



**HAL**  
open science

# Évaluation méthodologique du capteur WorldView-3 pour l'identification des arbres isolés du parc à Faidherbia albida de la zone de Bambey au Sénégal

Urcel Kalenga Tshingomba

► **To cite this version:**

Urcel Kalenga Tshingomba. Évaluation méthodologique du capteur WorldView-3 pour l'identification des arbres isolés du parc à Faidherbia albida de la zone de Bambey au Sénégal. Sciences de l'environnement. 2019. dumas-03383110

**HAL Id: dumas-03383110**

**<https://dumas.ccsd.cnrs.fr/dumas-03383110>**

Submitted on 18 Oct 2021

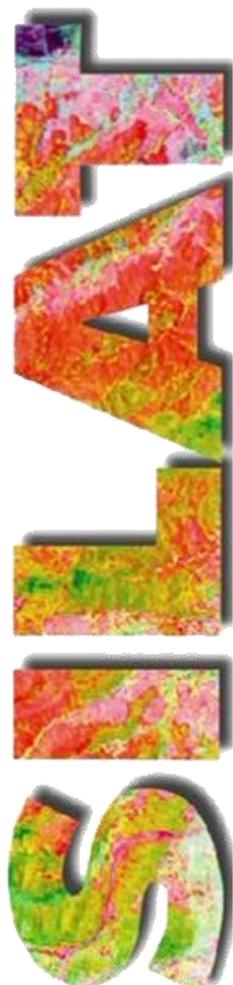
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

THESE PROFESSIONNELLE



"Evaluation méthodologique du capteur WorldView-3 pour l'identification des arbres isolés du parc à *Faidherbia albida* de la zone de Bambey au Sénégal"

**Urce! KALENGA TSHINGOMBA**

Directeurs du projet : Valérie SOTI (CIRAD - UR AÏDA)  
Camille LELONG (CIRAD - UMR TETIS)

Tuteur SILAT : Frédéric PORTET (AgroParisTech)

Rapporteurs Agnès BEGUE (UMR TETIS)  
Hassan BAZZI (UMR TETIS)

SYSTEMES  
D'INFORMATIONS  
LOCALISEES  
POUR  
L'AMENAGEMENT  
DES TERRITOIRES

**Octobre 2019**



Mastère Spécialisé de la Conférence des Grandes écoles



L'établissement AgroParisTech diffuse ce mémoire et se dégage de toute responsabilité sur le contenu proposé.

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition.

L'obligation de citation et de référencement lors de l'utilisation de ce document doit être respectée selon la licence Creative Commons apposée

Ce document est soumis aux conditions d'utilisation "Paternité – Pas d'utilisation commerciale – Pas de modification 4.0 France disponible en ligne : <https://creativecommons.org/licenses/by-nc-nd/4.0/>



## Résumé

La conservation des espèces arborées dans les systèmes agroforestiers constitue un enjeu important pour l'agriculture. Dans la zone d'étude qui nous intéresse située à Bambey à l'Est de Dakar au Sénégal, nous nous sommes intéressés au parc à *Faidherbia albida* où l'utilité de ses arbres est d'autant plus évidente, car l'écosystème y est fragile du fait de la présence de sols sableux peu fertiles et d'un climat semi-aride. D'où l'importance de développer des outils innovants pour inventorier les espèces existantes et les suivre dans le temps. Parmi ces techniques, nous proposons d'explorer les potentialités d'une image satellite WorldView-3 pour l'identification de sept espèces arborées, car elle combine à la fois une résolution spatiale sub-métrique et huit bandes spectrales sur un territoire avec une emprise au sol d'environ 20 km X 20 km. Pour ce faire, nous procédons en 4 étapes pour lesquelles plusieurs méthodes issues de la bibliographie ont été testées, puis comparées : 1) l'extraction des couronnes d'arbre par segmentation multirésolution, et par seuillage manuel du NDVI et du SAVI ; 2) l'identification des bandes et des indices radiométriques et texturaux permettant la discrimination des arbres par les méthodes d'ACP et AFD ; 3) la classification de l'image en comparant 3 algorithmes : le spectral angle mapper (SAM) ; le support vector machine (SVM) linéaire et le SVM radial ; et le random forest (RF) ; 4) enfin la mesure de la robustesse du meilleur algorithme choisi à la suite de la comparaison des précisions produites. Les résultats montrent que le seuillage de NDVI > 0,4 arrive à mieux cerner les couronnes d'arbres, contrairement à la segmentation multirésolution qui présente des débordements sur les contours des cimes, et que seuls les moyennes et les écarts-type des réflectances permettent de discriminer 4 espèces (*Azadirachta indica* ; *Balanites aegyptica* ; *Faidherbia albida* ; et *Tamarindus indica*) possédant des signatures spectrales typiques pouvant les distinguer des autres. Les espèces *Adansonia digitata*, *Anogeissus leiocarpus* et *Celtis integrifolia* ne possèdent pas de signatures spectrales spécifiques, nous obligeant à les regrouper en une classe « Autres espèces ». Dans l'ensemble, les meilleurs résultats ont été obtenus avec le SVM linéaire dont la classification finale faite à l'issue de l'entraînement sur 30 tirages aléatoires des données d'entraînement a produit de précisions globales et kappa respectives de 89 % et 0,71 pour le meilleur tirage. En perspective, afin d'améliorer ces travaux, il serait intéressant de tester l'apport complémentaire des données temporelles à haute résolution spatiale telles que PLANET et celles à très haute résolution spatiale de WV-3, en vue d'améliorer la possibilité de discrimination des espèces arborées dans une approche temporelle de l'analyse spectrale.

Mots clés : WorldView-3, Spectral Angle Mapper, Support Vector Machine, Random Forest

## Abstract

The conservation of tree species in agroforestry systems is an important issue for agriculture. In the study area located in Bambey, east of Dakar in Senegal, we focused on the park dominated by *Faidherbia albida* where the usefulness of its trees is all the more obvious because the ecosystem is fragile due to the presence of sandy soils with little fertility and a semi-arid climate. Hence the importance of developing innovative tools to inventory existing species and follow them over time. Among these techniques, we propose to explore the potential of a WorldView-3 satellite image for the identification of seven tree species, as it combines both a sub-metric spatial resolution and eight spectral bands over a territory, with a footprint of approximately 20 km X 20 km. To do this, we proceed in 4 steps for which several methods from the bibliography have been tested, then compared: 1) extraction of tree crowns by multiresolution segmentation, and by manual thresholding of NDVI and SAVI; 2) the identification of bands, and radiometric and textural indices allowing the discrimination of trees by PCA and LDA methods; 3) the classification of the image by comparing 3 algorithms: the spectral angle mapper (SAM); the linear support vector machine (SVM) and the radial SVM; and the random forest (RF); 4) finally the measurement of the robustness of the best algorithm chosen following the comparison of the precisions produced. The results show that the NDVI thresholding  $> 0.4$  is able to better define crowns of trees, unlike multiresolution segmentation, which has overflows on crown contours, and that only averages and standard deviations of reflectance allow to discriminate 4 species (*Azadirachta indica*, *Balanites aegyptica*, *Faidherbia albida*, and *Tamarindus indica*) with typical spectral signatures that can distinguish them from each others. The species *Adansonia digitata*, *Anogeissus leiocarpus* and *Celtis integrifolia* do not have specific spectral signatures, forcing them to be grouped into one class "Other species". The best results were obtained with the linear SVM whose final classification applied on 30 different data sets randomly sorted, produced respectively global and kappa accuracies of 0.89 % and 0.71 for the best data set. In perspective, in order to improve this work, it would be interesting to test the complementary contribution of high spatial resolution temporal data such as PLANET and those with very high spatial resolution of WV-3, in order to improve the possibility of discrimination of tree species in a temporal approach of spectral analysis.

Key Words: WorldView-3, Spectral Angle Mapper, Support Vector Machine, Random Forest

## **Acronymes et abréviations**

3D : Trois dimensions

ACP : Analyse en Composantes Principales

AFD : Analyse Factorielle Discriminante

DAPSA : Direction de l'Analyse, de la Prévision et des Statistiques Agricoles

GPS : Global Positioning System

IMD : Image Metadata

LIDAR : Light Detection And Ranging

NDVI : Normalized Difference Vegetation Index

OTB : OrfeoToolBox

PAN : Panchromatique

PIR1 : Proche infrarouge 1

PIR2 : Proche infrarouge 2

QGIS : Quantum GIS

RF : Random Forest

RPB : Radmin Phonebook Format

SAM : Spectral Angle Mapper

SAVI : Soil Adjusted Vegetation Index

SHP : Shapefile

SVM : Support Vector Machine

THRS : Très Haute Résolution Spatiale

TRECS : Apport de la Télédétection pour le Renforcement de la régulation Ecologique des ravageurs des Cultures de céréales sèches au Sénégal

WV-2 : WorldView-2

WV-3 : WorldView-3

XS : Multispectrale

## Table des matières

Résumé.....	I
Abstract .....	II
Lettre de mission.....	III
Acronymes et abréviations.....	IV
Table des matières .....	V
Liste des annexes.....	VII
Liste des figures.....	VIII
Liste des tableaux.....	IX
Introduction.....	1
1 Zone d'étude, données et logiciels utilisés .....	3
1.1 Zone d'étude .....	3
1.2 Données.....	3
1.2.1 Rasters.....	3
1.2.2 Vecteurs.....	5
1.3 Logiciels utilisés .....	5
2 Méthodologie .....	7
2.1 Prétraitement des images .....	7
2.2 Traitements et analyses d'images .....	8
2.2.1 Extraction des couronnes d'arbres.....	8
2.2.2 Descripteurs quantitatifs des arbres .....	10
2.2.3 Outils d'analyse de la séparabilité spectrale des espèces arborées .....	10
2.2.4 Paramétrage et choix des meilleurs algorithmes de classification .....	11
3 Résultats.....	13
3.1 Extraction des couronnes d'arbres.....	13
3.2 Radiométrie, indices spectraux et texturaux .....	14
3.3 Identification des facteurs discriminants .....	16
3.3.1 Analyse en composantes principales .....	16
3.3.2 Analyse factorielle discriminante .....	19
3.4 Validation et classification.....	22
3.5 Distribution spatiale des espèces.....	25
4 Discussion .....	26
4.1 Ressources techniques et logiciels .....	26
4.2 Délimitation des arbres .....	26
4.3 Descripteurs des classes d'arbres.....	27

4.4	Séparabilité des classes d'arbres.....	27
4.5	Sensibilité des classifications au jeu d'apprentissage .....	28
	Conclusion .....	30
	Bibliographie.....	32
	Annexes .....	- 1 -

## Liste des annexes

Annexe 1. Formules de conversion de la luminance absolue en réflectance au niveau de l’atmosphère .....	- 1 -
Annexe 2. Facteurs d’ajustement de radiométrie absolue et de conversion en réflectance TOA pour le capteur WV-3. Mise à jour du 06 juin 2017 par DigitalGlobe .....	- 2 -
Annexe 3. Indices spectraux calculés par utilisation de différentes bandes spectrales WV-3. ....	- 2 -
Annexe 4. Formulation mathématique des principaux descripteurs dérivés de la matrice de cooccurrence.....	- 3 -
Annexe 5. Description des algorithmes utilisés pour l’entraînement et la classification des espèces d’arbres de la zone de Bambey.....	- 4 -
Annexe 6. Moyennes et écarts-type des indices spectraux des espèces d’arbres de la zone de Bambey .....	- 5 -
Annexe 7. Boîtes à moustache des moyennes radiométriques des espèces d’arbres de la zone de Bambey .....	- 6 -
Annexe 8. Boîtes à moustache des écarts-type radiométriques des espèces d’arbres de la zone de Bambey .....	- 7 -
Annexe 9. Projection des axes 2-3 et 3-4 de l’ACP sur les réflectances des arbres .....	- 8 -
Annexe 10. Projection des axes 2-3 et 3-4 de l’AFD sur les réflectances des arbres .....	- 9 -
Annexe 11. Matrices de confusion produites par classification SVM linéaire sur 3 tirages différents des données d’entraînement, pour 5 et 6 classes d’espèces d’arbres. ....	- 10 -
Annexe 12. Importance des bandes spectrales dans la prédiction de la classification à partir de l’entraînement du RF sur 3 tirages différents des données d’entraînement. ....	- 12 -
Annexe 13. Comparaison des précisions apportées par les bandes multispectrales de WV-3 (16 variables) et ses bandes bleue, verte, rouge et PIR1 (8 variables) sur la classification de 30 tirages différents des données d’entraînement. ....	- 13 -

## Liste des figures

Figure 1. Localisation de la zone d'étude de Bambey, les deux scènes de l'image WV-3 couvrant la zone d'étude, et répartition des points GPS (en jaune) collectés lors des différentes missions de terrain. ....	4
Figure 2. Composition colorée infrarouge fausses couleurs des bandes spectrales WV-3 sur une portion de la zone d'étude.....	5
Figure 3. Segmentation multi-résolution sous eCognition (A) et ENVI (B), et seuillage de NDVI (C) pour l'extraction des couronnes d'arbre.....	13
Figure 4. Signatures spectrales des espèces majoritaires et minoritaires du parc à <i>Faidherbia albida</i> : A) les moyennes et B) les écarts-type des réflectances extraites des couronnes d'arbre. ....	15
Figure 5. Pourcentage d'inertie des composantes de l'ACP réalisée sur la radiométrie des espèces arborées. ....	16
Figure 6. Corrélation des moyennes et des écarts-type des réflectances des couronnes d'arbre. ....	17
Figure 7. Représentations des individus sur les plans factoriels de l'ACP. Axes 1 et 2 (A) et axes 2 et 3 (B).....	18
Figure 8. Pourcentage d'inertie des composantes de l'AFD réalisée sur la radiométrie des espèces arborées.....	19
Figure 9. Représentation des individus sur les plans factoriels de l'AFD. Axes 1 et 2 (A) et axes 1 et 3 (B).....	21
Figure 10. Précisions globales, coefficients Kappa, précisions producteur et utilisateur, produits à l'issue de la classification SVM linéaire appliquée à 30 tirages différents des données d'entraînement, pour 5 classes d'espèces d'arbres. ....	24

## Liste des tableaux

Tableau 1. Bandes spectrales et panchromatique du capteur WorldView-3. Source : (Michele A., 2017) .....	4
Tableau 2. Espèces inventoriées dans la zone d'étude. Source : (Thiaw et al., 2019) .....	9
Tableau 3. Proportion des espèces échantillonnées dans le Parc arborés de Bambey .....	14
Tableau 4. Précisions globales et coefficients Kappa sur les trois algorithmes appliqués à 3 tirages différents de données d'entraînement et de validation.....	22
Tableau 5. Matrice de confusion de la classification SVM linéaire du jeu de données ayant produit les meilleures précisions .....	24
Tableau 6. Effectif des classes d'espèces d'arbres du parc à Faidherbia albida de Bambey .....	25

## Introduction

Les services écosystémiques multiples et fondamentaux qu'offrent les arbres sont de plus en plus connus au travers de nombreux travaux (Reed *et al.*, 2017). Parmi ces multiples services, la présence des arbres et leur diversité, sont aussi connus pour favoriser la régulation naturelle des ravageurs de cultures, et devient donc une alternative agro-écologique efficace à l'utilisation des produits chimiques (Simon *et al.*, 2010 ; Brévault & Clouvel, 2019 ; Bagny-Beilhe *et al.*, 2019 ; Soti *et al.*, 2019). Des études récentes sur le Parc arboré à *Faidherbia albida* situé dans la zone de Bambey à l'Est de Thiès au Sénégal ont révélé les contributions importantes des arbres dans la pollinisation et la régulation des ennemis naturels du mil, la principale culture vivrière de cette zone (Soti *et al.*, 2019 ; Bagny-Beilhe *et al.*, 2019). Le maintien et la conservation de ces espèces arborées se heurtent aux différentes contraintes techniques permettant de connaître leur répartition et leur évolution, au-delà de simples techniques d'inventaire de terrain. Une maîtrise de la distribution spatiale de ces espèces permettrait de mettre en place des stratégies de gestion adaptées à différentes échelles.

Jusque récemment, l'identification des espèces arborées par télédétection a essentiellement été réalisée à l'aide de données issues de systèmes hyperspectraux aéroportés et LIDAR (Cho *et al.*, 2010 ; Clark *et al.*, 2005 ; Dalponte *et al.*, 2012 ; Féret & Asner, 2011 ; Feret & Asner, 2013 ; Naidoo *et al.*, 2012 ; Liu & Wu, 2018 ; Jakkula, 2006 ; Verlič *et al.*, 2014). Cependant, le coût et la difficulté de mise en œuvre des campagnes de vol en ont limité leurs applications. L'essor récent des données satellites à très haute résolution spatiale (THRS) a emmené les spécialistes et thématiciens de l'information spatiale à approfondir les analyses à de niveaux de détails plus fins sur des échelles d'observation relativement grandes. Avec des capteurs et des méthodes en plein développement, les perspectives d'application se multiplient à de nombreux domaines dont la cartographie des espèces arborées constituant l'objet d'étude de ce projet. En effet, les dernières générations de plateformes satellitaires, jumelant des constellations de capteurs à THRS, permettent désormais l'acquisition de plusieurs bandes spectrales avec une emprise au sol d'environ 20 km et peuvent ainsi constituer des alternatives moins coûteuses que les capteurs hyper spectraux pour l'identification des espèces arborées.

WorldView-2 (WV-2) est l'un de ces capteurs, qui s'est récemment montré prometteur pour l'identification de différents types d'écosystèmes forestiers que sont la forêt tempérée (Immitzer *et al.*, 2012 ; Waser *et al.*, 2014), la forêt urbaine (Verlič *et al.*, 2014 ; Rasieh *et al.*, 2016), la forêt de mangrove (Heenkenda *et al.*, 2014), les plantations forestières (Peerbhay *et al.*, 2014) et la savane arborée (Cho *et al.*, 2012). Malgré les avantages pratiques et économiques de cette source de données, peu d'études de ce type ont été réalisées dans les écosystèmes africains. On peut néanmoins noter

quelques travaux intéressants. Dans le Kruger, en Afrique du Sud, Cho *et al.* (2012) ont classés cinq espèces de savane en utilisant des images WV-2 avec une précision globale de 77%. Peerbhay *et al.* (2014) ont relevé une précision globale de 85,4% lors de la classification de six espèces arborées avec WV-2, au KwaZulu-Natal en Afrique du Sud. Adelabu & Dube (2015) ont rapporté une précision globale de 88,8% à l'aide d'images THRS QuickBird pour classifier cinq espèces d'arbres au Botswana. On remarque que ces recherches se sont concentrées exclusivement sur les régions du sud de l'Afrique. Enfin, une étude récente (Karlson *et al.*, 2016) a été réalisée avec des images WV-2 en Afrique de l'Ouest, dans un parc agroforestier du Burkina Faso et a permis d'identifier avec une précision globale de 83,4%, les cinq espèces arborées les plus fréquentes.

Ces résultats satisfaisants, nous ont encouragés à envisager la mise en place d'une méthode d'identification et de représentation spatiale des principales espèces végétales composant le parc arboré à *Faidherbia albida* situé dans la zone de Bambey, à partir des données satellites du capteur WorldView-3 (WV-3) acquises en décembre 2017.

L'objectif de cette étude est donc de définir une méthode de cartographie des espèces arborées en zone semi-aride, à partir d'une étude de cas dans le parc à *Faidherbia albida* de la zone de Bambey au Sénégal, à partir des images WV-3. Pour ce faire, on évaluera les potentialités offertes par les 8 bandes spectrales du visible et proche infrarouge, plus la bande panchromatique, du capteur WV-3, pour l'identification de différentes espèces présentes dans la zone d'étude. L'enjeu est de définir une démarche méthodologique et technique permettant de discriminer ces espèces, à l'échelle individuelle, avec la plus grande précision possible, pouvant en outre servir ultérieurement de référence pour une application dans d'autres écosystèmes similaires à notre zone d'étude.

Les étapes qui permettront d'atteindre cet objectif sont :

- Développer une technique appropriée d'extraction des couronnes d'arbres de la zone d'étude ;
- Analyser les propriétés spectrales et texturales des espèces d'arbres majoritaires et minoritaires afin d'établir un classement en fonction de meilleures propriétés discriminatoires ;
- Optimiser la procédure d'apprentissage permettant de maximiser la précision dans la reconnaissance des classes d'espèces arborées ;
- Evaluer différents algorithmes de classification ;
- Réaliser une cartographie valide des espèces arborées de la zone d'étude.

# 1 Zone d'étude, données et logiciels utilisés

## 1.1 Zone d'étude

La zone d'étude couvre une superficie de 400 km<sup>2</sup> (20 km X 20 km) centrée sur le village Dangalma appartenant au département de Bambey, dans la région de Diourbel au Sénégal. Elle est située à environ 100 km à l'Est de Dakar, entre les latitudes Nord de 14,6442° et 14,7187°, et les longitudes Ouest de 16,4629° et 16,6478° (Figure 1). Elle est caractérisée par un climat semi-aride avec une courte période pluvieuse de Juillet à octobre et des précipitations moyennes annuelles qui varient entre 400 et 600 mm (Badiane *et al.*, 2000). Son paysage est caractérisé par un parc arboré où le *Faidherbia albida* est majoritaire, associé aux cultures de mil et d'arachide. Caractérisé par de petites exploitations familiales pauvres dont la taille des parcelles agricoles excèdent rarement les 0,25 ha (DAPSA, 2014), les pratiques agricoles y sont peu mécanisées et l'utilisation des produits chimiques rare ou nulle. D'où l'importance de la conservation des arbres et de leurs biodiversités pour la régulation des ravageurs des cultures, mais aussi pour l'amélioration de la fertilité des sols, la limitation de l'érosion, et bien d'autres services écosystémiques.

Dans la zone d'étude de Bambey qui nous concerne, ces arbres majoritairement composés de *Faidherbia albida* sont isolés et tous âgés d'environ 100 ans, et leurs cimes sont facilement identifiables par l'œil humain en observant une composition colorée de l'image WV-3 (Figure 2).

## 1.2 Données

### 1.2.1 Rasters

Deux scènes d'archive WorldView-3 de notre zone d'intérêt ont été acquises le 04 décembre 2017 en saison sèche. En effet, durant cette période, la plupart des espèces arborées conservent encore leurs feuilles vertes, les nuages sont quasiment absents et l'atmosphère est généralement dégagée, permettant la prise d'images de qualité. Ces scènes comprennent 8 bandes multispectrales (visibles et proche infrarouges) et la bande panchromatique, dont les caractéristiques spatiales et spectrales sont détaillées dans le Tableau 1. Codés en 16 bits, la résolution spatiale des pixels est de 120 cm pour les bandes multispectrales (XS) et 30 cm pour la panchromatique (PAN).

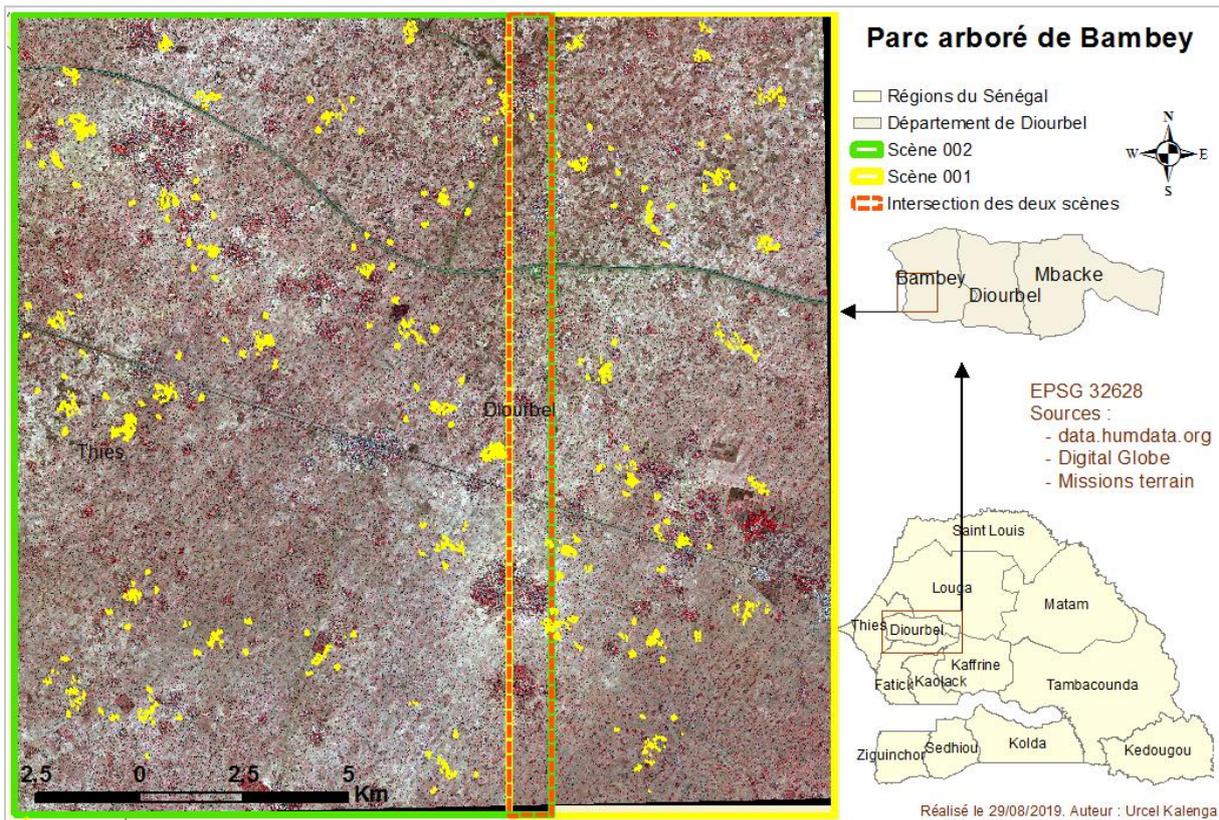


Figure 1. Localisation de la zone d'étude de Bambe, les deux scènes de l'image WV-3 couvrant la zone d'étude, et répartition des points GPS (en jaune) collectés lors des différentes missions de terrain.

Les produits WV-3 acquis sont de niveau 2A, les pixels radiométriquement corrigés en radiance relative (Michele A., 2017). Les fichiers des métadonnées de l'image (au format .IMD) et de la géométrie de la prise de vue (au format .RPB) du capteur ont été également fournis. Ils ont été utilisés principalement dans l'étape des prétraitements qui sera abordée dans la partie qui suit.

Tableau 1. Bandes spectrales et panchromatique du capteur WorldView-3. Source : (Michele A., 2017)

Bande spectrale	Longueur d'onde centrale (nm)	Largeur de bande spectrale (nm)	Taille du pixel (m)
Bleu « Côtier »	427,4	0,0405	1,20
Bleu	481,9	0,054	1,20
Vert	547,1	0,0618	1,20
Jaune	604,3	0,0381	1,20
Rouge	660,1	0,0585	1,20
Red-edge	722,7	0,0387	1,20
Proche infrarouge 1	824	0,1004	1,20
Proche infrarouge 2	913,6	0,0889	1,20
Panchromatique	649,4	0,2896	0,30



Figure 2. Composition colorée infrarouge fausses couleurs des bandes spectrales WV-3 sur une portion de la zone d'étude.

### 1.2.2 Vecteurs

Les missions de terrain effectuées dans le cadre des récentes études sur les arbres de la zone d'étude ont permis de collecter le positionnement géographique des arbres toutes les années de 2013 à 2018. Cette base de données a ensuite été complétée par des données collectées au cours de notre mission en juin 2019. Au total, 4869 points GPS d'arbres ont donc constitué notre base de données d'entraînement pour la zone d'étude de 400 km<sup>2</sup>. Ces points ont été collectés prioritairement autour de 45 champs dont on a mesuré pendant 4 années les taux de régulation naturelle des prédateurs des ennemis des cultures de mil, à travers un échantillonnage par transect basé sur un gradient de densité des arbres dans le paysage (Soti *et al.*, 2018). Les données géographiques au format shapefile (.shp) des limites administratives du Sénégal ont aussi été récupérées pour la représentation spatiale de la zone d'étude et la planification des missions de terrain.

### 1.3 Logiciels utilisés

Le choix des logiciels s'est basé prioritairement sur des produits libres permettant de réaliser les traitements envisagés. Les limites de ces logiciels libres, nous ont contraintes pour certains traitements d'utiliser certains logiciels payants.

Trois types de logiciels ont donc été utilisés :

- a. Les analyses et les traitements en mode raster ont été réalisés avec OrfeoToolBox (OTB) 6.6.1 libre d'accès, ENVI 5.5 et eCognition à licence payante. L'utilisation d'OTB, un logiciel libre, a été prioritaire pour les traitements en mode raster pour sa simplicité d'utilisation et notamment pour l'exécution facile de fonctions en ligne de commande. ENVI et eCognition ont été utilisés pour des fonctions spécifiques qui ne sont pas encore totalement implémentées dans OTB ;
- b. Les analyses statistiques et les traitements des données tabulaires ont été réalisés avec R-studio ;
- c. Enfin, les traitements vectoriels et les représentations cartographiques ont été réalisés avec QGIS 3.6 et ArcMap 10.4.1

## 2 Méthodologie

### 2.1 Prétraitement des images

Les images ont été livrées sans correction particulière, il a d'abord été nécessaire d'appliquer les prétraitements requis pour travailler avec une donnée parfaitement géoréférencée (donc superposable à des données de terrain ou tout autre données spatiales) et permettant d'analyser les réflectances spectrales caractéristiques des espèces d'arbres de façon reproductible. Pour ce faire, nous avons procédé en deux étapes :

- Nous avons d'abord appliqué une calibration radiométrique complète afin d'uniformiser les valeurs radiométriques entre les deux scènes livrées et permettre une reproductibilité des analyses sur d'autres images acquises à d'autres dates ou sur d'autres zones éco-géographiques similaires. Cette calibration consiste à convertir la valeur des pixels (initialement en luminance relative) en réflectance au niveau de l'atmosphère (TOA, pour Top Of Atmosphere) à partir des paramètres du capteur et des propriétés spectrales des bandes (Showengerdt, 2007), en appliquant successivement les deux équations explicitées en Annexe 1.
- Dans un second temps, nous avons procédé à une orthorectification des deux scènes afin de corriger les distorsions géométriques possibles dues aux conditions de prise de vue et aux effets du relief. Cette orthorectification a été faite sous ENVI 5.5, par rééchantillonnage bilinéaire sur la base des informations du modèle géométrique de prise de vue pour chacune des scènes et du modèle d'élévation SRTM à 30 m de maille.

Après avoir réalisé ces prétraitements, il a encore été observé une légère variation dans la réflectance des paires de pixels homologues sur la partie commune aux deux tuiles couvrant près de 20 km<sup>2</sup>. Un important travail de normalisation a alors été mené pour faire correspondre la réflectance de ces pixels homologues, en testant plusieurs options de régressions linéaires (à partir des moyennes, des médianes, de points invariants spectraux, etc.), mais cela n'a pas abouti à de bons résultats (améliorations locales, aggravations ailleurs, etc.). Nous avons donc fait le choix dans un premier temps de concentrer nos traitement sur la scène (002), celle couvrant la plus grande partie de la zone d'étude (environ 254 km<sup>2</sup>, soit près de 63% de l'étendue globale), et de garder la deuxième scène (001, d'environ 146 km<sup>2</sup>) pour tester la reproductibilité des chaînes de traitements développées.

## 2.2 Traitements et analyses d'images

Quatre étapes ont été réalisées pour l'identification des espèces arborées dans notre zone d'intérêt : l'extraction des couronnes d'arbres ; le calcul des descripteurs quantitatifs ; l'analyse de la séparabilité spectrale des différentes espèces arborées ; et le paramétrage et l'entraînement des algorithmes de classification.

### 2.2.1 Extraction des couronnes d'arbres

La détection et la délimitation des couronnes d'arbres reste encore aujourd'hui une étape délicate des applications de télédétection (Jakubowski *et al.*, 2013), car plutôt difficile à optimiser et pourtant cruciale pour toute analyse des objets individuels. Parmi les approches les plus efficaces, on retrouve l'utilisation de données 3D, comme par exemple dans les travaux d'Harikumar *et al.* (2019) qui proposent une approche de détection et de délimitation des couronnes d'arbres dans une forêt à dominance de conifères par l'utilisation des nuages LIDAR 3D, ou encore Naidoo *et al.* (2012b) ou Tiede *et al.* (2008) qui ont utilisé la combinaison des données optiques et LIDAR pour améliorer la précision dans l'extraction des couronnes d'arbres. Dans notre étude, nous ne disposons que de données optiques beaucoup moins riches en résolution spatiale que le LIDAR, mais nous avons la chance d'avoir un bon contraste radiométrique entre les arbres et leur environnement immédiat constitué majoritairement de sol nu. Nous avons donc plutôt testé différentes techniques de segmentation et de seuillage en intégrant l'information de contraste végétal grâce aux indices de végétation (NDVI et SAVI), dans l'esprit des travaux de Pu & Landry (2012), Singh *et al.*, (2012) et Hsieh *et al.* (2017), puis nous avons comparé les résultats de ces différentes méthodes pour valider le meilleur mode d'extraction des couronnes d'arbres pour la suite des traitements, sur des critères de précision de délimitation de l'emprise spatiale de l'arbre la plus proche de celle envisagée par la photointerprétation.

Nous avons d'abord appliqué une segmentation multi-résolution à l'échelle des couronnes d'arbre sous eCognition et ENVI, sur un fichier multi-bande (Layer Stack) composée des 8 bandes multispectrales et de la bande panchromatique de WV-3 ainsi que les deux indices spectraux, NDVI et SAVI très utilisés pour l'étude de la végétation. Afin d'obtenir le meilleur résultat de segmentation, plusieurs tests ont été appliqués en faisant varier les différents paramètres de segmentation multirésolution tels que les paramètres d'échelle, les poids apportés aux différents plans, l'influence de la radiométrie vs. la forme des objets... Pour eCognition et ENVI, les plages de valeurs optimales de forme étaient comprises entre 0,1 à 0,2 et 0,8 à 0,9 pour l'hétérogénéité. En parallèle, nous avons aussi testé sous OTB un seuillage manuel sur les valeurs des indices NDVI et SAVI pour extraire les couronnes

d'arbres. Onze seuillages ont été appliqués à chacun des indices spectraux, de 0,3 à 0,4 pour le NDVI et de 0,2 à 0,3 pour le SAVI. Les meilleurs résultats, permettant une extraction pertinente sur l'ensemble de la zone d'étude, ont été obtenus avec un seuillage des pixels ayant une valeur de NDVI supérieure à 0,35. Les meilleurs compromis entre disparition des petits arbres et segmentation de surfaces non arborées nous ont finalement conduits à retenir la valeur de 0,40 comme seuil pour le NDVI. Afin d'améliorer la délimitation des couronnes d'arbre, nous avons par la suite appliqué un filtre médian (Kernel = 3) pour régulariser le seuillage, puis un filtre morphologique d'ouverture (fenêtre = 3 X 3) pour séparer les polygones des arbres jointifs.

Enfin, les objets arbres ont été vectorisés, puis filtrés afin d'éliminer les surfaces cultivées ou de végétation non arborée, ou encore les massifs de plusieurs arbres jointifs impossibles à séparer. Pour ce faire, nous avons passé un filtre pour ne garder que les objets « arbre isolé », ayant une surface comprise entre 14 et 300 m<sup>2</sup> et un indice de compacité inférieur ou égal à 1,68.

Dans la suite de cette étude, nous avons donc procédé à des analyses orientées objet, où l'on ne considère pas les pixels individuellement, mais tous les pixels appartenant au polygone vectorisé comme faisant partie d'un tout. Les variables utilisées pour les différents calculs sont donc les valeurs globales (moyennes, et écarts-type) représentatives de l'objet « arbre individuel ». Une jointure spatiale a aussi été créée avec les données de terrain collectées lors des différentes missions dans la zone d'étude (géoréférencement par GPS), afin d'établir une correspondance entre les deux sources de données : image satellite vs. vérité-terrain. Seules les espèces majoritaires et minoritaires d'après l'inventaire réalisé par Thiaw *et al.* (2019) ont été retenues dans cette étude (Tableau 2).

Tableau 2. Espèces inventoriées dans la zone d'étude. Source : (Thiaw *et al.*, 2019)

Espèces	Nbre échantillonné	Pourcentage	Pourcentage groupe
Espèces majoritaires			
<i>Faidherbia albida</i>	4950	57,29	89,5
<i>Balanites aegyptica</i>	1850	21,41	
<i>Adansonia digitata</i>	930	10,76	
Espèces minoritaires			
<i>Tamarindus indica</i>	146	1,69	5,6
<i>Anogeissus leiocarpus</i>	120	1,39	
<i>Azadirachta indica</i>	116	1,34	
<i>Celtis integrifolia</i>	104	1,2	
Espèces rares			
<b>Toutes</b>	<b>423</b>	<b>4,9</b>	<b>4,9</b>

## 2.2.2 Descripteurs quantitatifs des arbres

### A. Réflectance et indices spectraux

Pour chaque arbre de la zone d'étude, nous avons calculé la moyenne et l'écart-type de sa réflectance dans chacune des 8 bandes spectrales (16 variables), ainsi que 14 indices spectraux susceptibles d'apporter une discrimination entre les arbres (Annexe 3). Chaque arbre est alors décrit potentiellement par 30 variables radiométriques explicatives.

### B. Analyses texturale

La matrice de co-occurrence des niveaux de gris est un outil statistique d'ordre 2 de description des textures d'images satellites, de référence pour beaucoup d'auteurs (Puissant *et al.*, 2005 ; Haralick *et al.*, 1973 ; Regniers, 2015 ; Marrakchi, 2012 ; Lefebvre, 2011), car elle prend en compte à la fois la distribution locale spatiale et spectrale des pixels au sein d'un voisinage ou d'un objet. Le calcul d'une matrice de co-occurrence s'applique à une image mono-bande, et consiste à repérer le nombre d'occurrences de paires de niveaux de gris séparés par une distance  $d$  dans une direction définie par un vecteur de déplacement  $(dx, dy)$  (Haralick *et al.*, 1973). Des indices texturaux sont ensuite dérivés de cette matrice afin de décrire quantitativement la texture (homogénéité, caractère périodique ou non, niveau de contraste, taille des éléments constitutifs, etc...), comme décrit par Maillard (2003).

Nous avons choisi d'appliquer ces calculs respectivement à la bande panchromatique et aux bandes - vert, rouge et PIR1- semblant présenter le plus de variations au sein d'un arbre. Etant donnée la finesse de la texture interne à un arbre (Patel *et al.*, 2008), nous avons considéré une distance de voisinage de 1 pixel (= fenêtre de 3 X 3 pixels ), pour trois orientations (0°, 45° et 90°). Quatre indices ont alors été calculés : l'entropie, le contraste, la corrélation et l'homogénéité (Annexe 4).

## 2.2.3 Outils d'analyse de la séparabilité spectrale des espèces arborées

Les signatures spectrales de chacune des 7 espèces majoritaires et minoritaires (Tableau 2) sont représentées dans la Figure 4. L'analyse de ces courbes doit permettre de vérifier si les classes d'arbre ont une signature suffisamment distincte pour être discriminable. Cette analyse doit tenir compte de la variabilité au sein de la classe, qui tend à confondre les enveloppes de classe les unes avec les autres. Il est probable que certaines classes doivent être fusionnées pour clarifier la séparabilité entre classes plus larges.

L'analyse en composantes principales (ACP) (Figure 7) projette une base de vecteurs (ici les descripteurs radiométriques des arbres) dans un nouvel espace sur la base de la diagonalisation de la

matrice de covariance des données. Elle permet donc de vérifier leur séparabilité sur des plans factoriels. Les axes définissant le mieux cette séparabilité définiraient les nouvelles variables à considérer pour la discrimination des espèces d'arbres. Comme l'ACP, l'analyse factorielle discriminante (AFD) projette les données dans un nouvel espace, dans le but de représenter les individus d'arbres dans un espace à dimensionnalité réduite maximisant le rapport entre la variance interclasse et la variance intra-classe des espèces, tout en garantissant le maximum de séparabilité de classes (Tharwat *et al.*, 2017 ; Ghogh et al., 2019). L'AFD a été appliquée sur trois jeux de données différents : les moyennes et écarts-type des 8 bandes spectrales ; les indices spectraux ; et la combinaison des deux.

#### **2.2.4 Paramétrage et choix des meilleurs algorithmes de classification**

Nous avons testé trois algorithmes de classification : le spectral angle mapper (SAM), le séparateur à vaste marge (SVM), et le random forest (RF), utilisés dans différents travaux (Liu & Wu, 2018 ; Madonsela *et al.*, 2017 ; Yonezawa, 2007 ; Cho *et al.*, 2010 ; Naidoo *et al.*, 2012 ; Immitzer *et al.*, 2012), et ayant produit des résultats satisfaisants. Notre démarche a été de tester ces trois algorithmes, décrits en Annexe 5, en les appliquant sous différents logiciels (OTB, R et ENVI), afin de trouver la méthodologie permettant d'obtenir les meilleures précisions tout en restant robuste sur des tests appliqués à différents jeux d'entraînement. Ces tests ont été créés sous R par tirages successifs des données d'entraînement, aléatoires et indépendants, afin de définir pour chacun des tests une base d'apprentissage d'une part (70% des arbres échantillonnées, pour chaque classe individuellement) et une base de validation d'autre part (30% des arbres échantillonnées, pour chaque classe individuellement). Dans un premier temps, les 3 algorithmes (SAM, SVM et RF) ont été testés sur 3 jeux de données issus de ce type de tirage, pour en comparer la qualité et la stabilité de précision de classification, ainsi que la sensibilité à la base d'apprentissage. Une fois la meilleure méthodologie identifiée, elle a été appliquée à nouveau à une trentaine de tirages différents du jeu d'apprentissage, afin de vérifier sa bonne stabilité et d'en sélectionner au final le jeu de données d'apprentissage permettant de générer le meilleur modèle à appliquer à l'ensemble des arbres de la zone d'étude.

Le SVM et le RF peuvent être appliqués sur des données vectorielles. Ils ont l'avantage de pouvoir contourner le problème de coût temporel de traitement lié aux données rasters à très haute résolution spatiale. De plus, ces deux algorithmes de classification sont adaptés aux traitements en mode objet et facilement mis en œuvre dans notre chaîne de traitements. Le SAM quant à lui s'applique qu'à des données raster. Il a donc fallu procéder à une conversion de notre base de données « arbres » en raster, puis procéder aux tests successifs qui ont été appliqués plus d'une dizaine de fois pour les différents jeux de données d'entraînement. Chacun de ces tests a pris en moyenne 13 heures de calcul

vu la lourdeur du jeu de données. Compte tenu de ces difficultés techniques, nous nous sommes fixés de ne retenir cette approche que si les précisions de classification obtenues sont nettement meilleures qu'avec les deux autres méthodes.

Pour la sélection des variables optimales du SAM, nous avons utilisé, sous ENVI, la technique « Band Add On » (BAO), qui implique d'appliquer l'algorithme sur deux bandes et d'ajouter itérativement les autres bandes spectrales (Madonsela *et al.*, 2017 ; Keshava, 2004), en vue d'obtenir le lot des variables permettant de minimiser les angles entre les espèces arborées. Le paramétrage du SVM a nécessité de calculer les bons paramètres, le « coût » et le « gamma » décrits en Annexe 5. Nous avons effectué des tests sous R afin d'identifier leurs valeurs optimales du nombre d'arbres et des variables à considérer à chaque arbre.

### 3 Résultats

#### 3.1 Extraction des couronnes d'arbres

Comme le montre la Figure 3, les techniques d'extraction de couronnes d'arbres choisies ont produits des résultats différents. La segmentation multi-résolution sous eCognition (Figure 3 A) et ENVI (Figure 3 B) présente des surfaces incluant plusieurs types d'objets. Quelques débordements sur les limites des couronnes d'arbres peuvent être observés. Le seuillage du SAVI n'a pas été plus effectif pour la mise en évidence des arbres à faible densité foliaires et de petite taille. Avec le seuillage du NDVI, les couronnes d'arbres sont plus ou moins bien cernées. Quelques individus de l'espèce *Adansonia digitata* n'ont pas été bien identifiés à cause de leurs faibles valeurs de NDVI qui sont pour certaines inférieures au seuillage fixé pour l'extraction des contours.

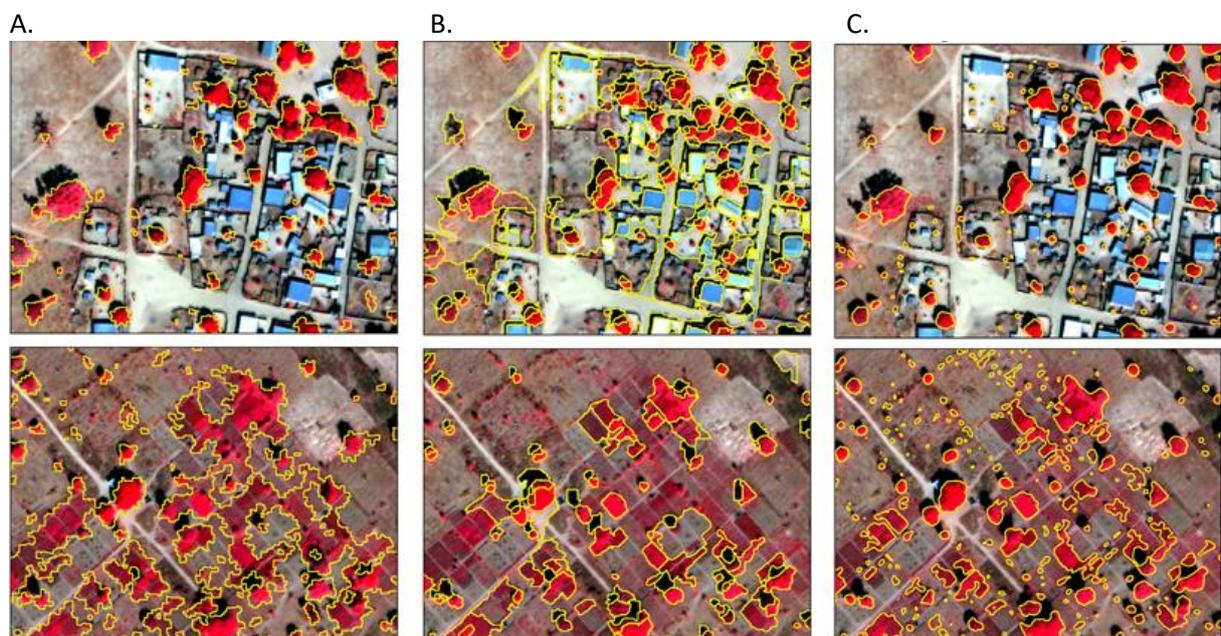


Figure 3. Segmentation multi-résolution sous eCognition (A) et ENVI (B), et seuillage de NDVI (C) pour l'extraction des couronnes d'arbre.

Appliqué à l'ensemble de la zone d'étude, le seuillage du NDVI suivi des filtres morphologique et médian ont permis d'extraire 134715 couronnes d'arbres dont les surfaces sont comprises entre 14 m<sup>2</sup> et 300 m<sup>2</sup>, après vectorisation et filtrage basés sur l'aire et l'indice de compacité inférieur ou égal à 1,68. Nous avons, par la suite, réalisé une jointure spatiale entre les couronnes d'arbre extraites et les points de géolocalisation des arbres inventoriés sur le terrain, afin de focaliser la suite des analyses radiométriques et texturales sur des objets arbres dont l'espèce était connue. En tout, 1121 couronnes d'arbre regroupant à la fois les espèces majoritaires et minoritaires ont été sélectionnées sur la tuile

02 de l'image WV-3 pour la suite des traitements, soit 0,83% de l'ensemble des couronnes de la zone d'étude, dont les proportions sont présentées dans le Tableau 3.

Tableau 3. Proportion des espèces échantillonnées dans le Parc arborés de Bambey

Espèces	Effectif	Entraînement (70 %)	Validation (30 %)
<i>Adansonia digitata</i>	41	29	12
Autres espèces	82	58	24
<i>Azadirachta indica</i>	21	15	6
<i>Balanites aegyptica</i>	131	92	39
<i>Faidherbia albida</i>	823	577	246
<i>Tamarindus indica</i>	23	17	6
<b>Total</b>	<b>1121</b>	<b>788</b>	<b>333</b>

### 3.2 Radiométrie, indices spectraux et texturaux

Les signatures spectrales entre les espèces d'arbres de Bambey sont très proches tant sur les moyennes que sur les écarts-type (Figure 4). Certes, les courbes des valeurs moyennes de ces deux paramètres se séparent dans certaines bandes spectrales, notamment le proche infrarouge, mais leurs écarts-type sont très larges et se recouvrent les uns les autres. Il en est de même pour les indices spectraux (NDVI et SAVI) calculés qui, bien que présentant des moyennes de valeurs différentes, ont une dispersion trop forte pour bien caractériser des classes spectrales (Annexe 6). Ces similitudes de signatures spectrales des espèces arborées montrent durant cette première étape d'analyse que les méthodes basiques de classification ne seront pas d'une grande efficacité.

Dans le détail (Figure 4), nous voyons que l'espèce *Azadirachta indica* se distingue spectralement des autres espèces dans les bandes du Red-edge, du PIR1 et du PIR2 par rapport à leurs écarts-type respectifs. Cette espèce pourrait donc être la plus facilement séparable.

Malgré les paramètres de traitement indiqués (cf. 2.2.2) pour discriminer les micro-textures au sein des couronnes d'arbre, la variabilité des textures d'un arbre à l'autre est beaucoup trop faible pour être instructive. La seule piste de solution consiste donc à approfondir cette analyse du point de vue spectral afin de différencier les groupes spectraux des espèces d'arbres de la zone d'étude.

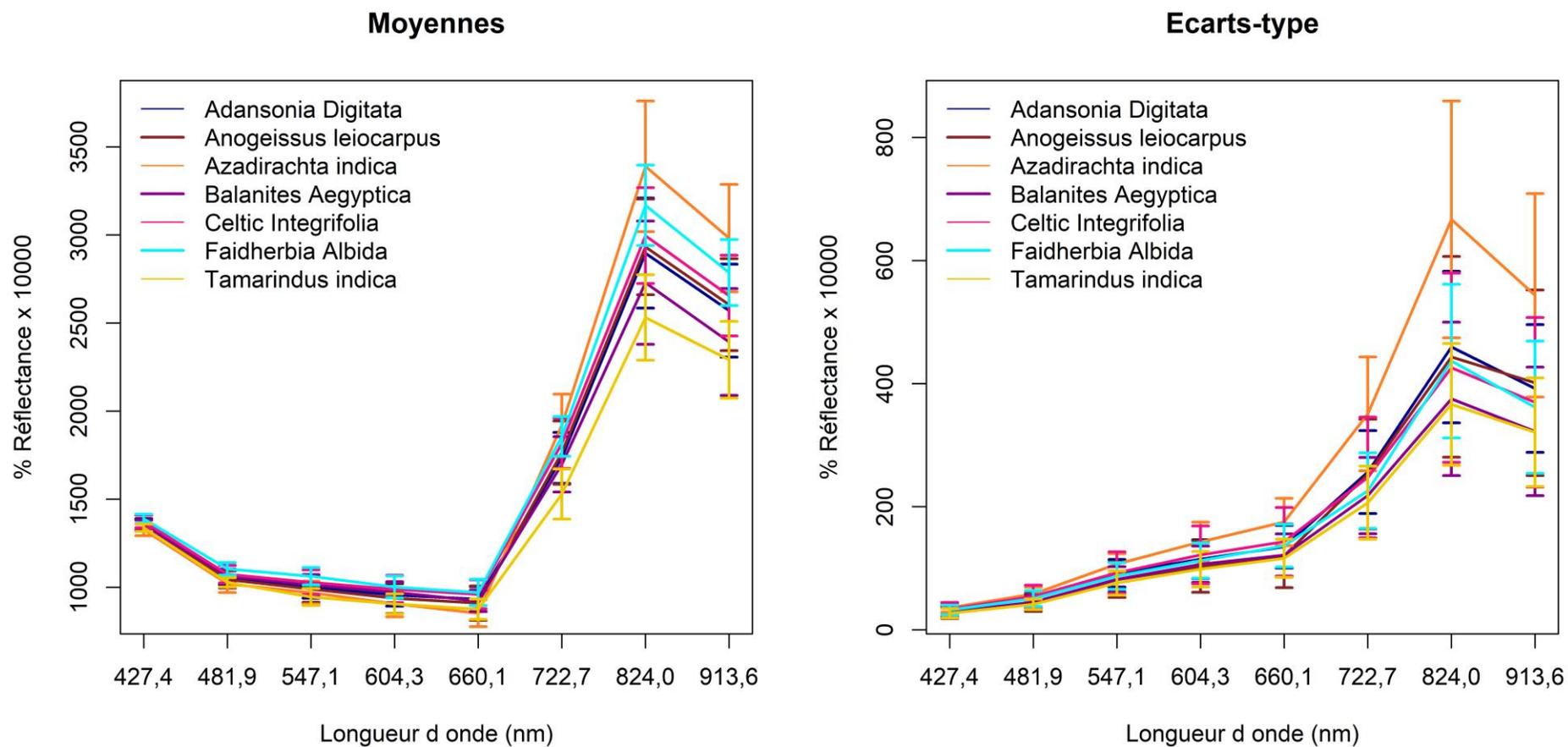


Figure 4. Signatures spectrales des espèces majoritaires et minoritaires du parc à Faidherbia albida : A) les moyennes et B) les écarts-type des réflectances extraites des couronnes d'arbre.

### 3.3 Identification des facteurs discriminants

#### 3.3.1 Analyse en composantes principales

Les deux premières composantes principales du jeu de 16 variables radiométriques (moyennes + écarts types des réflectances) ont des influences équivalentes dans ce jeu de données, représentant respectivement 41% et 37%, soit à elles deux 78%, de la variance globale (cf. Figure 5). Si le niveau de variance diminue ensuite très abruptement, il reste encore de l'ordre de 10% sur la troisième et 5% sur la quatrième composante principale, indiquant que ces deux composantes recèlent encore une part importante de la variance de la population d'arbres et doivent être largement prises en compte dans l'analyse de cette population. Enfin, la cinquième composante représente encore quelques 2,5% de la variance globale, la sixième 1,2%, puis cette contribution est à peu près distribuée sur les composantes suivantes. Il semble donc qu'il faille analyser les contributions des signatures radiométriques des espèces d'arbres jusqu'à la cinquième composante principale pour en appréhender 96% de variance.

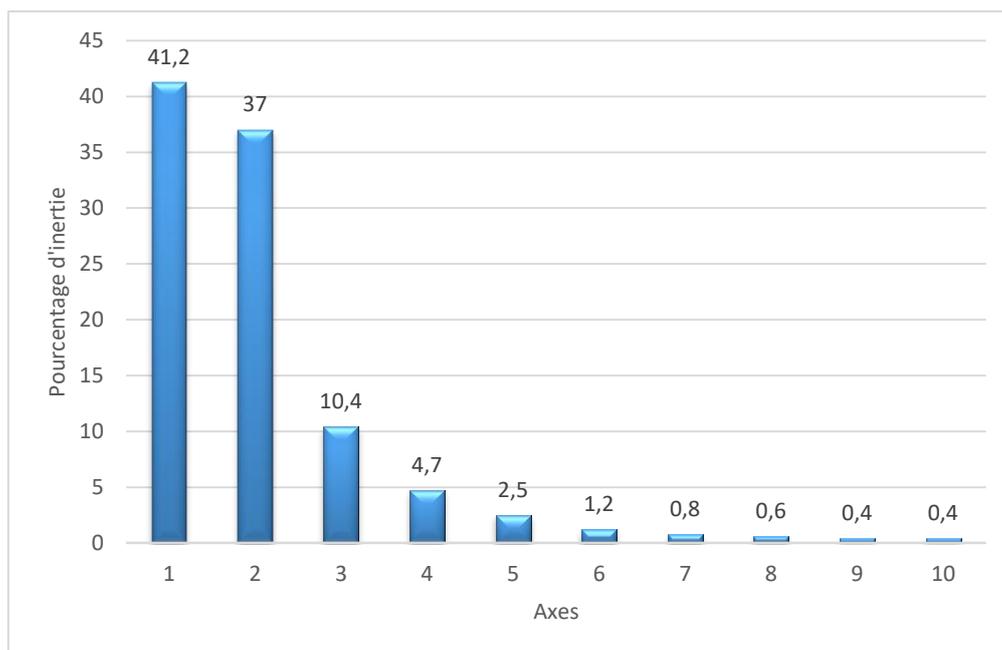


Figure 5. Pourcentage d'inertie des composantes de l'ACP réalisée sur la radiométrie des espèces arborées.

Pour ce qui est de la corrélation entre les variables et les dimensions de l'ACP (Figure 6), la dimension 1 est essentiellement liée aux écarts-types des radiométries et principalement à ceux des bandes verte et Red-edge, et la dimension 2 aux moyennes radiométriques des bandes verte et jaune. On note ici l'influence des nouvelles bandes spectrales disponibles dans WV-3 par rapport aux capteurs classiques (jaune et red-edge). On constate que les moyennes des réflectances des 5 premières bandes spectrales (visible) sont corrélées entre elles, d'une part, et les moyennes en bandes 7 et 8 (proche-infrarouge)

le sont aussi entre elles, d'autre part, indiquant clairement 3 domaines de variances décorrés : le visible, le red-edge, et le proche-infrarouge. Les écarts-type des réflectances des bandes 2 et 4 sont aussi fortement corrélés.

La projection de la population d'arbres sur les plans factoriels (Figure 7) montre globalement une agglomération des individus, toutes espèces confondues. Les espèces étudiées ne possèdent donc pas de signature spectrale caractéristique pouvant permettre de les distinguer les unes des autres. La variabilité spectrale au sein des espèces est plus grande que la variabilité spectrale entre les espèces, rendant impossible tout processus de classification sur l'ensemble de ces classes. Toutefois, l'espèce *Azadirachta indica* se distingue des autres espèces sur l'axe 3. Aussi, les espèces *Balanites Aegyptica*, *Tamarindus indica*, et *Anogeissus leiocarpus* semblent regroupées entre elles, mais séparables du reste du nuage d'arbres sur l'axe 2. Il semblerait donc qu'une analyse plus fine des classes, avec des possibilités de fusion de certaines classes, pourrait permettre d'identifier au moins des regroupements de différentes espèces d'arbres. L'analyse factorielle discriminante présentée au paragraphe suivant permettra d'approfondir cette hypothèse.

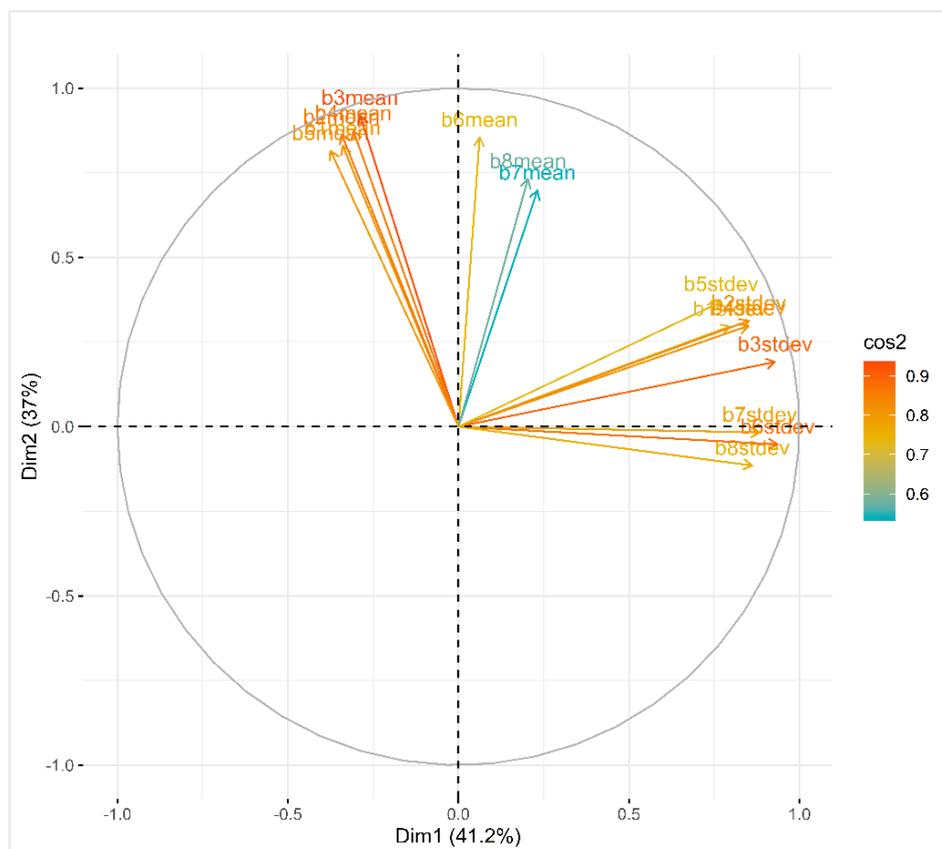


Figure 6. Corrélation des moyennes et des écarts-type des réflectances des couronnes d'arbre.

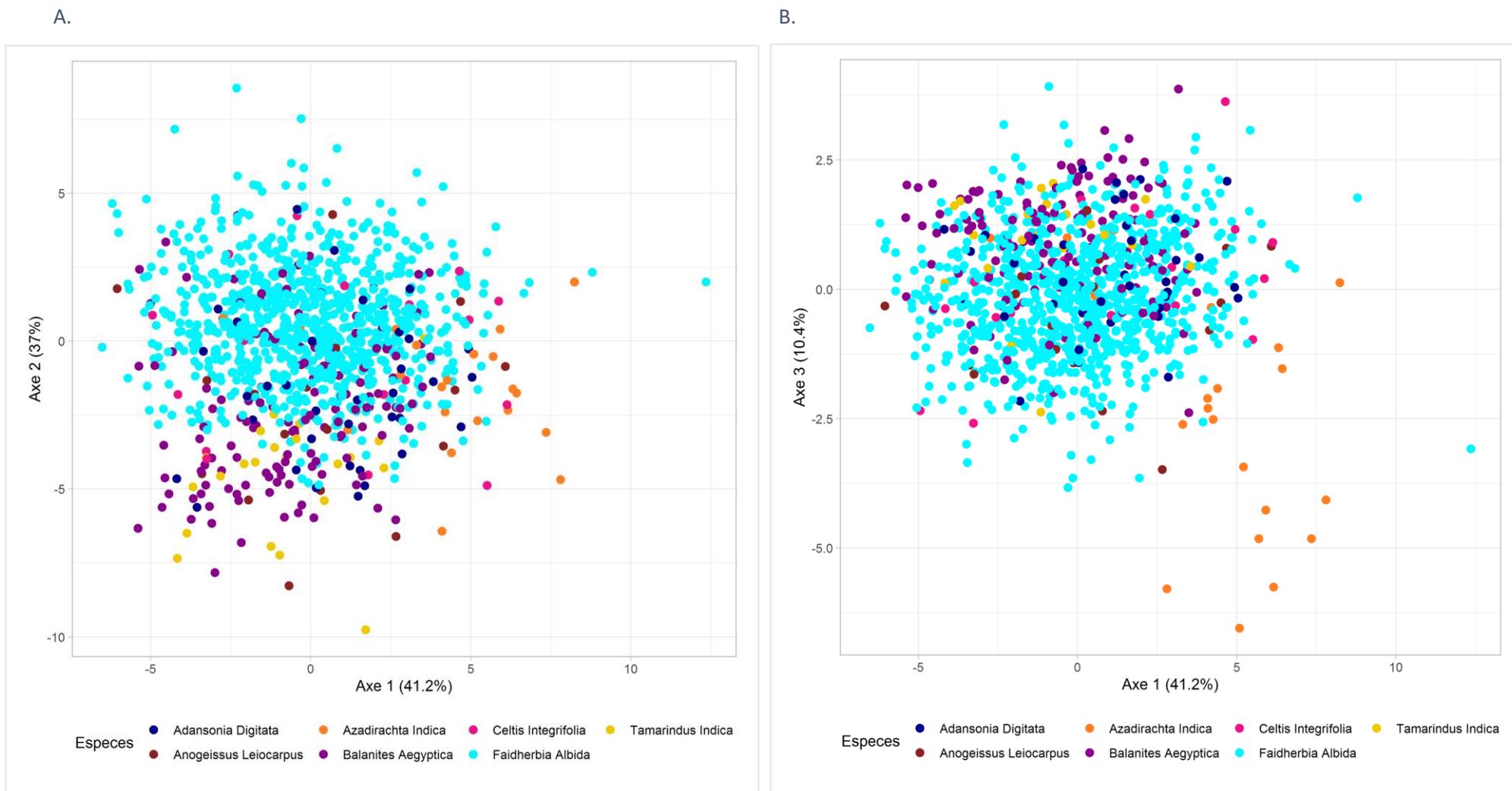


Figure 7. Représentations des individus sur les plans factoriels de l'ACP. Axes 1 et 2 (A) et axes 2 et 3 (B).

### 3.3.2 Analyse factorielle discriminante

L'AFD a été appliquée d'une part sur le jeu de variables comprenant les moyennes et les écarts-type des réflectances des arbres échantillonnés (16 variables), et d'autre part sur tous les indices spectraux calculés (14 variables, cf. Annexe 3). Les résultats sont totalement comparables dans les deux cas : les indices spectraux n'apportent donc pas d'information plus pertinente pour séparer les différentes classes d'arbres, et nous n'évoquerons ici que les résultats obtenus sur le jeu de variables issues de la réflectance. Les 3 premiers axes de l'AFD représentent près de 96 % de l'information sur les individus échantillonnés (Figure 8). Ils représentent respectivement 59,0%, 27,8%, et 8,8% de l'information de l'AFD. Ensemble, les 3 axes suivants de l'AFD représentent moins de 5% de niveau d'inertie.

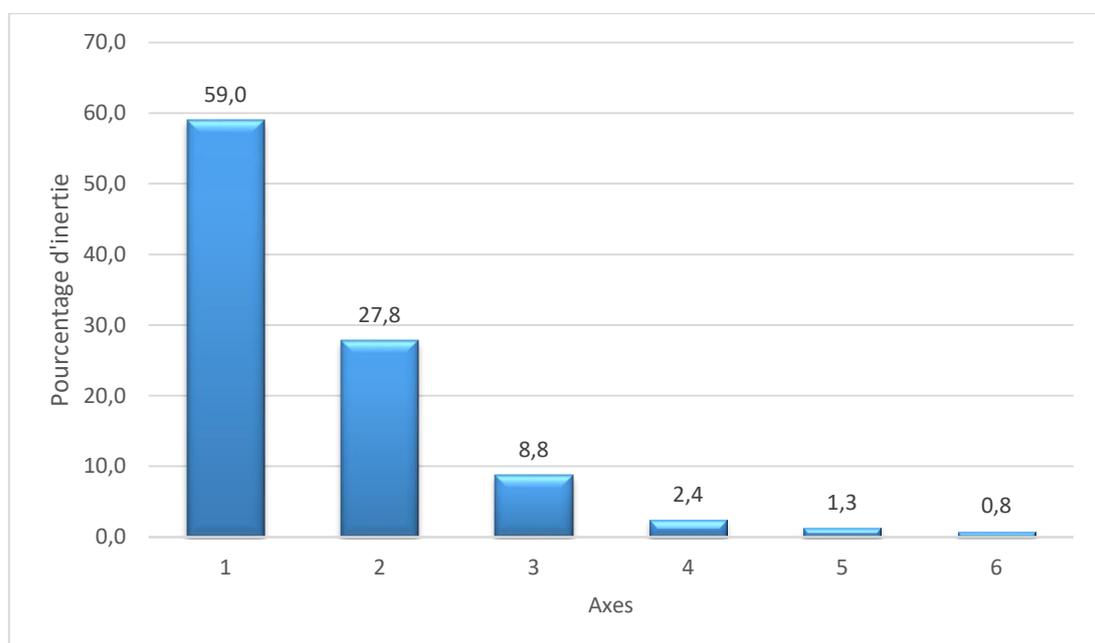


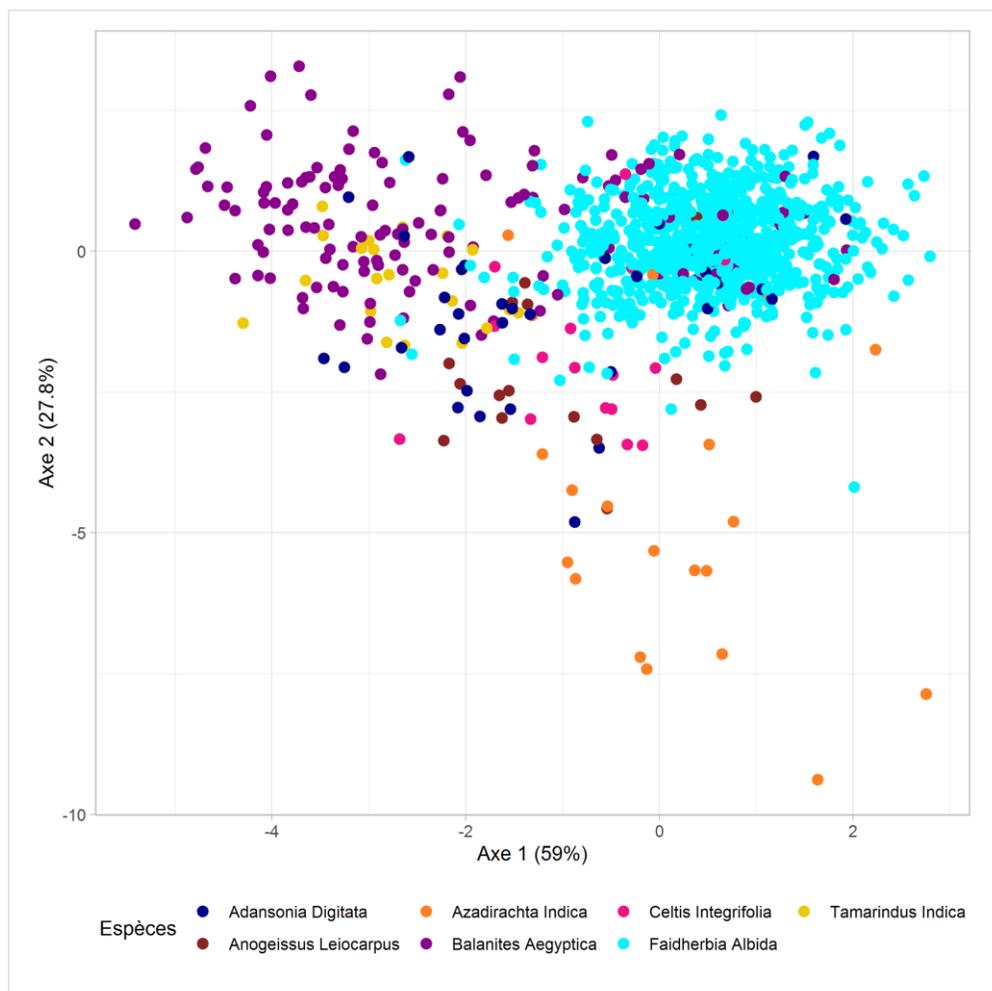
Figure 8. Pourcentage d'inertie des composantes de l'AFD réalisée sur la radiométrie des espèces arborées

En première observation de la répartition de la population dans l'espace défini par les composantes de l'AFD (Figure 9), il apparaît que les individus appartenant à l'espèce *Faidherbia albida* sont bien regroupés en un nuage assez compact, alors que ceux des autres espèces sont plutôt dispersés, y compris à l'intérieur du nuage des *Faidherbia albida*. Cela indique que cette classe sera très probablement facile à séparer des autres classes, avec des fausses détections possibles (i.e. arbres d'autres espèces identifiés comme appartenant à la classe *Faidherbia albida*). Toutefois, certaines tendances apparaissent dans la répartition des individus des autres espèces, représentées par des nuages étirés dans différentes directions, même si leur origine est commune au centre du nuage des *Faidherbia albida*. Les individus de l'espèce *Azadirachta indica*, en particulier, s'en écarte sur l'axe 2, donc devraient pouvoir être bien discriminés des autres espèces. *Balanites aegyptica* et *Tamarindus*

*indica* s'en séparent dans le plan axe 2 – axe 3, avec une partie de leurs individus respectifs qui se retrouvent mélangés, mais aussi un bon nombre d'individus qui se regroupent en deux nuages distincts, permettant d'envisager la potentialité de discriminer à minima ces deux espèces fusionnées, voire une bonne proportion des arbres de chacune de ces deux espèces. Enfin, les individus appartenant aux espèces *Adansonia digitata*, *Anogeissus leiocarpus* et *Celtic integrifolia* sont complètement dispersées au milieu de tous les autres arbres et ce sur tous les axes de l'AFD. On peut donc conclure, que les classes *Faidherbia albida* et *Azadirachta indica* devraient pouvoir être bien discriminées des autres espèces, *Balanites aegyptica* et *Tamarindus indica* aussi mais avec des réserves de séparabilité entre ces deux classes. En revanche, *Adansonia digitata*, *Anogeissus leiocarpus* et *Celtic integrifolia* doivent être regroupées en une seule classe, que nous appellerons « Autres espèces ».

Pourtant, l'espèce *Adansonia digitata* fait partie des espèces majoritaires et fournit de nombreux services à la population du milieu (alimentation humaine et animale, phytothérapie, divertissement, biodiversité, etc.) (Savard, 2003). Il est donc fort dommage de ne pas pouvoir l'isoler dans une classe à part. Nous nous proposons alors de tester l'influence qu'aurait cette classe sur la classification générale, en comparant les précisions obtenues dans les deux cas : avec la classe « *Adansonia digitata* » séparée et la classe « Autres espèces » (donc 6 classes en tout) d'une part, et avec uniquement la classe « Autres espèces » incluant les *Adansonia digitata* (donc 5 classes seulement) d'autre part.

A.



B.

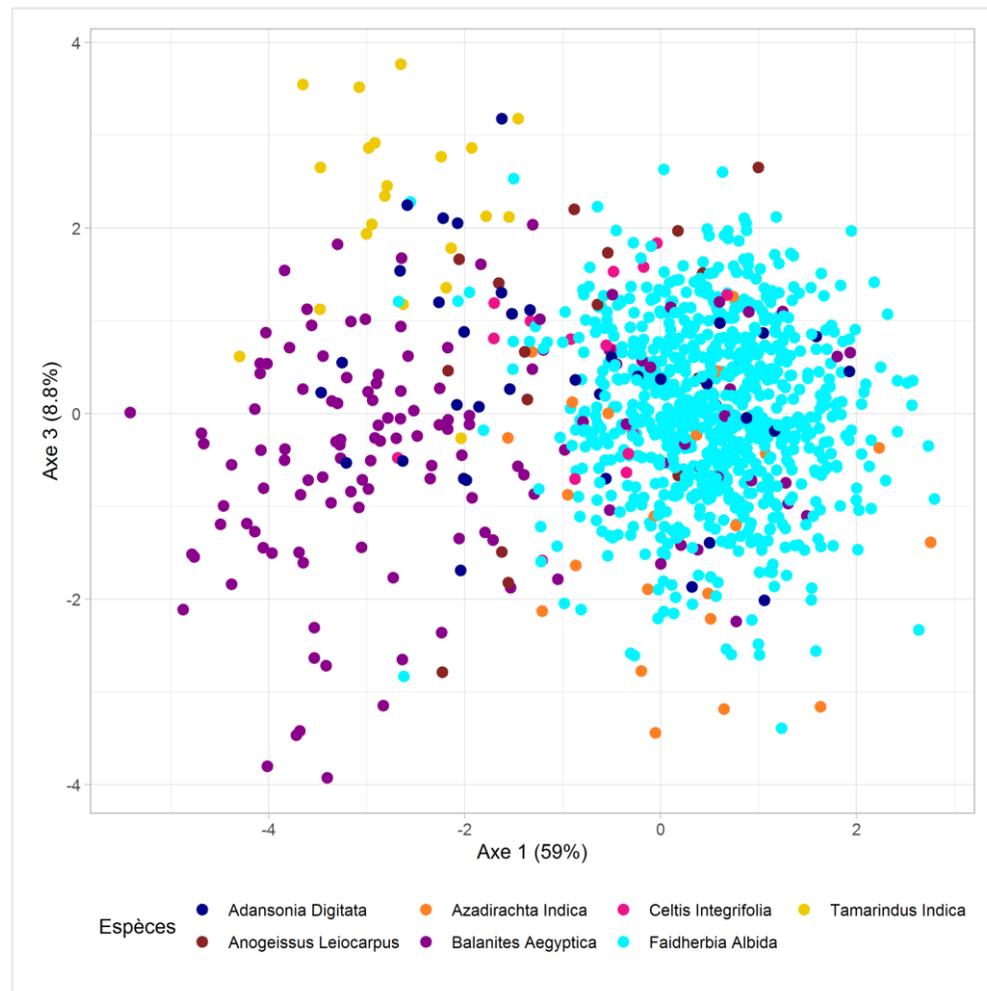


Figure 9. Représentation des individus sur les plans factoriels de l'AFD. Axes 1 et 2 (A) et axes 1 et 3 (B).

### 3.4 Validation et classification

Le choix des variables à considérer pour l'entraînement des algorithmes de classification dépend des caractéristiques particulières de ces derniers. Le SAM, le SVM et le RF étant des algorithmes permettant le traitement d'une grande quantité de variables, il n'y a donc pas de nécessité de réduction de la dimensionnalité du jeu de données. En revanche, pour assurer une bonne stabilité à l'algorithme, il est utile d'éliminer les variables n'apportant pas de surcroît d'information ou une ayant une information très corrélée. Les indices spectraux n'ont donc pas été pris en compte, et seules les moyennes et les écarts types des réflectances des 8 bandes spectrales ont été conservées pour la classification. Vu la lourdeur des données rasters contenant l'information des radiométries des arbres (16 images d'environ 80 giga-octets), les premiers tests du SAM ont mis plus de temps à être réalisés, produisant en plus de faibles précisions globales de moins de 50%. A cause de cette difficulté technique et des faibles précisions obtenues aux premiers tests, nous n'avons pas pu approfondir les analyses sur cet algorithme, et l'avons abandonné pour continuer avec le SVM linéaire, le SVM radial, et le RF. Le Tableau 4 montre les résultats des classifications produites par les différents tests (cf. point 2.2.4). Dans l'ensemble, il s'avère que R est plus efficient qu'OTB pour l'application d'un même algorithme, avec des précisions du même ordre de grandeur, mais toujours meilleures avec R. Tous les algorithmes testés sont assez performants, avec des précisions globales supérieures à 80%, mais on observe une grande variation dans les précisions par classe en fonction du jeu de données utilisé pour l'apprentissage (Annexe 11). Cette instabilité est bien entendu due à la faible séparabilité des classes et à la dispersion de certains arbres échantillonnés sur le terrain dans d'autres classes spectrales (cf. point 3.3.2), ainsi qu'au faible taux de fréquence des arbres dans les classes moins bien définies spectralement.

Tableau 4. Précisions globales et coefficients Kappa sur les trois algorithmes appliqués à 3 tirages différents de données d'entraînement et de validation.

Jeu	Précision	SVM_Linéaire R (%)	SVM_Linéaire OTB (%)	SVM_Radial R (%)	SVM_Radial OTB (%)	RF R (%)	RF OTB (%)
1	Globale 6sp	84	84	84	83	83	83
	Globale 5sp	87	87	85	84	82	83
	Kappa 6sp	55	57	55	50	51	53
	Kappa 5sp	65	64	58	56	51	53
2	Globale 6sp	84	84	82	83	80	80
	Globale 5sp	88	86	84	83	84	80
	Kappa 6sp	59	58	52	53	49	48
	Kappa 5sp	68	64	57	53	59	48
3	Globale 6sp	85	84	84	84	81	81
	Globale 5sp	87	86	86	85	83	81
	Kappa 6sp	59	57	56	54	48	47
	Kappa 5sp	66	61	63	59	55	48

En effet, l'ajout de la sixième classe « *Adansonia digitata* » influence bien négativement les degrés de précision globale autant que la précision de classification des autres classes, venant confirmer la pertinence supposée de regrouper cette espèce avec les « Autres espèces ». L'algorithme qui tendrait le mieux à identifier correctement cette classe « *Adansonia digitata* » sans trop influencer le reste des résultats serait le RF, mais il présente presque 10% de précision globale de moins que le SVM donc nous ne retiendrons pas cette option. Finalement, il résulte de ces différents tests (Tableau 4) que le meilleur algorithme pour classer les arbres de notre zone d'étude est le SVM linéaire appliqué aux 5 classes « *Faidherbia albida* », « *Azadirachta indica* », « *Balanites aegyptica* », « *Tamarindus indica* » et « Autres espèces ».

Compte tenu des variations notables dans les précisions et coefficients kappa des tests appliqués à différents jeux d'entraînement, il apparaît toutefois nécessaire de s'assurer de la stabilité et de la robustesse de cette classification, afin de pouvoir l'appliquer avec confiance à l'ensemble des arbres de la zone. Pour ce faire, nous avons appliqué à nouveau cet algorithme SVM linéaire aux 5 classes d'arbres sur 30 tirages aléatoires indépendants des individus échantillonnés pour l'apprentissage d'une part et la validation d'autre part. Tous les résultats sont présentés à la (Figure 10) et en Annexe 12. Les précisions globales varient entre 85% et 89%, et les coefficients kappa entre 0,55% et 0,71%, ce qui est hautement satisfaisant et relativement stable. Il nous semble toutefois préférable de sélectionner, pour l'appliquer à l'ensemble de l'image, l'algorithme qui laisse le moins la part au hasard, donc celui ayant un kappa le plus élevé, à savoir 0,71%. En outre, il se trouve que c'est aussi le test produisant la meilleure précision globale (89%). La matrice de confusion de cette classification et ses précisions par classe sont données dans le Tableau 5. On y observe que toutes les classes possèdent de meilleures précisions de l'utilisateur (supérieures à 73%). Les précisions du producteur des classes des espèces *Azadirachta indica* et *Tamarindus indica* sont faibles (respectivement 50% et 17%). En effet, *Azadirachta indica* est plus en confusion avec *Faidherbia albida*, et *Tamarindus indica* l'est plus avec les « Autres espèces » qui possèdent une précision de l'utilisateur intermédiaires de 61%. Les meilleures précisions du producteur concernent les espèces *Faidherbia albida* et *Balanites aegyptica* (respectivement 98% et 72%). En vue d'évaluer l'apport des bandes complémentaires (bleu « côtier », jaune, red-edge et PIR2) de WV-3 sur la précision de la classification, nous avons également réalisé une classification avec les moyennes et les écarts-type des 4 bandes spectrales classiques uniquement : le bleu ; le vert ; le rouge ; et le PIR1. D'après les résultats, on observe une diminution des niveaux de précisions globale et kappa compris respectivement entre 0,80% et 0,84%, et 0,45% et 0,57% (Annexe 11). Cela vient donc confirmer que ces bandes complémentaires apportent une information supplémentaire permettant d'augmenter la discrimination des espèces arborées de notre zone d'étude.

Tableau 5. Matrice de confusion de la classification SVM linéaire du jeu de données ayant produit les meilleures précisions

Classes		Prédites						% utilisateur
		Autres especes	A. indica	B. aegyptica	F. Albida	T. Indica	Total	
Observées	Autres especes	22,0	0,0	1,0	3,0	4,0	30,0	73,3
	A. indica	0,0	3,0	0,0	1,0	0,0	4,0	75,0
	B. aegyptica	2,0	0,0	28,0	0,0	1,0	31,0	90,3
	F. Albida	12,0	3,0	10,0	242,0	0,0	267,0	90,6
	T. Indica	0,0	0,0	0,0	0,0	1,0	1,0	100,0
	Total	36,0	6,0	39,0	246,0	6,0	333,0	
	% producteur	61,1	50,0	71,8	98,4	16,7		

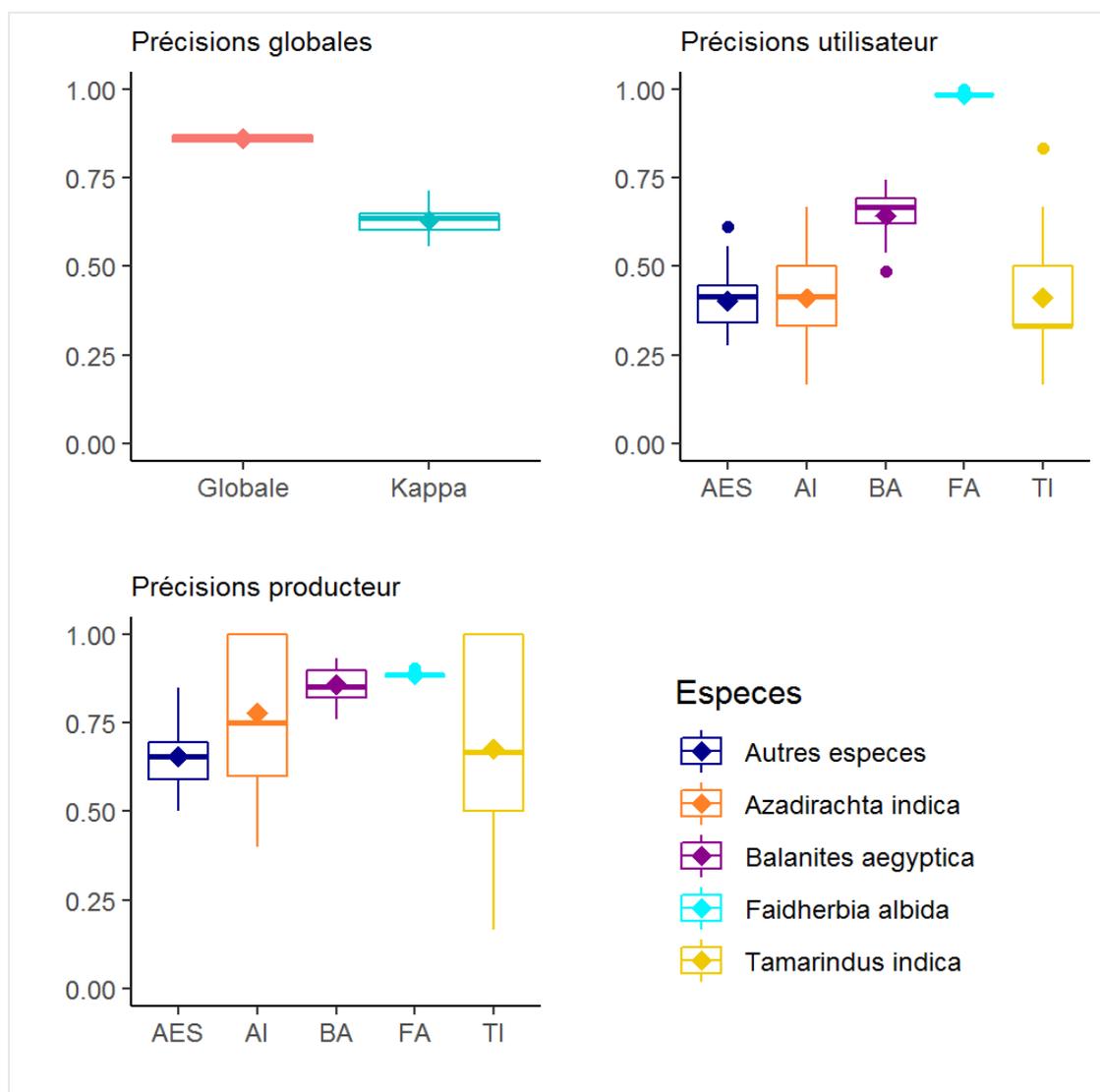


Figure 10. Précisions globales, coefficients Kappa, précisions producteur et utilisateur, produits à l'issue de la classification SVM linéaire appliquée à 30 tirages différents des données d'entraînement, pour 5 classes d'espèces d'arbres.

Les 134715 couronnes d'arbres extraites dans toute l'image ont été classées sur la base des moyennes et des écarts-type de radiométrie des 8 bandes spectrales du capteur WV-3 (Tableau 6). L'espèce *Faidherbia albida* est largement majoritaire avec plus de 75% de représentation sur la zone d'étude. A l'inverse, *Tamarindus indica* ne représente que 1,4% des arbres de la zone. Les espèces *Azadirachta indica* et *Balanites aegyptica*, ainsi que toutes les autres espèces confondues et incluant *Adansonia digitata* se retrouvent en proportions similaires, de l'ordre de 7 à 8 %. L'ordre des proportions des espèces issues de cette classification n'est pas comparable à celui de la base d'échantillonnage (Tableau 3), ni à celui des résultats de l'inventaire réalisé dans la zone d'étude (Tableau 2). Cela prouve sans doute le fait que certaines espèces ne sont pas régulièrement réparties sur la zone d'étude.

Tableau 6. Effectif des classes d'espèces d'arbres du parc à *Faidherbia albida* de Bambey

Espèces	Effectif	Pourcentage
Autres espèces	11815	8,77
<i>Azadirachta indica</i>	9822	7,29
<i>Balanites aegyptica</i>	9696	7,20
<i>Faidherbia albida</i>	101471	75,32
<i>Tamarindus indica</i>	1911	1,42
<b>Total</b>	<b>134715</b>	<b>100,00</b>

### 3.5 Distribution spatiale des espèces

La cartographie issue de ce processus de classification est présentée dans la carte au format A2, en fichier séparé accompagnant ce document. On remarque sur cette carte que la diversité en espèces arborée est faible vers le sud mais augmente progressivement en se déplaçant vers le Nord de la zone d'étude. Comme le montre le Tableau 6, l'espèce *Faidherbia albida* est la plus abondante et est répartie sur toute la zone d'étude. Certaines espèces par contre, sont généralement présentes aux alentours et à proximité des villages de la zone d'étude, c'est le cas d'*Azadirachta indica*, une espèce exotique d'origine d'Inde, dont la présence dans la zone est due à la plantation par les habitants des villages, pour différents usages (consommations, phytothérapie, etc.). Contrairement, l'espèce *Balanite aegyptica* est le plus souvent localisée dans la brousse où elle se trouve mélangée avec le *Faidherbia albida*. Enfin, les autres espèces sont très faiblement représentées et aléatoirement réparties sur la zone d'étude.

## 4 Discussion

### 4.1 Ressources techniques et logiciels

Les traitements sur les données raster du type imagerie multispectrale ou panchromatique WV-3 sont particulièrement lourds, nécessitant des performances puissantes et une grande capacité de mémoire vive. Les ordinateurs disponibles, par exemple, au CSE à Dakar (16 G de RAM) ont eu de grosses difficultés à prendre en charge ces traitements. Pour ne citer que le calcul des indices de texture, pourtant appliqués à une monobande, le calcul pouvait durer jusqu'à 2 jours sous ENVI, et n'a jamais pu tourner sous OTB malgré la réduction à des extraits de l'image.

L'approche orientée objet, permettant de stocker les variables associées aux objets d'étude (les arbres) sous forme vectorielle, a permis en revanche d'alléger considérablement les données de départ pour les traiter de manière statistique.

### 4.2 Délimitation des arbres

Les différentes analyses réalisées montrent que la précision de la caractérisation des propriétés spectrales des espèces d'arbre reste fortement liée à la précision de l'extraction des couronnes d'arbres. Ce constat a également été observé dans d'autres travaux démontrant l'impact de la qualité de la segmentation sur la qualité de la classification (Weidner, 2008 ; Neubert *et al.*, 2012). Sous eCognition et ENVI, la principale contrainte observée est le manque de précision dans la délimitation des couronnes d'arbres. Karlson *et al.* (2016) ont été confrontés aux mêmes difficultés pour identifier les couronnes d'arbre sur les données WV-2 sous eCognition. En revanche, dans notre zone d'étude, le seuillage manuel de l'indice du NDVI à 0,4 a permis de mieux cerner les couronnes d'arbres sans inclure d'autres espèces. Néanmoins pour l'espèce *Adansonia digitata* cette méthodologie de seuillage n'a pas bien fonctionné, à cause de ses faibles valeurs de NDVI (<0,3), dues à la morphologie et la phénologie de cette espèce qui est très hétérogène en terme de densité et de taux de couverture dans la zone d'étude, nous obligeant à exclure cette espèce arborée de la classification finale. Cette hétérogénéité est probablement due à la pression qu'elle subit face à l'exploitation de ses organes intéressants (branches, feuilles, fruits, etc.) pour de nombreuses utilisations (alimentation humaines et du bétail, phytothérapie, pratiques culturelles, etc.). En effet, l'élagage étant une pratique courante dans la zone de Bambey, la taille des couronnes des arbres et plus spécifiquement du *Faidherbia albida* est quasi homogène pour la majorité des arbres, rendant impossible la mise en relation de la taille de la couronne avec l'âge de l'arbre, mais aussi techniquement rendant son extraction plus difficile avec les outils de traitement des images satellites. L'enquête de terrain réalisée dans la zone

d'étude en 2015 dans le cadre du projet TRECS révèle que le Baobab (*Adansonia digitata*) et le Soump (*Balanites aegyptica*) sont les espèces les plus utilisées pour l'alimentation humaine. Les feuilles du baobab, réduites à l'état de poudre, et les fruits (pain de singe) entrent dans la composition du repas, et leur commercialisation occupe une place importante dans le revenu des ménages (Thiaw *et al.*, 2019).

### **4.3 Descripteurs des classes d'arbres**

Les propriétés spectrales des arbres ne sont pas suffisamment spécifiques, pour les espèces étudiées, rendant leur discrimination malaisée voire impossible. Plusieurs facteurs influencent la nature physico-chimique de ces espèces arborées. En effet, les facteurs climatiques conditionnent la physiologie, l'apparence des arbres et leurs propriétés radiométriques. Les symptômes de stress de la canopée des arbres varient en fonction de la durée et de l'intensité de la sécheresse, résultant en une réponse physiologique (Lee *et al.*, 2007). Au moment de l'acquisition de l'image, au troisième mois de la disparition des pluies, certaines espèces gardent encore leurs feuilles. L'espèce *Faidherbia albida* est caractérisée par une période de végétation en saison sèche, en l'absence de toute pluie, à une période où la température est élevée (Wood & Centre Technique Forestier Tropical, 1989). Nous n'avons pas recensé des études sur la phénologie des autres espèces, mais il est également important de signaler l'aspect optique caractéristique des espèces *Balanites aegyptica* et *Tamarindus indica* qui se font bien remarquer sur la composition colorée en vraie couleur.

### **4.4 Séparabilité des classes d'arbres**

Les analyses de l'ACP et de l'AFD ont montré la difficulté de discriminer les espèces sur la base de leurs radiométries, les indices spectraux ne donnant pas plus d'information que ces dernières, en raison de la grande variabilité des réflectances au sein d'une même espèce et de la superposition des nuages d'individus correspondant aux différentes espèces.

Les contributions à la construction des axes 1 et 2 de l'ACP respectivement par les écarts-type et les moyennes de radiométrie n'ont pas d'écart important (41% axe 1 et 37% axe2). Cela montre que les deux types de variables contiennent des informations quasi différentes et complémentaires, ce qui a dû également améliorer la discrimination des espèces. La moyenne ou l'écart-type des réflectances pris seul, ne permet pas de décrire complètement la radiométrie d'une couronne d'arbre, mais pris ensemble, ils permettent de mieux appréhender le comportement spectral de chaque arbre augmentant ainsi la possibilité de discriminer les espèces arborées.

L'AFD réalisé dans l'objectif de maximiser le rapport de la variance entre les espèces et la variance au sein de l'espèce a permis d'identifier 5 classes d'espèces ou groupement d'espèces potentiellement discriminables. Sa contribution pour la sélection des classes et des variables a également été prouvée par Waser *et al.* (2014) au cours d'un processus de classification de 7 espèces ligneuses dans la forêt de l'état de Mecklenburg–Vorpommern en Allemagne sur 1133 arbres. Cependant, une classification basée sur les principales composantes de l'AFD aurait certainement augmenté le potentiel de discrimination des espèces arborées, en raison de la possibilité qu'elle offre pour ressortir la variabilité interclasse des espèces arborées.

#### **4.5 Sensibilité des classifications au jeu d'apprentissage**

L'apprentissage réalisé premièrement sur le SVM linéaire, le SVM radial, et le RF, pour la sélection du meilleur algorithme, a été appliqué à partir des 16 variables retenues, les moyennes et les écarts-type des réflectances des 8 bandes spectrales pour chaque espèce arborée retenue. Le SVM et le RF possèdent des particularités différentes. En effet, le SVM peut être appliqué à plusieurs variables dans le cas de classifieurs non linéaires (Vapnik, 2000), tandis que la réduction du nombre de prédicteurs conduit souvent à une amélioration de la prédiction pour le RF (Karlson *et al.*, 2016 ; Naidoo *et al.*, 2012). Compte tenu de la complémentarité des variables relevée par les résultats de l'ACP et de l'AFD, nous avons choisis de maintenir les moyennes et les écarts-type des réflectances des arbres sans réduction. Sous R, il a été possible d'observer l'importance de chaque variable à l'issue de la prédiction du RF (Annexe 12). Sur des tests réalisés, la précision de la prédiction a diminué lorsque l'entraînement du RF n'a été réalisé que sur les 8 variables représentant les moyennes des réflectances.

Les résultats ont montré que les prédictions réalisées sur 6 espèces présentent des précisions inférieures à celles réalisées sur 5 espèces et cela lorsque la classe « *Adansonia digitata* » n'est pas prise en compte. La sensibilité des résultats avec ou sans la prise en compte de la classe *Adansonia digitata* est sans doute lié à l'hétérogénéité des signatures spectrales des espèces qui constituent la base d'apprentissage et de validation. Les écarts entre les précisions des deux algorithmes du SVM ne sont pas vraiment importants et le choix du SVM linéaire ne détermine pas nécessairement la grandeur de ses performances par rapport au SVM radial. En soi, le SVM radial est généralement adapté pour des traitements des données à grande dimensionnalité (Fung & Mangasarian, 2005), généralement aux données hyperspectrales. Liu & Wu (2018) ont produit un maximum de précision de 82% en appliquant du SVM radial pour la classification des espèces ligneuses par l'utilisation de données hyperspectrales et du LIDAR. Pour notre cas d'étude, les variables considérées ne sont pas du même ordre de grandeur que les données hyperspectrales, et sont tout de même bien gérées par le SVM

linéaire. Outre la principale contrainte du SAM, liée au poids des données rasters, il serait encore intéressant de s'attarder dans la sélection des variables, maximisant l'angle de séparation entre différentes espèces. Cette manière de procéder, conduirait certainement à une amélioration des précisions des prédictions. Madonsela *et al.* (2017), ont obtenus une précision globale d'environ 80% à l'issue de la classification des espèces arborées de la savane sud-africaine, par le SAM, après avoir sélectionné les bandes utiles selon le principe de « Band Add On » décrit en Annexe 5.

Les algorithmes de classification fonctionnent également sous un principe probabiliste. Au-delà des similitudes spectrales, les espèces minoritaires sous échantillonnées possèdent moins de chance d'être mieux classées que les espèces majoritaires. Cela a dû également contribuer aux confusions observées entre les classes des espèces minoritaires *Tamarindus indica* et *Azadirachta indica*, et les classes des *Faidherbia albida* et des « Autres espèces ». Ainsi donc, une augmentation de la base de données d'apprentissage et de validation des espèces minoritaires contribuerait sans doute à l'augmentation de leurs précisions à la suite d'une classification.

Dans cette étude où l'objectif majeur reste la discrimination des espèces, il conviendrait d'envisager d'autres approches complémentaires permettant d'analyser au cours du temps, les variations possibles dans le comportement spectral des arbres de la zone d'étude. Karlson *et al.* (2016) ont évalué les potentialités offertes par les données WV-2 sur deux périodes différentes (sèche et pluvieuse) pour la classification des espèces ligneuses en Afrique de l'ouest. Ils ont trouvé que la combinaison des deux images suivant une approche temporelle, a produit une meilleure précision (83%) que les deux prises individuellement (78 % et 68 %). Cependant, le coût d'acquisition élevé des données THRS de WV-3 pourrait constituer une contrainte pour une étude visant à analyser la variation temporelle des propriétés spectrales des espèces arborées. Alternativement, la donnée du capteur PLANET, intéressante du point de vue de la résolution spatiale (3 m) et d'une fréquence d'acquisition journalière et pouvant s'obtenir à un prix largement inférieur à celui des données WV-3, offre une bonne possibilité pour le suivi phénologique des espèces arborées, qu'il serait intéressant d'explorer. Il apparaît donc utile, de tester les potentialités offertes par la complémentarité de ces deux types de données (WV-3 et Planet) en vue de réunir à la fois la possibilité d'extraire les couronnes des arbres et de tracer leurs signatures spectrales sur des périodes différentes.

## Conclusion

Les résultats de cette étude, nous ont confirmé que la réussite dans l'identification et l'extraction des couronnes d'arbres des parcs agroforestiers en zone sahélienne est essentiellement conditionnée par la séparabilité des propriétés spectrales des espèces arborées. Les paramètres les plus descriptifs des espèces d'arbres se sont avérés être la moyenne et l'écart-type des réflectances dans les 8 bandes spectrales de WV-3, les indices spectraux (NDVI, SAVI, ...) n'y ajoutant pas d'information supplémentaire. Les indices de textures calculés à partir de la matrice de cooccurrence (entropie, contraste, corrélation et homogénéité) ne présentent quant à eux aucune variabilité utile au sein de la couronne des arbres.

Afin de pouvoir analyser et identifier chaque arbre individuellement, la meilleure méthode d'extraction des couronnes des arbres repose sur un seuillage du NDVI (supérieur à 0,4), permettant de circonscrire nettement les contours des arbres, mis à part pour l'*Adansonia digitata* présentant une morphologie atypique. Ce seuillage est notamment plus performant que la segmentation multi-résolution, en particulier grâce au fort contraste de densité de végétation entre les arbres et leur environnement en contexte sahélien.

L'analyse spectrale des différentes espèces en présence a finalement réduit à 4 les espèces potentiellement discriminables : l'*Azadirachta indica*, le *Faidherbia albida*, le *Balanites aegyptica* et le *Tamarindus indica*. L'espèce, pourtant majoritaire, *Adansonia digitata*, ainsi que les espèces minoritaires *Anogeissus leiocarpus* et *Celtis integrifolia* n'ayant pas de signatures spectrales spécifique permettant de les distinguer, ne peuvent pas être isolées des « Autres espèces d'arbres » regroupant toutes les autres espèces. Différents tests de classification ont confirmé cette caractéristique, les classifications visant à considérer l'*Adansonia digitata* comme espèce à part entière aboutissent même à de plus grandes confusions au sein des classes représentant les autres espèces majoritaires.

Un modèle de classification à la fois robuste et stable (89% de précision globale, kappa de 0,71 et forte reproductibilité aux tests appliqués sur un grand nombre de tirages aléatoires du jeu d'apprentissage), reposant sur un algorithme SVM linéaire, a finalement été sélectionné afin de l'appliquer à l'ensemble des arbres de la zone d'étude. Toutefois, ce modèle, comme tous les algorithmes testés dans notre étude, s'avère particulièrement sensible au jeu d'apprentissage. Cela implique que la crédibilité d'une telle classification repose sur un protocole de terrain en amont particulièrement soigné. On évoquera notamment la densité des effectifs au sein de chaque espèce, même minoritaire, ainsi que leur représentativité en matière de taille, âge, phénologie, et autres paramètres environnementaux. Aussi, il serait intéressant de tester l'apport des données multitemporelles à très haute résolution spatiale, par exemple en mobilisant des données Planet en complément des données WV3, pour inclure les

variations annuelles des différentes espèces, souvent décalées entre elles, dans les descripteurs discriminants, pour augmenter les performances de cette identification.

Nous avons donc, à l'issue de ce projet, pu produire une carte des espèces arborées sur toute la zone d'étude couverte par l'imagerie WV3, en les classant dans 4 espèces d'arbres identifiables plus une dernière classe « fourre-tout ». Cette carte permet notamment d'analyser la répartition des espèces sur l'ensemble de la zone, dans une approche de type « écologie du paysage ». La prépondérance de l'espèce *Faidherbia albida* s'y retrouve flagrante, l'espèce étant présente partout et souvent seule. Les autres espèces semblent plutôt cohabiter toutes ensemble, dans des « poches » à forte variété arborée, en proportions variables mais relativement similaires. Ces poches semblent plus nombreuses au Nord de la zone qu'au Sud, et plutôt regroupées aux alentours des villages et des croisements de routes. On notera aussi, par exemple, qu'*Azadirachta indica* est préférentiellement planté dans et autour des villages alors que le *Balanites aegyptica* se retrouve plus souvent associé au *Faidherbia albida* en brousse.

Cette carte fait donc clairement apparaitre de grandes tendances et des répartitions systématiques des différentes espèces d'arbres, qui pourraient être exploitées dans des domaines thématiques variés (lutte contre les bioagresseurs, aménagement du territoire, etc..).

## Bibliographie

- Adelabu S. & Dube T. 2015. Employing ground and satellite-based QuickBird data and random forest to discriminate five tree species in a Southern African Woodland. *Geocarto International*. 30(4), p. 457-471.
- Badiane A., Khouma M., & Sene M. 2000. Région de Diourbel : Gestion des eaux. , p. 15.
- Bagny-Beilhe L., Allinne C., Avelino J., Babin R., Brévault T., Gidoïn C., Ngo B., Marie-Ange, Motisi N., Soti V., & Ten H. G.M. 2019. Régulation des bioagresseurs des cultures dans les systèmes agroforestiers tropicaux, revue des approches.
- Brévault T. & Clouvel P. 2019. Pest management: Reconciling farming practices and natural regulations. *Crop Protection*. 115, p. 1-6.
- Cho M.A., Debba P., Mathieu R., Naidoo L., van Aardt J., & Asner G.P. 2010. Improving Discrimination of Savanna Tree Species Through a Multiple-Endmember Spectral Angle Mapper Approach: Canopy-Level Analysis. *IEEE Transactions on Geoscience and Remote Sensing*.
- Cho M.A., Mathieu R., Asner G.P., Naidoo L., van Aardt J., Ramoelo A., Debba P., Wessels K., Main R., Smit I.P.J., & Erasmus B. 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment*. 125, p. 214-226.
- Clark M., Roberts D., & Clark D. 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*. 96(3-4), p. 375-398.
- Dalponte M., Bruzzone L., & Gianelle D. 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment*. 123, p. 258-270.
- DAPSA. 2014. Rapport de présentation des résultats définitifs de l'enquête agricole 2013-2014.
- Féret J.-B. & Asner G.P. 2011. Spectroscopic classification of tropical forest species using radiative transfer modeling. *Remote Sensing of Environment*. 115(9), p. 2415-2422.
- Feret J.-B. & Asner G.P. 2013. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Transactions on Geoscience and Remote Sensing*. 51(1), p. 73-84.
- Fung G.M. & Mangasarian O.L. 2005. Multicategory Proximal Support Vector Machine Classifiers. *Machine Learning*. 59(1-2), p. 77-97.
- Ghojogh B., Samad M.N., Mashhadi S.A., Kapoor T., Ali W., Karray F., & Crowley M. 2019. Feature Selection and Feature Extraction in Pattern Analysis: A Literature Review. *arXiv:1905.02845 [cs, stat]*.
- Haralick R.M., Shanmugam K., & Dinstein I. 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-3(6), p. 610-621.
- Harikumar A., Bovolo F., & Bruzzone L. 2019. A Local Projection-Based Approach to Individual Tree Detection and 3-D Crown Delineation in Multistoried Coniferous Forests Using High-Density Airborne LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing*. 57(2), p. 1168-1182.

- Heenkenda M., Joyce K., Maier S., & Bartolo R. 2014. Mangrove Species Identification: Comparing WorldView-2 with Aerial Photographs. *Remote Sensing*. 6(7), p. 6064–6088.
- Hsieh H.-C., Chung C.-H., & Huang C.-Y. 2017. Using the NDVI and mean shift segmentation to extract landslide areas in the Lioukuei experimental forest region with multi-temporal FORMOSAT-2 images. *Taiwan Journal of Forest Science*. 32, p. 203–222.
- Immitzer M., Atzberger C., & Koukal T. 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*. 4(9), p. 2661–2693.
- Jakkula V. 2006. Tutorial on Support Vector Machine (SVM). , p. 13.
- Jakubowski M., Li W., Guo Q., & Kelly M. 2013. Delineating Individual Trees from Lidar Data: A Comparison of Vector- and Raster-based Segmentation Approaches. *Remote Sensing*. 5(9), p. 4163–4186.
- Karlson M., Ostwald M., Reese H., Bazié H.R., & Tankoano B. 2016. Assessing the potential of multi-seasonal WorldView-2 imagery for mapping West African agroforestry tree species. *International Journal of Applied Earth Observation and Geoinformation*. 50, p. 80–88.
- Keshava N. 2004. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Transactions on Geoscience and Remote Sensing*. 42(7), p. 1552–1565.
- Lee K.-S., Kook M.-J., Shin J.-I., Kim S.-H., & Kim T.-G. 2007. Spectral Characteristics of Forest Vegetation in Moderate Drought Condition Observed by Laboratory Measurements and Spaceborne Hyperspectral Data. *Photogrammetric Engineering & Remote Sensing*. 73(10), p. 1121–1127.
- Lefebvre A. 2011. Contribution de la texture pour l'analyse d'images à très haute résolution spatiale: application à la détection de changement en milieu périurbain. , p. 285.
- Liu H. & Wu C. 2018. Crown-level tree species classification from AISA hyperspectral imagery using an innovative pixel-weighting approach. *International Journal of Applied Earth Observation and Geoinformation*. 68, p. 298–307.
- Madonsela S., Cho M.A., Mathieu R., Mutanga O., Ramoelo A., Kaszta Ż., Kerchove R.V.D., & Wolff E. 2017. Multi-phenology WorldView-2 imagery improves remote sensing of savannah tree species. *International Journal of Applied Earth Observation and Geoinformation*. 58, p. 65–73.
- Maillard P. 2003. Comparing Texture Analysis Methods through Classification. *Photogrammetric Engineering & Remote Sensing*. 69(4), p. 357–367.
- Marrakchi O.C. 2012. Elaboration de signatures des textures d'images de télédétection. 11(2), p. 15.
- Michele A. K. 2017. Absolute Radiometric Calibration : 20016v0. , p. 8.
- Naidoo L., Cho M.A., Mathieu R., & Asner G. 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing*. 69, p. 167–179.

- Neubert M., Herold H., & Meinel G. 2012. Evaluation of remote sensing image segmentation quality – further results and concepts. , p. 7.
- Patel M.B., Rodriguez J.J., & Gmitro A.F. 2008. Effect of gray-level re-quantization on co-occurrence based texture analysis. Dans : *2008 15th IEEE International Conference on Image Processing. 2008 15th IEEE International Conference on Image Processing*. San Diego, CA, USA : IEEE, p. 585-588.
- Peerbhay K.Y., Mutanga O., & Ismail R. 2014. Investigating the Capability of Few Strategically Placed Worldview-2 Multispectral Bands to Discriminate Forest Species in KwaZulu-Natal, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 7(1), p. 307-316.
- Pu R. & Landry S. 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment*. 124, p. 516-533.
- Puissant A., Hirsch J., & Weber C. 2005. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing*. 26(4), p. 733-745.
- Rasieh S., Helmi Z.M.S., Shattri M., & Mohd H.I. 2016. The Use of WorldView-2 Satellite Data in urban Tree Species Mapping by Object-Based Image Analysis Technique.
- Reed J., van Vianen J., Foli S., Clendenning J., Yang K., MacDonald M., Petrokofsky G., Padoch C., & Sunderland T. 2017. Trees for life: The ecosystem service contribution of trees to food production and livelihoods in the tropics. *Forest Policy and Economics*. 84, p. 62-71.
- Regniers O. 2015. Méthodes d'analyse de texture pour la cartographie d'occupations du sol par télédétection très haute résolution: application à la forêt, la vigne et les parcs ostréicoles. , p. 164.
- Savard V. 2003. Évaluation du potentiel d'adoption des parcelles maraîchères de baobab (*adansonia digitata*) dans la région de ségou, au mali. , p. 139.
- Showengerdt R. 2007. *Remote Sensing. Models and methods for image processing*. Third edition. Etats Unis d'Amérique : Elsevier
- Simon S., Bouvier J.-C., Debras J.-F., & Sauphanor B. 2010. Biodiversity and pest management in orchard systems. A review. *Agronomy for Sustainable Development*. 30(1), p. 139–152.
- Singh D., Maurya R., Shukla A.S., Sharma M.K., & Gupta P.R. 2012. Building extraction from very high resolution multispectral images using NDVI based segmentation and morphological operators. Dans : *2012 Students Conference on Engineering and Systems. 2012 Students Conference on Engineering and Systems (SCES)*. Allahabad, Uttar Pradesh, India : IEEE, p. 1-5.
- Soti V., Lelong C., Goebel F.-R., & Brévault T. 2018. Designing a field sampling plan for landscape-pest ecological studies using VHR optical imagery. *International Journal of Applied Earth Observation and Geoinformation*. 72, p. 26-33.
- Soti V., Thiaw I., Debaly Z.M., Sow A., Diaw M., Fofana S., Diakhate M., Thiaw C., & Brévault T. 2019. Effect of landscape diversity and crop management on the control of the millet head miner, *Heliocheilus albipunctella* (Lepidoptera: Noctuidae) by natural enemies. *Biological Control*. 129, p. 115–122.

- Tharwat A., Gaber T., Ibrahim A., & Hassanien A.E. 2017. Linear discriminant analysis: A detailed tutorial. *AI Communications*. 30(2), p. 169-190.
- Thiaw I., Soti V., & Diakhate M. 2019. *Analyse des déterminants locaux et paysagers de la dégradation naturelle de l'espèce *Heliocheilus albipunctella* : Contribution à l'amélioration de la protection intégrée de la culture du mil à Bambey au Sénégal*. Dakar Sénégal : , 260 p.
- Tiede D., Lang S., & Hoffmann C. 2008. Domain-specific class modelling for one-level representation of single trees. Dans : Blaschke T., Lang Stefan, Hay G.J. (éd.). *Object-Based Image Analysis*. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 133-151.
- Vapnik V.N. 2000. *The nature of statistical learning theory*. 2<sup>e</sup> éd. New York : Springer
- Verlič A., Đurić N., Kokalj Ž., Marsetič A., Simončič P., & Oštir K. 2014. Tree Species Classification using WorldView-2 Satellite Images and Laser Scanning Data in a natural Urban Forest. *Šumarski list*. , p. 13.
- Waser L., Küchler M., Jütte K., & Stampfer T. 2014. Evaluating the Potential of WorldView-2 Data to Classify Tree Species and Different Levels of Ash Mortality. *Remote Sensing*. 6(5), p. 4515–4545.
- Weidner U. 2008. Contribution to the assessment of segmentation quality for remote sensing applications. , p. 6.
- Wood P.J. & Centre Technique Forestier Tropical, éd. 1989. *Faidherbia albida (Del.) A. Chev. (synonyme: Acacia albida Del.): a monograph*. Nogent-sur-Marne : , 66 p.
- Yonezawa C. 2007. Maximum likelihood classification combined with spectral angle mapper algorithm for high resolution satellite imagery. *International Journal of Remote Sensing*. 28(16), p. 3729–3737.

## Annexes

### Annexe 1. Formules de conversion de la luminance absolue en réflectance au niveau de l'atmosphère

#### Conversion en luminance absolue

$$L = GAIN * DN * \left( \frac{absacalfactor}{effctivebandwhith} \right) + OFFSET$$

Où :

- DN est le compte numérique du pixel
- Gain est le gain pour chaque bande spectrale, mis à jour par DigitalGlobe
- Offset est le décalage pour chaque bande spectrale, mis à jour par DigitalGlobe
- Abscalfactor est le facteur de calibration de la bande spectrale
- Effectivebandwith est la largeur de la bande spectrale

#### Conversion en réflectance

$$\rho(TOA)_\lambda = \frac{L_\lambda d^2 \pi}{E_\lambda \cos \theta_s}$$

Où :

- $L_\lambda$  est la luminance absolue du capteur pour la bande spectrale  $\lambda$  exprimée en  $W/m^2/\mu m /sr$  ;
- $d$  est la distance Terre-Soleil en unités astronomiques ;
- $E_\lambda$  est l'irradiance exoatmosphérique de la bande spectrale, en  $W/m^2/\mu m$  ; et
- $\theta_s$  l'angle zénithale solaire.

**Annexe 2. Facteurs d'ajustement de radiométrie absolue et de conversion en réflectance TOA pour le capteur WV-3. Mise à jour du 06 juin 2017 par DigitalGlobe**

Band	Gain	Offset	Abscalfact.1	Abscalfact.2	E. Bandwidth	E	d	teta
PAN	0,95	-3,629	0,053047	0,052734	0,2896	1574,41	0,985559	0,67222
Bleu C.	0,905	-8,604	0,014344	0,014329	0,0405	1757,89	0,985559	0,67222
Bleu	0,94	-5,809	0,011653	0,011608	0,054	2004,61	0,985559	0,67222
Vert	0,938	-4,996	0,008873	0,008857	0,0618	1830,18	0,985559	0,67222
Jaune	0,962	-3,649	0,006902	0,006843	0,0381	1712,07	0,985559	0,67222
Rouge	0,964	-3,021	0,010257	0,0102	0,0585	1535,33	0,985559	0,67222
Red-edge	1	-4,521	0,006232	0,00622	0,0387	1348,08	0,985559	0,67222
PIR 1	0,961	-5,522	0,011847	0,011797	0,1004	1055,94	0,985559	0,67222
PIR 2	0,978	-2,992	0,010651	0,010638	0,0889	858,77	0,985559	0,67222

**Annexe 3. Indices spectraux calculés par utilisation de différentes bandes spectrales WV-3.**

« ID bande » correspond aux bandes utilisées pour le calcul de l'indice, et IS1 et IS2 les deux groupes d'indices calculés à partir des deux types de formules différentes (type « NDVI » et type « SAVI »)

ID bande	Indice spectral 1 (IS1)	Indice spectral 2 (IS2)
4 et 5	$(b4-b5)/(b4+b5)$	$1.5*(b4-b5)/(b4+b5+0.5)$
6 et 4	$(b6-b4)/(b6+b4)$	$1.5*(b6-b4)/(b6+b4+0.5)$
7 et 4	$(b7-b4)/(b7+b4)$	$1.5*(b7-b4)/(b7+b4+0.5)$
7 et 5	$(b7-b5)/(b7+b5) = NDVI1$	$1.5*(b7-b5)/(b7+b5+0.5) = SAVI1$
7 et 6	$(b7-b6)/(b7+b6)$	$1.5*(b7-b6)/(b7+b6+0.5)$
8 et 5	$(b8-b5)/(b8+b5) = NDVI2$	$1.5*(b8-b5)/(b8+b5+0.5) = SAVI2$
8 et 6	$(b8-b6)/(b8+b6)$	$1.5*(b8-b6)/(b8+b6+0.5)$

#### Annexe 4. Formulation mathématique des principaux descripteurs dérivés de la matrice de cooccurrence

Source : (Regniers, 2015)

---

$$\text{Entropie} = \sum_i \sum_j P_{ij} (-\ln P_{ij})$$

$$\text{Corrélation} = \sum_i \sum_j P_{ij} \left[ \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right]$$

$$\text{Contraste} = \sum_i \sum_j P_{ij} (i - j)^2$$

$$\text{Homogénéité} = \sum_i \sum_j \frac{P_{ij}}{1 + (i - j)^2}$$

---

Où :

- $P_{ij}$  = probabilité d'occurrence de la paire de niveaux de gris  $i, j$  (GLCM normalisé) ;
- $\mu_i = \sum_i i P_{ij}$  , la moyenne des entrées normalisées pour le pixel de référence de valeur  $i$  ;
- $\mu_j = \sum_j j P_{ij}$  , la moyenne des entrées normalisées pour le pixel de référence de valeur  $j$  ;
- $\sigma_i = \sqrt{\sum_i P_{ij} (i - \mu_i)^2}$  , l'écart-type des entrées normalisées pour le pixel de référence de valeur  $i$  ; et
- $\sigma_j = \sqrt{\sum_j P_{ij} (j - \mu_j)^2}$  , l'écart-type des entrées normalisées pour le pixel de référence de valeur  $j$ .

## **Annexe 5. Description des algorithmes utilisés pour l'entraînement et la classification des espèces d'arbres de la zone de Bambe**

### **A. Spectral angle mapper (SAM)**

Le SAM est un algorithme de classification qui détermine la similarité spectrale entre deux objets en calculant l'angle entre leurs signatures spectrales dans un espace vectoriel de dimension égale au nombre de bandes (Liu & Yang, 2013). Plus l'angle entre les signatures spectrales de deux objets est faible, plus ils sont spectralement proches (Naidoo et al., 2012). L'étude de Keshava (2004) montre que l'angle entre deux signatures spectrales peut accroître ou décroître en fonction des bandes fournies pour le calcul du SAM. La technique « Band Add On » qui implique de partir sur deux bandes et d'ajouter itérativement les autres bandes (Madonsela et al., 2017 ; Keshava, 2004) a été testée sous ENVI sur les moyennes et les écarts-type des réflectances. Quelques tests ont été réalisés sous ENVI pour déterminer les bandes permettant de minimiser l'angle entre les signatures spectrales des espèces, permettant de les discriminer.

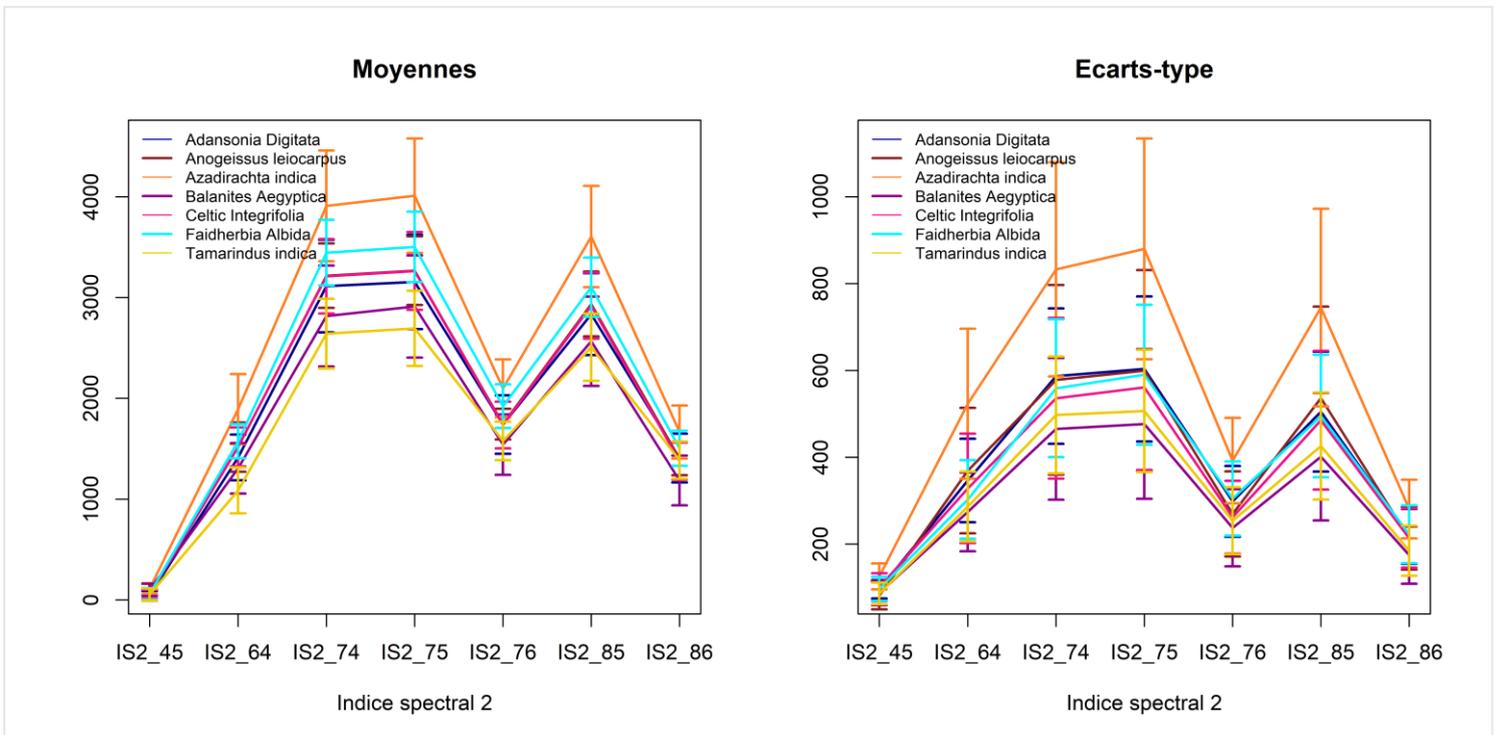
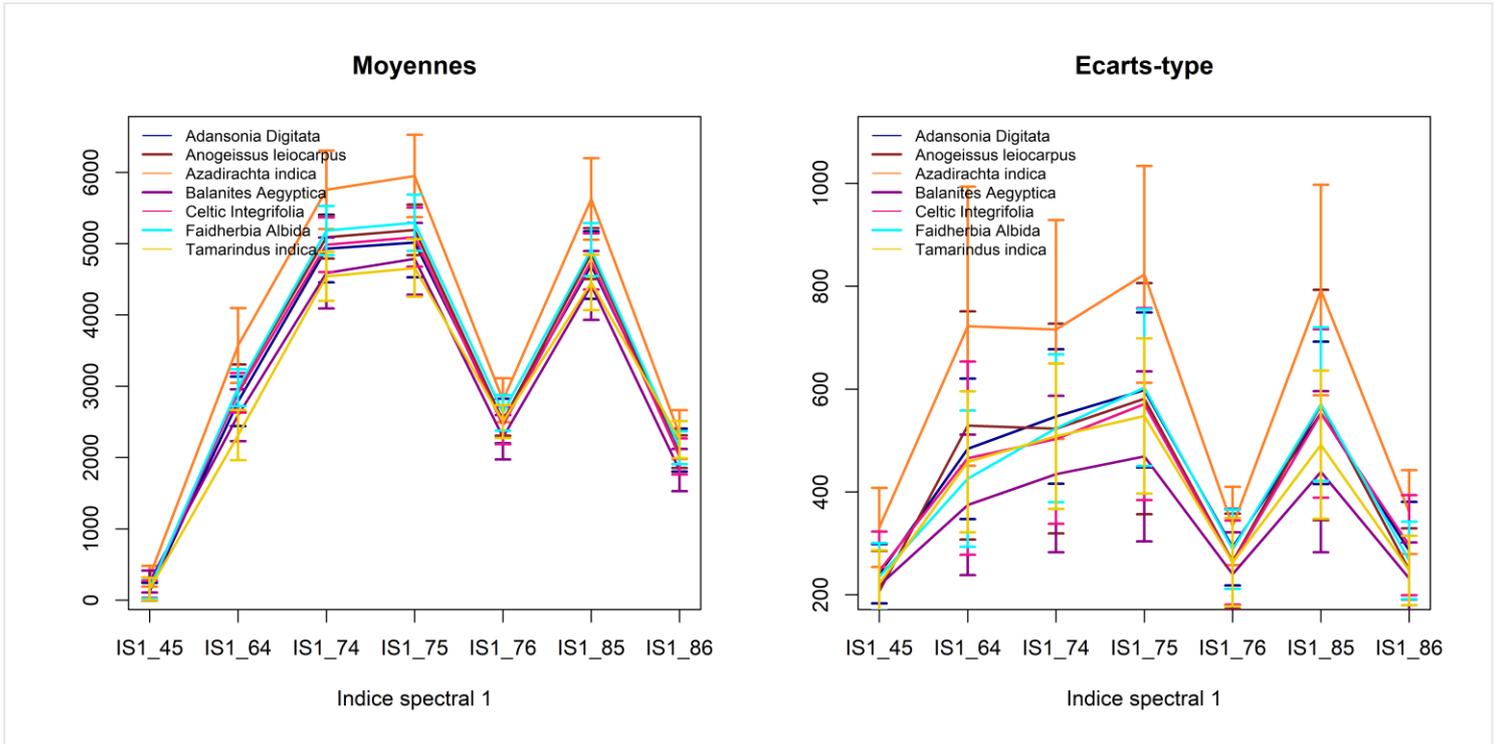
### **B. Support vector machine (SVM)**

Le SVM peut être défini comme un système utilisant l'hypothèse d'un espace à une fonction linéaire appliquée à un espace à grande dimensionnalité, à partir des données d'apprentissage (Jakkula, 2006). Il permet de classer 2 catégories de points en les assignant à l'un de deux espaces séparés dans un espace original en cas de classificateurs linéaires, ou dans un espace à plusieurs dimensions dans le cas de classificateurs non linéaires (Vapnik, 2000). Les algorithmes de SVM se différencient par les types de Kernels, qui sont des fonctions mathématiques utilisées dans les processus d'entraînement. Il existe différents types de fonctions Kernel : linéaire ; polynomial ; gaussien (radial basis function) ; et sigmoïde (Hsu et al., 2016). Nous avons utilisé les fonctions Kernel linéaire et gaussien pour ce projet. Les deux fonctions prennent en entrée le paramètre de « coût » représentant la largeur de l'hyperplan qui sépare les classes. En plus du « coût », le Kernel gaussien prend en compte le paramètre « gamma » qui correspond au degré de linéarité de l'hyperplan. Sous R, les valeurs optimales calculées de ces deux paramètres ont été de 1 pour le coût et 0,01 pour gamma.

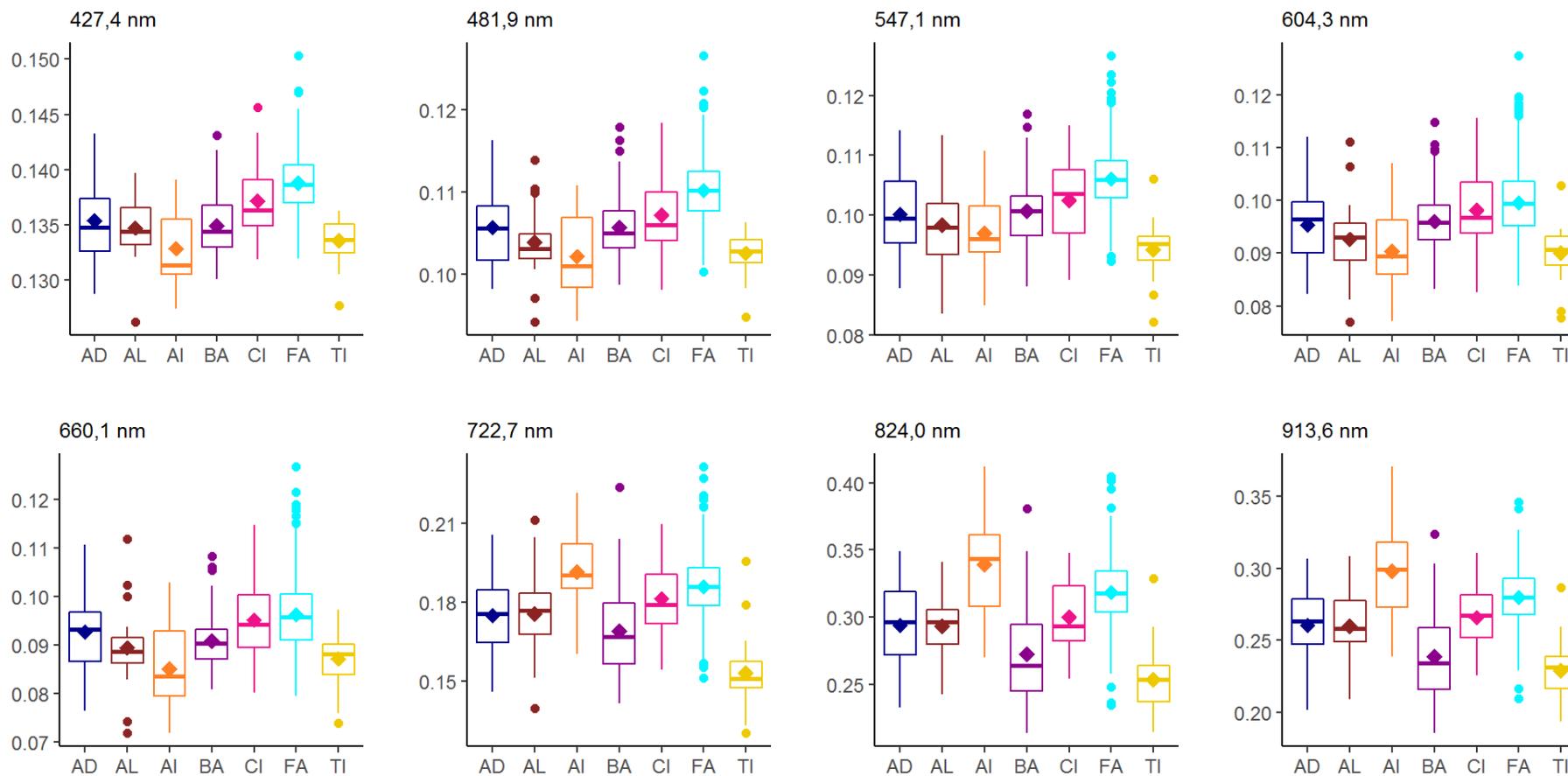
### **C. Random forest**

Random forest est un algorithme de classification et de régression. Il effectue un apprentissage sur des multiples arbres de décision à partir d'une sélection aléatoire sans remise des données d'entraînement au niveau de chaque arbre. La classification se fait par vote majoritaire pour les classes prédites au niveau de chaque arbre (Breiman, 2001 ; Fawagreh et al., 2014). Deux principaux paramètres sont nécessaires pour le random forest : le nombre d'arbres (généralement fixé à 500 par défaut) ; et le nombre de variables sélectionnées au niveau de chaque arbre. Les valeurs optimales de ces deux variables calculées après avoir réalisé plusieurs tests dans R sont respectivement de 500 et 16 (correspondant au nombre total des variables en entrée qui sont les moyennes et écarts-type des réflectances).

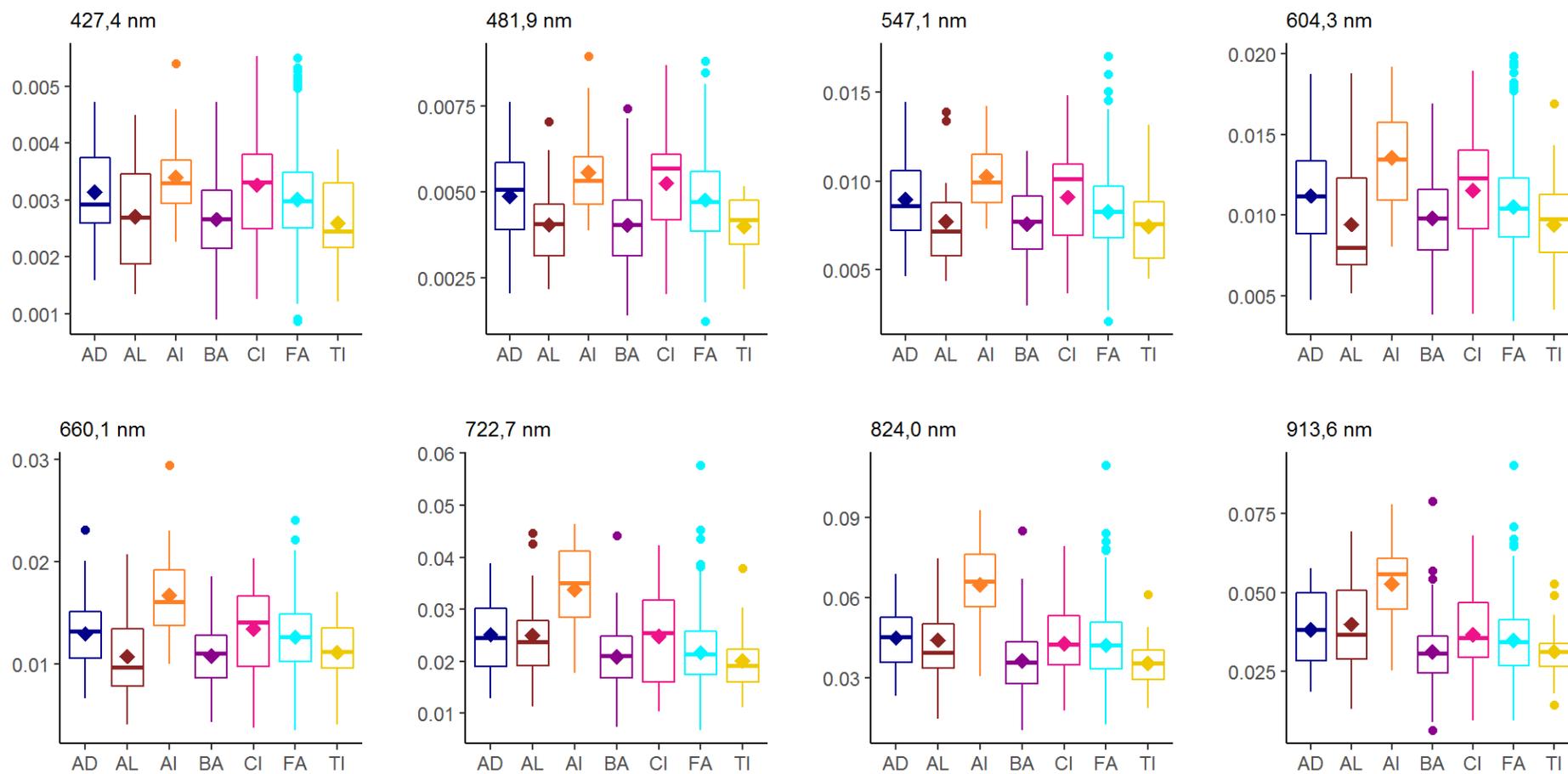
**Annexe 6. Moyennes et écarts-type des indices spectraux des espèces d'arbres de la zone de Bambey**



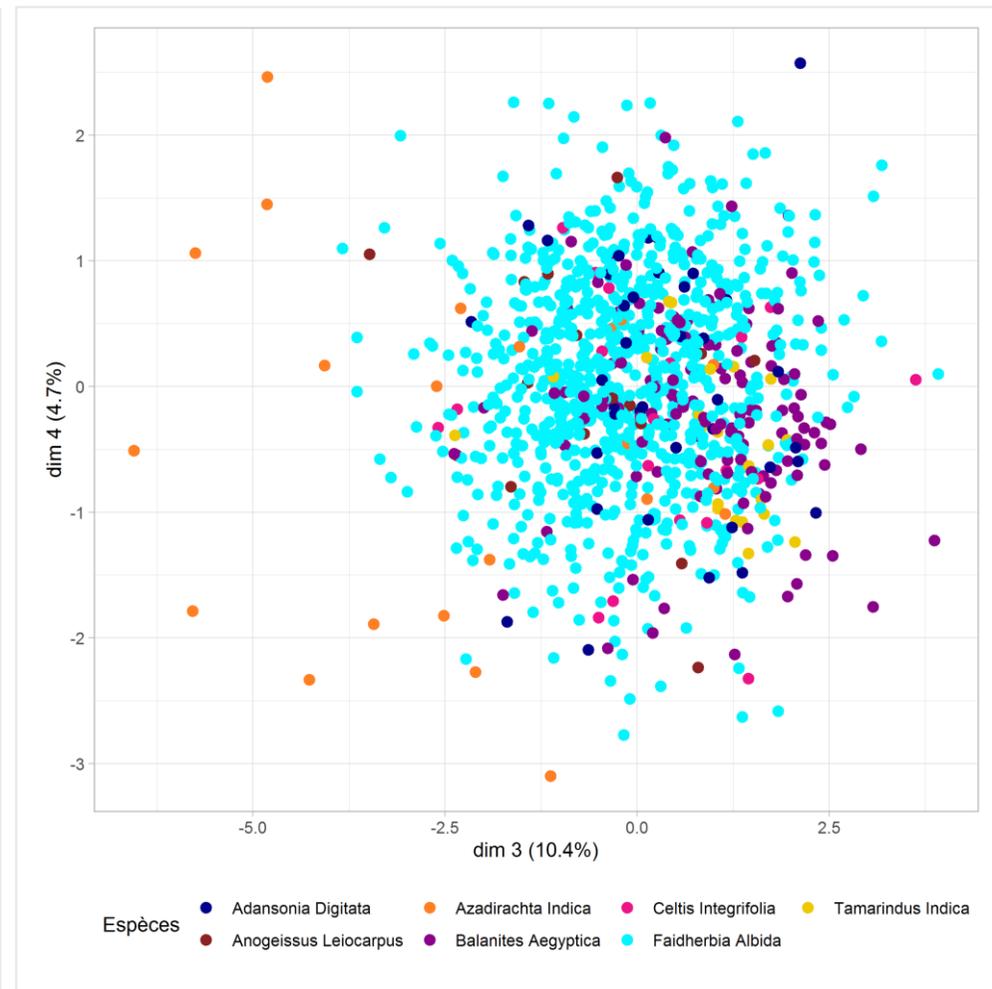
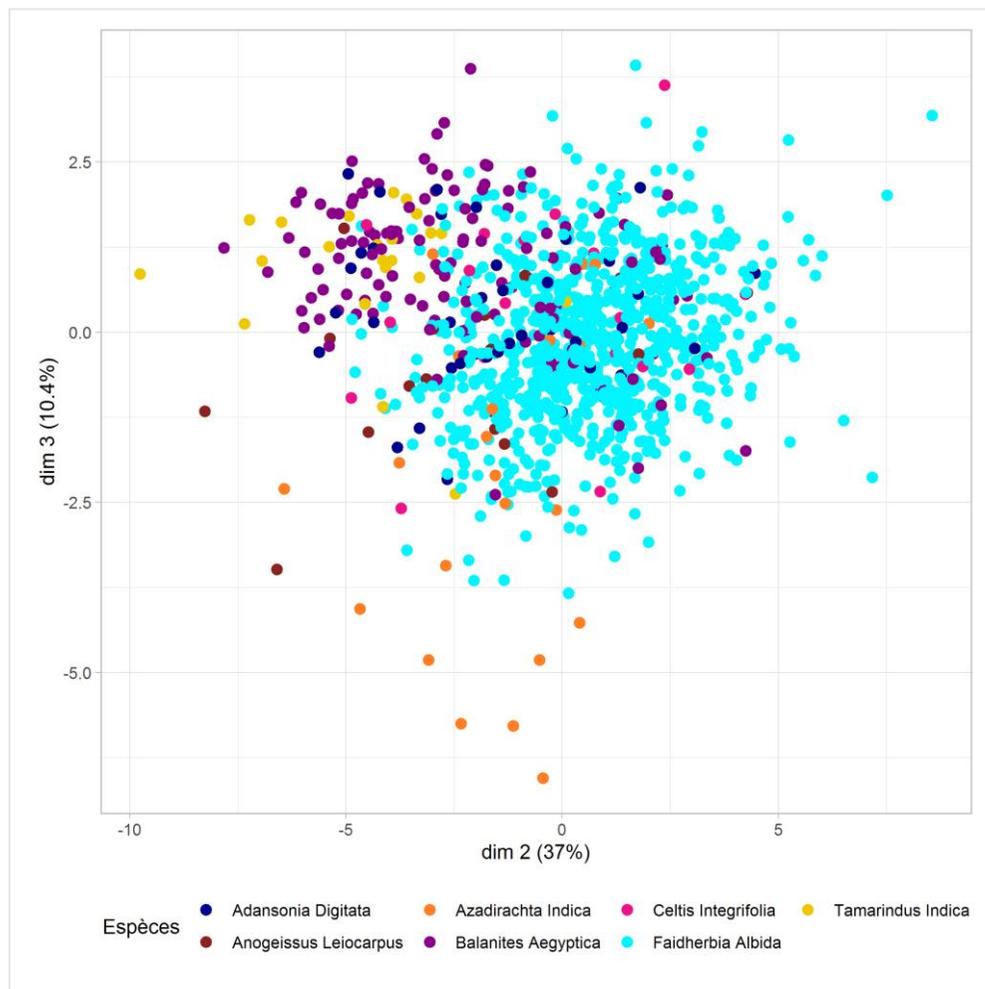
**Annexe 7. Boîtes à moustache des moyennes radiométriques des espèces d'arbres de la zone de Bambeby**



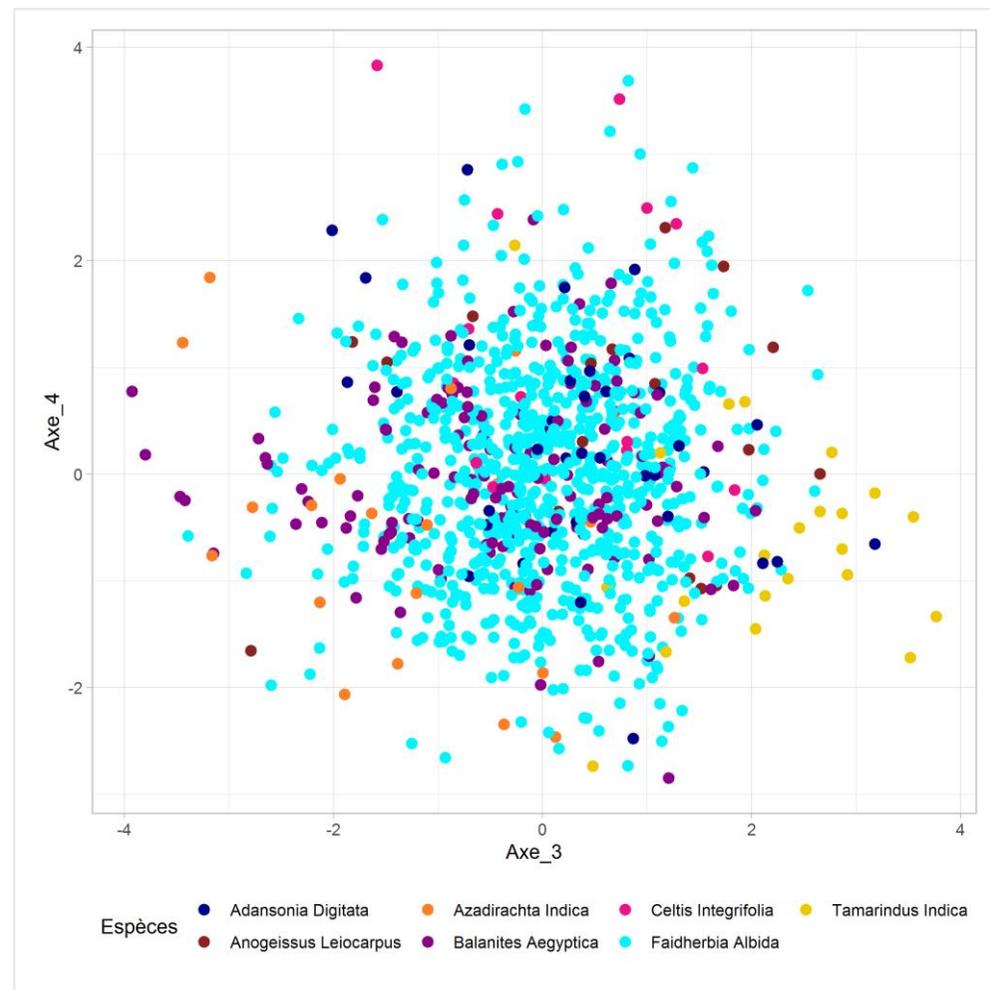
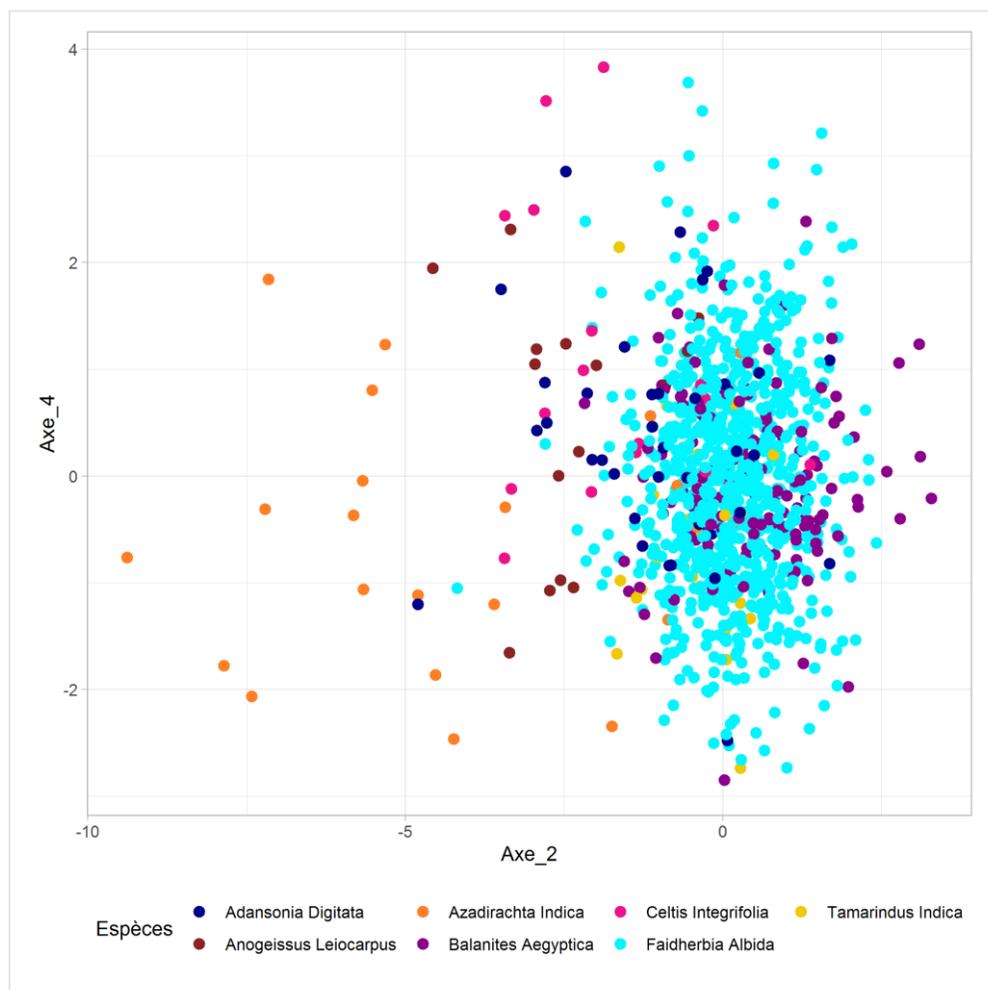
### Annexe 8. Boîtes à moustache des écarts-type radiométriques des espèces d'arbres de la zone de Bambeby



## Annexe 9. Projection des axes 2-3 et 3-4 de l'ACP sur les réflectances des arbres



Annexe 10. Projection des axes 2-3 et 3-4 de l'AFD sur les réflectances des arbres



**Annexe 11. Matrices de confusion produites par classification SVM linéaire sur 3 tirages différents des données d'entraînement, pour 5 et 6 classes d'espèces d'arbres.**

**Jeu de données 1**

**Classification avec 5 classes**

Classes		Prédites					Total	Precis. Util
		<i>Autres especes</i>	<i>A. indica</i>	<i>B. aegyptica</i>	<i>F. Albida</i>	<i>T. Indica</i>		
Observées	<i>Autres especes</i>	17	0	3	16	0	36	47,2
	<i>A. indica</i>	0	2	0	4	0	6	33,3
	<i>B. aegyptica</i>	0	0	25	14	0	39	64,1
	<i>F. Albida</i>	1	0	1	244	0	246	99,2
	<i>T. Indica</i>	2	0	2	0	2	6	33,3
	Total	20	2	31	278	2	333	
	Précis. prod	85,0	100,0	80,6	87,8	100,0		

**Classification avec 6 classes**

Classes		Prédites					Total	Precis. Util	
		<i>A. digitata</i>	<i>Autres especes</i>	<i>A. indica</i>	<i>B. aegyptica</i>	<i>F. Albida</i>			<i>T. Indica</i>
Observées	<i>A. digitata</i>	0	2	1	2	7	0	12	0,0
	<i>Autres especes</i>	0	6	0	2	15	1	24	25,0
	<i>A. indica</i>	0	1	2	0	3	0	6	33,3
	<i>B. aegyptica</i>	0	0	0	26	13	0	39	66,7
	<i>F. Albida</i>	0	0	0	2	244	0	246	99,2
	<i>T. Indica</i>	0	1	0	3	0	2	6	33,3
	Total	0	10	3	35	282	3	333	
	Précis. prod	-	60,0	66,7	74,3	86,5	66,7		

**Jeu de données 2**

**Classification avec 5 classes**

Classes		Prédites					Total	Precis. Util
		<i>Autres especes</i>	<i>A. indica</i>	<i>B. aegyptica</i>	<i>F. Albida</i>	<i>T. Indica</i>		
Observées	<i>Autres especes</i>	20	0	2	13	1	36	55,6
	<i>A. indica</i>	1	3	0	2	0	6	50,0
	<i>B. aegyptica</i>	1	1	25	12	0	39	64,1
	<i>F. Albida</i>	4	0	0	242	0	246	98,4
	<i>T. Indica</i>	4	0	0	0	2	6	33,3
	Total	30	4	27	269	3	333	
	Précis. prod	66,7	75,0	92,6	90,0	66,7		

### Classification avec 6 classes

	Classes	Prédites						Total	Precis. Util
		<i>A. digitata</i>	<i>Autres espèces</i>	<i>A. indica</i>	<i>B. aegyptica</i>	<i>F. Albida</i>	<i>T. Indica</i>		
Observées	<i>A. digitata</i>	1	2	0	1	8	0	12	8,3
	<i>Autres espèces</i>	1	9	1	0	13	0	24	37,5
	<i>A. indica</i>	0	1	3	0	2	0	6	50,0
	<i>B. aegyptica</i>	1	1	1	24	12	0	39	61,5
	<i>F. Albida</i>	1	2	0	0	243	0	246	98,8
	<i>T. Indica</i>	0	4	0	0	0	2	6	33,3
	Total	4	19	5	25	278	2	333	
Précis. prod	25,0	47,4	60,0	96,0	87,4	100,0			

### Jeu de données 3

#### Classification avec 5 classes

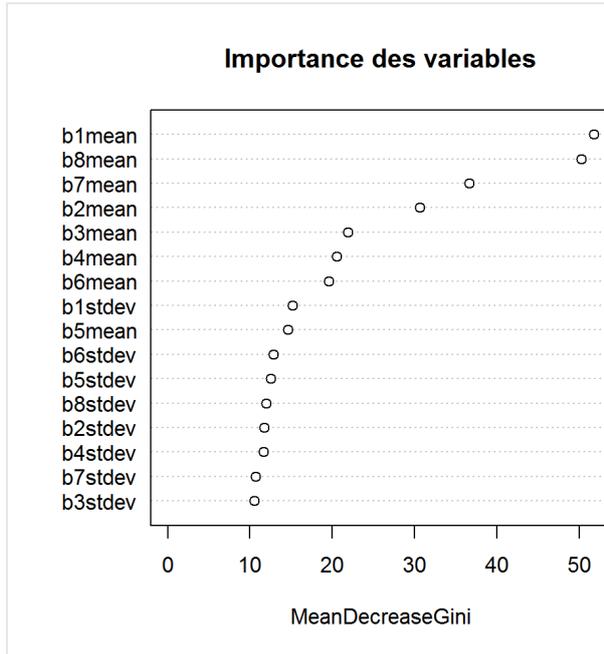
	Classes	Prédites					Total	Precis. Util
		<i>Autres especes</i>	<i>A. indica</i>	<i>B. aegyptica</i>	<i>F. Albida</i>	<i>T. Indica</i>		
Observées	<i>Autres especes</i>	16	0	1	16	3	36	44,4
	<i>A. indica</i>	1	3	0	2	0	6	50,0
	<i>B. aegyptica</i>	2	0	26	11	0	39	66,7
	<i>F. Albida</i>	0	1	2	243	0	246	98,8
	<i>T. Indica</i>	1	0	1	2	2	6	33,3
	Total	20	4	30	274	5	333	
	Précis. prod	80,0	75,0	86,7	88,7	40,0		

#### Classification avec 6 classes

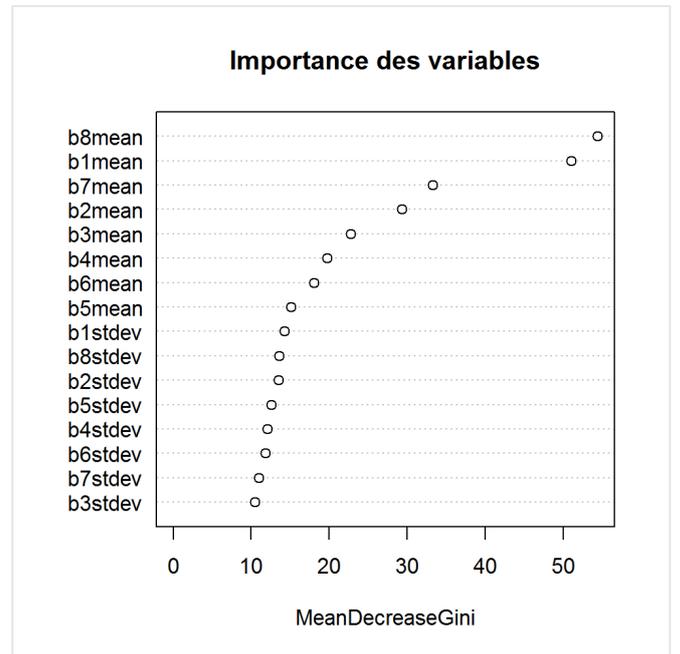
	Classes	Prédites						Total	Precis. Util
		<i>A. digitata</i>	<i>Autres espèces</i>	<i>A. indica</i>	<i>B. aegyptica</i>	<i>F. Albida</i>	<i>T. Indica</i>		
Observées	<i>A. digitata</i>	1	1	0	1	8	1	12	8,3
	<i>Autres espèces</i>	0	9	1	1	13	0	24	37,5
	<i>A. indica</i>	0	1	3	0	2	0	6	50,0
	<i>B. aegyptica</i>	0	3	0	23	11	2	39	59,0
	<i>F. Albida</i>	0	0	1	1	244	0	246	99,2
	<i>T. Indica</i>	0	0	0	0	3	3	6	50,0
	Total	1	14	5	26	281	6	333	
Précis. prod	100,0	64,3	60,0	88,5	86,8	50,0			

**Annexe 12. Importance des bandes spectrales dans la prédiction de la classification à partir de l'entraînement du RF sur 3 tirages différents des données d'entraînement.**

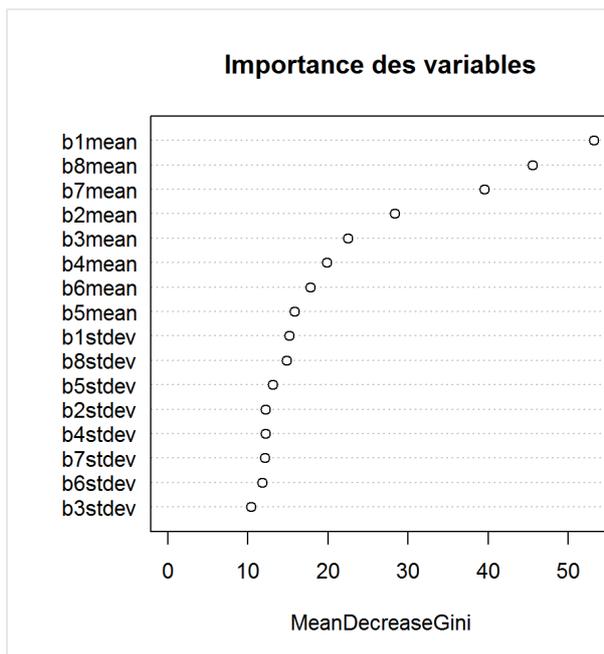
1.



3.



2.



**Annexe 13. Comparaison des précisions apportées par les bandes multispectrales de WV-3 (16 variables) et ses bandes bleue, verte, rouge et PIR1 (8 variables) sur la classification de 30 tirages différents des données d'entraînement.**

Jeu	16 variables		8 variables	
	Globale	Kappa	Globale	Kappa
1	0,87087087	0,64686298	0,83483484	0,52293506
2	0,87687688	0,67905501	0,82282282	0,50733469
3	0,87087087	0,6553044	0,82582583	0,49863718
4	0,85885886	0,61838929	0,82282282	0,51438529
5	0,86186186	0,63492064	0,84084084	0,55265759
6	0,86186186	0,63609151	0,81981982	0,50812408
7	0,87087087	0,65846154	0,84384384	0,55449213
8	0,86786787	0,63897102	0,83483484	0,53382712
9	0,85285285	0,60183016	0,81981982	0,49426684
10	0,87687688	0,67439365	0,81681682	0,48564266
11	0,87387387	0,66592619	0,83783784	0,53527679
12	0,86486487	0,64390105	0,82282282	0,50525043
13	0,84084084	0,57606111	0,81681682	0,50121547
14	0,85885886	0,62288564	0,82882883	0,5066282
15	0,86786787	0,64911272	0,82882883	0,5202093
16	0,84084084	0,5581343	0,81381381	0,46820184
17	0,84684685	0,58784128	0,82582583	0,5331061
18	0,84984985	0,6011785	0,84084084	0,56970451
19	0,84684685	0,58487937	0,82282282	0,49915876
20	0,86786787	0,64830417	0,83183183	0,52866242
21	0,87087087	0,64839779	0,84684685	0,57197943
22	0,85885886	0,61966902	0,82882883	0,5173913
23	0,85585586	0,62276975	0,82282282	0,50688954
24	0,85285285	0,59591382	0,82582583	0,50204966
25	0,85585586	0,60194247	0,81981982	0,47526001
26	0,84084084	0,57282893	0,80480481	0,45497809
27	0,88888889	0,71271684	0,82882883	0,53174956
28	0,87387387	0,66900959	0,81981982	0,51326464
29	0,86486487	0,64339259	0,82282282	0,52864546
30	0,86486487	0,64056991	0,82282282	0,49969442