition proportionnelle. Dans la répartition égale, la taille globale de l'échantillon est répartie de manière égale entre les classes. Dans la répartition proportionnelle, la taille globale de l'échantillon est distribuée proportionnellement selon la superficie des classes, de sorte que les classes rares reçoivent une petite proportion de la taille totale de l'échantillon. La stratification étant utilisée pour les classes rares, il est nécessaire de s'assurer qu'il y ait un nombre suffisant d'échantillons dans chaque classe. On s'assure donc que chaque classe possède un minimum de 20 à 100 échantillons (Congalton et Green, 2008).

Pour effectuer le calcul de la taille globale de l'échantillon, nous utilisons le tableur Excel proposé par la FAO (2016)⁵, qui intègre l'équation de Cochran (1977). La répartition de ces points au sein de chaque classe est réalisée de manière proportionnelle afin de considérer les classes rares.

Lorsque la taille et la distribution des échantillons ont été définis sur base des résultats de classification, la deuxième étape consiste à générer aléatoirement les points d'échantillonnage au sein de chaque classe. Cette étape est réalisée directement dans QGIS à partir de la couche raster, classe par classe.

III.4.3 Conception de la réponse

La conception de la réponse définit comment déterminer si la carte et les données de référence sont en accord. Elle consiste donc premièrement à définir des sources de données de référence à comparer avec les données cartographiques, en supposant que la classification de référence sera suffisamment plus précise que la classification qui aura été produite. Deux méthodes permettent de garantir cela : i) La source de données de référence doit être de meilleure qualité (image satellite ou aérienne de meilleure résolution spatiale ou radiométrique et/ou données de terrain) que la source de données utilisée pour la classification cartographique, ou ii) le processus de création de la classification de référence doit être plus précis si les mêmes données sources sont utilisées.

Dans notre cas, la disponibilité de sources de données de référence de plus haute résolution étant limitée, le processus d'obtention de la classification de référence sera plus précis en procédant à une évaluation visuelle par des opérateurs experts.

⁵ https://www.dropbox.com/s/wsihmldebjc024/sample_size_stratified_simple_random.xlsx?dl=0

Lorsque les points d'échantillonnage ont été générés et que les sources de données de référence ont été sélectionnées, l'étape suivante consiste à interpréter visuellement chaque échantillon de point afin de confronter ultérieurement les résultats aux résultats de classification. Pour réaliser l'interprétation visuelle de chaque point, ceux-ci sont intégrés dans Google Earth via Open Foris Collect Earth, utilisé par les techniciens du LOFM pour produire le niveau d'émission de référence pour Madagascar. Ce plugin Google Earth permet au praticien d'évaluer visuellement la couverture ou l'utilisation des terres des échantillons avec les données disponibles gratuitement dans Google Earth, Google Earth Engine et les cartes Bing. En cliquant sur un échantillon, une fenêtre s'ouvre permettant de sélectionner la classe, d'affecter le degré de certitude de l'utilisateur à la classification et d'ajouter un commentaire (facultatif). Lors de cette étape, lorsque les données de référence contiennent un mélange de classes, il conviendra d'adopter une approche cohérente pour labelliser les classes de référence, tout en respectant l'unité de cartographie minimale (MMU) de la classification produite.

III.4.4 Protocole d'analyse

Le protocole d'analyse est l'étape ultime qui consiste à traduire les informations contenues dans la comparaison des résultats d'interprétation visuelle des points d'échantillonnage et des résultats cartographiques en estimation de l'exactitude et des surfaces, ainsi que de l'incertitude associée. La plupart des calculs sont basés sur la matrice de confusion (également appelée matrice d'erreur).

La matrice de confusion est un tableau croisé simple des classes UOT attribuées par la classification issue des données de télédétection par rapport aux données de référence des points d'échantillonnage (Figure 16). Elle organise les données d'échantillon acquises d'une manière qui résume les résultats clés et facilite la quantification de l'exactitude et de la surface. La diagonale principale de la matrice (reprise en gras sur la figure) met en évidence les classifications correctes, tandis que les éléments hors diagonale indiquent les erreurs d'omission et de commission.

#Class	Code QA (photo-interprétation)																Total
	1	2	3	4	5	8	9	10	11	12	13	15	16	17			
1	123	1	0	2	0	0	0	0	0	0	0	0	0	0	0	126	
2	0	38	0	2	0	0	0	0	0	0	0	0	0	0	0	40	
3	0	0	238	27	4	1	0	4	4	1	0	0	0	0	0	279	
4	0	1	14	532	0	0	0	11	4	1	0	0	0	0	0	563	
5	0	0	9	1	47	2	0	0	0	0	0	0	0	0	0	59	
8	0	0	0	2	4	486	0	0	0	0	0	0	0	0	0	492	
9	0	0	0	0	0	0	41	0	0	0	0	0	0	0	0	41	
10	1	0	1	8	0	0	0	80	1	0	0	0	0	0	0	91	
11	0	0	0	1	0	0	0	3	36	0	0	0	0	0	0	40	
12	0	0	1	0	1	0	0	0	0	29	0	0	0	0	0	31	
13	0	0	5	0	0	0	0	0	1	0	35	0	0	0	0	41	
15	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	22	
16	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	21	
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	20	
Total	124	40	268	575	56	489	41	98	46	31	35	22	21	20	0	1866	

Figure 16 : Matrice de confusion issue de l'évaluation de l'exactitude des résultats cartographiques sur la zone PREAA

Les mesures de l'exactitude sont dérivées de la matrice d'erreur et rapportées avec leurs intervalles de confiance respectifs. Elles incluent généralement l'exactitude globale, l'exactitude de l'utilisateur et l'exactitude du producteur, mais elles peuvent également être complétée par d'autres indices tels que le coefficient Kappa, le « Recall » ou le F-score (voir III.3.2.3). L'exactitude globale est la proportion de la zone classée correctement, et fait donc référence à la probabilité qu'une zone sélectionnée de manière aléatoire sur la carte soit classée correctement. L'exactitude de l'utilisateur correspond à la proportion de la zone classée dans la classe i qui est également la classe i dans les données de référence. Il donne aux utilisateurs la probabilité qu'une zone particulière de la carte de la classe i soit également cette classe sur terrain. L'exactitude du producteur est la proportion de la zone qui est la classe de référence j et est également la classe j sur la carte. C'est la probabilité que la classe j sur le terrain soit cette même classe sur la carte.

Au-delà des mesures d'exactitude, il est également recommandé d'utiliser les informations de la matrice de confusion pour ajuster l'estimation de la surface des classes obtenue à partir de la carte. Il est recommandé de baser cette estimation sur $p_{.k}$, qui est le total des colonnes de la classe de référence k dans la matrice d'erreur (Équation 3).

$$\hat{p}_{.k} = \sum_{j=1}^q \hat{p}_{jk}$$

Équation 3 : Calcul de $p_{.k}$ pour l'ajustement de l'estimation de la surface des classes obtenue à partir des travaux cartographiques.

Ce protocole d'analyse est réalisé directement dans QGIS à l'aide d'un script développé spécifiquement à partir des algorithmes fournis par Olofsson et al. (2013) et Mas et al. (2014). A partir des résultats cartographiques sous format raster et de la couche vecteur contenant les points d'échantillonnage interprétés, le script génère automatiquement la matrice de confusion et en extrait les mesures d'exactitude, d'estimation des surfaces réelles ainsi que les incertitudes associées.

IV.1. Spécificités techniques

IV.1.1 Classes UOT retenues

Définies à partir du système de classification national adopté pour Madagascar⁶, le Tableau 4 reprend les 14 classes UOT retenues pour les travaux de cartographie de la zone PREAA, dont 5 classes de forêt (« Forêt secondaire », « Forêt dense humide », « Plantation », « Forêt de littorale » et « Mangrove »). Elles ont été sélectionnées sur base de leur présence dans la zone ainsi que sur la capacité à les discriminer avec un niveau de précision et d'exactitude suffisant (>75%) compte tenu des données disponibles. Parmi celles-ci, seule la classe « Plantation » s'écarte ou complète quelque peu la définition reprise par le système de classification. De fait, les palmeraies (hors raphière) n'étant pas identifiées explicitement dans le système, la décision a été prise d'intégrer ces zones dans la classe « Plantation ».

1	Zone nue
2	Zone habitée
3	Zone arbustive
4	Savane
5	Forêt secondaire
6	Forêt dense humide
7	Eau
8	Zone cultivée
9	Marais
10	Plantation
11	Forêt de littorale
12	Infrastructure
13	Zone minière
14	Mangrove

Tableau 4 : Classes UOT retenues pour les travaux de cartographie de la zone PREAA

IV.1.2 Données d'entrée

Afin de combler l'ensemble des zones nuageuses, l'utilisation de 35 images Sentinel-2 a été nécessaire pour produire la carte UOT sur la zone PREAA. La majorité des images sélectionnées ont été acquises en 2018, avec un écart de plus ou moins 5 mois pour les autres lorsque nécessaire.

⁶ https://www.bnc-redd.mg/images/documents/MNV/Liv7_UOT_30_05_2018.pdf

S2A_39KTA_20180605	S2A_39LVD_20181209	S2A_39KTB_20180605
S2B_39KTA_20180411	S2B_39LVD_20180809	S2A_39KTB_20180411
S2A_39LVC_20180605	S2B_39LVD_20190106	S2A_39KTV_20180605
S2A_39LVC_20170915	S2B_39LVD_20190317	S2B_39LUE_20180630
S2A_39LVC_20190327	S2A_39LUC_20171107	S2A_39LUE_20180715
S2B_39LVC_20180531	S2A_39LUC_20180605	S2B_39LUE_20180809
S2B_39LTE_20180630	S2A_39LUC_20190531	S2A_38KRE_20180605
S2B_39LVE_20180710	S2A_39LTC_20180605	S2A_39LTD_20180605
S2A_39KUB_20180605	S2B_39LTC_20190317	S2B_39LTD_20190317
S2A_39KUB_20190531	S2A_39KTU_20180605	S2B_39KUA_20180411
S2A_39KUV_20180605	S2A_39LUD_20180605	S2A_39KUA_20180605
S2A_38KRD_20180605	S2B_39LUD_20190317	

Tableau 5 : Liste des images Sentinel-2 utilisées pour la cartographie de la zone PREAA (nomenclature : Capteur_tuile_date d'acquisition)

A partir de ces 35 images, 17 223 zones d'entraînement ou ROI ont été numérisés pour entraîner l'algorithme de classification Random Forest.

IV.1.3 Unité cartographique minimale

L'unité cartographique minimale, c'est-à-dire la taille minimale des objets présents sur la carte, est de 50 pixels ou 0,5ha, correspondant au seuil minimal de 0,5ha pour la définition des forêts à Madagascar.

IV.2. Cartographie des UOT à l'échelle du PREAA

IV.2.1 Carte de la zone PREAA

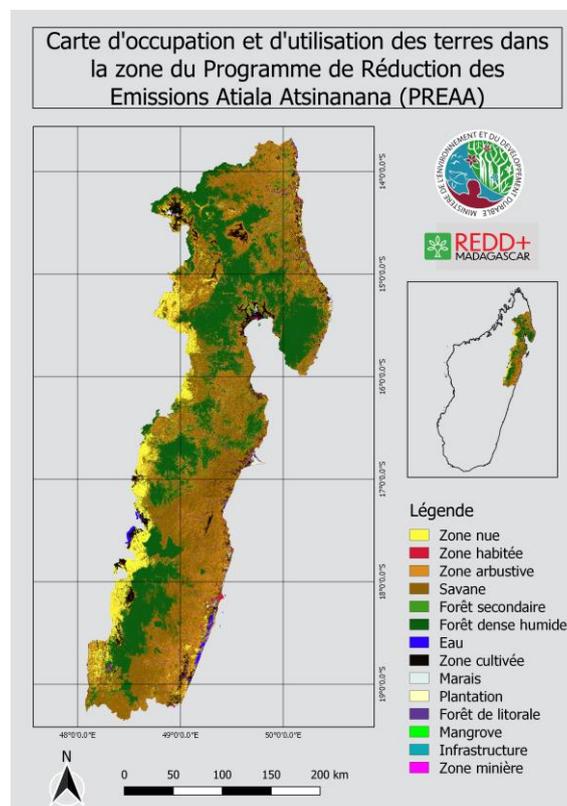


Figure 17 : Carte des UOT sur la zone PREAA

IV.2.2 Statistiques et évaluation de l'exactitude de la carte

IV.2.2.1 Plan d'échantillonnage et matrice de confusion

L'Annexe 3 fournit le détail des résultats du calcul du plan d'échantillonnage qui, selon les paramètres intégrés dans l'équation de Cochran, nous indiquait un minimum nécessaire de 1307 échantillons pour réaliser une analyse représentative. Afin de renforcer celle-ci, il a été décidé d'ajouter un minimum de 500 points supplémentaires. Ce sont finalement 1866 échantillons qui ont été générés de manière aléatoire selon le schéma stratifié repris à la Figure 18.

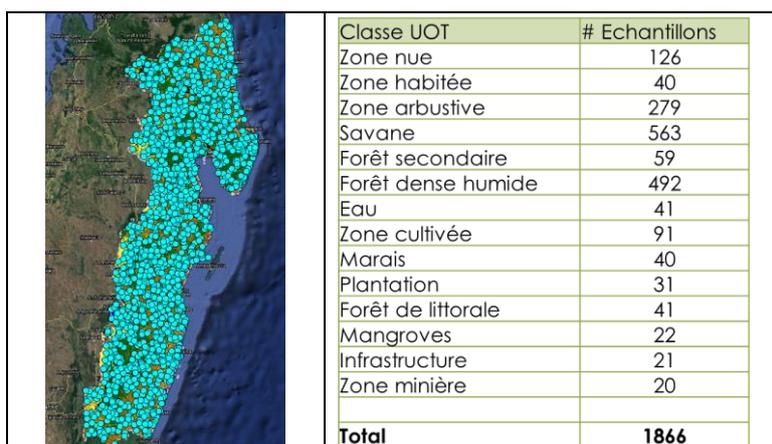


Figure 18 : Répartition des échantillons de points destinés à l'évaluation de la qualité de la carte PREAA

Après analyse de ces échantillons, la matrice de confusion suivante a pu être générée, nous permettant ensuite d'extraire les statistiques de qualité. Elle compare, en ligne, les résultats cartographiques et, en colonne, les résultats de l'analyse des échantillons par photo-interprétation.

#Class	Code QA (photo-interprétation)																Total
	1	2	3	4	5	8	9	10	11	12	13	15	16	17			
1	123	1	0	2	0	0	0	0	0	0	0	0	0	0	0	126	
2	0	38	0	2	0	0	0	0	0	0	0	0	0	0	0	40	
3	0	0	238	27	4	1	0	4	4	1	0	0	0	0	0	279	
4	0	1	14	532	0	0	0	11	4	1	0	0	0	0	0	563	
5	0	0	9	1	47	2	0	0	0	0	0	0	0	0	0	59	
8	0	0	0	2	4	486	0	0	0	0	0	0	0	0	0	492	
9	0	0	0	0	0	0	41	0	0	0	0	0	0	0	0	41	
10	1	0	1	8	0	0	0	80	1	0	0	0	0	0	0	91	
11	0	0	0	1	0	0	0	3	36	0	0	0	0	0	0	40	
12	0	0	1	0	1	0	0	0	0	29	0	0	0	0	0	31	
13	0	0	5	0	0	0	0	0	1	0	35	0	0	0	0	41	
15	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	22	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	21	
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	20	
Total	124	40	268	575	56	489	41	98	46	31	35	22	21	20		1866	

Tableau 6 : Matrice de confusion, comparant, en ligne, les résultats cartographiques et, en colonne, les résultats de l'analyse des 1866 échantillons par photo-interprétation

IV.2.2.2 Indicateurs statistiques de la qualité des résultats

Sur base de l'analyse de la matrice de confusion, il ressort que la carte possède une **exactitude globale (« Global accuracy » en anglais) de 93,7% (+/- 1,1%)** et un **indice Kappa de 0,92**. Le Tableau 7 fournit pour chaque classe les indicateurs de qualité suivants : Précision (« Precision » en anglais), Rappel (« Recall » en anglais) et F-score. Le Tableau 8 détaille également pour chaque classe l'exactitude du producteur et de l'utilisateur, ainsi que l'erreur estimée sur ces mesures.

Classe	Precision of class [x] vs all	Recall of class [x] vs all	F-score of class [x] vs all
Zone nue	97,6%	99,2%	98,4%
Zone habitée	95,0%	95,0%	95,0%
Zone arbustive	85,3%	88,8%	87,0%
Savane	94,5%	92,5%	93,5%
Forêt secondaire	79,7%	83,9%	81,7%
Forêt dense humide	98,8%	99,4%	99,1%
Eau	100,0%	100,0%	100,0%
Zone cultivée	87,9%	81,6%	84,7%
Marais	90,0%	78,3%	83,7%
Plantation	93,5%	93,5%	93,5%
Forêt de littorale	85,4%	100,0%	92,1%
Mangrove	100,0%	100,0%	100,0%
Infrastructure	100,0%	100,0%	100,0%
Zone minière	100,0%	100,0%	100,0%

Tableau 7 : Précision (« Precision » en anglais), Rappel (« Recall » en anglais) et F-score mesurés pour les différentes classes UOT cartographiées sur la zone PREAA

Classe	Producer Accuracy (PA)	PA Uncertainty	User Accuracy (UA)	UA Uncertainty
	(100% - Omission)	(%)	(100% - Commission)	(%)
1 Zone nue	99,30	1,36	97,62	2,66
2 Zone habitée	71,32	28,32	95,00	6,75
3 Zone arbustive	91,04	3,11	85,30	4,15
4 Savane	93,18	1,87	94,49	1,88
5 Forêt secondaire	83,43	9,61	79,66	10,27
8 Forêt dense humide	99,44	0,62	98,78	0,97
9 Eau	100,00	0,00	100,00	0,00
10 Zone cultivée	81,56	7,42	87,91	6,70
11 Marais	37,77	15,27	90,00	9,30
12 Plantation	54,69	34,37	93,55	8,65
13 Forêt de littorale	100,00	0,00	85,37	10,82
15 Mangrove	100,00	0,00	100,00	0,00
16 Infrastructure	100,00	0,00	100,00	0,00
17 Zone minière	100,00	0,00	100,00	0,00

Tableau 8 : Précision (« Precision » en anglais), Rappel (« Recall » en anglais) et F-score mesurés pour les différentes classes UOT cartographiées sur la zone PREAA

Ces résultats montrent que la carte réalisée présente un niveau de précision élevé tant au niveau global qu'au niveau de chaque classe, avec des valeurs généralement situées entre 80% et 100%. On notera toutefois les faibles valeurs de l'exactitude du producteur (« Producer accuracy ») des classes « Marais » et « Plantation », avec respectivement ~38% et ~55%. Bien que le niveau d'incertitude sur ces valeurs soit important, celles-ci proviennent d'un niveau d'omission élevé comparé à la taille de l'échantillon global pour chaque classe. Ces classes étant sous-représentées,

les quelques erreurs détectées impactent de manière conséquente la précision de cet indice. Il en ressort donc que ces classes sont probablement sous-évaluées.

IV.2.2.3 Mesure et estimation de la couverture des terres

La zone du PREAA couvre 6 914 742 ha (69 147 km²), à savoir environ 11,8% de la surface de Madagascar. Elle est dominée à environ 95% par les 5 classes suivantes : « Savane » (~35%), « Forêt dense humide » (~30%), « Zone arbustive » (~17%), « Zone nue » (~8%) et « Zone cultivée » (~5%).

La **Erreur ! Source du renvoi introuvable.** et le Tableau 9 fournissent la distribution de la couverture des différentes classes UOT, où sont repris les surfaces mesurées à partir de la carte et les surfaces « réelles », estimées statistiquement à partir des informations extraites de la matrice de confusion. Dans l'ensemble, vu le niveau de précision élevé de la carte, les surfaces « réelles » estimées sont proches des surfaces mesurées, excepté pour la classe « Marais » dont la surface estimée est plus de deux fois supérieure à celle mesurée.

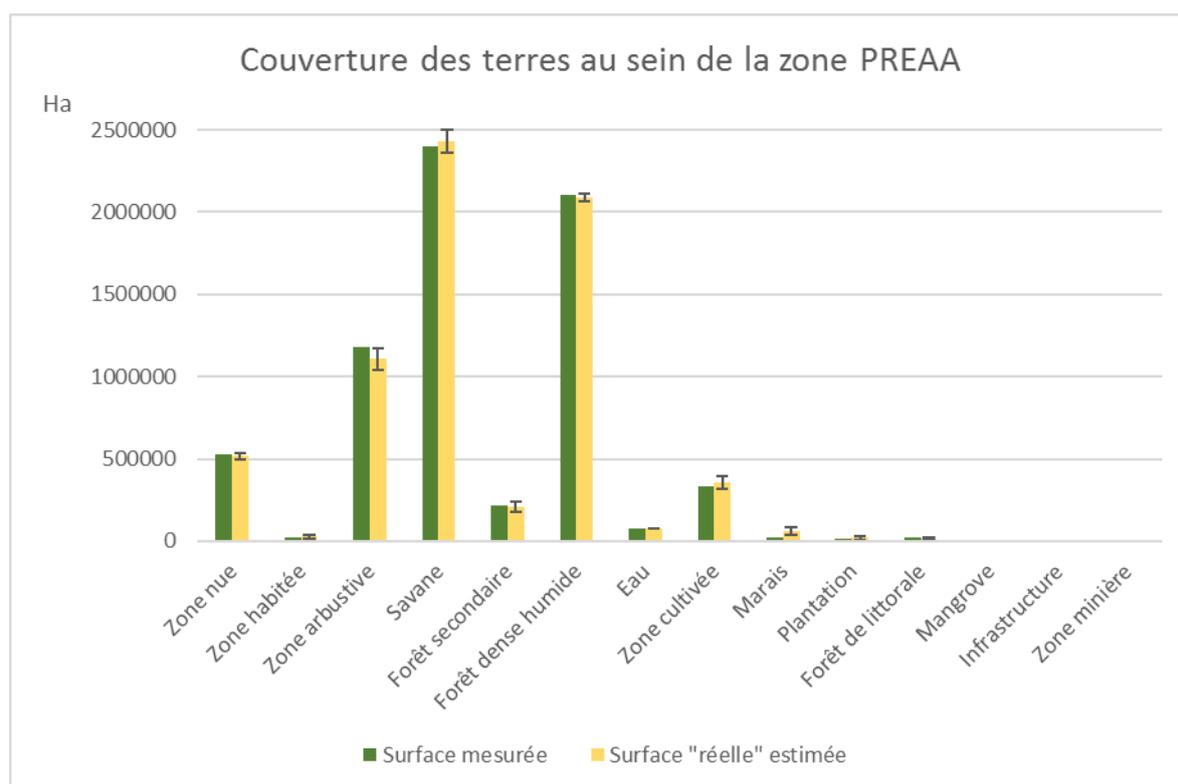


Figure 19 : Distribution des classes UOT au sein de la zone PREAA (Surfaces mesurées à partir de la carte et surfaces « réelles », estimées statistiquement)

Classe	Surface (ha)	% couverture	Surface estimée (ha)	% couverture estimée	Incertitude sur la surface estimée (ha)
Zone nue	524 993	7,59%	516 126	7,46%	16 056
Zone habitée	22 055	0,32%	29 378	0,42%	12 015
Zone arbustive	1 181 612	17,09%	1 107 186	16,01%	63 162
Savane	2 397 504	34,67%	2 431 218	35,16%	67 942
Forêt secondaire	217 409	3,14%	207 585	3,00%	33 214
Forêt dense humide	2 103 390	30,42%	2 089 344	30,22%	24 753
Eau	76 286	1,10%	76 286	1,10%	0
Zone cultivée	330 591	4,78%	356 340	5,15%	39 856
Marais	25 691	0,37%	61 218	0,89%	25 146
Plantation	10 958	0,16%	18 745	0,27%	12 052
Forêt de littorale	20 063	0,29%	17 127	0,25%	2 242
Mangrove	981	0,01%	981	0,01%	0
Infrastructure	794	0,01%	794	0,01%	0
Zone minière	2 414	0,03%	2 414	0,03%	0

Tableau 9 : Mesure cartographique des surfaces et estimation statistique des surfaces « réelles » des classes UOT de la zone PREAA

IV.3. Cartographie des UOT à l'échelle régionale

IV.3.1 Atlas cartographique régional



La zone PREAA est couverte par 5 régions :

- Région Sava
- Région Sofia
- Région Analanjirofo
- Région Atsinanana
- Région Alaotra Mangoro

Un atlas cartographique régional en version [A4](#) et [A3](#), annexe à ce document, illustre les résultats de classification pour chaque région.

IV.3.2 Couverture des terres par région

La Figure 20 et le Tableau 10 fournissent la distribution de la couverture des différentes classes UOT au sein des 5 régions présentes dans la zone PREAA. Le Tableau 11 illustre quant à lui la distribution de chaque classe UOT entre les différentes régions, exprimée en pourcentage de couverture des terres. On notera que la région d'Analanjirofo est la plus grande région de la zone, avec 31% de couverture des terres. C'est également la région où la forêt est la plus représentée avec la région de Sava. Ces deux régions sont couvertes respectivement par environ 36% et 35% de forêt

dense humide. Ensembles, elles englobent ainsi environ les deux tiers de la forêt dense humide de la zone PREAA (respectivement 37% et 28% - Tableau 11).

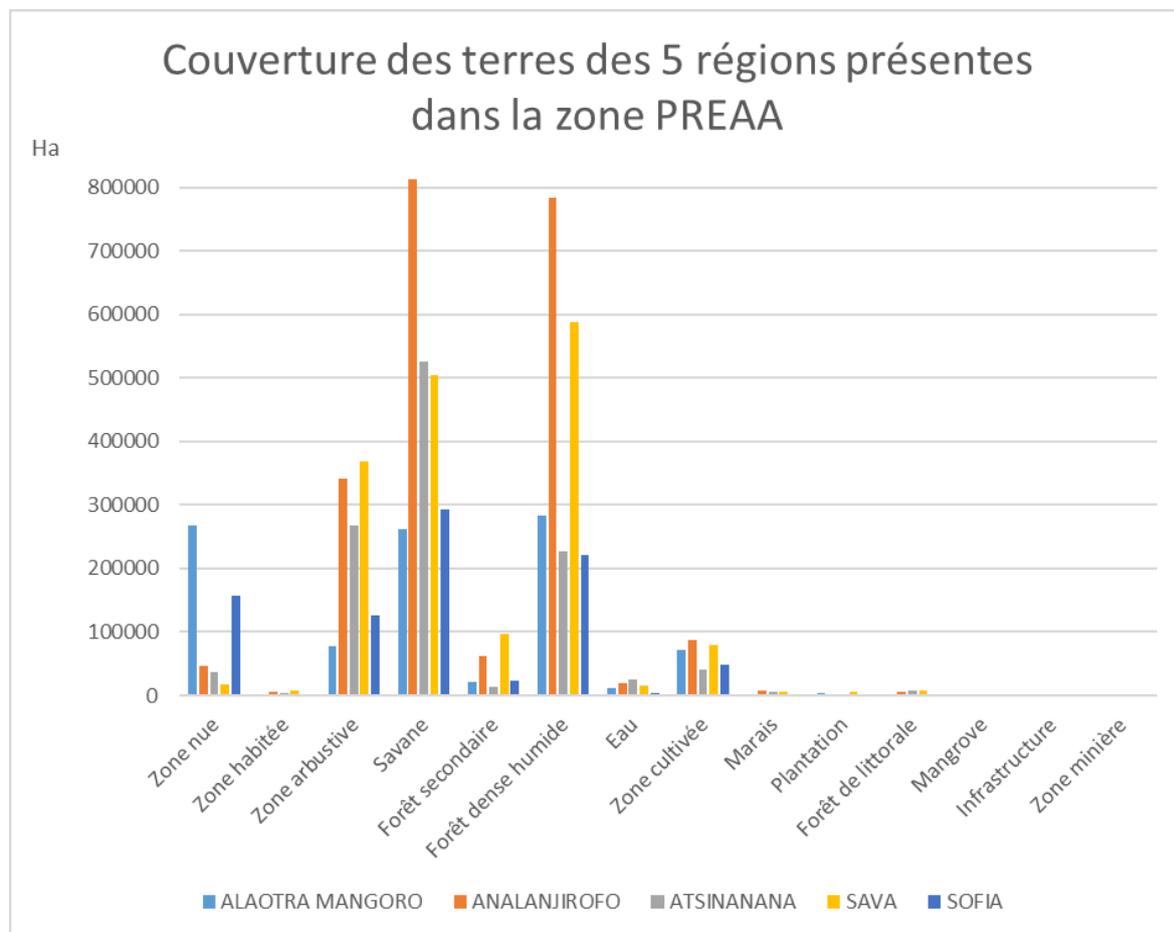


Figure 20 : Distribution des classes UOT au sein des 5 régions de la zone PREAA (Surfaces mesurées à partir de la carte)

Classe	ALAOTRA MANGORO		ANALANJIROFO		ATSINANANA		SAVA		SOFIA	
	Ha	% couverture	Ha	% couverture	Ha	% couverture	Ha	% couverture	Ha	% couverture
Zone nue	268 465	26,74%	45 797	2,11%	36 068	3,12%	17 746	1,04%	156 917	17,88%
Zone habitée	2 479	0,25%	5 791	0,27%	4 607	0,40%	7 778	0,46%	1 400	0,16%
Zone arbustive	77 171	7,69%	342 146	15,74%	267 779	23,15%	369 259	21,69%	125 259	14,27%
Savane	261 245	26,02%	812 134	37,36%	526 255	45,50%	505 136	29,67%	292 736	33,36%
Forêt secondaire	22 154	2,21%	61 709	2,84%	12 735	1,10%	97 507	5,73%	23 306	2,66%
Forêt dense humide	282 916	28,18%	784 053	36,06%	226 347	19,57%	588 286	34,55%	221 789	25,28%
Eau	11 098	1,11%	20 179	0,93%	24 644	2,13%	16 193	0,95%	4 173	0,48%
Zone cultivée	71 306	7,10%	88 287	4,06%	41 799	3,61%	79 708	4,68%	49 492	5,64%
Marais	2 678	0,27%	8 628	0,40%	6 259	0,54%	5 708	0,34%	2 418	0,28%
Plantation	3 183	0,32%	0	0,00%	1 368	0,12%	6 408	0,38%	0	0,00%
Forêt de littorale	0	0,12%	5 259	0,24%	6 986	0,60%	7 818	0,46%	0	0,00%
Mangrove	0	0,00%	87	0,00%	0	0,00%	894	0,05%	0	0,00%
Infrastructure	0	0,00%	8	0,00%	644	0,06%	142	0,01%	0	0,00%
Zone minière	1 190	0,12%	0	0,00%	1 212	0,10%	11	0,00%	0	0,00%
Total (Ha)	1 003 884		2 174 076		1 156 701		17746		877489	

Tableau 10 : Surfaces mesurées et couverture des terres pour les cinq régions couvrant la zone PREAA

PRE AA		ALAOIRA MANGORO	ANALANJIROFO	ATSINANANA	SAVA	SOFIA
Classe	Ha	Distribution par région des classes UOT de la zone PREAA (%)				
Zone nue	524 993	51,1%	8,7%	6,9%	3,4%	29,9%
Zone habitée	22 055	11,2%	26,3%	20,9%	35,3%	6,3%
Zone arbustive	1 181 612	6,5%	29,0%	22,7%	31,3%	10,6%
Savane	2 397 504	10,9%	33,9%	22,0%	21,1%	12,2%
Forêt secondaire	217 409	10,2%	28,4%	5,9%	44,8%	10,7%
Forêt dense humide	2 103 390	13,5%	37,3%	10,8%	28,0%	10,5%
Eau	76 286	14,5%	26,5%	32,3%	21,2%	5,5%
Zone cultivée	330 591	21,6%	26,7%	12,6%	24,1%	15,0%
Marais	25 691	10,4%	33,6%	24,4%	22,2%	9,4%
Plantation	10 958	29,0%	0,0%	12,5%	58,5%	0,0%
Forêt de littorale	20 063	5,9%	26,2%	34,8%	39,0%	0,0%
Mangrove	981	0,0%	8,8%	0,0%	91,2%	0,0%
Infrastructure	794	0,0%	1,0%	81,1%	17,9%	0,0%
Zone minière	2 414	49,3%	0,0%	50,2%	0,5%	0,0%
Total	6 914 742	14,5%	31,4%	16,7%	24,6%	12,7%

Tableau 11 : Répartition par région des classes UOT de la zone PREAA

IV.4. Cartographie des UOT à l'échelle des districts

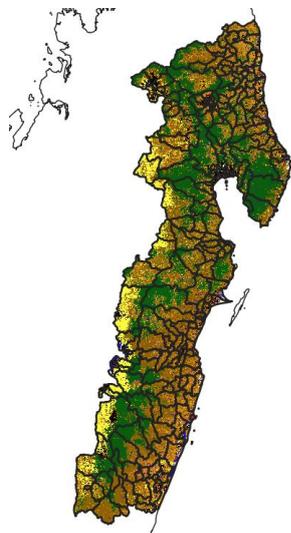
IV.4.1 Atlas cartographique des districts



La zone PREAA est couverte par 20 districts. Un atlas cartographique des districts en version [A4](#) et [A3](#), annexe à ce document et incluant les [statistiques de surfaces pour chaque classe et leur couverture des terres exprimée en pourcent](#), illustre les résultats de classification détaillés pour chaque district.

IV.5. Cartographie des UOT à l'échelle communale

IV.5.1 Atlas cartographique communal



La zone PREAA est couverte par 251 communes. Un atlas cartographique communal en version [A4](#) et [A3](#), annexe à ce document et incluant les [statistiques de surfaces pour chaque classe et leur couverture des terres exprimée en pourcent](#), illustre les résultats de classification détaillés pour chaque commune. Les sont également fournies pour chaque district dans un document annexe.

V | Conclusion

Suite à une période de formation spécifique sur les techniques et outils de télédétection, les techniciens du Laboratoire d'Observation des Forêts de Madagascar ont pu mettre en pratique leurs connaissances acquises dans le but de cartographier l'utilisation et l'occupation des terres (UOT) de la zone du Programme de Réduction des Emissions Atiala-Atsinanana (PREAA).

Une méthode semi-automatique orientée pixel, utilisant l'algorithme de classification Random forest, a été appliquée avec succès sur un total de 35 images Sentinel-2. Ces images, acquises à plus ou moins un an autour de la période de référence 2018, ont été sélectionnées sur base de critères de qualité, dont la couverture nuageuse afin d'analyser l'ensemble de la zone. La carte a été produite à partir d'un total de 17 223 zones d'entraînement ou ROI distribués sur les différentes images pour entraîner l'algorithme de classification et de post-traitement manuels. Compte tenu de la résolution des données Sentinel-2 (10m) et du seuil minimal défini pour la forêt à Madagascar, l'unité cartographique minimale a été fixée à 0,5ha. Quatorze classes UOT sont représentées, sélectionnées suivant leur présence dans la zone (confirmée par les reconnaissances terrain et les ateliers de validation) et la capacité à les discriminer avec un niveau de confiance suffisant à partir des outils et données disponibles. Ces dernières se composent de 5 classes de forêt (« Forêt secondaire », « Forêt dense humide », « Plantation », « Forêt de littorale » et « Mangrove ») et de 9 classes de non forêt (« Zone nue », « Zone habitée », « Zone arbustive », « Savane », « Eau », « Zone cultivée », « Marais », « Infrastructure » et « Zone minière »), toutes définies par le système de classification national adopté par Madagascar.

L'évaluation de la qualité de la carte réalisée à partir d'un plan d'échantillonnage aléatoire stratifié de 1 866 points montre que la carte finale possède une exactitude globale de 93,7% (+/- 1,1%). Chaque classe possède une précision globale et un indice F-score supérieur ou égal à 80%. Ces résultats d'analyse de précision et d'exactitude nous permettent de valider la carte puisque celle-ci possède des seuils de qualité supérieurs aux recommandations générales pour la REDD+ (>75%). Par ailleurs, ces travaux cartographiques ont également été validés à l'échelle régionale et nationale, à travers la tenue d'ateliers spécifiques rassemblant les experts et principales parties prenantes impliqués dans le processus REDD+. Ces résultats nous permettent également de valider la procédure utilisée tant au niveau technique qu'au niveau de la méthode de formation des agents. En effet, la méthode du « learn by doing » appliquée, où les techniciens sont formés aux techniques et procédures dans un premier temps pour ensuite mettre en pratique leurs acquis, avec l'appui du consultant, à travers des travaux de production est particulièrement efficace. Bien que plus

chronophage au départ (délai nécessaire à la formation, à l'appréhension des outils et à la production), elle s'avère être la méthode la plus efficace car elle s'inscrit dans la durée et assure une complète autonomie au niveau local (à condition bien sûr de conserver les compétences dans le temps).

Finalement, il en ressort des résultats cartographiques que la zone du PREAA couvre 6 914 742 ha, à savoir près de 12% de la surface de Madagascar, et que celle-ci est dominée à environ 95% par les 5 classes suivantes : « Savane » (~35%), « Forêt dense humide » (~30%), « Zone arbustive » (~17%), « Zone nue » (~8%) et « Zone cultivée » (~5%). Si l'on considère l'ensemble des classes forestières, celles-ci couvrent environ 34% de la zone contre 66% pour les classes non forestières. Comme indiqué en introduction, l'ensemble des informations fournies par cette carte constitue un outil de base nécessaire à la planification des activités à entreprendre dans la mise en œuvre du Programme de réduction des émissions. Ces données sont également cruciales pour la mise en œuvre de la REDD+ dans la zone PREAA car elles appuient la mise en place des Schémas d'aménagement des terres.



Figure 21 : Photos de groupe des ateliers de validation de la carte UOT du PREAA à l'échelle régionale (à gauche) et nationale (à droite) réalisés respectivement à Tamatave et Ankerana en Janvier 2020

Annexe 1 : Résultats des tests réalisés sur l'intégration de la bande SRTM comme variable d'entrée

Après analyse bibliographique des travaux de cartographie similaires réalisés dans la région, il apparaît que l'ajout des données d'élévation issues du SRTM améliorerait les performances du processus de classification. Par conséquent, afin de s'assurer d'une telle plus-value dans le cadre de notre étude, des tests ont été réalisés sur une tuile Sentinel-2 en comparant les performances de l'algorithme de classification Random Forest avec et sans l'ajout de cette variable en entrée. Cinq opérateurs ont donc réalisé ce test comparatif de manière indépendante à partir de zones d'entraînement (ROI) différentes destinées à discriminer 14 classes UOT. Pour chaque test, les mêmes ROI ont été utilisées pour comparer les performances des deux images classifiées, à savoir (i) l'image 1 composée des bandes spectrales Sentinel-2 et de l'indice de végétation NDVI et (ii) l'image 2, de même composition que l'image 1 mais complétée par la bande d'information SRTM.

Les résultats présentés ci-dessous à la Figure 22 confirment la plus-value de l'intégration de la bande SRTM sur l'index de performance globale Kappa, dont la valeur se situe entre 0 et 1 (du moins au plus performant). Pour les cinq tests réalisés indépendamment, la valeur de l'index Kappa est significativement plus élevée pour les modèles réalisés à partir de l'image 2 intégrant la donnée SRTM. En moyenne, la performance des modèles augmente de 16% avec l'ajout du SRTM.

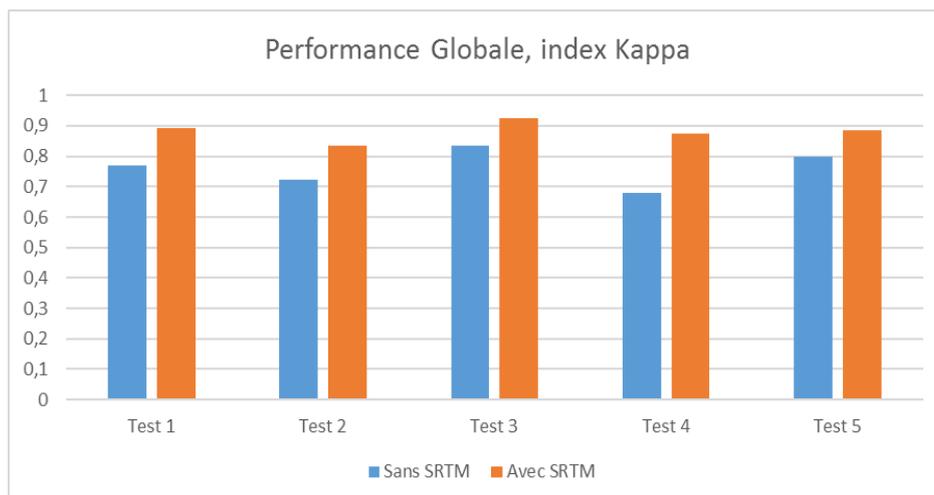


Figure 22 : Comparaison des résultats des tests réalisés sur la performances globale (index *Kappa*) de l'algorithme Random Forest, produit à partir d'une tuile multispectrale Sentinel-2 avec et sans l'ajout des données SRTM

Annexe 2 : Extrait du catalogue d'images Sentinel-2 et outil de suivi de la production

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	Pré-traitement des données																			Tra	
2	Dossier de destination					/	/	01_Input_data02_S2_L2A	02_Output_01_Preprocessing01_Stack_32739	02_Output_01_Preprocessing02_Clip_ERPAA	02_Output_01_Preprocessing03_Mosaic	02_Output_01_Preprocessing04_Cloudmask_32739	02_Output_01_Preprocessing05_Cloudmask_1emp	02_Output_01_Preprocessing06_Cloudmask_spatial	02_Output_01_Preprocessing07_Cloudmask_spatial	02_Output_01_Preprocessing08_Cloudmask_spatial	02_Output_01_Preprocessing09_Cloudmask_spatial	02_Output_01_Preprocessing10_Cloudmask_spatial	02_Output_01_Preprocessing11_Cloudmask_spatial	02_Output_01_Preprocessing12_Cloudmask_spatial	02_Output_01_Preprocessing13_Cloudmask_spatial
3	Sentinel ID	Capteur	Tile ID	Date	Cloud cover	QC des données brutes	Niveau de priorité	NDVI	Layer stack (projection originale)	Projection 32739 (UTM39S)	Georéférencement	Clip Stack_32739 (Float32) sur emprise ERPAA	Mosaic	Extraction cloudMask	Projection 32739	Editer CloudMask (v_finale)	Compilation temporelle des masques	Compilation spatiale des masques	PCI	Stats image	
4	S2A MSILIC 20180605T070231 N0206 R1 S2A	38KRD	20180605	0	0	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
5	S2B MSILIC 20180411T070229 N0206 R1 S2B	38KRD	20180411	0,77	0,77	ok	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
6	S2B MSILIC 20180918T070219 N0206 R1 S2B	38KRD	20180918	0	0	ok	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
7	S2B MSILIC 20180531T070229 N0206 R1 S2B	38KRE	20180531	0,02	0,02	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
8	S2A MSILIC 20180605T070231 N0206 R1 S2A	38KRE	20180605	0	0	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
9	S2B MSILIC 20180620T070229 N0206 R1 S2B	38KRE	20180620	0,1	0,1	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
10	S2A MSILIC 20180804T070231 N0206 R1 S2A	38KRE	20180804	0,02	0,02	ok	P2	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
11	S2B MSILIC 20180809T070219 N0206 R1 S2B	38KRE	20180809	0,11	0,11	ok	P2	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
12	S2B MSILIC 20181127T070219 N0207 R1 S2B	38KRE	20181127	2,13	2,13	ok	P2	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
25	S2B MSILIC 20180411T070229 N0206 R1 S2B	39KTA	20180411	0,05	0,05	ok	P2	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
26	S2A MSILIC 20180605T070231 N0206 R1 S2A	39KTA	20180605	0,05	0,05	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
27	S2B MSILIC 20190317T070229 N0207 R1 S2B	39KTA	20190317	6,01	6,01	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
28	S2B MSILIC 20180411T070229 N0206 R1 S2B	39KTB	20180411	0,01	0,01	ok	P2	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
29	S2A MSILIC 20180605T070231 N0206 R1 S2A	39KTB	20180605	0,03	0,03	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
30	S2A MSILIC 20190322T070231 N0207 R1 S2A	39KTB	20190322	2,71	2,71	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
31	S2B MSILIC 20180411T070229 N0206 R1 S2B	39KTU	20180411	0	0	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
32	S2A MSILIC 20180605T070231 N0206 R1 S2A	39KTU	20180605	0	0	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
33	S2B MSILIC 20180531T070229 N0206 R1 S2B	39KTV	20180531	0,01	0,01	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
34	S2A MSILIC 20180605T070231 N0206 R1 S2A	39KTV	20180605	0	0	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
35	S2B MSILIC 20180620T070229 N0206 R1 S2B	39KTV	20180620	1,17	1,17	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
36	S2A MSILIC 20180804T070231 N0206 R1 S2A	39KTV	20180804	0,79	0,79	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
37	S2B MSILIC 20180809T070219 N0206 R1 S2B	39KTV	20180809	1,39	1,39	ok	P1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok

Annexe 3 : Définition du plan d'échantillonnage pour l'analyse de l'exactitude des résultats

	1	2	3	4	5	8	9	10	11	12	13	15	16	17	Total
	Zone nue	Zone habitée	Zone arbustive	Savane	Forêt secondaire	Forêt dense humide	Eau	Zone cultivée	Marais	Plantation	Forêt de littorale	Mangrove	Infrastructure	Zone minière	Total
Area in pixels	52499256	2205542	1,18E+08	2,4E+08	21740946	2,1E+08	7628553	33059141	2569089	1095791	2006331	98104	79434	241362	691474167
<i>Wi (Mapped proportion)</i>	0,08	0,00	0,17	0,35	0,03	0,30	0,01	0,05	0,00	0,00	0,00	0,00	0,00	0,00	1,00
<i>Ui (Expected user's accuracy)</i>	0,90	0,90	0,80	0,80	0,75	0,90	0,90	0,80	0,80	0,90	0,75	0,90	0,90	0,90	
<i>Si (Standard deviation)</i>	0,30	0,30	0,40	0,40	0,43	0,30	0,30	0,40	0,40	0,30	0,43	0,30	0,30	0,30	
<i>Wi*Si</i>	0,02	0,00	0,07	0,14	0,01	0,09	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,36
											<i>SE overall accuracy</i>			0,01	
											<i>Total number of samples</i>			1307	
											<i>Total number of samples (adjusted)</i>			1800	
	<i>Sample size per stratum</i>														
<i>equal</i>	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	128,57	1800
<i>proportional</i>	136,66	5,74	307,59	624,10	56,59	547,54	19,86	86,06	6,69	2,85	5,22	0,26	0,21	0,63	1800
<i>A13</i>	126	40	279	563	59	492	41	91	40	31	41	22	21	20	1866

Références

- Belgiu, M., Dragut, L., 2016. RandomForest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-4), 239-258.
- Blaschke, T., Hay, G.J., 2001. Object-oriented image analysis and scale-space: Theory and methods for modeling and evaluating multi-scale landscape structure. *International Archives of Photogrammetry and Remote Sensing* 34 (Part 4/W5), 22-29.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Breiman, L., 2001. Random Forests, *Machine Learning*, 45(1): pp 5-32
- Bylander, T. 2002. Estimating generalization error on two-class datasets using out-of-bag estimates. *Machine Learning*, 48: pp 287–297.
- Cakir, H. I., Khorram, S., & Nelson, S. A.C. (2006). Correspondence analysis for detecting land cover change. *Remote Sensing of Environment*, 102, 306–317.
- Cochran, W. G., 1977. *Sampling techniques*. John Wiley & Sons.
- Cohen, J., 1960. A coefficient of agreement for nominales scales, *Educational and Psychological Measurement*, vol. 20, 1: pp 37 – 46
- Congalton, R. G. & Green, K. 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Second Edition. CRC Press
- Darwish, A., Leukert, K., Reinhardt, W., 2003. Image segmentation for the purpose of object-based classification. In: *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. 2003 IEEE International (3)*, 2039-2041.
- Dorren, L.K., Maier, B., Seijmonsbergen, A.C., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management* 183 (1-3), 31-46.
- Duveiller, G., Defourny, P., Desclée, B., Mayaux, P., 2008. Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. *Remote Sensing of Environment* 112 (5), 1969-1981.

European Space Agency, 2015. Sentinel-2a MSI Spectral Responses.xlsx, https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiSaR/content/sentinel-2a-spectral-responses

FAO. (2016). Map Accuracy Assessment and Area Estimation: A practical guide. Rome: National forest monitoring assessment. Working paper No.46/E

Flanders, D., Hall-Beyer, M., Pereverzoff, J., 2003. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing* 29 (4), 441-452.

Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 61, 399–409.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R., 2006. Random forests for land cover classification. *Pattern Recognition Letters*, 27: pp 294–3

Hansen, M., Dubayah, R., DeFries, R., 1996. Classification trees: an alternative to traditional land cover classifiers. *Int. J. Remote Sens.* 17, 1075–1081.

Hay, G.J., Castilla, G., 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: Blaschke, T., Lang, S., Hay, G. (Eds.), *Object Based Image Analysis*. Springer, Heidelberg, Berlin, New York, pp. 93-112.

Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, 114, 183–198.

Baisantry, Munmun & Shukla, Dericks & Bansal, Geeta. (2017). Evaluation of Object-based Image Analysis Softwares for segmentation in Remote Sensing Applications. The 20th AGILE International Conference on Geographic Information Science, At Wageningen, Netherlands.

Jobin, B., Labrecque, S., Grenier, M., Falardeau, G., 2008. Object-based classification as an alternative approach to the traditional pixel-based classification to identify potential habitat of the grasshopper sparrow. *Environmental Management* 41 (1), 20-31.

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. *Remote Sens. Environ.* 177, 89–100.

Mas, J. F., Perez-Vega, A., A. Ghilardi, S. Martinez, J. Octavio Loya-Carrillo, and E. Vega, A Suite of Tools for Assessing Thematic Map Accuracy, *Geography Journal*, vol. 2014, Article ID 372349, 10 pages, 2014. doi:10.1155/2014/372349

Mayaux, P., Eva, H., Gallego, J., Strahler, A. H., Herold, M., Agrawal, S., et al. (2006). Validation of the Global Land Cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1728–1739.

Olofsson, P., Foody, G., Herold, M., Stehman, S., Woodcock, C., & Wulder, M. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 42–57.

Olofsson, P., Foody, G.M., Stehman, S.V., & Woodcock, C.E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy, area, and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122–131

Olofsson, P., Kuemmerle, T., Griffiths, P., Knorn, J., Baccini, A., Gancz, V., et al. (2011). Carbon implications of forest restitution in post-socialist Romania. *Environmental Research Letters*, 6, 045202.

Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 2016, 187, 156–168.

Rakotomala FA, Rabenandrasana JC, Andriambahiny JE, Rajaonson R, Andriamalala F, Burren C, et al. Estimation de la déforestation des forêts humides à Madagascar utilisant une classification multitempore d'images Landsat entre 2005, 2010 et 2013. *Revue Française de Teledetection et de Photogrammetrie*. 2015; 211–212:11–23.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS Symp.*, Washington, NASA 309–317.

Sundaresan A, Varshney PK, Arora MK (2007). Robustness of change detection algorithms in the presence of registration errors. *Photogramm. Eng. Remote Sens.*, 73(4): 375-383.

Szuster, B.W., Chen, Q., Borger, M., 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography* 31, 525–532.

van der Werff, H.M.A., van der Meer, F.D., 2008. Shape-based classification of spectrally identical objects. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (2), 251-258.