Reconnaissance spécifique des arbres de la canopée en forêt tropicale par fusion de données lidar et hyperspectrales

Anthony Laybros

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En écologie et biodiversité

École doctorale : GAIA – Biodiversité, Agriculture, Alimentation Environnement, Terre, Eau

Unité de recherche : UMR AMAP, Montpellier

RECONNAISSANCE SPECIFIQUE DES ARBRES DE LA CANOPEE EN FORET TROPICALE PAR FUSION DE DONNEES LIDAR ET HYPERSPECTRALES

Présentée par Anthony LAYBROS
Le 4 mars 2021

Sous la direction de Grégoire VINCENT

Devant le jury composé de

Nesrine CHEHATA, Maître de conférences, EA G&E, Bordeaux INP	Rapporteur
Audrey MINGHELLI, Maître de conférences, UMR LIS, Université de Toulon	Rapporteur
Jean-Phillipe GASTELLU - ETCHEGORRY, Professeur, CESBIO, Université de Toulouse III	Examinateur
Jocelyn CHANUSSOT, GIPSA-Lab, Grenoble INP	Examinateur
Raphael PELISSIER, Directeur de recherche, UMR AMAP, IRD	Examinateur
Xavier BRIOTTET, Directeur de recherche, ONERA	Invité
Olivier BRUNAUX, Responsable de pôle RDI de Cayenne, ONF	Invité
Grégoire VINCENT, Chargé de recherche, UMR AMAP, IRD	Directeur

UNIVERSITÉ DE MONTPELLIER

I. Remerciements

Tout d'abord, je tiens à remercier Grégoire Vincent, mon directeur de thèse de m'avoir permis de réaliser cette thèse. Merci de m'avoir fait confiance pour mener à bien ce travail et de m'avoir accompagné tout au long de cette tâche. Merci de m'avoir enrichi de par tes connaissances et de m'avoir transmis ton intérêt pour ce biome particulier qu'est la forêt tropicale.

Merci au CNES d'avoir entrepris, sur différents sites en forêt tropicale avec une multitude de capteurs, des acquisitions qui forment un jeu de données exceptionnel, précieux et rare. Merci à Marie-José Lefèvre qui a porté le projet d'acquisition de ces données sur ces beaux sites d'études au sein du CNES. J'espère que le CNES continuera ses efforts pour la création de satellites hyperspectraux.

Merci à l'ONF, pour avoir soutenu la création de cette thèse. Plus particulièrement, merci à Laurent Descroix, qui a soutenu le projet au sein de l'ONF, qui a vu en l'hyperspectral et au LiDAR la capacité d'améliorer la gestion et l'aménagement de l'exploitation forestière en Guyanes française, dont l'ONF à la charge. Merci à Olivier Brunaux et Caroline Bedeaux de continuer à soutenir l'hyperspectral appliqué à la foresterie tropicale au sein du pôle R&D Guyane de l'ONF.

Merci à l'ANRT (Association Nationale Recherche Technologie), qui au nom du MESRI (Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation) a soutenu financièrement cette étude.

Merci à ces petites mains qui ont de façon méticuleuses identifiées, répertoriées les espèces des arbres sur les sites expérimentaux. Grâce à leurs efforts et leurs patiences, j'ai pu profiter d'un inventaire riche et exceptionnel. Merci à Jean-Louis Smock, Chantal Geniez, Carla Baltzer, Raphaël Dutrieux, Vincent Bézard, Daniel Sabatier,…

Merci à mon comité de suivi pour leurs aides et leurs temps précieux afin que je puisse mener au mieux mes travaux et ma première expérience dans le monde de la recherche. Merci donc à Jocelyn Chanussot, Anne Jolly, Raphaël Pelissier, Jean-François Molino, Laurent Descroix, Jean-Baptiste Féret et bien sûr Agnès Begue. La pluridisciplinarité de cette équipe m'a permis de considérer de multiple aspect, bien que malheureusement tous n'ont pas été explorés.

Merci à Nesrine Chehata et Audrey Minghelli d'avoir accepté d'être les rapporteurs de mon travail de thèse. Merci aux membres du jury, Jean-Phillipe Gastellu, Xavier Briottet, Raphael Pelissier et Jocelyn Chanussot, d'avoir pris du temps pour évaluer ce travail. Ainsi que merci à Olivier Brunaux d'avoir accepté d'être membre invité du jury.

Merci à l'UMR – AMAP, représenté par Thierry Fourcaud de m'avoir accueilli dans ses locaux, dans un cadre interdisciplinaire et convivial. Merci à Valérie Roinel pour la gestion de la partie administrative entre l'IRD, CIRAD et l'ONF.

Merci à Sandrine Poittevin, secrétaire générale à l'ONF, et Sonia Martinez, gestionnaire RH, pour m'avoir facilité la partie administrative au sein de l'ONF. Merci à Cendrine Jay-Allemand, assistance responsable de l'école doctorale GAIA qui accompagne les doctorants tout le long de leur cursus doctoral, pour sa réactivité et son efficacité.

Merci à Marc Bouvy, responsable de l'école doctorale GAIA et à l'ONF qui m'ont accordé et ont justifié une demande de prolongation de 3 mois afin que je puisse finir de rédiger cette thèse dans les meilleures conditions.

Merci à tous les collègues AMAPiens et ami(e)s, toujours présent à la machine à café, qui raccourcissent et adoucissent les journées. Merci donc à Colette, Claudia, François(s) P. et G., Yohann, Claire, Alain, Nicolas Be. (dit l'indécis), Antoine, Camille PS-1, Dimitri, Sue, Jeanne, Olivier, Rhaphael, Camille PS-2, Hervé, Mathias, Julien, Guillaume, Alice, Philippe B., Philippe V., Christophe, Pierre, Cédric, Nicolas Ba., Gilles, Maxime, Amaury, Maximilien, Boris, Tom, Colin, Guanqgi, Dav, Emilie, Begum, Florian, Maïri, Mégane,Alex, Eliott…, Merci à Florence avec qui nous avons gravi la TAF, passé un bon séjour à Paracou, régalé par la cuisine du Chef Grégoire. Cela compensait les « tuc » et les maquereaux sauce escabèche.

Merci aux coureurs, pour les sorties autour du laboratoire, Ghislain, Philippe V., Christophe. Il ne faut pas négliger cette détente simple et efficace pour se relâcher quelques instants. Un merci tout particulier à Mélaine, co-résidente dans le bureau intitulé « Ne pas entrer, ici rien ne marche ». Merci pour ton aide et ton soutien durant cet exercice de thèse. Rammstein est un bon exutoire quand rien ne

marche. Merci aussi de m'avoir incité à faire du Trail et d'avoir joué le rôle de lièvre durant 2 années. En espérant avoir l'opportunité de courir de nouveau ensemble dans les montagnes sur de longues distances et qui sait, un jour je serais devant toi et je t'entendrais respirer.

Merci à Bastien B. de m'avoir fait découvrir la recherche et plus particulièrement la nutrition végétale. « Si toi et moi on avait un enfant, ce serait un génie » H.

Un grand merci à ma famille. Merci à mon grand-père pour tous ces moments, parti cette année, avant que je n'ai pu l'impressionner avec mon diplôme. Merci Papa et Maman pour m'avoir toujours soutenu durant mon parcours scolaire, d'avoir accepté de me déménager aux 4 coins de la France. Grand-frère, petite-sœur et Maxime, nous sommes voués à moins se voir alors j'attends vos coups de fils ! Merci à Gérard, Sophie et Alizée pour m'avoir accueilli au sein de votre famille. Marie, un grand merci pour tout. Ton soutien permanent, ton enthousiasme et ton bonheur contagieux ont été précieux pour parcourir ces 3 années. Notre couple a grandement mûri durant cette ardente patience. La fin de cette thèse est donc l'achèvement de nos années séparées, du célibat géographique mais le début, enfin, d'une vie commune.

II. Résumé

Les forêts tropicales, représentant 6.4% de la surface terrestre, abritent la plus grande biodiversité des écosystèmes terrestres et jouent un rôle fondamental dans le cycle du carbone à l'échelle mondiale. La durabilité de l'exploitation des forêts tropicales est un enjeu fondamental tant du point de vue de la conservation de la biodiversité que de la réduction des émissions liées à la déforestation et à la dégradation des forêts (REDD +). L'Office National des Forêts (ONF) est chargé de la conservation et de la gestion de 6 millions d'hectares de forêts du domaine privé de l'Etat en Guyane française.

La donnée d'inventaire spatialisé à l'échelle du paysage contribuerait à faire progresser les connaissances fondamentales sur ce biome complexe et menacé et aiderait à sa gestion durable. Les cartes de distribution d'espèces peuvent en effet être croisées avec les facteurs environnementaux et fournir ainsi des clés d'interprétation des schémas d'organisation des peuplements forestiers. Du point de vue de la gestion, les cartes de distribution des espèces permettent une rationalisation de l'exploitation forestière. La cartographie des espèces commerciales pourrait ainsi favoriser des pratiques forestières minimisant l'impact environnemental de l'exploitation. L'identification des espèces permettrait de prioriser les zones particulièrement riches en espèces commerciales, tout en évitant d'ouvrir des pistes d'exploitation dans les zones à faible niveau de ressources exploitables. La télédétection offre également la possibilité de surveiller l'extension des espèces proliférantes, telles que les lianes.

La possibilité de cartographier les espèces dans la canopée par télédétection est donc d'un intérêt évident, tant d'un point de vue de la gestion que de la connaissance scientifique.

Des capteurs hyperspectraux et LiDAR ont été utilisés à bord d'un avion pour identifier les espèces dans les forêts tropicales guyanaises. Une large gamme spectrale issue des capteurs hyperspectraux (400–2500 nm) est mesurée permettant d'avoir de nombreux descripteurs. Le LiDAR embarqué offre une description fine de la structure du couvert, facilitant la segmentation des houppiers. La fusion de ces deux informations améliore la caractérisation de la ressource.

Afin de tirer le meilleur parti des données hyperspectrales, différents prétraitements radiométriques ont été évalués. Le lissage spatial et le filtrage des ombres sont les principaux facteurs qui améliorent la discrimination des espèces. L'utilisation de la gamme spectrale complète est également bénéfique. Ces résultats de classification ont été obtenus sur un groupe 20 espèces abondantes. L'identification de ces mêmes espèces en mélange au sein d'un peuplement hyperdivers a constitué la deuxième étape de ce travail.

Nous avons évalué le niveau d'information nécessaire et le degré de confusion tolérable dans les données d'apprentissage afin de retrouver une espèce cible dans une canopée hyperdiverse. Une méthode de classification spécifique a été mise en œuvre pour être insensible à la contamination entre classes focales/non focales. Même dans le cas où la classe non focale contient jusqu'à 5% de pixels de la classe focale (espèce à identifier), les classifieurs se sont révélés efficaces.

La troisième étape aborde le problème de la transposabilité des classifieurs d'une acquisition à une autre. La caractérisation des conditions d'acquisition et la prise en compte de leurs effets sont nécessaires pour convertir les données de radiance en réflectance de surface. Cependant cette opération de standardisation reste une étape extrêmement délicate au vue des nombreuses sources de variabilité : état de l'atmosphère, géométrie soleil-capteur et conditions d'éclairement. Nous évaluons en comparant des vols répétés sur le même site, la contribution des diverses caractéristiques d'acquisition à la divergence spectrale entre dates. Ce travail vise à proposer des pistes pour développer des méthodes de reconnaissance d'espèces qui soient plus robustes aux variations des caractéristiques d'acquisition.

Mots clés : Biodiversité, Forêt tropicale, Gestion forestière, Hyperspectral, LiDAR, Segmentation et identification des arbres, Apprentissage automatique, Prétraitement, Correction atmosphérique, Variations spectrales.

III. Abstract

Tropical forests, representing 6.4% of the Earth's surface, host the greatest biodiversity of any terrestrial ecosystem and play a fundamental role in the carbon cycle on a worldwide scale. The sustainable use of tropical forests is a fundamental issue from both the point of view of biodiversity conservation and the reduction of emissions from deforestation and forest degradation (REDD+). The Office National des Forêts is responsible for the conservation and management of 6 million hectares of forests in French Guiana. The possibility of mapping species in the canopy by remote sensing is of obvious interest. For both applied and scientific purposes, the use of airborne observation measurements can enable local field information that is difficult to collect over large areas of tropical forest to be extrapolated.

The specific spatialized inventories at the landscape scale would contribute to advancing fundamental knowledge of this complex and threatened biome and assist in its sustainable management. Maps of species distribution can in fact be cross-referenced with maps of environmental factors and thus provide keys for interpreting the organization patterns of forest stands. From a management point of view, species distribution maps are an help to the rationalization of forestry operations. The mapping of commercial species could promote forestry practices that minimize the environmental impact of logging. The identification of species would in particular enable priority to be given to areas that are particularly rich in commercial species, while avoiding the opening up of exploitation tracks in areas with low levels of exploitable resources. Remote sensing also offers the possibility of monitoring the spread of pervasive species, such as lianas.

Hyperspectral imagers and LiDAR sensor have been used on board an aircraft to identify species in the Guyanese tropical forests. A wide spectral range from hyperspectral sensors (400-2500 nm) is measured allowing to have many descriptors. LiDAR provides a detailed description of canopy structure and facilitates the segmentation of canopies. The fusion of these two types of information improve the characterization of the resource.

In order to make the most of the hyperspectral data, different radiometric preprocessing has been evaluated. Spatial smoothing and shadow filtering are the main factors that improve species discrimination. The full spectral range rather than only the visible-near-infrared region (400-1000nm) is also beneficial. These classification results were obtained on a group of 20 abundant species. The identification of these same species in a mixture within a hyperdiverse stand was the second step of this work.

We thus assessed the level of spectral information required and the degree of confusion tolerable in the learning data when the task is to find a target species in a hyperdiverse canopy. A special classification method was implemented in order not to be sensitive to contamination between focal/non-focal classes for training. Even in the case where the non-focal class contains up to 5% of pixels of the focal class (species to be identified), the classifiers developed proved to be efficient.

The third step deals with the problem of transposability of the classifiers from one acquisition to another. The characterization of the acquisition conditions and the consideration of their effects are necessary to convert the radiance data into surface reflectance. However, this standardization operation remains an extremely delicate step given the many variability sources to be considered: state of the atmosphere, sun-sensor geometry and illumination conditions. By comparing repeated flights on the same site, we evaluate the contribution of the various acquisition characteristics to the spectral divergence between dates. This work aims to propose ways to develop species recognition methods that are more robust to variations in acquisition characteristics.

Keywords: Biodiversity, Tropical Forest, Forest Management, Hyperspectral, LiDAR, Tree Segmentation and Identification, Machine Learning, Preprocessing, Atmospheric Correction, Spectral Variations.

Contents

IV. Scientific contributions

The main contributions are described below:

1. Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J., and Vincent, G. (2019). Across Date Species Detection Using Airborne Imaging Spectroscopy. Remote Sens. 11, 789.
2. Laybros, A., Aubry-Kientz, M., Féret, J.-B., Bedeau, C., Brunaux, O., Derroire, G., and Vincent, G. (2020). Quantitative Airborne Inventories in Dense Tropical Forest Using Imaging Spectroscopy. Remote Sens. 12, 1577.

The other contributions are described below:

1. Tusa, E., Laybros, A., Monnet, J.-M., Dalla Mura, M., Barré, J.-B., Vincent, G., Dalponte, M., Féret, J.-B., and Chanussot, J. (2020). Fusion of hyperspectral imaging and LiDAR for forest monitoring. In Data Handling in Science and Technology, (Elsevier), pp. 281–303.
2. Aubry-Kientz, M., Laybros, Anthony, Lefèvre-Fonollosa, Marie-José, Vincent, Grégoire, 2019. Towards high throuput inventory of tropical forest using aerial LiDAR and multispectral data fusion. Presented at the WHISPERS, Amsterdam.
3. Aubry-Kientz, M., Laybros, A., Weinstein, B.G., Ball, J., Jackson, T., Coomes, D., and Vincent, G. (Submitted). Multi-sensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. J. Sel. Top. Appl. Earth Obs. Remote Sens.

V. Introduction générale (Version française)

Les forêts tropicales abritent des écosystèmes complexes qui ne sont que partiellement explorés. Elles abritent une grande diversité de formes de vie et contribuent au bien-être humain à l'échelle mondiale grâce à la régulation du climat (Lewis et al., 2015) et aux cycles biogéochimiques (Powers and Marín-Spiotta, 2017). Par conséquent, leurs fonctions fournissent de multiples services écosystémiques tels que définis par (Alcamo, Bennett, et Millennium Ecosystem Assessment (Program) 2003). Cependant, l'étendue de la forêt tropicale mature est de plus en plus réduite en raison de la pression humaine (Barlow et al., 2018; Wright, 2010), quel que soit le continent considéré. Entre 1980 et 2000, 28 % des nouvelles terres agricoles ont été conquises sur des forêts matures (Gibbs et al., 2010). Entre 2000 et 2005, environ 1,4 % des forêts tropicales humides ont subi une déforestation pour l'extraction de ressources en bois et/ou pour l'agriculture (Foley et al., 2005). Cependant, (Song et al., 2018) ont montré qu'entre 1982 et 2006, la superficie forestière mondiale a augmenté de 7,1 %. La perte dans les tropiques a été largement compensée par le gain dans les zones extratropicales, telles que les systèmes de montagne, comme l'a indiqué (Piao et al., 2015). La situation des forêts dans le monde (2020) fait état d'un bilan mitigé. Une augmentation des zones forestières a été signalée en Asie, en Océanie et en Europe entre 2010 et 2020. Le continent américain (nord, centre et sud) et le continent africain ont observé une tendance inverse, avec une perte significative, en particulier pour le continent africain qui a subi à une perte croissante entre 1990 et 2020. Dans la forêt tropicale guyanaise, la perte de surface forestière a été estimée à 0,28 % entre 2011 et 2014 (Pickering et al., 2019). Selon (FAO, 2020), à plus grande échelle, entre 1990 et 2020, la forêt tropicale guyanaise a perdu environ 1,51 % de sa superficie.

D'une part, plus de 60 millions de personnes dépendent entièrement des forêts (Dieterle, 2009) et 1,2 milliard de personnes tirent leur subsistance des produits forestiers (Vantomme, 2011). En 2012, INTERPOLE a estimé que le commerce illégal du bois représentait entre 50 et 150 milliards de dollars par an, soit environ 30 % du commerce total du bois, avec une demande mondiale croissante de ressources en bois et une capacité insuffisante pour contrôler ce commerce illégal. D'autre part, l'importance de la conservation des écosystèmes est clairement identifiée. La dégradation des forêts tropicales entraîne des modifications dans divers domaines, notamment : l'hydrologie (Eshleman, 2004; Giambelluca, 2002; Likens et al., 1978; Whitehead and Robinson, 1993) (modification des cours d'eau, du ruissellement et des réservoirs d'eau naturels, augmentation / diminution des épisodes de pluie, érosion et lessivage des sols), la climatologie (Wang et al., 2009) (régime des précipitations, échanges thermiques), les flux biogéochimiques (Kremen et al., 1999; F. E. Putz et al., 2008) (stockage du carbone, augmentation des micronutriments dans les sols), la santé humaine (McMichael et al., 2006; Patz et al., 2005; Patz and Norris, 2004) (vecteurs de maladies liés aux changements d'habitat, qualité de l'air) et la biodiversité (F. E. Putz et al., 2008; Steege et al., 1996; Turner, 1996; Watson et al., 2018) (perte et fragmentation de l'habitat, altération de la chaîne de prédation, réduction des espèces végétales et animales).

Pour réduire l'exploitation non durable des bois tropicaux, des systèmes de certification du bois tels que le Forest Stewardship Council (FSC) ou le Programme de reconnaissance des certifications forestières (PEFC) ont été mis en place. Les entreprises forestières se voient délivrer des certificats garantissant l'origine légale du bois ainsi que le maintien des fonctions écologiques des forêts exploitées. En contrepartie, elle améliore l'accès au marché et donne accès à des primes (Rametsteiner and Simula, 2003), pour préserver les zones de prédation et pour conserver la faune locale (Polisar et al., 2017) ainsi que pour préserver les droits des populations indigènes. Cette certification peut être encore plus engagée. Les plantations de nouvelle génération (PNG), ont une approche plus impliquée en identifiant, gérant et restaurant des zones de conservation sensibles (Silva et al., 2019). Une analyse détaillée, prenant en compte des variables environnementales, économiques et sociales a été réalisée par (Burivalova et al., 2017), comparant les méthodes de gestion des forêts tropicales. Ils ont montré que la certification du bois est souvent moins rentable, mais que les populations locales et la biodiversité sont bien mieux loties.

La ressource en bois utilisée dans le monde provient en grande partie de la forêt tropicale (Hari Poudyal et al., 2018). Il y a une pression croissante pour que l'exploitation des forêts, et plus particulièrement des forêts tropicales, soit réalisée en utilisant des techniques d'exploitation à faible

impact environnemental (RIL ou Reduced Impact Logging). Bien que non rentable pour le premier cycle d'exploitation (Boltz et al., 2003), le RIL est considéré comme le critère minimum pour une bonne gestion forestière. Ces contraintes d'exploitation forestière semblent bénéfiques pour la réduction des perturbations et des émissions de carbone (E. A. Ellis et al., 2019; P. W. Ellis et al., 2019; Griscom et al., 2019). Cependant, la minimisation des dommages ne peut être maintenue lorsque l'intensité d'exploitation est trop élevée. (Francis E Putz et al., 2008). Les dommages résiduels de l'exploitation forestière persistent, comme la construction de routes, de sentiers, le stockage et l'empiètement des machines minières (Jackson et al., 2002; Matangaran et al., 2019; Pereira et al., 2002; Sist and Ferreira, 2007). L'impact de l'exploitation forestière sur la diversité des espèces (en termes de richesse et d'abondance) et la conservation de la biodiversité est sujet à débat. (Xu et al., 2015) ont observé un rétablissement rapide du nombre d'espèces après une coupe à blanc. Sur un autre site, (Clark and Covey, 2012) ont observé une réduction significative du nombre d'espèces d'arbres. (Carreño-Rocabado et al., 2012) ont montré que les différents traitements des perturbations n'ont pas affecté les espèces et la diversité fonctionnelle. En plus d'avoir un effet sur la richesse des espèces, certains auteurs ont constaté un effet sur la composition des espèces (de Avila et al., 2015; Hari Poudyal et al., 2018; Hu et al., 2018; Xu et al., 2015). Dans des écosystèmes très spécifiques, tels que la forêt tropicale humide de Madagascar, la perte de la richesse en espèces était irréversible après l'exploitation, malgré un temps de récupération de 150 ans (Brown and Gurevitch, 2004). Il convient donc de mieux comprendre la régénération des espèces d'arbres dans les forêts tropicales.

La population de la Guyane française n'a cessé d'augmenter, passant de 157 000 habitants en 1999 à 268 700 habitants en 2017 et devrait doubler d'ici 2030. Cette croissance démographique s'est accompagnée d'une augmentation du taux de chômage, qui semble se poursuivre (Marie and Rallu, n.d.). La forêt, qui couvre 96% du territoire, pourrait contribuer au développement économique de ce territoire : les bois durs trouvent un intérêt pour la construction, tandis que les déchets de bois et les bois plus tendres sont utilisés comme biomasse pour la production d'énergie électrique. L'augmentation de la quantité de bois extrait, dans le cadre d'une gestion durable de la ressource, et sa valorisation à travers différentes étapes seraient bénéfiques pour le territoire. Cependant, 40% de la forêt guyanaise a été attribuée au parc amazonien de Guyane en février 2007, limitant fortement l'exploitation économique de la forêt.

L'ONF a décidé de mener un projet visant à évaluer le potentiel des techniques de télédétection pour l'identification des espèces sur de grandes surfaces dans les forêts tropicales. L'objectif général de ce projet de l'ONF était d'explorer la capacité des méthodes de télédétection à cartographier les ressources en bois (en termes de volume marchand par classe de bois commercial).

A. Introduction

1. Contexte de l'étude et problématisation

Nos yeux nous permettent de percevoir les différentes couleurs d'un paysage résultant des interactions entre le rayonnement solaire et les éléments de ce paysage, de la surface terrestre à l'atmosphère. Les yeux humains sont sensibles à la lumière entre 400 nm et 650 nm, qui définit le domaine visible, grâce à un ensemble de cellules photoréceptrices répondant à la lumière de différentes longueurs d'onde. Toutefois, le soleil émet des rayonnements électromagnétiques dans un spectre beaucoup plus large. Le principe de l'imagerie hyperspectrale consiste à capturer une gamme de longueurs d'onde bien supérieure à nos capacités sensorielles (400 nm à 2500 nm). Une autre caractéristique de l'imagerie hyperspectrale est le nombre de bandes spectrales échantillonnées. En général, une image est qualifiée d'"hyperspectrale" lorsque le nombre de fréquences électromagnétiques mesurées est supérieur à cent, continue et étroite (Goetz et al., 1985).

La télédétection utilisant l'imagerie hyperspectrale continue d'évoluer, à la fois en termes de mode d'exploitation de ces informations et de technologie elle-même (Liu et al., 2017; Wang et al., 2010). De nombreux programmes spatiaux transportant des capteurs hyperspectraux voient le jour, tout comme la création de petits capteurs hyperspectraux transportables par drone (Adão et al., 2017). La technologie hyperspectrale a déjà démontré, par de multiples études, son potentiel pour caractériser les surfaces terrestres ou même martiennes (Bernard-Michel et al., 2009). La réflexion du spectre électromagnétique sur les surfaces est en effet riche. Elle permet de connaître la teneur en différents minéraux d'un sol

(Asner, 2008), la qualité de l'eau, de caractériser les types de surfaces urbanisées (Alonzo et al., 2014), de connaître les caractéristiques biophysiques des plantes (Feret et al., 2008a) et même d'identifier des espèces végétales (Clark et al., 2005). Cette technologie offre un grand potentiel pour la caractérisation des forêts tropicales.

L'identification de différentes espèces de forêts tropicales à l'aide de l'imagerie hyperspectrale a déjà été réalisée auparavant. Dans un article fondateur (Clark et al., 2005), ils ont étudié l'utilité des données hyperspectrales à haute résolution spectrale et à différentes résolutions spatiales pour l'identification de sept espèces d'arbres dans les forêts tropicales. Ils ont évalué la contribution de différentes plages spectrales pour la discrimination d'espèces spécifiques à l'aide de méthodes de classification, à l'échelle des feuilles et de la couronne des arbres. Ils ont constaté que la sélection de la bande optimale diffère selon l'échelle. À l'échelle de la couronne, ils ont montré une contribution particulièrement importante du domaine visible (437 - 700 nm) et du domaine infrarouge court (1994 - 2435 nm) pour la discrimination des espèces. En travaillant à l'échelle du pixel et de l'objet, la précision était respectivement de 100% et 92% en utilisant l'analyse discriminante linéaire (LDA) comme méthode de classification. Ils ont été suivis par différents auteurs qui ont réussi à identifier des espèces d'arbres dans les forêts tropicales en utilisant l'imagerie hyperspectrale (Baldeck and Asner, 2014; Ballanti et al., 2016; Feret and Asner, 2013; Ferreira et al., 2016a). Ces études menées sur différents sites d'étude ont confirmé les capacités de cette technologie pour l'identification spécifique des espèces. Cependant, l'imagerie spectroscopique nécessite une étape de prétraitement, comme les corrections atmosphériques. Les études citées ci-dessus utilisent des données qui ont subi un traitement lié à la composition atmosphérique ou un traitement statistique. Cependant, la contribution de chacun des prétraitements dans un processus d'identification des espèces n'a pas été clairement identifiée.

Une autre technologie de télédétection également utilisée dans le présent travail est le LiDAR (Light Detection and Ranging) à faible encombrement. Le principe du LiDAR est relativement simple. Des impulsions de lumière sont émises à très haute fréquence. Lorsqu'une impulsion de lumière rencontre un obstacle sur son chemin, le signal rebondit et retourne à la source émettrice. En connaissant les propriétés de la propagation d'une onde lumineuse dans un milieu défini, il est possible d'estimer la distance entre la source émettrice et l'obstacle. Combinée à un positionnement (Differential Global Positioning System) et une orientation (Inertial Measurement Unit) précis du capteur, cette technologie permet de créer un nuage de points, dont chacun des points est positionné dans l'espace en coordonnées cartésiennes. Le LiDAR permet de connaître la topographie de la surface étudiée et est déterminant pour la cartographie de la biomasse à grande échelle (Réjou-Méchain et al., 2015; Rocha de Souza Pereira et al., 2018), l'estimation des stocks de carbone (Bazezew et al., 2018), la hauteur des arbres (Dalponte and Coomes, 2016; St-Onge et al., 2015) et la structure des forêts (Ferraz et al., 2020, 2016a; Hamraz et al., 2017; Vincent et al., 2012; Williams et al., 2020). L'ONF mobilise cette technologie depuis des années pour mieux évaluer la ressource en bois et optimiser la gestion de la forêt dont il a la charge. Pour l'identification des espèces dans les forêts tropicales, la caractérisation de la couronne de l'arbre améliore l'identification des arbres. Au sein de l'objet caractérisé (couronne de l'arbre), un vote majoritaire de pixels est calculé pour augmenter la fiabilité des prédictions du classificateur (Clark et al., 2005). Cependant, la délimitation précise des couronnes d'arbres sur l'imagerie hyperspectrale n'est pas une tâche simple (Aubry-Kientz et al., 2019).

Les botanistes déterminent le genre et l'espèce des arbres, en se basant sur la morphologie (forme et couleur) des fleurs, des feuilles, des troncs, de l'écorce et des fruits. L'article de synthèse de (ter Steege et al., 2013) estime que la forêt amazonienne est dominée par 227 espèces d'arbres, alors que le nombre total d'espèces d'arbres serait d'environ 16 000. À titre de comparaison, les forêts tempérées européennes abritent environ 124 espèces indigènes (Slik et al., 2015). Cette hyper-diversité augmente la difficulté d'identification par télédétection. Identifier un élément parmi un nombre restreint de propositions est plus facile que d'identifier ce même élément parmi un très grand nombre. De plus, les variations hyperspectrales intra-spécifiques et intra-couronnes peuvent compliquer encore plus la discrimination des espèces (Zhang et al., 2006).

En outre, notre connaissance des écosystèmes tropicaux repose en grande partie sur des sites expérimentaux, qui représentent une petite proportion de la superficie totale de la forêt tropicale.

Depuis 1950, la forêt tropicale amazonienne a été inventoriée à l'aide de 413 parcelles réparties dans 9 pays et couvrant 404,6 hectares avec des périodes de revisite incertaines (Mitchard et al., 2014; Saatchi et al., 2015). (de Lima et al., 2015) rapportent que 0,01% des vestiges de la forêt atlantique brésilienne ont été inventoriés en 70 ans. Ces deux exemples illustrent la difficulté d'inventorier la forêt tropicale en raison de leur étendue - souvent importante -, de leur grande diversité floristique et de leur accès généralement difficile.

À cela s'ajoute un élément perturbateur dans la couronne des arbres. Les lianes, bien qu'elles soient une composante importante de la biodiversité tropicale (Schnitzer and Bongers, 2011), sont des sources de perturbation de la signature spectrale. Les lianes sont en concurrence avec leurs hôtes pour capter les ressources en eau et en lumière (Y.-J. Chen et al., 2015; Martínez-Izquierdo et al., 2016). Bien qu'une différence spectrale ait été démontrée entre un petit échantillon de lianes et d'arbres (Castro-Esau, 2004), il est d'abord nécessaire d'identifier la présence de ces derniers au sein des couronnes. Cette évaluation à très haute résolution spatiale s'avère délicate et fastidieuse (Waite et al., 2019), d'autant plus que le taux de présence de lianes dans les couronnes n'est pas directement lié à la présence de tiges de lianes au sol
(Cox et al., 2019).

Cette tâche difficile d'identification des arbres est rendue plus compliquée par la variabilité spécifique à chaque site. La distribution spatiale des arbres montre des distributions disparates entre les sites d'étude (Condit et al. 2002 ; Friis et al. 2005 ; Legendre, Borcard, et Peres-Neto 2005). La composition floristique (liste des espèces et abondance relative) est très variable entre les sites parmi les arbres émergeant dans la canopée (Lugo and Helmer, 2004). La distribution des espèces est multi factorielle. L'histoire liée à l'utilisation passée des terres, la dispersion des graines liée à l'activité animale (Trolliet et al., 2017), la géomorphologie (Koponen et al., 2004), la climatologie (Miles et al., 2004) et le comportement spécifique à l'espèce (Traissac and Pascal, 2014) sont quelques déterminants de la distribution des espèces d'arbres. Ainsi, des sources d'information auxiliaires (en dehors des données hyperspectrales et LiDAR) pourraient être utilisées pour l'identification des espèces, mais leur utilité est jusqu'à présent limitée par l'absence d'inventaires à grande échelle.

VI. General introduction (English version)

Tropical forests are home to complex ecosystems which are only partially explored. They harbor a great diversity of life forms, and contribute to human well-being at global scale through climate regulation (Lewis et al., 2015) and biogeochemical cycles (Powers and Marín-Spiotta, 2017). Therefore, their functions provide multiple ecosystem services as defined by (Alcamo, Bennett, and Millennium Ecosystem Assessment (Program) 2003). However, the extent of mature tropical forest is increasingly reduced due to human pressure (Barlow et al., 2018; Wright, 2010), regardless of the continent considered. Between 1980 and 2000, 28% of new agricultural land was conquered over mature forest (Gibbs et al., 2010). Between 2000 and 2005, about 1.4% of tropical rainforests underwent deforestation for the extraction of timber resources and / or for farming (Foley et al., 2005). However, (Song et al., 2018) showed that between 1982 and 2006, the global forest area increased by 7.1%. The loss in the tropics was largely offset by the gain in extra-tropical areas, such as mountain systems, as reported by (Piao et al., 2015). The State of the World's Forests (2020) highlights a mixed record. Increased forested areas were reported in Asia, Oceania, and Europe between 2010 and 2020. The American continent (North, Center and South) and the African continent observed a reverse trend, with a significant loss, in particular for the African continent which has been sugjected to increasing loss since 1990 and 2020. In the Guyanese rainforest, the loss of Forest area was estimated at 0.28% between 2011 and 2014 (Pickering et al., 2019). According to (FAO, 2020), on a larger timescale, between 1990 and 2020, the Guyanese rainforest experienced an area loss of about 1.51%.

On the one hand, over 60 million people are completely dependent on forests (Dieterle, 2009) and 1.2 billion people derive their livelihoods from forest products (Vantomme, 2011). In 2012, INTERPOLE estimated illegal timber trade to be worth between US$50 billion and US$150 billion annually, accounting for around 30% of the total timber trade, with increasing global demand for timber resource and insufficient capacity to control this illegal trade. On the other hand, the importance of ecosystem conservation is clearly identified. The degradation of tropical forests leads to modifications in various fields, including : hydrology (Eshleman, 2004; Giambelluca, 2002; Likens et al., 1978; Whitehead and Robinson, 1993) (modification of watercourses, runoff and natural water reservoirs, increase / decrease in rainfall episodes, soil erosion and leaching), climatology (Wang et al., 2009) (the rainfall regime, thermal exchanges), bio-geo-chemical flows (Kremen et al., 1999; F. E. Putz et al., 2008) (carbon storage, increased micronutrients in soils), human health (McMichael et al., 2006; Patz et al., 2005; Patz and Norris, 2004) (habitat change disease vectors, air quality) and biodiversity (F. E. Putz et al., 2008; Steege et al., 1996; Turner, 1996; Watson et al., 2018) (habitat loss and fragmentation, altered predation chain, reduction of plant species and animals).

To reduce unsustainable logging of tropical timber, timber certifications schemes such as Forest Stewardship Council (FSC) or the Program for the Endorsement of Forest Certification (PEFC) have been put in place. Forestry companies are issued certificates guaranteeing the legal origin of the wood as well as the maintenance of the ecological functions of the exploited forests. In return, it improves access to the market and gives access to bonuses (Rametsteiner and Simula, 2003), to preserve areas of predation and to conserve local fauna (Polisar et al., 2017) as well as to preserve the rights of indigenous populations. This certification can be even more committed. New Generation Plantations (NGP), have a more involved approach by identifying, managing and restoring sensitive conservation areas (Silva et al., 2019). A detailed analysis, taking into account environmental, economic and social variables was carried out by (Burivalova et al., 2017), comparing the management methods of tropical forests. They showed that wood certification is often less profitable, but that local populations and biodiversity are much better off.

The wood resource used in the world comes largely from tropical forest (Hari Poudyal et al., 2018). There is a growing pressure for the exploitation of forests, and more particularly tropical forests, to be carried out using low environmental impact exploitation techniques (RIL or Reduced-Impact Logging). Although not profitable for the first cycle of logging (Boltz et al., 2003), RIL is considered the minimum criterion for good forest management. These logging constraints seem beneficial for the reduction of disturbance and carbon emissions (E. A. Ellis et al., 2019; P. W. Ellis et al., 2019; Griscom et al., 2019).

However, damage minimization cannot be maintained when the operating intensity is too high. (Francis E Putz et al., 2008). Residual damages from logging persist, such as road construction, pathways, storage and encroachment of mining machinery (Jackson et al., 2002; Matangaran et al., 2019; Pereira et al., 2002; Sist and Ferreira, 2007). Impact of forest exploitation on species diversity (in terms of richness and abundance) and conservation of biodiversity is open to debate. (Xu et al., 2015) observed a rapid recovery in the number of species after a clear cut. On another site, (Clark and Covey, 2012) observed a significant reduction in the number of tree species. (Carreño-Rocabado et al., 2012) showed that the different disturbance treatments did not affect species and functional diversity. In addition to having an effect on species richness, some authors found an effect on species composition (de Avila et al., 2015; Hari Poudyal et al., 2018; Hu et al., 2018; Xu et al., 2015). In very specific ecosystems, such as the tropical rainforest of Madagascar, the loss of species richness was irreversible after exploitation, despite 150 years recovery time (Brown and Gurevitch, 2004). Understanding the regeneration of tree species in tropical forests is therefore to be improved.

The population of French Guiana has been steadily increasing, from 157,000 inhabitants in 1999, to 268,700 inhabitants in 2017 and is projected to double by 2030. This population growth was accompanied by an increase in unemployment rate, which seems to continue (Marie and Rallu, n.d.). The forest, which covers 96% of the territory, could contribute to the economic development of this territory: hardwoods find an outlet for construction, while wood waste and softer woods are used as biomass for the production of electrical energy. The increase in the quantity of wood extracted, as part of a sustainable management of the resource, and its valuation through different stages would be beneficial to the territory. However, 40% of the Guyanese forest was assigned to the Guyana Amazonian Park in February 2007, strictly limiting economic exploitation of the forest.

ONF decided to carry out a project to assess the potential of remote sensing techniques for species identification over large areas in tropical forests. The wider objective of this ONF project was to explore capability of remote sensing methods to map wood resource (in terms of merchantable volume per commercial timber class).

A. Introduction

1. Context of the study and problematization

Our eyes allow us to perceive the different colors of a landscape resulting from the interactions between solar radiation and the elements of this landscape, from Earth surface to atmosphere. Human eyes are sensitive to light between 400 nm and 650 nm, which defines the visible domain, thanks to a set of photoreceptor cells responding to light of different wavelengths. However, the sun emits electromagnetic radiation in a much wider spectrum. The principle of hyperspectral imaging is to capture a range of wavelengths much greater than our sensory capacities (400 nm to 2500 nm). Another characteristic of hyperspectral imagery is the number of spectral bands sampled. Generally, an image is characterized as "hyperspectral" when the number of electromagnetic frequencies measured is greater than a hundred, continuous and narrow (Goetz et al., 1985).

Remote sensing using hyperspectral imagery continues to evolve, both in terms of how this information is exploited and in the technology itself (Liu et al., 2017; Wang et al., 2010). Numerous space programs carrying hyperspectral sensors are emerging, as is the creation of small hyperspectral sensor transportable by drone (Adão et al., 2017). Hyperspectral technology has already demonstrated through multiple studies its potential to characterize terrestrial or even Martian surfaces (Bernard-Michel et al., 2009). The reflection of the electromagnetic spectrum on surfaces is indeed rich. It allows to know the content of different minerals in a soil (Asner, 2008), the quality of the water, to characterize the types of urbanized surfaces (Alonzo et al., 2014), to know the biophysical characteristics of plants (Feret et al., 2008a) and even identify plant species (Clark et al., 2005). This technology offers great potential for the characterization of tropical forests.

The identification of different tropical forest species using hyperspectral imagery has been carried out before. In a seminal paper (Clark et al., 2005) investigated the utility of hyperspectral data with high spectral resolution and different spatial resolutions for the identification of seven tree species in tropical forests. They assessed the contribution of different spectral ranges for the discrimination of specific

species using classification methods, at both leaf and tree crown scale. They found that the optimal band selection differs depending on the scale. At crown scale, they showed particularly important contribution from the visible domain (437 - 700 nm) and the short infrared domain (1994 - 2435 nm) for species discrimination. Working at pixel scale and object scale the accuracy was respectively 100% and 92% using Linear Discriminant analysis (LDA) as classification method. They were followed by various authors who succeeded in identifying tree species in tropical forests using hyperspectral imagery (Baldeck and Asner, 2014; Ballanti et al., 2016; Feret and Asner, 2013; Ferreira et al., 2016a). These studies carried out at different study sites confirmed the capabilities of this technology for species-specific identification. However, imaging spectroscopy requires preprocessing step, such as atmospheric corrections. The studies cited above use data that have undergone processing related to atmospheric composition or statistical processing. However, the contribution of each of the pretreatments in a species identification process was not clearly identified.

Another remote sensing technology also used in the present work is small footprint LiDAR (Light Detection and Ranging). The principle of LiDAR is relatively simple. Pulses of collimated light are emitted at very high frequency. When a pulse of light encounters an obstacle on its path, the signal bounces back and goes back to the emitting source. Knowing the properties of the propagation of a light wave in a defined medium, it is possible to estimate the distance between the emitting source and the obstacle. Combined with a precise positioning (Differential Global Positioning System) and orientation (Inertial Measurement Unit) of the sensor, this technology makes it possible to create a point cloud, each of the points of which are spatially positioned in Cartesian coordinates. LiDAR makes it possible to know the topography of the surface studied and is decisive for biomass mapping at large scales (Réjou-Méchain et al., 2015; Rocha de Souza Pereira et al., 2018), carbon stock estimation (Bazezew et al., 2018), tree height (Dalponte and Coomes, 2016; St-Onge et al., 2015) and forest structure (Ferraz et al., 2020, 2016a; Hamraz et al., 2017; Vincent et al., 2012; Williams et al., 2020). ONF has been mobilizing this technology for years to better assess the wood resource and optimize the management of the forest for which it is responsible. For the identification of species in tropical forests, the characterization of the crown of the tree improves the identification of trees. Within the characterized object (crown of the tree), a predicted majority vote of pixels is calculated to increase the reliability of classifier predictions (Clark et al., 2005). However, accurately delineating tree crowns on hyperspectral imagery is not a simple task (Aubry-Kientz et al., 2019).

Botanists determine trees Genus and Species, based on the morphology (shape and color) of flowers, leaves, trunks, bark and fruits. The review article by (ter Steege et al., 2013) estimates that the Amazon rainforest is dominated by 227 tree species, while the total number of tree species would be about 16,000. By way of comparison, European temperate forests host around 124 native species (Slik et al., 2015). This hyper-diversity increases the difficulty of identification by remote sensing. Identifying an element among a restricted number of propositions is an easier task compared to identifying this same element among a very large number. In addition, intra-specific and intra crown hyperspectral variations may further complicate species discrimination (Zhang et al., 2006).

In addition, our knowledge of tropical ecosystems relies largely on experimental sites, which represent a small proportion of the total area of tropical forest. Since 1950, the Amazon rainforest has been inventoried using 413 plots spread over 9 countries covering 404.6 hectares with uncertain revisit times (Mitchard et al., 2014; Saatchi et al., 2015). (de Lima et al., 2015) report that 0.01% of the Brazilian Atlantic forest relicts have been inventoried in 70 years. These two examples illustrate the difficulty of inventorying the tropical forest due to their – often - large extent, their high floristic diversity and their usually difficult access.

Added to this is a disruptive element within the crowns of trees. Lianas, although they are an important component in tropical biodiversity (Schnitzer and Bongers, 2011), are sources of disturbance of the spectral signature. Lianas compete with their hosts in order to capture water and light resources (Y.-J. Chen et al., 2015; Martínez-Izquierdo et al., 2016). Although a spectral difference has been shown between a small sample of lianas and trees(Castro-Esau, 2004), it is first necessary to identify the presence of the latter within the crowns. This very high spatial resolution assessment turns out to be

delicate and tedious (Waite et al., 2019), especially since the rate of presence of lianas in the crowns is not directly related to the presence of lianas stems (Cox et al., 2019)

This difficult task of identifying trees is made more complicated by site-specific variability. The spatial distribution of trees shows disparate distributions between the studies sites (Condit et al. 2002; Friis et al. 2005; Legendre, Borcard, and Peres-Neto 2005). The floristic composition (list of species and relative abundance) is very variable between the sites among trees emerging in canopy (Lugo and Helmer, 2004). The distribution of species is multi factorial. History linked to past land use, seed dispersal linked to animal activity (Trolliet et al., 2017), geomorphology (Koponen et al., 2004), climatology (Miles et al., 2004) and the behavior specific to the species (Traissac and Pascal, 2014) are some determinants of tree species distribution. So ancillary sources of information (apart from hyperspectral and LiDAR data) could be used for species identification but there usefulness is so far limited by the lack of large scale inventories.

2. Objective and hypotheses

The main objective of this thesis is to propose operational methods for the fine mapping of the wood resource by fusion of LiDAR and hyperspectral data. The specific hypotheses we shall be testing along the way are the following 1) the electromagnetic spectrum reflected by the canopy is sufficiently rich to discriminate between species in tropical rainforest 2) The fusion of hyperspectral imagery with LiDAR derived structural information can be achieved accurately and consistently 3) The surface reflectance data is robust to the acquisition conditions.

The specific objectives of this thesis are as follows:

1. Evaluate the capacity of hyperspectral measurements for the identification of trees in Guyanese tropical forests,
2. Propose an operational method for the identification of the wood resource from the fusion of hyperspectral and LiDAR data,
3. Evaluate the conditions of transferability of hyperspectral data for large-scale applications.

3. Manuscript organization

The first chapter explores which steps in the data processing workflow are critical for discriminating the 20 most abundant species in the database which was set-up in view of this project (Baltzer, 2015; Dutrieux, 2018). We focus on the Paracou experimental site and consider only the VNIR spectral range in this study. We evaluate the contribution of various atmospheric corrections and various (spectral and spatial) filtering procedures to the improvement of species classification accuracy.

The second chapter evaluates the additional information brought by the inclusion of the SWIR region. The task considered here differs from the previsous chapter as we move from a task involving discriminating a few species against one another to a task involving the retrieval of those species from within a hyperdiverse canopy. We illustrate the results of this section by applying the binary classifiers developed to produce a map of timber resource of the Paracou experimental site (115 ha).

The third chapter is a first attempt to understand the causes of spectral divergence observed in repeat overflights. Most of the analysis builds on the Nouragues site that was imaged twice on successive afternoons with a different flight plan. This created large variations in illumination conditions which have impacted spectral consistency. We evaluate how much of the spectral divergence can be attributed to different factors (atmosphere disturbance, solar position, solar-sensor geometry) in interaction with topography [in progress].

B. National Forestry Office

The ONF (Office National des Forêts) is a public industrial and commercial establishment in charge of the management of public forests, under the supervision of the Ministry of Agriculture, Agri-Food and Forestry and the Ministry of 'Ecology, Sustainable Development and Energy. Its missions are numerous: to manage, protect and develop French forests. The ONF aims to be a major player in ecological transition and sustainable development with the ambition of "making forests and natural

spaces an essential lever for regional development and contributing to the local economic growth". This establishment is in charge of more than 11 million hectares of forests. 4.6 million hectares concern mainland France and comprise 1.7 million hectares of state forests and 2.9 million hectares of communal forests. The overseas territories represent 6.4 million hectares, of which 85% are located in French Guiana.

C. Tropical forest

1. Study sites description

French Guiana is an overseas department and region of France. It is located on the northern Atlantic coast of South America in the Guiana shield. It borders Brazil and Suriname. The climate is equatorial with a seasonal regime under the influence of intertropical convergence zones which oscillates along the north-south. Four periods can be distinguished: the period of light rains (December to January), the short dry season or summer from February to March, the period of long rains from April to June and the dry season from July to November. Precipitations can reach 400 mm/month during the rainy season and 100 mm/month during dry seasons.

The maximum annual precipitation is recorded near the Atlantic coast and can reach 4000 mm. The precipitation gradient is strong, and varies from 2000 mm / year in the southwest to more than 3500 mm / year in the northeast. The average temperature is 26 ° C with very little variations during the year (temperature amplitude ~ 1° C).

French Guyana belongs to the Guyana Shield, which is a vast geological complex limited to the North by the Atlantic Ocean and to the South by the Amazon Basin. It stretches 900 km wide from north to south and 1800 km from east to west. The oldest rocks were formed 2.2 billion years ago, linked to the opening of an ocean, separating the Archean shields of the Amazon and West Africa. Supergene weathering phenomena (very specific to tropical climates) shape the Guyanese landscapes mainly in the peneplain.

The sites studied in this thesis are Paracou, Montagne-Tortue and Nouragues. On these three sites forest permanent plots are managed by public institutes in order to study ecological processes occurring in tropical forests.

Paracou site:

The Paracou site is located in the north of French Guiana, south of the town of Sinnamary and 15 km from the nearest coastline (5 ° 15'N, 52 ° 55'W). This site is easily accessible via the national 1. The average temperature is 26 ° C with a variation of + -1 ° C. Rainfall is highest in May and lowest in September, typical of a tropical climate. The rainfall is around 3000 + - 700 mm per year. The Paracou site is located in the northernmost part of the hilly area, on a formation called the "Armina series", characterized by shales and sandstones and locally crossed by veins of pegmatite, aplite and quartz. The hydrographic system is generally oriented SW - NE. The relief of the site consists of small elliptical hills separated by narrow sandy water beds (<5 m wide). The elevation varies from 5 m to about 45 m above sea level.

The Paracou plot network covering 118.75 ha is the largest in French Guiana. It was setp in 1982 to acquire scientific data for forest management in order to find the best compromise between forest productivity and the maintenance of ecosystem services in tropical forests. CIRAD is a research centre in charge of this experimental setup.

Montagne-Tortue site :

The Montagne-Tortue site is located a few kilometers west of Regina, on the edge of the Bélizon track (4 ° 19'N, 52 ° 14'W). The average temperature is similar at the Paracou site, averaging 26 ° C. Grande Montagne Tortue covers an area of 4373 ha with an altitude varying between 114 m and 438 m.

The Montagne Tortue site is located on a set of tabular massifs, culminating at 483 m, called Grande Montagne Tortue. Its peculiarities make the Grande Montagne Tortue a geomorphological and ecological unit attached to a group consisting mainly of the Cocoa Mountain, Kaw Mountain, Maripa Mountain and Observatory Mountains. This is the Septentrional Range, a vast set of volcano-

sedimentary rocks, one of the three major geomorphological regions of the highlands of Guyana. The tabular form of the site testifies to the existence of a lateritic cuirass covering its summit. The topography and geology have made this place a refuge area for fauna and flora, recognized as ancient centers of speciation, having developed a remarkable rate of endemism and biodiversity. For example, *Guatteria elegans* (*Annonaceae*), *Licaria rufotomentosa (Lauraceae), Coussarea granvillei (Rubiaceae*) are rare trees endemic to this site. *Passiflora trialata (Passifloraceae)* is a liana only known worldwide from this locality.

Permanent plots (including logged over plots) cover 16 ha and are managed by ONF.

Nouragues site:

The Nouragues site is located in the Center-East of French Guiana, in the south-west and 50 km from Regina, the nearest town and 100 km from the coast (4 ° 05'N, 52 ° 40'W). The site is only accessible by helicopter or canoe. Its remoteness results in minimum anthropogenic disturbance. The average temperature is similar to the Paracou site with a high average relative humidity ranging between 80 and 90% all year long. The average rainfall is 2990 mm / year. The relief is made up of hills and plateaus. The elevation of the site varies from 60 m to 420 m at the top of the inselberg. The Nouragues site is located on a geological substrate belonging to the "Lower Paramaca" formation, comprising volcanic and sedimentary rocks, crossed by "Guyanese" granites.

The Nouragues experimental site was set up in 1986 with the aim to study the functioning of tropical forest and its biodiversity. This research station is managed by the CNRS. It has three large permanent plots. The two main plots to the north of the site (Camp Inselberg) are made up to the west of 12 1ha-plots (Petit plateau) and to the east (Grand plateau) of 10 1ha plots. The third area to the south of the site (Saut Pararé) includes six 1ha plots.

2. Floristic composition

Tropical forests covered about 17% of the earth's surface in 1983 (Olson et al., 1983) and represent a large reservoir of carbon stock. In 1993, (Solomon et al., 1993) estimated that 28% of terrestrial carbon was conserved in tropical forests and woody areas. The Guyanese forests are mainly composed of a dense primary forest and humid tropical plains (Sabatier and PRÉVOST, 1990).

The census of species in French Guiana began in 1775 by Jean Baptiste Christian Fusée-Aublet. Since then, this enumeration has been constantly increasing. On three plots of 1 hectare each, inventorying trees with a dbh greater than 20 cm (Black and Pavan 1950) identified 79 species for 230 individuals (terra firma near Téfé – Brazil), 41 species for 134 individuals (flooded forest, near Guama River – Brazil) and 62 species for 195 individuals (terra firma, near Belém – Brazil). At the tree study sites, located in Brazil, some species were found in one, two or three sites. (Sabatier and PRÉVOST, 1990) counted 68 families and more than 1050 species with a DBH greater than 30 cm and nearly 350 species with a DBH greater than 60 cm, identified into French Guyana. Currently, thanks to the establishment of several experimental site, it is possible to identify and better understand the new species. Recently, (ter Steege et al., 2013) recorded over half a million trees across 1170 plots through international inventory efforts and information sharing. They estimated about 16,000 tree species for lowlands of the Amazon, 1.4% (227) of which being considered as "hyperdominant". To date, 1769 tree species have been identified and identified across French Guiana (Molino et al., In prep.) and about 230 tree species have not been assigned to any taxon.

From the censuses conducted in 2016 of tree with a diameter at breast height (DBH) greater than 10 cm the following table was produced:

Table 1 : Quantification of the species present at each of the study sites. The average DBH and the associated standard deviation are given.

Site	stem number	Genus number	Species number	stem not determined to species	DBH (cm)
Paracou	75251	253	743	5934	67.5 (38.8)

Montagne Tortue	7600	183	273	442	69.0 (46.5)
Nouragues	14680	283	510	4604	72.0 (50.0)

3. Forest management

The tropical forest of French Guiana is an extremely rich ecosystem in terms of fauna and flora. This forest has great potential for both French Guiana and France, but also for the global scientific community. Through the various experimental study sites, the tropical forest is a fertile ground for the exploration of forest management methods. Currently, the following 4 points can define the management carried out by the ONF:

- To economically and sustainably develop forests, while preserving its natural functioning to meet the folloing objectives : supply of the wood sector, ecotourism development, use of forest products by the populations
- Preserving ecological roles: with a strict conservation network sampling the different ecosystems and natural habitats over large areas,
- Unite Guyanese society around the forest: involve the population and elected officials for a better guarantee of sustainable management,
- Continue the effort to acquire basic knowledge about this biome that is still so poorly known.

VII. Chapitre 1 : Capacity of hyperspectral data for species characterization

A. Introduction

This first chapter aims to establish the feasability of discrimination of common tree species using hyperspectral data. The assessment of hyperspectral abilities for tree species discrimination is a topic of interest for scientific community, due to the evolution and comparison of classification algorithms as well as the evolution of hyperspectral technology. So far, few studies had compared performances for species discrimination obtained with sensor radiance with those obtained when using bottom of atmosphere reflectance computed after application of atmospheric corrections on radiance data. Radiance is influenced by multiple factors, including solar illumination, atmospheric properties and ground properties. The objective of the correction of atmospheric disturbances is to remove distorsions due to the atmospheric composition in order to convert the radiance measurements into reflectance values. Although the studies for the identification of trees in tropical forests generally used data corrected for atmospheric effects, we wanted to estimate the contribution of atmospheric corrections using these same data without corrections. We were particularly interested in assessing the gain in robustness atmospheric corrections might bring whenapplying classifiers across dates.

We also wanted to evaluate the effect of various statistical treatments applied before or after atmospheric corrections on species classification accuracy. Indeed, before developing classifiers, it seemed preferable, to enhance the spectral information available in the data. The role of illumination quality (intensity and homogeneity characteristics) was examined. Indeed the spectral signature of an object masked by a shadow (due to solar geometry) will be very different from an unshaded one. In addition, the normalization of the spectral signature by the mean spectral value (radiance or reflectance) suppresses variations in spectra due only to intensity.

B. Article: Across Date Species Detection Using Airborne Imaging Spectroscopy

remote sensing

MDPI

Across Date Species Detection Using Airborne Imaging Spectroscopy

Anthony Laybros [1,*] , Daniel Schläpfer [2], Jean-Baptiste Féret [3] , Laurent Descroix [4] , Caroline Bedeau [4], Marie-Jose Lefevre [5] and Grégoire Vincent [1]

1 AMAP, IRD, CNRS, INRA, Univ Montpellier, CIRAD, 34 000 Montpellier, France; gregoire.vincent@ird.fr
2 ReSe Applications LLC, CH-9500 Wil, Switzerland; daniel@rese.ch
3 TETIS, Irstea, AgroParisTech, CIRAD, CNRS, Université Montpellier, 500 Rue Jean-François Breton, 34000 Montpellier, France; jean-baptiste.feret@teledetection.fr
4 ONF-Guyane, 97300 Cayenne, France; laurent.descroix@onf.fr (L.D.); caroline.bedeau@onf.fr (C.B.)
5 CNES, DCT/SI/AP, 18 Avenue Edouard Belin, BPI 1219, 31401 Toulouse CEDEX 9, France; marie-jose.lefevre@cnes.fr
* Correspondence: anthony.laybros@etu.umontpellier.fr; Tel.: +33-467-614-40

Received: 22 February 2019; Accepted: 29 March 2019; Published: 2 April 2019

Abstract: Imaging spectroscopy is a promising tool for airborne tree species recognition in hyper-diverse tropical canopies. However, its widespread application is limited by the signal sensitivity to acquisition parameters, which may require new training data in every new area of application. This study explores how various pre-processing steps may improve species discrimination and species recognition under different operational settings. In the first experiment, a classifier was trained and applied on imaging spectroscopy data acquired on a single date, while in a second experiment, the classifier was trained on data from one date and applied to species identification on data from a different date. A radiative transfer model based on atmospheric compensation was applied with special focus on the automatic retrieval of aerosol amounts. The impact of spatial or spectral filtering and normalisation was explored as an alternative to atmospheric correction. A pixel-wise classification was performed with a linear discriminant analysis trained on individual tree crowns identified at the species level. Tree species were then identified at the crown scale based on a majority vote rule. Atmospheric corrections did not outperform simple statistical processing (i.e., filtering and normalisation) when training and testing sets were taken from the same flight date. However, atmospheric corrections became necessary for reliable species recognition when different dates were considered. Shadow masking improved species classification results in all cases. Single date classification rate was 83.9% for 1297 crowns of 20 tropical species. The loss of mean accuracy observed when using training data from one date to identify species at another date in the same area was limited to 10% when atmospheric correction was applied.

Keywords: tropical forest; atmospheric correction; hyperspectral; linear discriminant analysis

1. Introduction

The Amazon forest, the largest tropical forest basin on earth, covers an area of 5.5 million km^2 and harbours an estimated 16,000 tree species [1]. It plays a major role in global climate regulation, particularly through the cycling and storage of carbon [2] and it constitutes an extraordinary terrestrial reservoir of biodiversity [1]. However, the Amazon faces degradation threats [3] from unsustainable logging [4], climate change [5], land use change [6], agricultural [7] and other human activities [8]. While the threats are increasing, our knowledge about tropical forest composition and its ecological functioning progresses at a slow pace. The total above ground carbon in the Amazon may still be poorly estimated [9]. Similarly, the mere number of tree species in the Amazon is a matter of debate [10].

Large-scale tree species inventory capability would be instrumental in advancing fundamental knowledge on this complex and threatened biome and in helping with its sustainable management. Mapping species distribution at large scale in relation to environmental constraints is likely to provide novel insights to the long-standing question of why tropical forests are so species rich. From a management point of view, high-throughput airborne species mapping would help rationalizing logging. In particular, being able to map targeted commercial species could foster environmentally benign forestry practices. It would help prioritize areas that are rich in economically valuable species, while avoiding opening logging tracks in areas with low resources. It also has potential for monitoring invasive species [11].

Imaging spectroscopy holds great promise to map canopy species by air, as has already been proven by a series of seminal studies [12–15]. However, to date only few case studies have been published, and to our knowledge there is no clear evidence that airborne imaging spectroscopy is sufficiently reliable and transferable to achieve operational tree species mapping in the absence of a large ground sampling effort to build a training data set for every new site to be mapped. Practical solutions for enhanced reliability and operationality of airborne tree species mapping include the collection of spectral libraries focusing on hyperspectral signatures from different tree species, which could be stored and organized in shared spectral databases like the Specchio database [16,17]. A prerequisite would then be to capture the representative spectral features independently of the particular acquisition settings (i.e., time and date, atmospheric composition, solar irradiance angles, topographic conditions, observation angle, and sensor characteristics).

A large number of pre-processing options have been identified through state-of-the-art techniques, aiming at improving the capabilities of hyperspectral data for tree species identification. The main goal is to reduce possible sourcesof signal variation that are extrinsic to vegetation, including sensor noise, illumination conditions and atmospheric composition. The first pre-processing step commonly used in order to reduce the noise is spatial filtering [18,19]. It consists in an homogenization of spectral information of each pixel based on the averaging of its reflectance and its neighbor's reflectance. The second pre-processing step deals with the illumination variation. Clarck and Roberts [20] suggested that illumination variation is not a relevant factor for species discrimination. They showed that excluding shadowed pixels leads to improved classification accuracy. The masking of shadowed pixel [14,15] or their correction by different methods [21] should be considered, especially at very high spatial resolution [22]. Finally, atmospheric composition affects scattering and absorption of sunlight and thereby affects the reflected signal recorded by the sensor. Atmospheric correction then appears as an important step towards standardisation of spectral signature and applicability of classifiers from one image to another acquired at a different time. This atmospheric correction aims at converting at-sensor radiance into surface reflectance. The high spectral resolution offered by imaging spectroscopy is a particular advantage for achieving physically-based fine atmospheric corrections as the signal measured at particular wavelengths or combination of wavelengths may inform about atmospheric characteristics [23,24].

In this study, we evaluate the impact of various levels of image pre-processing steps on the classification accuracy of 20 tree species of tropical forest using Linear Discriminant analysis (LDA). We evaluate two different settings: the first setting focuses on single date imagery while the second setting uses training data from one date and testing data from a second flight date to appraise transferability of the classifier across dates.

2. Materials and Methods

1.1. Study Site

The study was conducted at Paracou (51°8′N, 52°53′W), an experimental site in the North of French Guiana set-up by CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement) in 1982. The local topography is gently rolling. The site is covered with

lowland terra firme rain forests. Mean annual rainfall was 2875 ± 510 mm over the 1986–2005 period with a 3-month dry season from mid-August to mid-November [25]. All stems above 10 cm diameter at breast height have been censured every 1–2 years for more than 35 years. Nine of the plots were selectively logged in 1986. More than 750 tree species have been inventoried on the site, dominated by *Leguminosae-Caesalpinioidae*, *Lecythidaceae*, *Chrysobalanaceae*, and *Sapotaceae*. A detailed description of the site and experimental design can be found in [26].

For the purpose of this study, a field survey was conducted to build a large ground truth dataset. Easily discernible crowns were first delineated manually on the Canopy Height Model derived from LiDAR (Light Detection And Ranging) data with the help of a high resolution (10 cm) RGB (Red, Green, Blue) mosaic. The correct delineation of these Individual Tree Crowns (ITC) was then validated in the field and the corresponding species ascertained. The 20 most represented species i.e., those with at least 24 individuals tagged, were selected for the purpose of the present study, totalling 1297 trees (Table 1, Figure 1). On the first flight date, all crowns were imaged twice (once each of two neighbouring flight lines) and the cumulated crown area considered for the single date analysis was 112,313 m^2. On the second flight date, only a subset of the crowns was imaged (Table 1).

Table 1. List of species used and their associated crown and pixel numbers. The entire site was covered on date 1 and a subset was imaged again on date 2.

Species	Date 1			Date 2			Proportion of Area Covered on Date 2 Set (%)
	Crown Image Segments	Area Covered (m^2)	Mean Crown Area (m^2) (SD)	Crown Image Segments	Area Covered (m^2)	Mean Crown Area (m^2) (SD)	
Bocoa prouacensis	24	1319	54.9 (35.8)	8	448	66.9 (40.2)	34.0
Couratari multiflora	49	2701	55.1 (33.8)	11	386	29.7 (14.7)	14.3
Dicorynia guianensis	108	11090	102.7 (66.8)	36	3746	109.7 (68.2)	33.8
Eperua falcata	106	7599	71.7 (41.3)	48	3193	70.4 (38.0)	42.0
Eperua grandiflora	74	6457	87.3 (46.2)	13	958	88.2 (45.4)	14.8
Eschweilera sagotiana	139	6824	49.1 (29.0)	65	2818	46.6 (25.9)	41.3
Goupia glabra	25	3343	133.7 (77.3)	3	214	117.5 (72.8)	6.4
Inga alba	26	2113	81.3 (58.7)	0	0	-	-
Jacaranda copaia	24	970	40.4 (22.7)	8	292	33.0 (13.1)	30.1
Licania alba	46	2161	47.0 (18.4)	10	443	49.5 (27.2)	20.5
Licania heteromorpha	27	1087	40.3 (21.7)	9	296	34.5 (18.5)	27.2
Moronobea coccinea	27	1858	68.8 (36.7)	19	1067	60.0 (29.6)	57.4
Pradosia cochlearia	164	23330	142.3 (122.5)	40	4640	128.8 (101.5)	19.9
Qualea rosea	206	22548	109.5 (59.4)	10	821	95.0 (34.6)	3.6
Recordoxylon speciosum	69	4802	69.6 (26.2)	28	1947	71.8 (25.9)	40.5
Sextonia rubra	32	3791	118.5 (99.3)	10	682	75.7 (38.2)	18.2
Symphonia sp1	34	1708	50.2 (20.1)	16	735	46.8 (21.1)	43.0
Tachigali melinonii	51	5415	106.2 (67.1)	23	985	86.6 (27.7)	18.2
Tapura capitulifera	32	975	30.5 (12.2)	19	668	36.0 (27.7)	68.5
Vouacapoua americana	34	2222	65.4 (34.0)	8	400	43.03 (22.8)	18.0

1.2. Hyperspectral Data

Imaging spectroscopy was acquired with a Hyspex VNIR-1600 (Hyspex NEO, Skedsmokorset, Norway) sensor mounted alongside a Riegl LMSQ780 laser scanner (See Appendix D). The 160 spectral bands cover the range from 414 nm to 994 nm (i.e., visible to near infrared) with a spectral sampling distance of 3.64 nm. The entire spectral range was used in this study (no spectral masking). The King Air B200 airplane flew at an average altitude of 920 m. The two flights took place on cloudless days on the 19^{th} (from 15:00 to 17:00, solar time) and the 20^{th} (from 16:00 to 17:00, solar time) September 2016. The second flight covered approximately one third of the area imaged on the first date. Images were orthorectified and georeferenced (level L1b see below) at 1 m spatial resolution with the PARGE software [27] using the canopy DSM (Digital Surface Model) produced from the LiDAR point cloud. The DSM was created from the point cloud by selecting point of maximum height on a 1-m resolution grid.

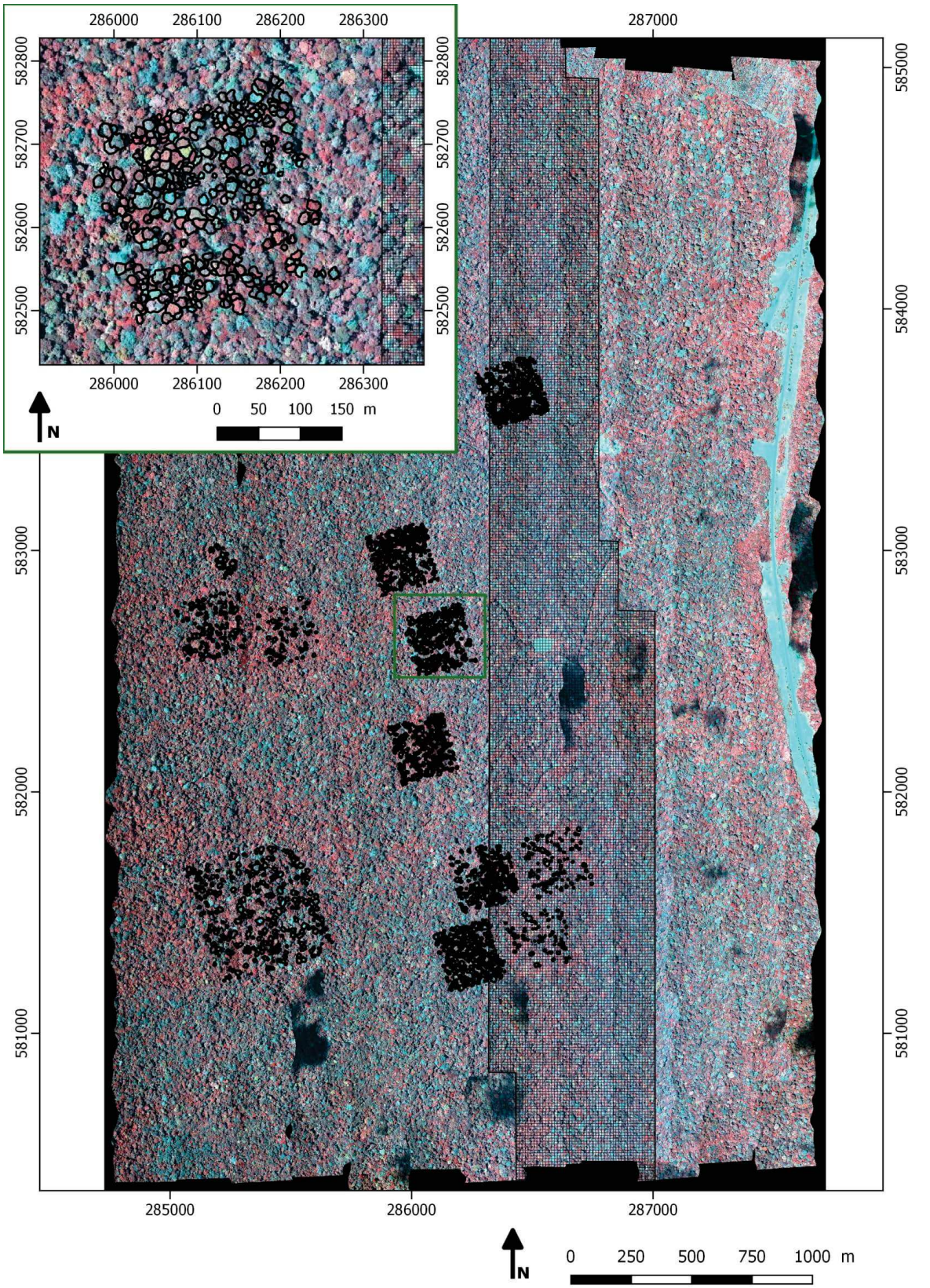

Figure 1. Hyperspectral image (false colour, R: 775 nm, G: 637 nm, B: 426 nm) at sensor radiance, captured on 19 September. Black segments feature the tree crowns used in this study. The grid pattern in black represents the area imaged on 20 September. In the zoomed map (green box), crown segments are delineated in black too. Projection used is WGS 84 UTM 22N (EPSG: 32622).

1.3. Statistically Based Spectral Data Pre-Processing

In this section, we present the various statistical data pre-processing steps that were tested as part of the different processing pathways evaluated (Figure 2, Table 2).

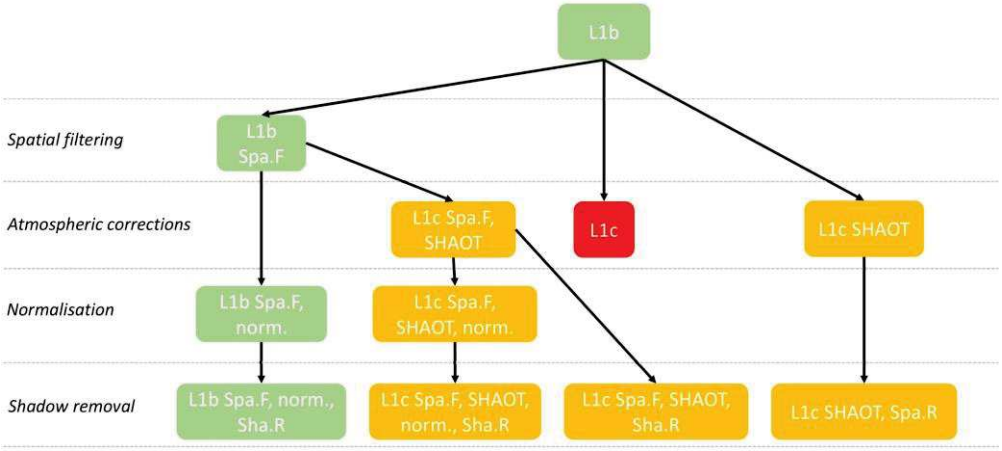

Figure 2. Pathways of the processing workflows presented in this paper. Green, red and orange correspond to georeferenced (at sensor radiance), basic atmospheric correction and enhanced atmospheric correction using SHAOT data, respectively.

Table 2. Nomenclatures for each processing.

Nomenclature	Processing
L1b	At sensor radiancegeo-referenced L1cAtmospheric correction
Spa.F	A spatial mean filter is applied
SHAOT	Variable AOT is considered for atmospheric correction and aerosols are not considered as constant.
Sha.R	Shadow pixels are removed
Norm.	Division by spectrum mean

Python programming language was used to compute mean filtering and spectrum normalisation using "Numpy" package.

1.3.1. Mean Filtering

In addition to atmosphere composition and illumination conditions, hyperspectral images may be affected by instrumental noise. Spatial filtering decreases the local noise on each pixel and may improve the separability of objects in hyperspectral data [28]. In several of the workflows explored in this study, a spatial filter (mean of a 3×3 moving window) was applied (Figure 2).

1.3.2. Spectrum Normalisation

A radiance/reflectance spectrum normalization was applied aside from atmospheric corrections. It was previously noted [29] that a simple normalization could significantly improve segmentation from imaging spectroscopy, even more so than complex atmospheric corrections. The normalization consisted in dividing the spectrum of a pixel by its mean (Figure 2).

$$S_{\lambda_i norm} = \frac{S_{\lambda_i}}{\sum_{\lambda i, i=1}^{n=160} S_i} \quad (1)$$

S_{λ_i} stands for the spectral signal value (radiance or reflectance) at wavelength λ_i. n = 160, is the number of spectral bands.

1.4. Physically Based Spectral Data Pre-Processing

In this section, we present the physically based pre-processing steps which were combined to statistical pre-processing into different pathways (Figure 2).

1.4.1. Atmospheric Corrections

Flight lines were clipped to 1 km long strips. For each strip, the mean sensor altitude was retrieved (ranging between approximatively 840 m and 940 m). Clouds contribution was negligible over the area of interest (Figure 1) and no cloud shadow correction or cloud masking nor haze removal were applied.

Contribution of oxygen and nitrogen, the major contributors to molecular scattering, to atmospheric optical thickness were inferred from atmospheric pressure. Water vapour was retrieved using the APDA (Atmospheric Precorrected Differential Absorption) algorithm implemented in ATCOR-4 [30]. The APDA algorithm is based on the depth of the absorption features derived from three specific spectral bands: one spectral band is centred at 820 nm and the other reference spectral bands correspond to neighboring spectral bands (just before and just after 820 nm), which are less affected by water vapour. The depth of the absorption feature at 820 nm is then directly related to the water vapour column under consideration of the aerosol path scattering effect. The AOT (Aerosol Optical Thickness) was either considered constant or allowed to vary spatially. AOT represents the amount of aerosols in the entire column of the atmosphere. We used the shadow-based AOT retrieval method (SHAOT) proposed in [31] to map the spatial variability of this parameter. This method builds on an idea proposed in [32] to use shadows in urban areas. It was optimized for all kinds of shadows and refined for hyperspectral images with a spatial resolution less than or equal to 5 m [31,33]. The method for detecting the effects of aerosol scattering relies on the comparison between well illuminated vegetation and shaded vegetation based on the identification of the shadows. The method is based on two main hypotheses: (i) diffuse irradiance, as the main contributor to the illumination of shaded areas, depends on aerosol distribution and multiple scattering effects, and (ii) the average adjacent shaded and directly illuminated pixels have the same reflectance level. First, only the four spectral bands (450, 530, 670, and 780 nm) not impacted by water vapour absorption are used to compute shadows maps. The aerosol amount is tuned using an iterative procedure to adjust the diffuse irradiance onto the cast shadow areas until the areas have reached the same reflectance characteristics as the well-illuminated area (0.05% difference in average reflectance). A moving window smoothing is applied to derive the AOT distribution. Then, in the process of atmospheric correction, the atmospheric compensation model uses AOT value for each pixel. This approach produces a map of AOT distribution.

1.4.2. Shadow Removal

Shadow-induced illumination variations among tree crowns affect spectral information. Several studies reported improved tree classification accuracy when selecting sunlit pixels only [14,15,34]. Consequently, before training the classifier, the dark pixels were removed to evaluate the ensuing gain in classification accuracy. For this step, we used the shadow detection method described in [33] to compute the illumination ratio and the pixels with a scaled shadow fraction value lower than 0.6 were removed. The number of pixels used in the classification was reduced by 20% after this step.

1.4.3. Bidirectional Reflectance Distribution Function

Bidirectional Reflectance Distribution Function (BRDF) correction can be applied to reduce spectrum variation caused by acquisition conditions. The moderate change in sun-sensor angles (variation in scan angle ±6.5 deg.; variation in sun zenith angles: 6.5–10.0 deg.; variation in sun azimuth angle 285.9–267.7 deg.) in the present settings suggested that BRDF effects were of moderate intensity. Therefore, no BRDF correction was applied. First order BRDF effects were considered to stem from tree-tree shadowing. Little or no projected shadows are visible when sun and view directions are aligned and maximum shadows are apparent when sun and view directions are opposite. Such BRDF effects were minimized by shadow masking.

1.4.4. Impact of Flight Line Overlap

Individual flight lines were not mosaicked prior to crown extraction. In this way the full spectral variability of the crowns species in the images was preserved (and notably variability among flight lines induced by changes in the atmospheric properties and solar angle). Hence, due to a 50% overlap between flight lines on date 1, most of the crowns were extracted from two neighbouring flight lines. On date 2, only about 40% of the crowns were imaged twice. The effect of this lower redundancy on date 2 imagery is later considered.

We specifically evaluated the effect of prior mosaicking for an arbitrarily selected processing pathway (L1c Spa.F, SHAOT). To this end, we mosaicked the strips using center cropped method in the PARGE software. Hence, mosaicking was done by selecting the most central pixel of any two overlapping flight lines and not by averaging reflectance values of co-occuring pixels.

1.5. Data Analysis

1.5.1. Variance Analysis

For each wavelength, one-way analyse of variance (ANOVA) with species identity as the unique factor was run to assess how much variance in the reflectance at a particular wavelength could be ascribed to species. We compared effectiveness of various pre-processing steps by analyzing the R^2 averaged across the spectral domain. We interpret an increase in R^2 as an enhancement of species-specific spectral discrimination [35].

1.5.2. Classification

We used a LDA classification algorithm that is well adapted to classification problems in high dimension feature spaces provided that the number of observation is larger than the number of feature in the least abundant class [36]. LDA also has the advantage of not being affected by the possible inconsistent relative abundance of classes in training set and test set, provided prior probabilities of class membership are set to 1 for all classes as done in the present case. This method maximizes the ratio of between-class variance to the within-class variance. The "Mass" package version 7.3-50 [37] of R language [38] was used to compute LDA method. Good performance of LDA has been reported for tropical tree species classification from hyperspectral data in various studies [14,15,20]. In the first approach, the LDA classifier was trained using a random selection of 70% of the tree crowns (Table 1) using hyperspectral data from the 19 September for all processing levels. For every tree included in the training set both image segments (from neighbouring flight lines) were selected if available. The remaining 30% of crown segments were used to evaluate the performance of the classifier. This procedure was repeated 20 times to limit the impact of the random draw of the crowns used in training and testing sets. Classification accuracy was evaluated for all image-processing pathways using the same random sampling. The choice of using LDA can be questioned, as many alternative machine learning algorithms can be used as well. These state of the art classification algorithms include Random Forest (RF) [39], Support Vector Machine (SVM) [40] or neural network. The goal of our study was not to compare of the performances of the different algorithms available for

tree species discrimination, or to identify the most suitable as proposed by [15,20,41]. We compared the performances of LDA with a "state of the art" machine learning algorithm, RF. The RF classification algorithm was applied, following the exact same training and validation scheme as defined for LDA. The RandomForest R package [42] was used for this task, and 1000 trees were defined in the forest during the training stage. The results obtained with the RF algorithm are described in Appendix C Table A2.

In a second approach (cross date validation), the classifier was trained using hyperspectral data from 19 September, excluding any crown imaged on 20 September and the prediction was applied to crowns imaged on 19 September (single date case) and on 20 September (multidate case).

1.5.3. Classification Strategy

Numerous studies concluded on the gain in performances obtained when using object oriented approaches based on ITCs, combining pixel-wise classification with a majority vote rule to decide on the species to be assigned to individual tree crowns [20,41,43]. For each processing level, we compared classification accuracy at pixel level and crown level. Table 1 summarizes the crowns and pixels available for classification. The column named "proportion of area covered on date 2" corresponds to proportion of crown area viewed on date 2 which was viewed on date 1 (double counting any pixel viewed twice on a given date).

1.5.4. Spectral Stability Analysis

We computed various measures of spectral stability based on spectral correlation matrices in order to explore the sensitivity of the various processing pathways to changes in acquisition parameters induced by the multi date classification scenario.

For every species, we computed indices of species spectral signature stability over time, based on crowns viewed on both dates. Firstly, we computed the Pearson correlation between spectral correlation matrices at different dates. Secondly, we computed species specific F-measures for single date and multiple date cases.

$$\text{F measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{2}$$

$$\text{Precision} = \frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}} \tag{3}$$

$$\text{Recall} = \frac{\sum \text{True positive}}{\sum \text{Condition positive}} \tag{4}$$

Thirdly, we computed a "between dates spectrum distortion rate" by taking for each spectral band the difference between date 1 and date 2 of normalized values of signal divided by values at date 1. Then, the average (over all spectral bands) of the absolute value of that difference was computed per species. Both indices were computed for radiance spectrum (i.e., prior to atmospheric correction) and reflectance spectrum (post atmospheric correction).

$$\text{Distorsion rate} = \sum_{\lambda i, i=1}^{n=160} \frac{|S_{norm_{\lambda i,T1}} - S_{norm_{\lambda i,T2}}|}{S_{norm_{\lambda i,T1}}} \tag{5}$$

T_1 and T_2 represent the first and the second date respectively.

The contribution of each wavelength to the LDA was evaluated by summing absolute scaling values (coefficient of each wavelength to each linear discriminant function) after centring and reducing the data.

3. Results

1.6. Variance Analysis

We report the effect of various pre-processing steps on species specific signal noise ratio (SNR) as captured by the R^2 of the ANOVA of the different wavelengths.

Spatial filtering significantly increased the spectral variance part explained by species for radiance and reflectance values (from 12.1% to 14.2% for L1b and 12.3% to 16.1% for L1c, see Table 3). An even more significant surge in R^2 followed from normalisation of spectra (14.2% to 32.3% for L1b and 16.1 to 29% for L1c, see Table 3). This was not expected to be the case for L1c reflectance data, since it was implemented primarily to correct illumination variation affecting the whole spectrum.

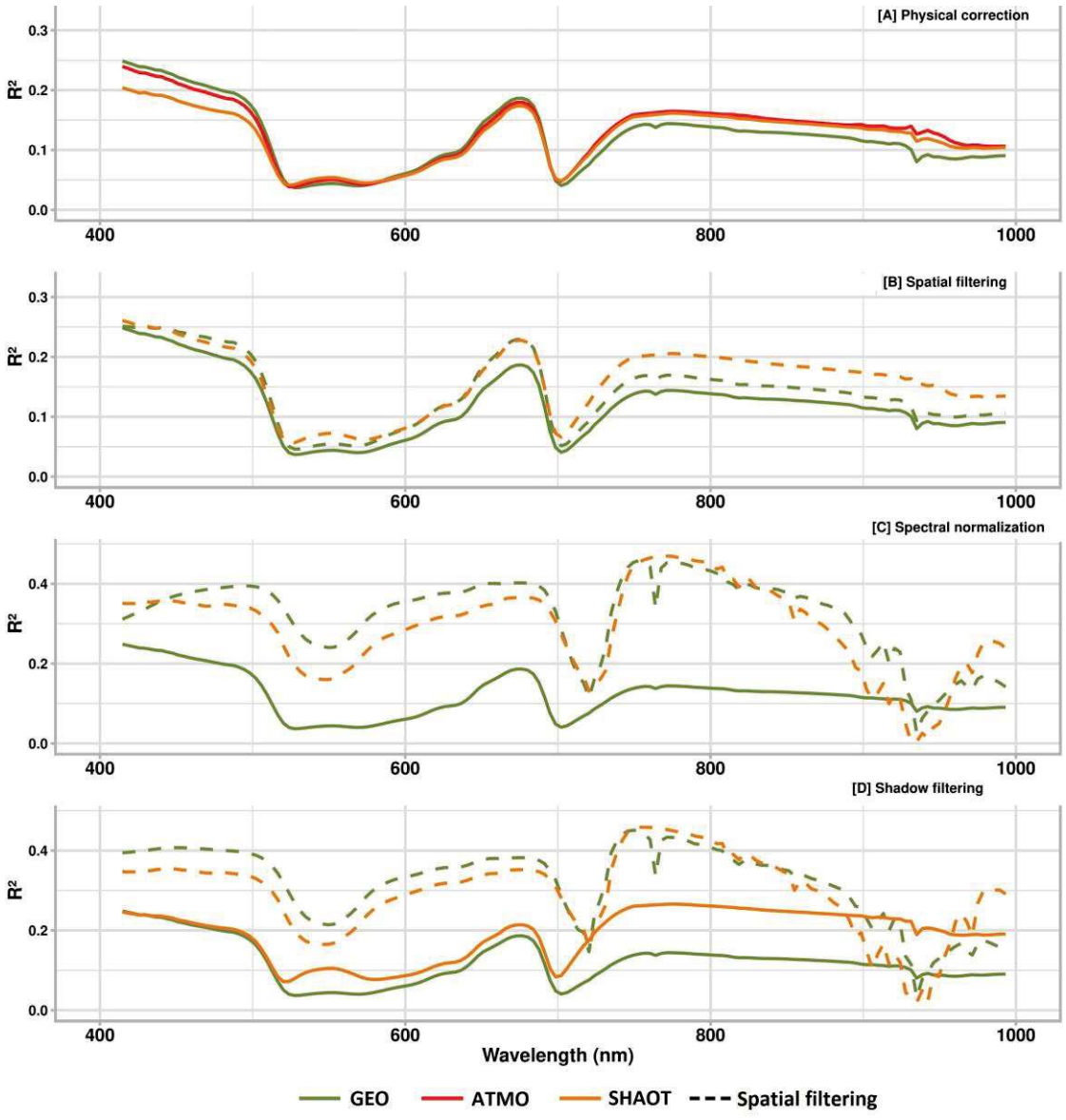

Figure 3. Proportion of variance explained by tree species identity (20 species) for each wavelength (data acquired on 19 September). L1b (GEO in green solid line) is plotted on each graphic as a reference.
(**A**) After atmospheric corrections only. (**B**) After spatial filtering (either alone or in addition to atmospheric corrections). (**C**) After normalisation applied to spatially filtered data. (**D**) After additional shadow removal.

Overall, ANOVA following spectral normalization of spatially filtered radiance data achieved the highest R^2 (32.3%). ANOVA following atmospheric correction, spatial filtering, SHAOT and shadow exclusion also had a high R^2 (29.5%). Removal of shadowed pixels ("Sha.R") had a slightly positive impact on L1c data only.

The atmospheric correction taking into account the spatial variation of AOT led to a reduction of R^2 in the spectral domain from 400 nm to 550 nm when compared to the standard correction (Figure 3), resulting in a slight overall decrease in R^2 (Table 3). Shadow removal mostly increased R^2 between 500 nm and 1000 nm. Spectral normalisation decreased R^2 both for L1b and L1c spectral data in the region from 902 nm to 956 nm due to very low signal and low SNR in the raw data. Not surprisingly, R^2 increase was found to have some similarities with the SNR evolution along the spectral variation following the different processing pathways (see Appendix B, Figure A1). Atmospheric correction alone (L1c with SHAOT or L1c without SHAOT) barely increased R^2 compared to radiance (L1b) values (Table 3).

Table 3. Mean proportion of variance explained by the tree species identity (mean R^2 over the entire spectrum) for each processing level (data acquired on the 19 September). The treatments are listed in order of application.

Treatments	Mean R^2 (%) over Wavelength
L1b	12.1
L1b, Spa.F	14.2
L1b, Spa.F, norm.	32.3
L1b, Spa.F, norm., Sha.R	31.9
L1c	13.0
L1c SHAOT	12.3
L1c, Spa.F, SHAOT	16.1
L1c SHAOT, Sha.R	19.0
L1c, Spa.F, SHAOT, norm.	29.0
L1c, Spa.F, SHAOT, Sha.R	21.8
L1c, Spa.F, SHAOT, norm.,Sha.R	29.5

1.7. Discriminant Analysis

Results are given for a subset of pre-processing pathways at pixel and crown level.

1.7.1. First Setting (Single Date)

Analysis of the impact of atmospheric correction was evaluated on all processing levels for a single data set to check the consistency of the methods. Standard deviation of accuracy at pixel level computed for the 20 runs was typically around 1–1.5% giving a standard deviation of the reported mean accuracy of 0.2–0.3%.

The pre-processing pathways that yielded the most accurate classifications (both at pixel and pixel-majority level) were L1b Spa.F, norm., Sha.R followed by L1c Spa.F, SHAOT, norm.,Sha.R (Table 4). Figure 4 shows a scatter plot of R^2 from the analysis of variance against the LDA accuracy. A slight positive correlation can be observed.

Table 4. Summary of mean accuracies and kappa results from data acquired on 19 September using the 20 most abundant species. The highest scores are highlighted in bold.

Treatments	Accuracy (%)		Kappa (%) Pixel	
	Pixel	Object	Pixel	Object
L1b	64.2	75.5	48.4	70.6
L1b Spa.F	73.8	81.7	66.5	78.9
L1b Spa.F, norm.	75.6	83.0	69.4	80.5
L1b Spa.F, norm., Sha.R	**76.9**	**83.3**	**71.2**	**80.9**
L1c without SHAOT	63.1	74.3	46.1	68.9
L1c with SHAOT	63.6	75.3	47.2	70.3
L1c SHAOT, Spa.F	73.4	81.9	65.9	79.1
L1c Spa.F, SHAOT	73.4	82.0	65.9	79.3
L1c SHAOT, Sha.R	66.9	76.5	54.1	71.9
L1c Spa.F, SHAOT, norm.	75.1	82.7	68.5	80.1
L1c Spa.F, SHAOT, Sha.R	74.7	82.6	68.1	80.0
L1c Spa.F, SHAOT, norm.,Sha.R	**76.5**	**83.2**	**70.7**	**80.8**

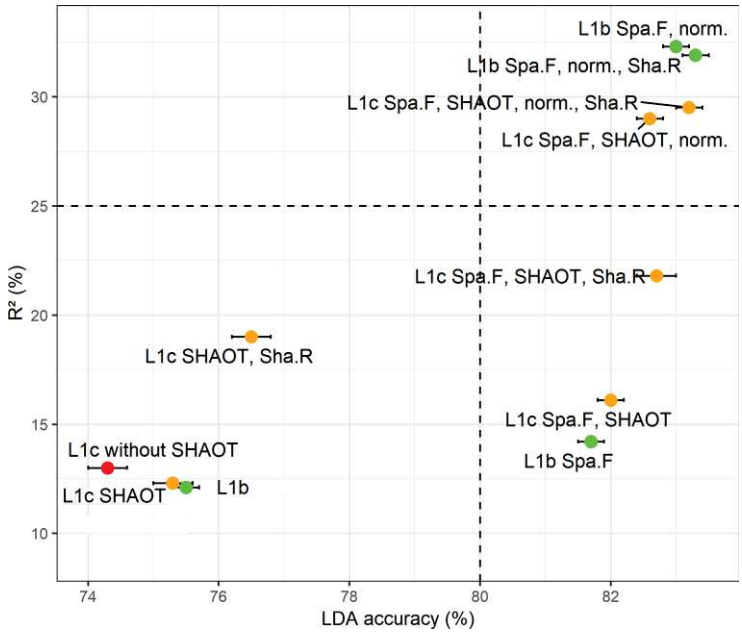

Figure 4. Treatments performance ranking by the mean R^2 (%) from the analysis of variance and LDA (%) accuracy of the object based on spectrum. Error bars represent ± standard error of the mean of 20 replicate runs. The legends and the colors refer to Figure 2.

The majority vote approach systematically outperformed the pixel classification. The majority vote was more efficient to retrieve species identity than a classifier based on a mean crown spectrum (see Appendix A, Table A1). Moreover, the standard deviation of accuracy using the mean crown spectrum computed for the 20 runs was around 4% which is higher than through majority vote procedure. The spatial filtering improved classification, especially at pixel level (from 64.2% to 73.8% for L1b and from 63.6% to 73.4% for L1c).

The confusion matrix (Table 5) indicated that not all species were recognized with equal accuracy. *Qualea rosea* was the best-identified species followed by *Tachigali melinonii*. While some species were poorly represented like *Tapura capitulifera*, their F-measure was still higher than 80.0%. By contrast, *Licania heteromorpha* was poorly identified and was mixed with an extremely abundant species *Eschweilera sagotiana*.

Table 5. Confusion matrix from L1b Spa.F, norm. at the crowns scale for the twenty most abundant species.

Predicted \ True	B. prouacensis	C. multiflora	D. guianensis	E. falcata	E. grandiflora	E. sagotiana	G. glabra	I. alba	J. copaia	L. alba	L. heteromorpha	M. coccinea	P. cochlearia	Q. rosea	R. speciosum	S. rubra	S. sp.1	T. melinonii	T. capitulifera	V. americana	Recall (%)	Precision (%)	F-measure (%)
B. prouacensis	60	2	0	0	0	6	0	0	0	0	4	0	0	0	0	0	0	2	0	8	73.2	42.9	54.1
C. multiflora	0	203	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	7	94.9	67.7	79.0
D. guianensis	15	36	565	25	0	34	0	6	0	3	13	0	10	7	13	0	22	3	11	5	73.6	88.3	80.3
E. falcata	35	3	14	504	5	6	7	0	4	24	0	11	7	0	19	0	15	0	18	15	73.4	78.8	76.0
E. grandiflora	0	15	5	15	420	15	0	0	15	8	7	0	2	0	0	0	0	0	4	7	81.9	95.5	88.1
E. sagotiana	19	0	2	11	0	739	0	0	0	9	76	5	0	0	0	7	2	0	0	15	83.5	88.0	85.7
G. glabra	0	2	3	9	0	8	133	0	55	6	0	0	0	0	0	7	0	0	0	0	59.6	95.0	73.3
I. alba	0	6	0	0	0	0	0	118	4	0	0	0	0	0	0	0	0	0	0	0	92.2	73.8	82.0
J. copaia	0	0	0	0	0	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0	100	31.4	47.8
L. alba	1	0	0	0	6	6	0	0	0	173	0	0	4	0	2	0	4	6	0	0	85.6	62.7	72.4
L. heteromorpha	0	2	0	0	0	2	0	0	0	0	34	0	1	0	0	0	0	0	0	0	87.2	21.3	34.2
M. coccinea	0	0	0	0	0	0	0	0	0	0	0	85	0	0	0	0	0	0	0	0	100	53.1	69.4
P. cochlearia	0	18	40	20	9	9	0	0	6	37	6	26	954	1	14	0	8	4	15	3	81.5	97.4	88.7
Q. rosea	0	1	0	27	0	5	0	0	12	15	16	4	0	1232	0	13	10	0	0	10	91.6	99.4	95.3
R. speciosum	0	0	0	0	0	0	0	2	0	0	0	0	0	0	369	0	0	0	0	0	99.5	87.9	93.3
S. rubra	1	12	8	8	0	0	0	0	0	0	0	4	2	0	2	173	0	0	0	4	80.8	86.5	83.6
S. sp.1	0	0	1	1	0	6	0	0	0	1	4	19	0	0	0	0	127	0	8	0	76.1	63.5	69.2
T. melinonii	0	0	2	7	0	0	0	34	0	0	0	1	0	0	0	0	6	285	0	0	85.1	95.0	89.8
T. capitulifera	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	144	0	98.6	72.0	83.2
V. americana	9	0	0	11	0	4	0	0	0	0	0	1	0	0	1	0	6	0	0	126	79.8	63.0	70.4

Additionally, an extra set of processing options not detailed here was tested. Spectrum derivation following spectral smoothing were applied and spectrum derivative used instead of raw spectrum. Excluding the noisy region between 900 and 950 nm prior to applying the LDA was also tested. Finally, we tested whether including the pixel spectral mean as an additional feature to the standardized spectrum would make a significant difference. None of those variants yielded a significant improvement of the best classification results (improvement <1%).

1.7.2. Second Setting (Cross Date Training and Validation)

In this operational situation training is done on a well-known scene at first and the training results are transferred to the second date of data acquisition.

One species was absent from the area covered on the second date. The 19 species left were trained on date 1 imagery and predicted on date 2 imagery (multidate case). For comparison purposes, we report the performance of the LDA classifier trained and predicted on date 1 (single date case) using the same subset of crowns in both training and testing sets as used in cross date validation. The same combination of processing steps was considered in this setting and their effectiveness is compared in Table 6.

Regardless of the processing pathway considered, the transition from a single date to a multiple date case induced a marked decrease in the accuracy of the classification especially at object level. Nonetheless, the pathway including atmospheric correction (L1c, Spa.F, SHAOT, norm.) clearly outperformed (by more than 8 percentage points at object level) any filtering/normalizing strategy when training and testing data came from different dates. The best pathway is the L1c, Spa.F, SHAOT, norm., Sha.R (pixel illumination was considered) which uses 20% less data.

Remarkably the gain in accuracy at crown level (as compared to pixel level) was not significant in the multidate scenario for most processing pathways whereas it typically improved accuracy by c. 10% in the single date scenario.

Table 6. Summary of the results from the comparison between two dates using 19 species with respect of each treatment.

Treatments	Accuracy (%)		Kappa (%)	
	Pixel	Object	Pixel	Object
Single date case				
L1b	55.0	65.4	48.5	61.4
L1b, Spa.F	65.5	73.6	60.8	70.9
L1b, Spa.F, norm.	67.5	76.4	63.3	73.9
L1b, Spa.F, norm., Sha.R	69.4	76.6	65.3	74.2
L1c SHAOT	54.3	65.4	47.6	61.5
L1c, Spa.F, SHAOT	64.6	72.4	59.5	69.4
L1c, Spa.F, SHAOT, norm.	67.8	76.6	63.5	73.9
L1c, Spa.F, SHAOT, norm., Sha.R	**69.7**	**78.2**	**65.5**	**75.9**
Multidate case				
L1b	39.7	39.20	32.0	34.6
L1b, Spa.F	53.0	53.3	46.2	48.7
L1b, Spa.F, norm.	54.7	54.9	49.0	50.8
L1b, Spa.F, norm., Sha.R	61.2	60.3	55.0	56.6
L1c SHAOT	46.5	50.2	39.4	45.6
L1c, Spa.F, SHAOT	58.6	61.5	52.8	57.7
L1c, Spa.F, SHAOT, norm.	60.2	66.1	55.2	62.9
L1c, Spa.F, SHAOT, norm., Sha.R	**67.0**	**68.6**	**61.7**	**65.6**

Table 7 presents classification accuracy (at pixel and crown level) for one particular processing pathway L1c, Spa.F, SHAOT when either training or prediction applies to mosaicked data or raw flight lines.

Table 7. Summary of the results comparing learning and prediction when multi-flight lines or mosaicked data were applied in the classification process. Standard error of the mean (SEM) is given at pixel and object level. Pixel and object express the accuracy rate. * 60% of the data is not viewed twice.

	Learning Data	Mosaicked		Multi Flight Lines	
	Predict Data	Pixel (%) (SEM)	Pixel-Majority (%) (SEM)	Pixel (%) (SEM)	Pixel-Majority (%) (SEM)
First setting	Mosaicked	71.9 ± 0.4	77.8 ± 0.4	72. ± 0.3	78.1 ± 0.2
	Multi flight lines	-	-	73.4 ± 0.4	82.0 ± 0.2
Second setting with single date case	Mosaicked	63.7	64.8	64.4	69.1
	Multi flight lines	-	-	64.6	72.4
Second setting with multidate case *	Mosaicked	51.4	50.3	57.1	58.9
	Multi flight lines	-	-	58.6	61.5

Pixel accuracy was higher when training and prediction sets used multiple flight lines. The gain in accuracy ensuing a majority vote (i.e., going from pixel to pixel-majority level accuracy) is larger in the single date settings and also larger when multiple flight lines are used (both for training and testing) instead of mosaic. Mosaicked imagery is not only on average slightly less effective but it is also more variable at crown level.

1.8. Comparing ANOVA and LDA Results

Figure 3 indicates that while normalization improved ANOVA R^2 very significantly (10–18 percentage points) it had only a modest effect on LDA pixel classification accuracy (1–2 percentage points). Conversely, spatial filtering brought significant improvement to LDA accuracy (6–10 percentage points) irrespective of the processing level considered and also improved R^2 albeit less significantly (2–3 percentage points). Spatial filtering reduced noise and increased both separability and signal to noise ratio in each band as expected [44]. Normalisation mostly affected ANOVA results by smoothing out effects of variable illumination between pixels, which was confirmed by a slight increase using L1c Spa.F, SHAOT and norm., Sha.R data. LDA benefitted less from normalisation

probably because it is sensitive to covariation between wavelengths, which may not be as sensitive to illumination level.

Although we focused on mean R^2 in ANOVA analysis, the improvement was not homogeneous across wavelengths. In some extreme cases processing increased R^2 overall while decreasing it in some particular spectral window (e.g., Figure 3, normalisation method from 900 nm to 950 nm). This may further blur the relationship between mean R^2 from single band ANOVA analysis and LDA since the net effect of a global increase in R^2 associated with a decrease in a particular spectral window may affect LDA classification accuracy either way.

1.9. Spectral Stability Analysis

Species were differently affected by the date transfer. The most impacted species were *Qualea Rosea*, *Licania heteromorpha*, *Vouacapoua americana* and *Eperua grandiflora*, with a loss in F-measure of 41%, 31%, 27% and 20% respectively (Table 8). Removal of shadow pixels increased classification accuracy at pixel level more than at crown level.

Table 8. Person's correlation between spectral correlation matrices at each date for two processing pathways, corresponding classification F-measure * and spectrum distortion rates (multiple date scenario, see text). L1b and L1c refer to L1b Spa.F, norm., Sha.R and to L1c Spa.F, norm., Sha.R data respectively.

Species	Person's correlation		Species Classification F-Measure (%)						Distortion Rate (%)		Segment Number
			Single Date		Multi Date		Delta				
	L1b	L1c	L1b	L1c	L1b	L1c	L1b	L1c	L1b	L1c	
B.prouacensis	0.79*	0.98	90.9	75.0	0	75.0	−90.9	0	4.5	4.2	8
C.multiflora	0.98	0.97	20.0	36.4	50.0	36.5	30	0.1	5.2	3.7	11
D.guianensis	0.71*	0.97	78.9	72.7	60.2	68.2	−18.7	−4.5	2.2	1.4	36
E.falcata	0.99	0.98	82.6	85.4	4.5	65.7	−78.1	−19.7	3.9	2.1	48
E.grandiflora	0.99	0.93	61.1	66.7	72.7	72.0	11.6	5.3	6.1	2.1	13
E.sagotiana	0.99	0.92	86.2	85.2	71.0	73.3	−15.2	−11.9	5.6	2.4	65
G.glabra	0.97	0.96	100	57.1	21.1	40.0	−78.9	−17.1	8.7	9.5	3
J.copaia	0.90	0.90	57.1	57.1	57.1	57.1	0	0	4.7	3.5	8
L.alba	0.99	0.93	55.6	62.5	62.5	66.7	6.9	4.2	2.5	1.1	10
L.heteromorpha	0.99	0.98	16.7	30.8	30.8	0	14.1	-30.8	3.3	1.9	9
M.coccinea	0.90	0.95	45.5	45.5	30.0	45.5	−15.5	0	5.6	3.4	19
P.cochlearia	0.99	0.99	78.8	76.5	60.9	74.3	−17.9	−2.2	3.6	1.1	40
Q.rosea	0.95	0.97	77.8	70.0	46.7	29.2	−31.1	−40.8	5.4	3.1	10
R.speciosum	0.66*	0.98	91.7	91.7	84.4	91.7	−7.3	0	2.2	1.4	28
S.rubra	0.99	0.98	94.1	94.1	66.7	77.8	−27.4	−16.3	4.7	3.8	10
S.sp.1	0.99	0.99	64.3	71.4	45.5	60.9	−18.8	−10.5	4.1	1.8	16
T.melinonii	0.77*	0.94	90.0	90.0	90.0	90.0	0	0	5.6	3.1	12
T.capitulifera	0.98	0.92	75.9	80.0	50.0	66.7	−25.9	−13.3	6.3	7.8	19
V.americana	0.92	0.91	44.4	72.7	20.0	46.2	−24.4	−26.5	6.0	5.4	8
Global	**0.84**	**0.97**					**−20.4**	**−9.7**	**4.8**	**3.3**	

LDA scaling values (applied to reduced and centered data from date 1) are plotted in Figure 5.

Figure 5 reveals a close correlation between scaling values with and without atmospheric correction. However, the contribution of the 800–1000-nm region is much lower in atmospherically corrected (L1c Spa.F, norm., Sha.R) data, especially relative to the visible range.

Species-specific distortion rates between dates are reported in Table 8. Those rates were higher for normalized radiance values than for normalized reflectance (i.e., corrected for atmospheric effects) in 17 out of 19 species. Remarkably, the distortion rate was contrasted across species (ranging from 1% to 9.3%). There was a systematic and almost complete reduction of distortion in the 700–900 nm region following atmospheric correction, as illustrated for three arbitrarily chosen species in Figure 6.

Table 8 shows how species signature and discrimination rate (F-measure) were affected by change in dates between training and testing data sets for the two best pre-processing pathways. The correlation between spectral correlation matrices of date 1 and date 2 are also reported.

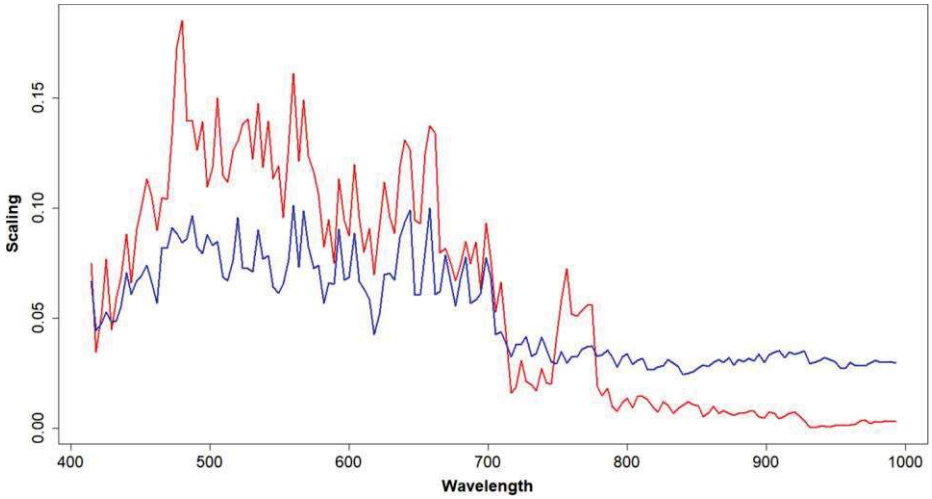

Figure 5. Sum of absolute scaling values over the spectrum from linear combinations. Blue and red lines represent L1b Spa.F, norm., Sha.R and L1c Spa.F, norm., Sha.R data respectively.

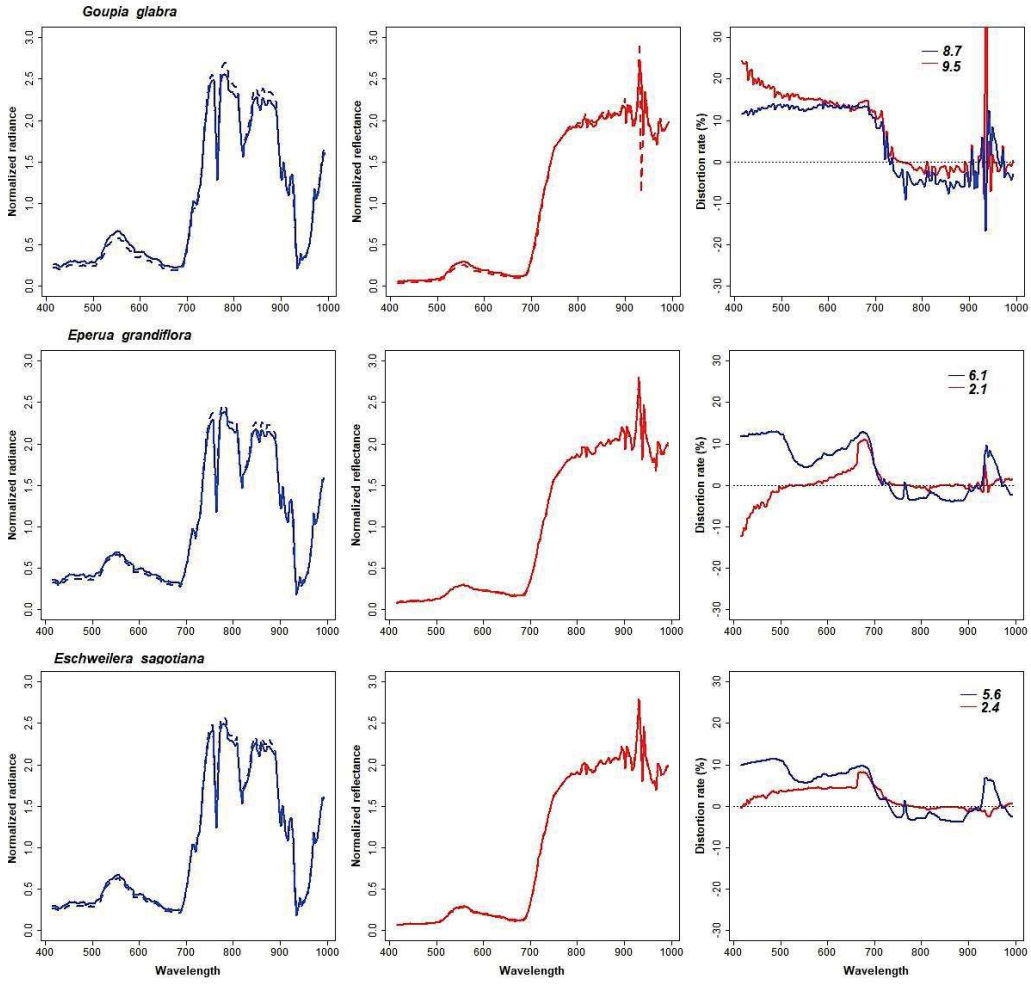

Figure 6. Mean normalized radiance (first column) and mean normalized reflectance (second column) for 3 arbitrarily selected species. Solid line stands for date 1, dotted line for date 2. Third column represents normalized difference of average spectrum using the two acquisition dates. Blue line is for L1b filtered and normalized spectrum and red line for L1c SHAOT, filtered and normalized spectrum. The mean distortion rate (over all wavelengths) is given in the upper right corner for each curve.

The global correlation between spectral correlation matrices at date 1 and date 2 is lower for L1b Spa.F, norm., Sha.R data than to L1c Spa.F, norm., Sha.R. For a majority of species, the correlation between dates is similar in both processing pathways. For four species this correlation is more than 10 percentage points lower (marked with an asterisk in Table 8). Species classification F-measure are globally more affected using L1b Spa.F, norm., Sha.R data than to L1c Spa.F, norm., Sha.R, and spectrum distortion rates are larger. Large differences across species in distortion rates and change in distortion rates from L1b to L1c are striking.

Correlation coefficients, distortion rates and change in F-measure are not strongly related. However, in some extreme cases the indicators do appear to vary consistently. For instance *Bocoa prouacensis*, has a much lower correlation coefficient across dates for L1b data than L1c data, it also suffers from a stronger reduction in F-measure and a higher distortion rate.

4. Discussion

1.10. LDA Classification Accuracy

The impact of different preprocessing pathways on species separability was evaluated both at pixel and crown scale. Classification accuracies tend to decrease with the number of classes [45]. The classification of 20 tree species is a task of slightly higher complexity than what had been previously attempted in studies concerning tree species classification in tropical forests. References [14,20] classified seven species and achieved an accuracy of 86% and 87.4% respectively with a lower spatial resolution (1.6 m) using HYDICE (HYperspectral Digital Imagery Collection Experiment) airborne sensor, which measures radiance between 400 nm and 2500 nm in 210 discrete bands. Ferreira et al. [41] classified eight tropical tree species using the full spectrum (400–2400 nm) using airborne AisaEAGLE and AisaHAWK instruments, achieving 84.9% overall accuracy of crowns. Feret et al. [15] classified 17 pure species with SVM (Support Vector Machine) algorithm using RBF (Radial Basic function) kernel reaching an overall accuracy around 72%. A pure species was determined by similarity measure based on the difference in amplitude and spectral angle, which is a specific approach taking into account seasonal and environmental factors. In the above-mentioned studies, there are no species in common with those studied here. In our case (Table 4), the accuracy of crown scale classification was 83.3% via L1b Spa.F, norm., Sha.R data, and 83.2% for L1c Spa.F, SHAOT, norm., Sha.R, comparable to the results of the above-mentioned studies.

On date 1 crowns were imaged twice and all views were used in the classification thereby preserving a greater variability. This proved beneficial for the identification of trees at the crown scale compared to the use of a mosaic image (77.8% vs. 82.0%, tested on L1c Spa.F, SHAOT at crown level cf Table 7). Moreover, once the classifier is trained on a low variability (mosaicked data) and used for prediction on multi flight lines data, the results are similar (77.8% and 78.1% on the object scale). Increasing the variability in the learning step improved the performance of the classifier. The gain was mostly achieved at crown level after application of the majority filter. Therefore, we found that increasing the size of the training set increases the classifier's performance even if the variability in the validation data set increases in parallel. Increasing the size (and variability) of the training set data more than compensated for the correlative increase in the size of the data set to be classified.

Despite the admittedly small sample of species examined, it is worth noting that confusion between phylogenetically close species was not systematically higher than between more distant species. In particular, species from the same genus (*L. heteromorpha* and *L. alba* on one side and *E. grandiflora* and *E. falcata* on the other) were well separated.

Regarding the wavelength contributing most to the LDA (Figure 5), we found similarities with the results of [46] who used a partial least squares discriminant analysis (PLS-DA) for the identification of tree species on a site in South Africa (KwaZulu-Natal). Große-Stoltenberg et al. [47] also report a major contribution of the visible range (400–700 nm) to the discrimination of vegetation sampled with hand

held spectrometer within a Mediterranean dune ecosystem. They also found a large contribution of SWIR range (not included in this study).

1.11. Simple Methods

Although physical approaches have a role to play in identifying trees in the canopy through atmospheric corrections, other methods that are easier to implement have previously been shown to perform well. Shahriari et al. [48] explored Gaussian filtering applied to raw spectrum prior to atmospheric correction. Classification accuracy was not systematically improved by filtering. Results depended on the smoothing window size and the atmospheric corrections made (either using FLAASH or ATCOR). Nevalainen et al. [49] tested normalization (division by the sum of all bands) which improved slightly the classification accuracy. In the present study, a spatial filter improved the classification in all cases (single date and multi date) as opposed to atmospheric correction alone. The average filtering reduced the spatial noise and homogenized the spectral signature of trees. The results suggest that noise in the signal remained after atmospheric correction which was later reduced using a simple spatial filtering.

Another part of our study focused on the improvement brought by the removal of shadow (low illumination pixels). Lopatin et al. and De Sa et al. [22,50] used an UAV (unmanned aerial vehicles) for the classification of invasive species, and reported that shadows have a high negative effect on classification. Lopatin et al. [22] identify shadows by histogram thresholding, as also proposed by [51,52] at a higher spatial resolution. In the present study, removal of shadowed pixels increased the separability both at the pixel and object-crown level. Crown level separability was less significantly improved than pixel level suggesting that majority filter contributed to sieve out many shadowed pixel in the first place. However, this filtering decreased the number of pixels by c. 20% in the present case. Nagendra, H. [53] shows that a classification can be more efficient by smoothing shadowed and well illuminated pixels. This may partly explain the accuracy increase when the data is filtered by mean filtering, reducing intra-class variability. De-shadowing would allow preserving shadowed pixels but may be more difficult to achieve in a consistent way. As high radiometric sensitivity is required to analyse shaded pixels radiometrically, such approaches require careful further analysis.

1.12. Operational Setting

In an operational perspective, the comparison made between dates is more relevant to consider, as this setting is a standard use case. Selecting one part of the imaged area to predict the other (both at the same date) yielded a decrease in tree identification rate of about 5% compared to random selection of training and testing of individuals (compare Table 4 and Table 6). This was probably due to the unbalanced training versus testing set. The split between training and testing sets of the crowns was imposed by the actual spatial distribution of the species and resulted in a sub-optimal design (Table 1, last column).

Training on one date and predicting on another date (in a same area) yielded a further decrease in tree identification rate of c. 10% point (Table 6). The benefits of atmospheric correction data became obvious as full atmospheric correction outperformed simple statistical procedures by 8% (Table 6). Hence, despite very similar conditions (single site, same flight time, same cloudiness, and only 24 h delay between successive flights) neglecting atmospheric corrections degraded classification accuracy very significantly. In a multiple site context, simple standardization that does not consider atmospheric variations may not allow effective species identification using a classifier trained on a single site. Normalisation of hyperspectral signal through atmospheric correction seems to be required to effectively transfer a classifier from one site to another. It should be stressed that other parameters affect transferability and should be taken into account. These include phenological changes and notably seasonality (transferring from one date to another), degree of similarity of the species communities (transferring from one site to another), technical specifications of imager and acquisition parameters (all cases) [54].

In addition, the use of multiple flight lines should be preferred over the use of a mosaic image, both for the learning and prediction steps (Table 7). The lower accuracy observed in date 2 prediction is partly due to the lower redundancy in training data with only 40% of the crowns being viewed twice.

Intriguingly, species were not equally affected by change in date, and their sensitivity depended both on the processing pathway considered and the wavelength considered (Figure 6).

In our case, the use of ATCOR-4 was primarily motivated by its ability to derive spatially explicit AOT estimates (SHAOT method). The SHAOT method for AOT recovery had not been previously tested for tree species classification purposes. SHAOT correction did not increase ANOVA R^2 and notably reduced the variance explained by species identity in the 400–550 nm range (Figure 3, red and orange lines). However, the SHAOT method included in the atmospheric correction slightly improved tree species discrimination achieved by the LDA, at pixel (0.5%) and crown level (1.0%).

Overall, the results of this study are very encouraging as a number of improvements are foreseen for the near future. First, improved matching of tree crowns (mostly delineated on the LiDAR derived canopy height) and pixels may be achieved by better data co-alignment between LiDAR and spectral data as shown in [55]. Second, only VNIR data (400 nm to 1000 nm) were used. Ferreira et al. [41] reported that the addition of SWIR (Short-Wave Infrared) data improved the identification of species of interest by 13% in their study. SWIR data was acquired simultaneously and will be used to enrich species spectral characterization. Third, LiDAR derived features such as tree height [56], crown density and crown shape may constitute complementary features to include in the species classification [57]. Additional information related to pulse distribution or other features extracted from the LiDAR wave form have also proven to be useful [58]. Ultimately, more advanced classification methods such as SVM or Convolutional Neural Networks may provide some additional improvement in classification accuracy.

Discriminating trees species becomes more difficult as the number of species increases [45]. Detecting targeted species within a large set of species, many of which not being identified, is a slightly different problem from the one addressed in this study, in which a predetermined subset of species had to be sorted out. The former problem is particularly challenging in the case of hyperdiverse canopies [13]. Yet it is an unavoidable step to be taken before a reference hyperspectral database for the vast number of tropical species becomes available. The extreme species richness of tropical forests makes recognition of all species in the canopy a very challenging enterprise, even if those species constitute only a subset of the tree species diversity. Futher complication comes from the fact that species abundance is extremely variable and that some species are globally or locally rare [1]. However mapping diversity patterns does not necessarily rely on species identification. Spatial patterns of species diversity can correlate to patterns of spectral diversity [59,60].

The majority filter per crown used here requires a pre-segmentation of individual crowns. In the present case, crown segmentation was performed manually on high-resolution images and validated in the field. However, automatic segmentation of tree crowns from point clouds is becoming effective even in dense tropical forests [61] and can be further refined using hyperspectral data [62]. Therefore high throughput mapping of selected species in complex tropical forest canopies seems to be a realistic goal for the near future. One limitation to automatic detection of targeted species may stem from locally abundant lianas and epiphytes, which can strongly affect or even mask out the tree's specific spectral signature. Detection of lianas and characterization of their abundance at crown level seems to be tractable [63] but deserves further study.

BRDF effects caused by anisotropic scattering property, related to view and illumination geometry during images acquisition [64], may affect the mapping of tree species. Colgan et al. [43] reported a slight improvement using SVM classification after applying a MODIS derived BRDF model to airborne hyperspectral data. The spectrum range most impacted by this effect seemed to be in the visible range. Correction of such bi-directional reflectance effects may also be required to fully standardize species spectral signature, in particular in cases of large variation in Viewing Zenith Angle (VZA).

5. Conclusions

This study confirms earlier ones showing that the discrimination and identification of tree species in a hyper-diverse canopy by means of imaging spectroscopy is possible with a high accuracy, even for as many as 20 species. A major contribution of our study was to clearly show that the best pre-processing pathway will likely depend on the particular settings of the study.

We found that atmospheric correction did not improve the classifier's accuracy in a single date-single site setting (as compared to more readily applicable statistical procedures). However, when different dates were considered (the classifier trained on one date and tested on another date using different trees in training and testing sets) the atmospheric correction provided an obvious benefit, improving classification accuracy by 8% at crown level.

We, therefore, recommend to systematically conduct atmospheric compensation, in order to gradually build databases of standardized species spectral signatures. Such databases will be required to effectively move towards high throughput species mapping of tropical forest canopies, by allowing to circumvent or significantly alleviate the site specific calibration stage of the classifier. Finally, our results showed that the use of raw flight lines images improved both training and classifying over the use of a mosaic image. We recommend to systematically examine the benefit of using overlapping images for species recognition.

Author Contributions: G.V., J.-B.F., L.D. and A.L. designed the study, A.L. and G.V. conducted the analysis, D.S. provided expert advice on atmospheric correction methodology, A.L. and G.V. wrote the original draft, all authors contributed to final review and editing.

Funding: This research was funded by the French Ministry of Agriculture, grant number FSFB-2017. This study benefited from funding by the Centre National d'Etudes Spatiales (CNES Tosca Leaf EXPEVAL). This is a publication of Laboratoire d'Excellence CEBA (ANR-10-LABX-25).

Acknowledgments: Field support from Jean-Louis Smock (AMAP-IRD), Chantal Geniez (AMAP-IRD), Carla Baltzer (AMAP-IRD) and Raphaël Dutrieux (AMAP-IRD) is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

6. Abbreviations

The following abbreviations are used in this manuscript:

CIRAD	Centre de coopération internationale en recherche agronomique pour le développement
LiDAR	Light Detection And Ranging
RGB	Red, Green, Blue
ITC	individual tree crowns
DSM	Digital Surface Model
WGS	World Geodetic System
UTM	Universal Transverse Mercator
EPSG	European Petroleum Survey Group
APDA	Atmospheric Precorrected Differential Absorption
AOT	Aerosol Optical Thickness
SHAOT	shadow-based AOT
BRDF	bidirectional reflectance distribution function
SNR	Signal to Noise Ratio
LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
RBF	Radial Basic function
SWIR	Short-Wave Infrared

7. Appendix A

Table A1 shows the discrepancy between two strategies: mean spectrum and majority vote by crown. The mean spectrum was computed using the multi flight line. The training was realized as described in the classification section. The standard deviation of accuracy using the mean crown spectrum computed for the 20 runs was around 4% which it is higher than using majority vote procedure (1–1.5%). This is probably due to the much lower number of elementary observations when mean crown spectra are used.

Table A1. Summary of mean accuracies and kappa results from data acquired on the 19 September using the 20 most abundant species and comparing two strategies: object and majority vote.

Treatment	Spectral Average (%)		Majority Vote Accuracy (%)	Kappa
		Accuracy (%)	Kappa (%)	
L1b	79.4	76.2	75.5	70.6
L1b Spa.F	79.6	76.4	81.7	78.9
L1b Spa.F, norm.	81.7	79.0	83	80.5
L1b Spa.F, norm., Sha.R	81.7	79.0	83.3	80.9
L1c without SHAOT	79.1	75.7	74.3	68.9
L1c with SHAOT	79.2	75.9	75.3	70.3
L1c SHAOT, Spa.F	79.2	75.9	81.9	79.1
L1c Spa.F, SHAOT	79.4	76.2	82.0	79.3
L1c SHAOT, Sha.R	79.7	76.5	76.5	71.9
L1c Spa.F, SHAOT, norm.	81.3	78.5	82.7	80.1
L1c Spa.F, SHAOT, Sha.R	79.7	76.5	82.6	80.0
L1c Spa.F, SHAOT, norm., Sha.R	81.4	78.6	83.2	80.8

8. Appendix B

Figure A1 shows the similarities with the ANOVA results.

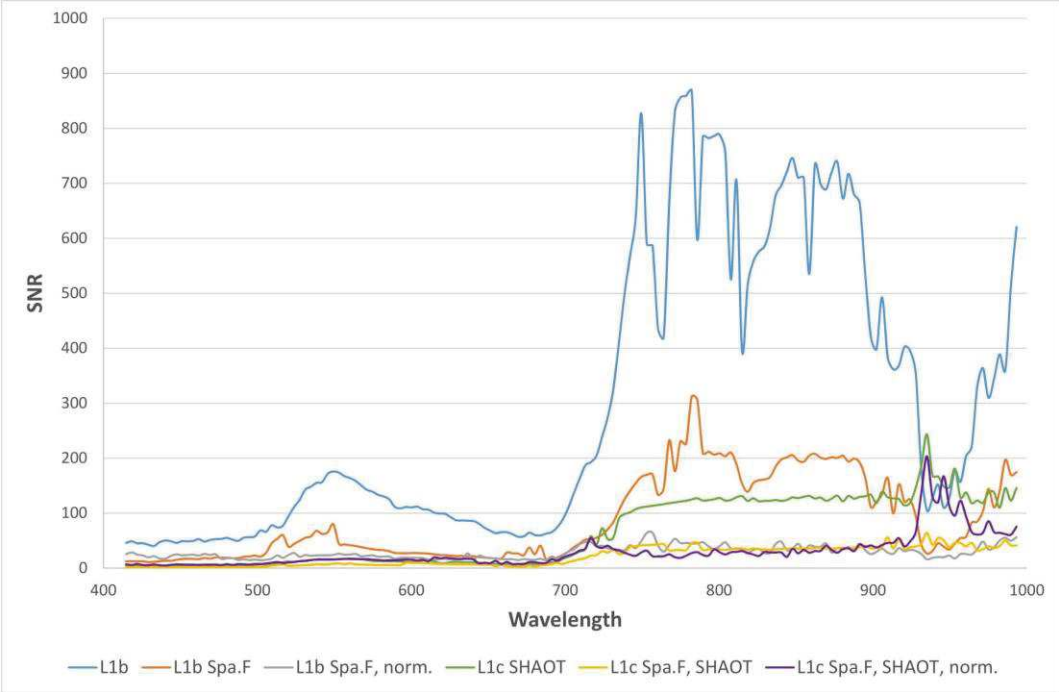

Figure A1. SNR (Signal to Noise Ratio) computed for one flight line.

The SNR was computed on one cloudless flight line based on the PARGE method. The algorithm applies a high pass filtering on a dark homogeneous ROI (Region Of Interest) and works as follows:

- Search for the 7 × 7 pixels patch with smallest noise throughout the whole image in each band.

- Calculate the mean of the whole image and the mean of the patch
- Calculate the noise in the found patch after high pass filtering
- Obtain SNR values as mean reflectance divided by the noise in the patch

The smallest noise is obtained as the pixel of the minimum squared deviation between of high-pass filtered image (with 7 pixel size kernel). The obtained SNRs are a best-guess estimate and should not be taken as absolute quality measure for an imaging system. More details are given in PARGE documentation [27].

9. Appendix C

A Random Forest classifier was used with 1000 trees repeated 20 times while keeping the same methodology used for the LDA classifier (see Table A2). The atmospheric correction and the suppression of shadowed pixels have an impact on classification accuracies. Nevertheless, these accuracies are lower than using a LDA classifier (about 18% difference, see Table A2 and Table 4).

Table A2. Summary of mean accuracies and kappa results from data acquired on the 19 September using the 20 most abundant species based on a Random Forest classifier.

Treatment	Pixel		Object	
	Accuracy (%)	Kappa (%)	Accuracy (%)	Kappa (%)
L1b Spa,F, norm,	56.4		21.5	59.0	40.7
L1c Spa,F, SHAOT, norm,	58.3		28.2	61.7	46.8
L1b Spa,F, norm., Sha,R	57.3		25.4	60.3	44.0
L1c Spa,F, SHAOT, norm., Sha,R	59.3		32.6	62.4	48.6

10. Appendix D

The airborne LiDAR scanning data were acquired at the same time as the hyperspectral data, using a LMS Q780 RIEGL. The scan frequency was 400 khz and the final point density was around 33 pts/m^2. The scan angle was between ±30°. The point cloud was processed to provide geo-referenced 3D point cloud. The position and orientation of the platform were given by on-board GPS/IMU measurements. These parameters provided a point cloud in the WGS84/UTM zone 22N coordinate system. Point cloud filtering was done to remove non-valid points.

11. References

1. Ter Steege, H.; Pitman, N.C.A.; Sabatier, D.; Baraloto, C.; Salomao, R.P.; Guevara, J.E.; Phillips, O.L.; Castilho, C.V.; Magnusson, W.E.; Molino, J.F.; et al. Hyperdominance in the Amazonian Tree Flora. *Science* **2013**, *342*, 1243092. [CrossRef]
2. Brienen, R.J.W.; Phillips, O.L.; Feldpausch, T.R.; Gloor, E.; Baker, T.R.; Lloyd, J.; Lopez-Gonzalez, G.; Monteagudo Mendoza, A.; Malhi, Y.; Lewis, S.L.; et al. Long-term decline of the Amazon carbon sink. *Nature* **2014**, *519*, 344–348. [CrossRef]
3. Barlow, J.; França, F.; Gardner, T.A.; Hicks, C.C.; Lennox, G.D.; Berenguer, E.; Castello, L.; Economo, E.P.; Ferreira, J.; Guénard, B.; et al. The future of hyperdiverse tropical ecosystems. *Nature* **2018**, *559*, 517–526. [CrossRef]
4. DRYFLOR; Banda, K.; Delgado-Salinas, A.; Dexter, K.G.; Linares-Palomino, R.; Oliveira-Filho, A.; Prado, D.; Pullan, M.; Quintana, C.; Riina, R.; et al. Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* **2016**, *353*, 1383–1387. [CrossRef]
5. McDowell, N.; Allen, C.D.; Anderson-Teixeira, K.; Brando, P.; Brienen, R.; Chambers, J.; Christoffersen, B.; Davies, S.; Doughty, C.; Duque, A.; et al. Drivers and mechanisms of tree mortality in moist tropical forests. *New Phytol.* **2018**, *219*, 851–869. [CrossRef]
6. O'Neill, D.W.; Fanning, A.L.; Lamb, W.F.; Steinberger, J.K. A good life for all within planetary boundaries. *Nat. Sustain.* **2018**, *1*, 88–95. [CrossRef]
7. Deininger, K.; Byerlee, D. *Rising Global Interest in Farmland: Can It Yield Sustainable and Equitable Benefits?* The World Bank: Washington, DC, USA, 2011. [CrossRef]

8. Alamgir, M.; Campbell, M.J.; Sloan, S.; Goosem, M.; Clements, G.R.; Mahmoud, M.I.; Laurance, W.F. Economic, Socio-Political and Environmental Risks of Road Development in the Tropics. *Curr. Biol.* **2017**, *27*, R1130–R1140. [CrossRef]
9. Mitchard, E.T.A.; Feldpausch, T.R.; Brienen, R.J.W.; Lopez-Gonzalez, G.; Monteagudo, A.; Baker, T.R.; Lewis, S.L.; Lloyd, J.; Quesada, C.A.; Gloor, M.; et al. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites: Divergent forest carbon maps from plots & space. *Glob. Ecol. Biogeogr.* **2014**, *23*, 935–946. [CrossRef]
10. Cardoso, D.; Särkinen, T.; Alexander, S.; Amorim, A.M.; Bittrich, V.; Celis, M.; Daly, D.C.; Fiaschi, P.; Funk, V.A.; Giacomin, L.L.; et al. Amazon plant diversity revealed by a taxonomically verified species list. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10695–10700. [CrossRef]
11. Somers, B.; Asner, G.P. Hyperspectral Time Series Analysis of Native and Invasive Species in Hawaiian Rainforests. *Remote Sens.* **2012**, *4*, 2510–2529. [CrossRef]
12. Baldeck, C.; Asner, G. Improving Remote Species Identification through Efficient Training Data Collection. *Remote Sens.* **2014**, *6*, 2682–2698. [CrossRef]
13. Baldeck, C.A.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Knapp, D.E.; Kellner, J.R.; Wright, S.J. Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS ONE* **2015**, *10*, e0118403. [CrossRef] [PubMed]
14. Clark, M.; Roberts, D.; Clark, D. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* **2005**, *96*, 375–398. [CrossRef]
15. Feret, J.B.; Asner, G.P. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 73–84. [CrossRef]
16. Hueni, A.; Nieke, J.; Schopfer, J.; Kneubühler, M.; Itten, K. The spectral database SPECCHIO for improved long-term usability and data sharing. *Comput. Geosci.* **2009**, *35*, 557–565. [CrossRef]
17. Bojinski, S.; Schaepman, M.; Schläpfer, D.; Itten, K. SPECCHIO: A spectrum database for remote sensing applications. *Comput. Geosci.* **2003**, *29*, 27–38. [CrossRef]
18. Chen, C.; Li, W.; Tramel, E.W.; Cui, M.; Prasad, S.; Fowler, J.E. Spectral–Spatial Preprocessing Using Multihypothesis Prediction for Noise-Robust Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1047–1059. [CrossRef]
19. Hively, W.D.; McCarty, G.W.; Reeves, J.B.; Lang, M.W.; Oesterling, R.A.; Delwiche, S.R. Use of Airborne Hyperspectral Imagery to Map Soil Properties in Tilled Agricultural Fields. *Appl. Environ. Soil Sci.* **2011**, *2011*, 1–13. [CrossRef]
20. Clark, M.L.; Roberts, D.A. Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sens.* **2012**, *4*, 1820–1855. [CrossRef]
21. Friman, O.; Tolt, G.; Ahlberg, J. Illumination and shadow compensation of hyperspectral images using a digital surface model and non-linear least squares estimation. In *Proceedings of SPIE—The International Society for Optical Engineering*; SPIE: Bellingham, WA, USA, 2011; p. 81800Q. [CrossRef]
22. Lopatin, J.; Dolos, K.; Kattenborn, T.; Fassnacht, F.E. How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. *Remote Sens. Ecol. Conserv.* **2019**. [CrossRef]
23. Gao, B.C.; Montes, M.J.; Davis, C.O.; Goetz, A.F. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sens. Environ.* **2009**, *113*, S17–S24. [CrossRef]
24. Thompson, D.R.; Guanter, L.; Berk, A.; Gao, B.C.; Richter, R.; Schläpfer, D.; Thome, K.J. Retrieval of Atmospheric Parameters and Surface Reflectance from Visible and Shortwave Infrared Imaging Spectroscopy Data. *Surv. Geophys.* **2018**. [CrossRef]
25. Wagner, F.; Hérault, B.; Stahl, C.; Bonal, D.; Rossi, V. Modeling water availability for trees in tropical forests. *Agric. For. Meteorol.* **2011**, *151*, 1202–1213. [CrossRef]
26. Gourlet-Fleury, S.; Guehl, J.M.; Laroussinie, O. (Eds.) *Ecology and Management of a Neotropical Rainforest: Lessons Drawn From Paracou, a Long-Term Experimental Research Site in French Guiana*; Elsevier: Paris, France, 2004.
27. Richter, R.; Schlapfer, D. PARametric GEocoding: Orthorectification for Airborne Scanner Data. User Manual, Version 3.4; 2018. Available online: http://dev.rese.ch/software/parge/index.html (accessed on 1 April 2019).
28. Kang, X.; Li, S.; Benediktsson, J.A. Feature Extraction of Hyperspectral Images With Image Fusion and Recursive Filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3742–3752. [CrossRef]

29. Dalponte, M.; Ørka, H.O.; Ene, L.T.; Gobakken, T.; Næsset, E. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sens. Environ.* **2014**, *140*, 306–317. [CrossRef]
30. Schläpfer, D.; Borel, C.C.; Keller, J.; Itten, K.I. Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. *Remote Sens. Environ.* **1998**, *65*, 353–366. [CrossRef]
31. Schläpfer, D.; Richter, R. Atmospheric correction of imaging spectroscopy data using shadow-based quantification of aerosol scattering effects. *EARSeL eProc.* **2017**, *16*, 21.
32. Thomas, C.; Briottet, X.; Santer, R. Remote sensing of aerosols in urban areas from very high spatial resolution images: Application of the OSIS code to multispectral PELICAN airborne data. *Int. J. Remote Sens.* **2013**, *34*, 919–937. [CrossRef]
33. Schläpfer, D.; Hueni, A.; Richter, R. Cast Shadow Detection to Quantify the Aerosol Optical Thickness for Atmospheric Correction of High Spatial Resolution Optical Imagery. *Remote Sens.* **2018**, *10*, 200. [CrossRef]
34. Shen, X.; Cao, L. Tree-Species Classification in Subtropical Forests Using Airborne Hyperspectral and LiDAR Data. *Remote Sens.* **2017**, *9*, 1180. [CrossRef]
35. Fisher, R. Statistical Methods and Scientific Induction. *J. R. Stat. Soc. Ser. B (Methodol.)* **1955**, *17*, 69–78. [CrossRef]
36. Yu, H.; Yang, J. A direct LDA algorithm for high-dimensional data—With application to face recognition. *Pattern Recogit.* **2001**, *34*, 2067–2070. [CrossRef]
37. Venables, W.N.; Ripley, B.D.; Venables, W.N. *Modern Applied Statistics with S*, 4th ed.; Statistics and Computing; Springer: New York, NY, USA, 2002.
38. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2000. [CrossRef]
41. Ferreira, M.P.; Zortea, M.; Zanotta, D.C.; Shimabukuro, Y.E.; de Souza Filho, C.R. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sens. Environ.* **2016**, *179*, 66–78. [CrossRef]
42. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
43. Colgan, M.; Baldeck, C.; Féret, J.B.; Asner, G. Mapping Savanna Tree Species at Ecosystem Scales Using Support Vector Machine Classification and BRDF Correction on Airborne Hyperspectral and LiDAR Data. *Remote Sens.* **2012**, *4*, 3462–3480. [CrossRef]
44. Vidal, M.; Amigo, J.M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 138–148. [CrossRef]
45. Cohen, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213–220. [CrossRef]
46. Peerbhay, K.Y.; Mutanga, O.; Ismail, R. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 19–28. [CrossRef]
47. Große-Stoltenberg, A.; Hellmann, C.; Werner, C.; Oldeland, J.; Thiele, J. Evaluation of Continuous VNIR-SWIR Spectra versus Narrowband Hyperspectral Indices to Discriminate the Invasive Acacia longifolia within a Mediterranean Dune Ecosystem. *Remote Sens.* **2016**, *8*, 334. [CrossRef]
48. Shahriari Nia, M.; Wang, D.Z.; Bohlman, S.A.; Gader, P.; Graves, S.J.; Petrovic, M. Impact of atmospheric correction and image filtering on hyperspectral classification of tree species using support vector machine. *J. Appl. Remote Sens.* **2015**, *9*, 095990. [CrossRef]
49. Nevalainen, O.; Honkavaara, E.; Tuominen, S.; Viljanen, N.; Hakala, T.; Yu, X.; Hyyppä, J.; Saari, H.; Pölönen, I.; Imai, N.; Tommaselli, A. Individual Tree Detection and Classification with UAV-Based Photogrammetric Point Clouds and Hyperspectral Imaging. *Remote Sens.* **2017**, *9*, 185. [CrossRef]
50. De Sá, N.C.; Castro, P.; Carvalho, S.; Marchante, E.; López-Núñez, F.A.; Marchante, H. Mapping the Flowering of an Invasive Plant Using Unmanned Aerial Vehicles: Is There Potential for Biocontrol Monitoring? *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
51. Richter, R.; Müller, A. De-shadowing of satellite/airborne imagery. *Int. J. Remote Sens.* **2005**, *26*, 3137–3148. [CrossRef]

52. Adeline, K.; Chen, M.; Briottet, X.; Pang, S.; Paparoditis, N. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 21–38. [CrossRef]
53. Nagendra, H. Using remote sensing to assess biodiversity. *Int. J. Remote Sens.* **2001**, *22*, 2377–2400. [CrossRef]
54. Vanden Borre, J.; Spanhove, T.; Haest, B. Towards a Mature Age of Remote Sensing for Natura 2000 Habitat Conservation: Poor Method Transferability as a Prime Obstacle. In *The Roles of Remote Sensing in Nature Conservation*; Díaz-Delgado, R., Lucas, R., Hurford, C., Eds.; Springer: Cham, Switzerland, 2017; pp. 11–37. [CrossRef]
55. Valbuena, R.; Mauro, F.; Arjonilla, F.J.; Manzanera, J.A. Comparing airborne laser scanning-imagery fusion methods based on geometric accuracy in forested areas. *Remote Sens. Environ.* **2011**, *115*, 1942–1954. [CrossRef]
56. Große-Stoltenberg, A.; Hellmann, C.; Thiele, J.; Werner, C.; Oldeland, J. Early detection of GPP-related regime shifts after plant invasion by integrating imaging spectroscopy with airborne LiDAR. *Remote Sens. Environ.* **2018**, *209*, 780–792. [CrossRef]
57. Yao, W.; Krzystek, P.; Heurich, M. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sens. Environ.* **2012**, *123*, 368–380. [CrossRef]
58. Bruggisser, M.; Roncat, A.; Schaepman, M.E.; Morsdorf, F. Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification. *Remote Sens. Environ.* **2017**, *196*, 28–41. [CrossRef]
59. Féret, J.B.; Asner, G.P. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecol. Appl.* **2014**, *24*, 1289–1296. [CrossRef] [PubMed]
60. Vaglio Laurin, G.; Chen, Q.; Lindsell, J.A.; Coomes, D.A.; Frate, F.D.; Guerriero, L.; Pirotti, F.; Valentini, R. Above ground biomass estimation in an African tropical forest with LiDAR and hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 49–58. [CrossRef]
61. Ferraz, A.; Saatchi, S.; Mallet, C.; Meyer, V. LiDAR detection of individual tree size in tropical forests. *Remote Sens. Environ.* **2016**, *183*, 318–333. [CrossRef]
62. Tochon, G.; Féret, J.; Valero, S.; Martin, R.; Knapp, D.; Salembier, P.; Chanussot, J.; Asner, G. On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images. *Remote Sens. Environ.* **2015**, *159*, 318–331. [CrossRef]
63. Marvin, D.C.; Asner, G.P.; Schnitzer, S.A. Liana canopy cover mapped throughout a tropical forest with high-fidelity imaging spectroscopy. *Remote Sens. Environ.* **2016**, *176*, 98–106. [CrossRef]
64. Susaki, J.; Hara, K.; Kajiwara, K.; Honda, Y. Robust estimation of BRDF model parameters. *Remote Sens. Environ.* **2004**, *89*, 63–71. [CrossRef]

C. Conclusion of the chapter

This study is in agreement with previous studies, showing the feasibility of species identification. In our case, more than 20 species were successfully separated. However, we have highlighted the need for preprocessing in order to improve this identification. Comparison between corrected and uncorrected data for atmospheric effects revealed that at a site scale atmospheric corrections are not the most important corrections. The LDA classifier[1] maintained very good performance when developed with and applied to radiance spectra. However, when using new acquisitions, atmospheric corrections were necessary. We recommend to systematically apply atmospheric corrections in order to build a database of species spectral signatures. On the other hand, we found that the use of overlapping images (flight strips rather that mosaic) improved classifiers significantly.

In this first analysis, we used only hyperspectral data using the spectral range between 500nm and 1000nm. The spectral range between 1000 nm and 2500 nm was not used. This additional information is rich and has already enabled more efficient classifications to be made (Clark et al., 2005).

Another limitation of this work is that it discriminated 20 species amongst one another. However, there are hundereds of species in a single ha of tropical forest. The task of identifying trees in tropical forests among a great variety of plants is a much more complex challenge which will be dealt with in the next chapter.

[1] There was a mistake in the parameterizing of the LDA classifier in this study. The prior probabilities used were defined based on the species relative abundance (instead of setting them to a constant value). After noticing the error, we rerun the calculations which produced numerically diferent results but did not affect the results qualitatively.

VIII. Chapter 2: Specific identification of plant species in tropical environments in a hyper-diverse context

A. Introduction:

The identification of tropical forest tree species on spectral iagery is feasible. As shown in the previous chapter, the results are acceptable if we try to identify one species among twenty other species even when restricting the spectral range to 500 nm to 1000 nm.

In an operational configuration, the managers of Guyanese forest areas must identify a species from among a wide variety of flora. The problem is much more delicate, because the classifier must identify what discriminates a particular species from a whole set. To take a comparison builing on object recognition, the simple case would amount to finding the car brand among all cars and the complicated case would be to find the car brand among all motor vehicles. Thus, the objective is to find an approach which would make it possible to identify a species of commercial interest on the entire image acquired. To improve our chances of discriminating one species from the others, we used the entire spectral range (from 500 nm to 2500 nm). We test the efficacy of our classifiers by predicting stocking per psecies per plot in the Paracou experimental. For this step we used LiDAR based segmentation of individual crowns.

B. Article: Quantitative Airborne Inventories In Dense Tropical Forest Using Imaging Spectroscopy

remote sensing

MDPI

Article

Quantitative Airborne Inventories in Dense Tropical Forest Using Imaging Spectroscopy

Anthony Laybros [1,2,*] , Mélaine Aubry-Kientz [1] , Jean-Baptiste Féret [3] , Caroline Bedeau [2] ,Olivier Brunaux [2] , Géraldine Derroire [4] and Grégoire Vincent [1]

1 AMAP, IRD, CNRS, INRA, Université Montpellier, CIRAD, 34000 Montpellier, France; melaine.aubry-kientz@cirad.fr (M.A.-K.); gregoire.vincent@ird.fr (G.V.)

2 Pôle RDI, ONF Guyane, 97300 Cayenne, French Guiana, France; caroline.bedeau@onf.fr (C.B.); olivier.brunaux@onf.fr (O.B.)

3 TETIS, INRAE, University of Montpellier, 500 rue François Breton, 34093 Montpellier CEDEX 5, France; jean-baptiste.feret@teledetection.fr

4 Cirad, UMR EcoFoG (AgroParistech, CNRS, INRAE, Université des Antilles, Université de la Guyane), 97379 Kourou, French Guiana, France; geraldine.derroire@cirad.fr

* Correspondence: anthony.laybros@onf.fr

Received: 22 April 2020; Accepted: 12 May 2020; Published: 15 May 2020

Abstract: Tropical forests have exceptional floristic diversity, but their characterization remains incomplete, in part due to the resource intensity of in-situ assessments. Remote sensing technologies can provide valuable, cost-effective, large-scale insights. This study investigates the combined use of airborne LiDAR and imaging spectroscopy to map tree species at landscape scale in French Guiana. Binary classifiers were developed for each of 20 species using linear discriminant analysis (LDA), regularized discriminant analysis (RDA) and logistic regression (LR). Complementing visible and near infrared (VNIR) spectral bands with short wave infrared (SWIR) bands improved the mean average classification accuracy of the target species from 56.1% to 79.6%. Increasing the number of non-focal species decreased the success rate of target species identification. Classification performance was not significantly affected by impurity rates (confusion between assigned classes) in the non-focal class (up to 5% of bias), provided that an adequate criterion was used for adjusting threshold probability assignment. A limited number of crowns (30 crowns) in each species class was sufficient to retrieve correct labels effectively. Overall canopy area of target species was strongly correlated to their basal area over 118 ha at 1.5 ha resolution, indicating that operational application of the method is a realistic prospect (R^2 = 0.75 for six major commercial tree species).

Keywords: tropical forest; species diversity; hyperspectral; LiDAR

1. Introduction

Tropical forests are a major terrestrial plant biodiversity reservoir [1]. The preservation of this biome is therefore globally important. Persistent deforestation [2] is resulting in drastic declines in biodiversity. From 2000 to 2010, logging in natural forests removed approximately 5% of the world's forest area [3]. While the biggest drivers for biodiversity change [4] have been shown to be land use change and climate change, logging impact on the biodiversity of various taxonomic groups (mammals, birds, amphibians) is also well documented [5]. Logging impact is not restricted to the removal of a few commercial stems per ha, but also includes damage associated with opening tracks to access to the logging area. Untargeted trees can be wounded in the process of timber felling and hauling, contributing to carbon release [6] and possibly slowing forest recovery rate [7] by favouring fast-growing species such as lianas [8,9]. Logging may therefore profoundly modify the floristic

composition of the exploited plot [10] while disturbing wildlife [11,12]. Proper management of tropical forests is therefore crucial to mitigate the erosion of biodiversity [13].

Remote sensing technologies can provide valuable information for the preservation of tropical forests. Large-scale species distribution maps may provide valuable insights into species autecology and species population dynamics. Imaging spectroscopy has proven useful in monitoring the development of invasive species [14–16]. From a forest management perspective, evaluating abundance and spatial distribution of commercial species will help prioritize areas to be set aside for regeneration purposes and biodiversity conservation, or simply because there is not enough commercially valuable resource for logging operations. Accurate mapping of species prior to opening roads and tracks can therefore support the development of reduced impact logging techniques by minimizing the detrimental impact on soil and forest cover and contributing to protect and maintain forest diversity [17]. This technology is particularly well adapted to tropical forestry, since trees of commercial interest are typically large, upper canopy trees (dominant or emergent) but making up a small part of the entire tree community.

The efficacy of tree species discrimination in tropical forests via airborne imaging spectroscopy has been demonstrated [18–22]. At leaf scale, part of the SWIR1 domain (1467 nm to 1771 nm) was found to support accurate species identification. The SWIR2 domain (1994 nm to 2435 nm) was also found to be important in discriminating species at crown level [23]. A study conducted in semi-deciduous tropical forest [24] reported that including SWIR information (1045 nm to 2395 nm) improved their classification accuracy from 64.2% to 79.8%. While extending the spectral range increases the discriminative power of hyperspectral imagery, it has known drawbacks. Sensors limited to the VNIR domain are lower cost, more widespread, more easily miniaturized to mount on UAV platforms and have better intrinsic performance (higher signal-to-noise ratio). Hence, the balance between costs and benefits of increasing the spectral range should be assessed with respect to the specific task considered.

Improved accuracy has been reported from performing classification at the crown level as opposed to the pixel level [22,25,26]. However, segmentation of individual tree crowns (ITCs) in dense canopies remains a difficult task. Light detection and ranging (LiDAR) can help considerably. A recent benchmark study comparing crown segmentation methods using the LiDAR point cloud of a dense tropical canopy [27] identified AMS-3D (Adaptive Mean Shift in 3 Dimensions) as the most effective method. LiDAR and hyperspectral data fusion may further improve our ability to distinguish individual crowns [28].

Most studies in tropical forest so far have focused on discriminating a limited number of species against one another. In the present study, we are interested in retrieving a few target species amongst a very large number of non-focal species. To this end, one might consider focussing the annotation effort on the well-known more easily recognized focal species. This raises the question of the definition of the non-focal species class. A random sample of pixels of the area may capture the diversity of species in the background but will most likely include pixels of the focal species as well. Standard measures of accuracy such as F-measure then become inappropriate during the training stage due to this possible source of bias. A new unbiased criterion was proposed to replace the F-measure in such settings [29] It is based on the evaluation of the focal class (given by the recall metrics) divided by the probability that a sample is classified in the focal class among the non-focal class. This methodology was applied to identify five species on Barro Colorado Island (BCI) [19]. In the latter study, the authors did not specifically study the sensitivity of the classifier to the level of bias (impurity of the non-focal class).

Finally, we go beyond individual tree classification accuracy evaluation by exploring how well one can predict the cumulative basal area of commercial species per unit area from imaging spectroscopy to truly tackle the issue of quantitative forest resource mapping.

Our study starts by examining the benefits of using a SWIR sensor (1000 nm to 2500 nm) in conjunction with a VNIR sensor (400 nm to 1000 nm) for tree species discrimination. We then move on to the core of our study, which is to evaluate the potential of hyperspectral imagery in an operational setting. We first explore for 20 locally abundant species how much the retrieval rate may depend on the diversity of the set of species among which the target species are mixed. We compare the

performance of different binary classifiers developed to retrieve target species from wide views of the forest canopy after delineating individual tree crowns using LiDAR data. We further examine the impact of mislabeling of the focal and non-focal classes on the different classifiers' performance. We also briefly examine the impact of the size of the training set representing the species of interest. We finally compare hyperspectral predictions of basal area per species to a large ground inventory database.

2. Materials and Methods

1.1. Study Site

The study site of Paracou (5°18′N, 52°53′W) was established during the 1980s in French Guiana. The main objective of the experimental set-up was to document the impacts of different silvicultural practices on the dynamics of a tropical rainforest in order to provide guidelines for its sustainable exploitation [30]. Seventy-six 125 × 125 m plots (118.75 ha) are regularly monitored (stem girth and vital status being recorded). All trees with a diameter at breast height (DBH) higher than 10 cm have been identified to species level (or higher order taxon level if species was unknown). At the onset of the experiment, all the trees were geo-located using a meter tape after materializing a 12.5 × 12.5 m grid on the ground using strings. Tree location accuracy was estimated to be +/− 2 m.

1.2. General Methodology Outline

The general methodology is outlined in Figure 1. LiDAR data was used to segment crowns (Section 2.4.1) over the entire area of interest (118.75 ha). Manual segmentation (Section 2.3.3) and ground referencing was used to build a reference spectral signature database and species-specific allometric models. Inventory data (Section 2.4.2) were then paired with the automatically segmented (Section 2.4.1) crowns based on size and geolocation congruence.

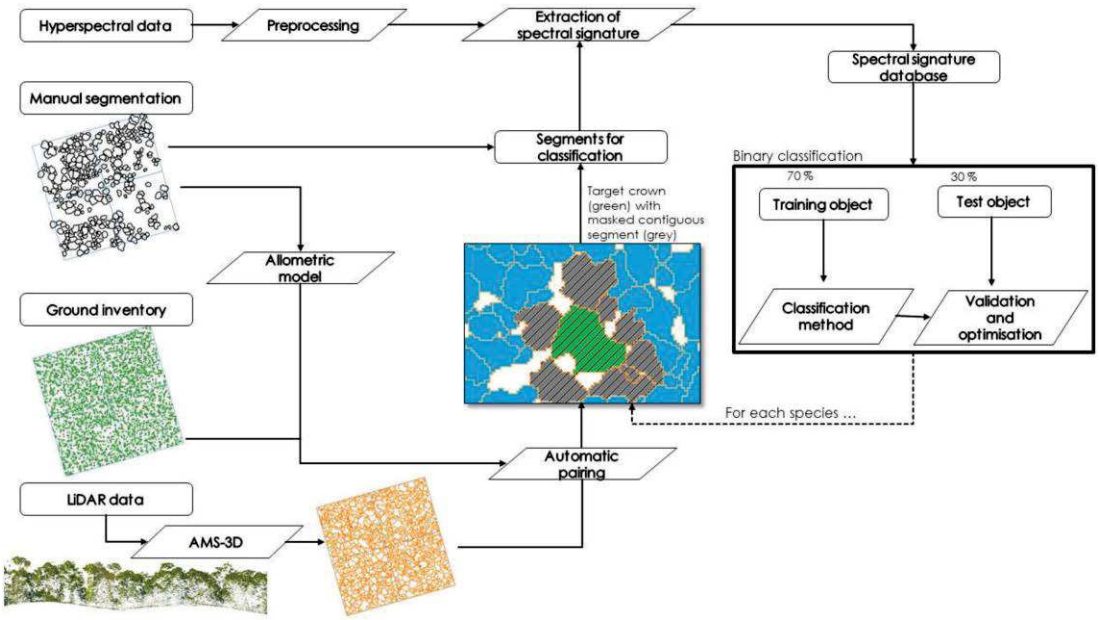

Figure 1. General methodological outline illustrated for a sample plot. The inventory data (green points) are paired with the automatically segmented crown (orange outline). Pairing is based on an allometric model developed from manually segmented crowns (black outline). In the segmented and colorized image, the green segment is one potential crown of the focal species; in grey, the neighbouring crowns that are also masked to make sure all pixels of the focal class are removed (striped segments) from the data set; in blue, the crowns preserved.

The spectral signature database was used to train the binary classifiers, developed for each focal species (Section 2.5). The automatically delineated and labeled segments were used to mask the focal species across the whole area of the permanent plots to ensure a high level of purity of the non-focal class. To ensure that the non-focal class contained as few pixels as possible from the focal class, we considered a buffer around the crowns determined by the automatic assignment as belonging to the focal class. All contiguous segments and the segment itself were excluded from the training and testing sets (Figure 1).

1.3. Remote Sensing Data

1.3.1. Hyperspectral Imaging

Imaging spectroscopy data was acquired with a Hyspex VNIR-1600 and a Hyspex SWIR-384 (Hyspex NEO, Skedsmokorset, Norway) sensor. The VNIR-1600 sensor covers the range from 414 nm to 1000 nm with 160 spectral bands and a spectral sampling distance of 3.64 nm. The SWIR-384 sensor covers the range from 1000 nm to 2500 nm with 288 spectral bands and a spectral sampling distance of 5.45 nm. The entire spectral range of the VNIR sensor was used. The information acquired in the SWIR domain is sensitive to atmospheric water absorption, so the spectral information corresponding to the low signal-to-noise ratio domains was discarded. Finally, the following spectral windows were kept: for the SWIR domain: 1009 nm to 1318 nm, 1454 nm to 1796 nm and 1964 nm to 2458 nm. Hence, 195 SWIR bands were retained. The King Air B200 airplane flew at an average altitude of 920 m. The flight took place on a cloudless day (from 15:00 to 17:00, solar time) on the 19th September 2016. Images were orthorectified and georeferenced with the PARGE software [31] using the canopy DSM (digital surface model) produced from the LiDAR point cloud (see next section). The DSM was created from the point cloud by selecting point of maximum height on a 1-m resolution grid. The spatial resolutions of the VNIR and SWIR images were 1 m and 2 m respectively. In order to merge the data without degrading the spatial resolution of the VNIR imagery, we resampled the SWIR imagery at 1 m using the nearest neighbor method. In addition, the fields of view of the two sensors are not identical (17° and 16°). We constrained the spectral information to the narrowest field of view corresponding to the one of the SWIR sensor. In the end, we obtained VSWIR imagery covering the entire spectrum (414 nm to 1963 nm) corresponding to 355 bands. Spectral information used to train the species classifiers was extracted from flight lines rather than from a mosaic, as it was previously shown to be more effective [22]. Although the training was carried out on the flight lines, the prediction was done on the mosaic where the pixels with viewing angle closest to vertical were kept ("center cropped" option of PARGE software [32]).

Atmospheric correction was applied using ATCOR-4 software [31]. The water vapor was retrieved using the APDA (atmospheric pre-corrected differential absorption) algorithm. The aerosol optical thickness (AOT) represents the amount of aerosols in the entire column of the atmosphere. It was considered constant. Non-vegetated pixels were identified based on NDVI (normalized difference vegetation index) thresholding, following recommendations for wavelength selection given in [33].

$$NDVI = \frac{\lambda_{774} - \lambda_{676}}{\lambda_{774} + \lambda_{676}} \quad (1)$$

Pixels with NDVI values higher than 0.4 were considered vegetation. Then, we applied a spatial mean filtering and a spectrum normalization by the mean as described in [22]. The illumination ratio was used to remove pixels with a scaled shadow fraction value higher than 0.6.

1.3.2. LiDAR Data and RGB Imagery

Airborne laser scanning data and RGB imagery were acquired during the same flight as the hyperspectral data. The LiDAR system was an LMS Q780 RIEGL operated with the following characteristics: a scan frequency of 400 kHz, a scan angle of +/−30°. Neighbouring strips overlapped

by 90%. Final mean pulse density was 33 pulses·m^{-2}. In addition, RGB (red, green, blue) imagery was acquired during the same flight using an iXu 180 Phase One camera. The final spatial resolution of the RGB orthorectified mosaic was 10 cm.

1.3.3. Ground Reference Data

A field survey was conducted to build a large ground reference dataset distributed over 56 of the 76 plots monitored on the study site (87.5 ha). Easily discernible crowns were first delineated manually based on the canopy height model (CHM), derived from LiDAR data with the help of the high-resolution RGB mosaic. The correct delineation of these ITCs was then validated in the field and the corresponding species ascertained. The 20 most abundant species in the delineated ITC dataset represented 1297 ITCs with a minimum of 24 individuals per species. Another 949 crowns were delineated, corresponding to 226 species. In total, the labeled database contained 2246 trees (Table 1). Most crowns were imaged twice by hyperspectral sensors because of the overlapping flight lines. The normalized mean spectra and its standard deviation for the most abundant tree species is illustrated in Figure 2.

Table 1. List of target species, and corresponding number of crowns and total number of pixels retained for analysis (each crown was imaged twice, but some pixels were removed due to low illumination or NDVI values). Sample representation refers to the proportion of manually segmented crowns of a given species. Mean crown area followed by standard deviation in brackets. Commercial species are underlined.

Species (Acronyms)	Number of Crowns	Number of Pixels	Mean Crown Area (m2) (SD)	Sample Representation (%)
Qualea rosea (Q.r.)	206	27,828	109.5 (59.4)	9.2
Pradosia cochlearia (P.c.)	164	38,349	142.3 (122.5)	7.3
Eschweilera sagotiana (E.s.)	139	12,559	49.1 (29.0)	6.2
Dicorynia guianensis (D.g.)	108	18,589	102.7 (66.8)	4.8
Eperua falcata (E.f.)	106	15,355	71.7 (41.3)	4.7
Eperua grandiflora (E.g.)	74	10,859	87.3 (46.2)	3.3
Recordoxylon speciosum (R.s.)	69	7944	69.6 (26.2)	3.1
Tachigali melinonii (T.m.)	51	6745	106.2 (67.1)	2.3
Couratari multiflora (C.m.)	49	4850	55.1 (33.8)	2.2
Licania alba (L.a.)	46	3894	47.0 (18.4)	2.0
Symphonia sp.1 (S.s.)	34	3355	50.2 (20.1)	1.5
Vouacapoua americana (V.a.)	34	3218	65.4 (34.0)	1.5
Sextonia rubra (S.r.)	32	4070	118.5 (99.3)	1.4
Tapura capitulifera (T.c.)	32	1224	30.5 (12.2)	1.4
Licania heteromorpha (L.h.)	27	1437	40.3 (21.7)	1.2
Moronobea coccinea (M.c.)	27	3355	68.8 (36.7)	1.2
Inga alba (I.a.)	26	3846	81.3 (58.7)	1.2
Goupia glabra (G.g.)	25	4998	133.7 (77.3)	1.1
Bocoa prouacensis (B.p.)	24	2375	54.9 (35.8)	1.1
Jacaranda copaia (J.c.)	24	1705	40.4 (22.7)	1.1
Others	949	84,713	80.8 (71.6)	42.3

In the 20 plots where no manual segmentation of ITC was done, 532 species were recorded for a total of 19,379 trees.

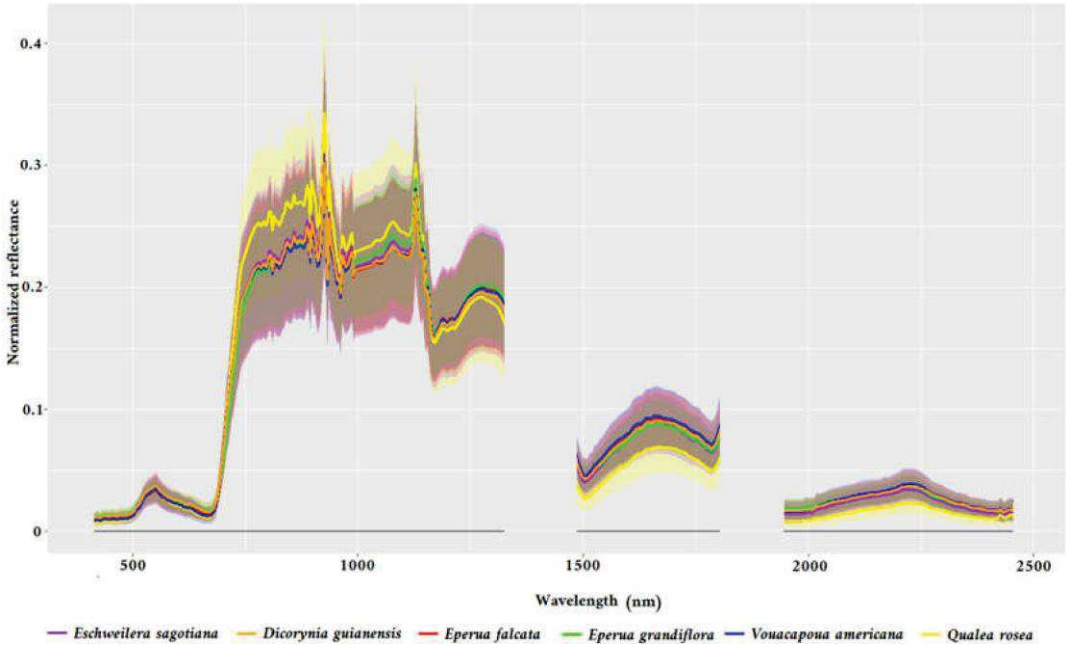

Figure 2. Normalized mean reflectance spectra (and standard deviation) for six focal species estimated on manually segmented crowns.

1.4. Automatic Segmentation Method

1.4.1. AMS-3D Method

We applied AMS-3D [34] to the LiDAR data for ITCs segmentation. The AMS-3D algorithm considers the point cloud as a multi-modal 3D distribution where each mode corresponds to the location of a tree crown. To find the modes, a Pollock function was used as bandwidth [35], with parameters changing with points' height to allow higher crowns to be bigger [34]. Points converging to the same mode were attributed the same ITC. The 3D clusters corresponding to each crown were then projected vertically to obtain polygons. The point cloud segmentation was conducted with the Computree platform (http://computree.onf.fr/), and the projection with the R package lidR [36].

1.4.2. Correspondence between ITC and Inventory Data

The automatically segmented crowns were then paired with the inventory data, using a pairing algorithm based on allometric relation and distance between the crown and the stem [27]. First, using data from the manually segmented crowns, we modelled each tree DBH as a function of crown diameter, height and taxon:

$$ln(DBH_i) = \alpha_{sp_i} + \beta_{sp_i} ln(H_i \times CD_i) + \varepsilon \quad (2)$$

$$\varepsilon \sim \mathcal{N}(0, \sigma_i^2) \quad (3)$$

where DBH_i is the DBH (cm) of tree i of species sp_i, H_i is its height, CD_i its crown diameter, and ε the error term. We then computed the distance between each manual segmented crown and all trunks of the inventory, computed the distance ranking of the trunk associated with the crowns, and fitted an exponential distribution on the ranks:

$$d_{i,j} = \alpha e^{-\alpha r_{i,j}} \quad (4)$$

where $r_{i,j}$ is the rank of tree j associated with crown i, and α is the parameter of the exponential distribution. Then, we developed an algorithm to optimally pair a crown with a tree from the inventory using the model previously parameterized. This algorithm takes two types of information into account:

(i) the distance between the crown center (computed as the 2D-centroid of crown projection) and the trunk location in the inventory should be close, and (ii) the tree should obey the allometric relation between the crown size and the trunk diameter. We considered those two rules to apply independently, neglecting a possible dependency of crown trunk distance on tree size. The pairing algorithm pairs each crown with the tree with highest:

$$p_{i,j} = d_{i,j}.g_{i,j} \quad (5)$$

where $g_{i,j}$ is the density function of the allometric model. Stems were ordered by decreasing DBH, and they were paired one after another in that order, with the segmented crown having the largest $d_{i,j}$ in the neighbourhood of the stem (15 m). When a crown was allocated to a stem, it was no longer available for others. Details of the algorithm can be found in [27].

1.5. Classification Method

We compared three classification methods for tree species identification: linear discriminant analysis (LDA), regularized discriminant analysis (RDA) and logistic regression (LR). We used the python language and the sklearn library [37]. The optimization of hyperparameters for the LR required considerable computing power. We used a Singularity container [38] and ran the code on a cluster of computers.

The LDA classification algorithm is well adapted to classification problems in high-dimension feature spaces provided that the number of observations is larger than the number of features in the least abundant class [39]. The principle of LDA consists in calculating new variables called discriminating canonical variables as linear combinations of the initial variables. It is calculated to maximize the ratio of the inter-group variance to the intra-group variance.

RDA is a compromise between LDA and quadratic discriminant analysis (QDA). QDA is a variant of LDA, the former being better adapted for non-linear separation. LDA assumes multivariate normal distribution of features with a common covariance matrix and different mean vector for each class. As shown by [40], LDA is not sensitive to the normal distribution assumption. Nevertheless, when the assumption of common covariance matrix is not satisfied, the individual covariance matrix for each group should preferably be estimated. This specific step characterizes QDA. The discriminating boundaries of the QDA are quadratic curves. The intermediate method between the two gives the RDA method proposed by [41] and makes it possible to regularize group covariance matrices, while preserving its performance in multi-dimensionality [42].

The assumptions of multivariate normality and equal variance–covariance matrices between groups are required before proceeding with LDA. On the contrary, in LR, these assumptions are not made. The authors of [43] showed in a bivariate case that in binary classification LR would be more efficient as soon as the normality assumption would be violated. A regularization term penalizing large values of the parameters was considered to avoid possible overfitting during the training step. A penalty parameter controlling the relative cost of errors occurring in the two classes was also tuned. We used a grid search to find the best combination of hyper-parameters C (regularization) and w_c (relative focal class weight) over the values C $\{e^0, e^1, e^2, \ldots, e^6\}$ and $w_c\{1,2, \ldots ,10\}$. In case of bias (impure non-focal class), we used the criterion of [29], otherwise we used the F-measure metric (see below).

1.5.1. Classifier Evaluation Criteria

To assess the reliability of the binary classifier, the F-measure was the chosen accuracy measurement. It relies on the estimation of the precision and recall of the class of interest. F-measure was used instead of accuracy due to the difference in proportion between the classes. The precision is the number of correct positive results divided by the number of all positive results returned by the classifier,

and the recall is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

$$\mathrm{F-measure} = \frac{2 \times \mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \quad (6)$$

$$\mathrm{Precision} = \frac{\sum \mathrm{TP}}{\sum (\mathrm{TP} + \mathrm{FP})} \quad (7)$$

$$\mathrm{Recall} = \frac{\sum \mathrm{TP}}{\sum (\mathrm{TP} + \mathrm{FN})} \quad (8)$$

TP, FP and FN refer to true positive, false positive and false negative, respectively.

For comparison between LDA, RDA and LR, we used the exact same randomly sampled training and testing sets.

1.5.2. Optimization Step

All classifiers used in this study output a probability for a pixel of belonging to the focal class, which is defined by the maximum likelihood. The threshold used for assignment to the focal class can be refined by using a specified criterion, to improve the classification performance. We used different criteria for the standard (unbiased) case and when the non-focal class was impure. In the standard case, we sought to maximize the pixel level F-measure. When dealing with the mislabeled non-focal class (i.e., when the non-focal class contained pixels of the focal class), true precision was unknown, so we sought to maximize the following quantity [29]:

$$\frac{Recall^2}{P(f(x) = 1)} \quad (9)$$

where the *Recall* is recall of the accuracy and $P[f(x) = 1]$ is the probability that a sample is classified in the focal class.

1.6. Experimental Set-Up

We present four different experiments. All classifiers considered in this study were binary classifiers; i.e., a particular focal species is classified against a large set of non-focal species. As a rule, we separated the data into 70% of the crowns used for the training set and 30% for the test set (Figure 3) repeating the operation 30 times. We provide the mean and the standard deviation of the F-measure. We first compared classification performances obtained when using spectral information from either VNIR or VSWIR on the 20 most abundant species only (Experiment 1). Then, we explored how the performance of the classifier would be affected by gradually increasing the number of species among which our focal species was retrieved (Experiment 2). In a third experiment, we explored the sensitivity of the classifiers to different types of errors affecting the training data (labeling errors) (Figure 3, Experiment 3). In a final test (Experiment 4), we predicted the basal area of focal species per 1.56 ha plot from the hyperspectral data.

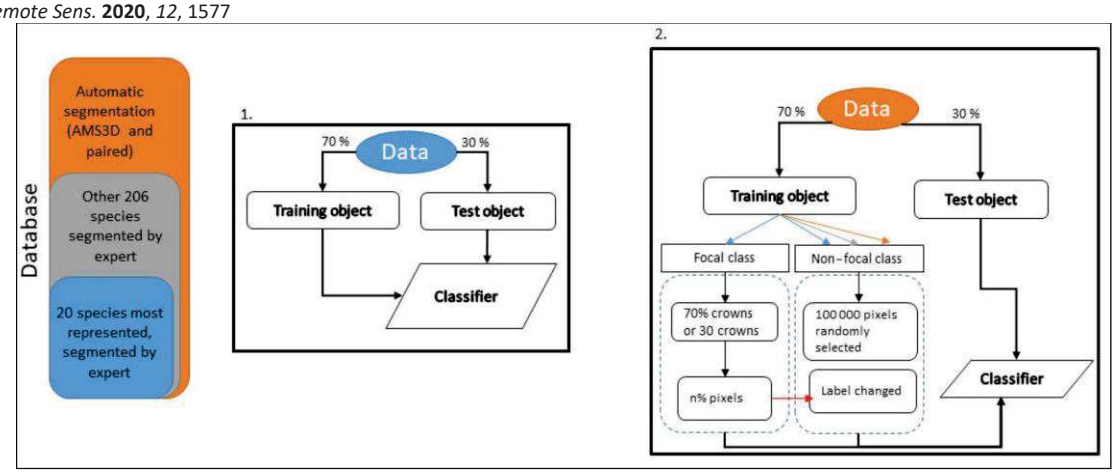

Figure 3. Nested datasets are used as illustrated in the leftmost chart. Flowchart 1 refers to Experiment 1. Flowchart 2 illustrates how classifiers were penalized by contamination of the non-focal class (Experiment 3).

1.6.1. Experiment 1—VNIR versus VNIR+SWIR

We assessed the effectiveness of adding SWIR bands to the VNIR spectrum. We used manual segmentation data and we selected the 20 most abundant tree species (Figure 3, 1). The same pixels were used to build training and test sets (70%–30%) for VNIR and VSWIR data. For this first experiment, none of the three classification methods were subject to secondary optimization.

1.6.2. Experiment 2—Increasing Background Spectral Diversity

The same 20 focal species were considered for this experiment, but we gradually added more species to the non-focal class. We first added the crowns of the most abundant species present in the manually segmented crowns database, up to 99 different species. We then added all the remaining crowns from the manual segmentation database (each crown being the sole representative of its species) in arbitrary order. Thereby, the 364 different species were opposed to each of the 20 species alternately. For this first experiment, LDA was used, without an optimization step.

1.6.3. Experiment 3—Effect of Noise in the Training Data

In an operational case, the inventory effort will likely be focused on few target species. Thus, the other non-focal species will not be systematically inventoried. Therefore, to train the classifier, the non-focal species class will be represented by a random selection of pixels over the entire study area (excluding crowns known to belong to the focal species). However, the probability for a pixel of the focal class of being selected and contributing to the non-focal class is dependent on the distribution of this species in the study area, viewed at canopy level.

To assess the impact of a given level of mislabeling of the non-focal pixels, we artificially biased the learning data set (Figure 3, 2). We randomly selected 100,000 pixels from the background, excluding the focal species to create the training set for the non-focal species. We simulated three level of bias in the training data by successively replacing 1%, 2% and 5% of the pixels labeled as belonging to the non-focal class with pixels taken from crowns of the focal species class.

We also explored the potential impact of mislabeling of the focal pixels, reckoning that such errors may occur regardless of the efforts made in the field (identification errors, matching errors, etc.). We used Mahalanobis distance to detect outliers among the 20 most abundant species in the manually segmented crown database. For each species, we computed this multi-dimensional distance for each pixel, we then computed the mean Mahalanobis distance per crown. Crowns beyond two standard deviations were considered as pseudo-outliers and removed from the database. Binary

classifications using all the information (including pseudo outliers) were then compared to those excluding pseudo-outliers.

Mahalanobis distance is a multi-dimensional distance measure between a point and a distribution. This distance $d(x)$ can also be defined as a dissimilarity measure between a vector and a mean vector from the same distribution:

$$d(\vec{x}) = \sqrt{(\vec{x} - \vec{u})^T \sum^{-1} (\vec{x} - \vec{u})} \quad (10)$$

where x is a spectrum, u the mean spectrum, Σ is the variance–covariance matrix.

1.6.4. Experiment 4—Predicting Basal Area per Species per Plot

The ground inventory data comprised 76 plots of 1.56 hectares. For each plot, we compared the canopy surface predicted per species (at crown and pixel level) by the best-performing classification method and the basal area of the species using a linear regression. The learning step used all the information on the focal species which was available in the manually segmented crown database to build the focal class and 100,000 pixels randomly selected from the same 56 plots. Hence, 20 plots (c. 25% of the area) were set aside to be used for independent validation. The learning step used 100,000 pixels from flight lines, while the prediction was realized on the mosaic.

We compared prediction of the three classifiers applied with and without an optimization step and with or without majority vote at crown level. The majority vote rule was applied to decide on the species label to be assigned to ITCs, which corresponded to a filling rate higher than 50% of the pixels. We present the results obtained with the LDA only. The basal area considered here is the basal area of all trees with DBH greater than 10cm. It may be noted that smallest trees will not be visible in the areal imagery.

3. Results

1.7. Experiment 1

The combination of VNIR and SWIR information strongly improved the performances of all classifiers compared to VNIR information only, with an increase in F-measure of 23.8%, 27.3% and 19.4% for LDA, RDA and LR respectively. The contribution of SWIR was very important for all species. The difference between LDA and RDA was not significant in this case. All species were rather well identified (F-measure never lower than 59%). LR did not outperform LDA nor RDA.

Moderate differences in performance were observed between classifiers when only the VNIR information was used. LR performed slightly better than the other two methods except for *Eschweilera sagotiana* (70% vs. 79%) and RDA appeared to perform slightly worse, particularly at pixel level (Appendix A). The change from pixel to object level was confirmed to be very beneficial. LDA gained the most from the majority vote, with an increase of 15.5% in F-measure using the complete VSWIR information (Table 2).

Table 2. Comparison of the F-measure at object level (and standard deviation) obtained after classification with the three methods when either VNIR or VSWIR information are used. Commercial species are underlined. The best prediction for each species with either VNIR or VSWIR information is in bold.

	VNIR			VSWIR		
Species	**LDA**	**RDA**	**LR**	**LDA**	**RDA**	**LR**
B.p.	25.4 (±11.6)	22.3 (±8.5)	**26.4** (±7.2)	66.9 (±11.3)	**83.3** (±10.4)	61.0 (±18.0)
C.m.	66.7 (±13.2)	64.8 (±13.5)	**67.5** (±11.0)	75.0 (±10.7)	67.9 (±9.9)	**77.0** (±8.4)
D.g.	61.2 (±4.5)	61.3 (±4.7)	**69.2** (±7.0)	88.6 (±2.5)	86.3 (±3.6)	**90.3** (±2.4)
E.f.	**46.9** (±5.5)	7.3 (±2.5)	29.6 (±9.6)	70.0 (±5.8)	72.9 (±8.2)	**73.9** (±6.7)
E.g.	63.1 (±6.4)	61.8 (±7.4)	**79.4** (±5.0)	82.1 (±5.3)	87.6 (±4.4)	**89.7** (±3.6)

E.s.	**79.0** (±3.3)	72.8 (±5.3)	70.6 (±7.3)	**89.6** (±2.8)	89.3 (±1.9)	86.4 (±2.1)
G.g.	44.3 (±8.9)	63.3 (±12.1)	**67.6** (±12.9)	**84.1** (±7.3)	83.9 (±12.0)	80.4 (±12.7)
I.a.	44.4 (±14.2)	49.4 (±15.6)	**62.7** (±17.2)	**77.6** (±10.0)	76.3 (±8.5)	66.0 (±16.0)
J.c.	58.4 (±16.7)	**59.2** (±12.6)	57.1 (±16.6)	59.2 (±19.1)	58.2 (±18.1)	**70.2** (±16.8)
L.a.	55.9 (±10.0)	**62.9** (±10.7)	62.6 (±11.4)	78.1 (±4.5)	**79.6** (±9.5)	74.6 (±9.1)
L.h.	**22.2** (±8.0)	16.7 (±0.0)	20.0 (±0.0)	**65.2** (±11.3)	60.7 (±13.7)	49.0 (±14.6)
M.c.	**64.5** (±11.4)	52.7 (±14.3)	64.1 (±13.3)	**82.8** (±10.9)	74.7 (±17.4)	79.7 (±12.7)
P.c.	80.5 (±4.2)	83.6 (±3.8)	**88.4** (±3.3)	93.5 (±1.9)	93.6 (±1.9)	**94.1** (±1.6)
Q.r.	94.8 (±1.6)	94.6 (±1.5)	**95.6** (±1.2)	96.9 (±1.1)	96.2 (±1.4)	**97.2** (±1.1)
R.s.	**84.9** (±6.0)	81.4 (±5.0)	**90.0** (±4.7)	**91.4** (±3.3)	87.1 (±6.1)	90.6 (±3.8)
S.r.	**55.0** (±9.2)	50.6 (±14.5)	53.7 (±13.2)	77.6 (±11.5)	82.9 (±8.4)	**85.2** (±7.3)
S.s.	**24.0** (±0.0)	0.0 (±0.0)	11.3 (±0.0)	64.0 (±11.7)	**68.9** (±12.1)	50.5 (±13.7)
T.m.	73.3 (±7.7)	79.2 (±6.0)	**84.1** (±5.6)	**93.5** (±4.1)	93.3 (±4.3)	93.4 (±4.1)
T.c.	35.4 (±8.7)	17.1 (±4.0)	**38.2** (±14.1)	82.6 (±8.9)	72.7 (±11.8)	**83.5** (±9.8)
V.a.	48.7 (±13.8)	**52.8** (±13.7)	43.0 (±14.7)	**85.1** (±8.3)	83.8 (±10.0)	77.1 (±10.1)
Mean F-measue	56.4	52.7	**59.1**	**80.2**	80.0	78.5

1.8. Experiment 2

We gradually added crowns from new species to increase the spectral variability of the non-focal class. We used an LDA binary classifier to retrieve each of the 20 most abundant species. No optimization was applied. We present the results at object level (following majority vote).

The overall trend in Figure 4 highlights that increasing the number of species degraded the classifier's performance. Some species were highly affected (e.g., *Eschweilera sagotiana*); others were only slightly affected (e.g., *Qualea rosea*). Increasing the number of species from 99 to 364 significantly affected discrimination of all target species. F-measure for *Qualea rosea*, the best-identified species, decreased from 88.6% to 84.9%, a loss of about 4% points.

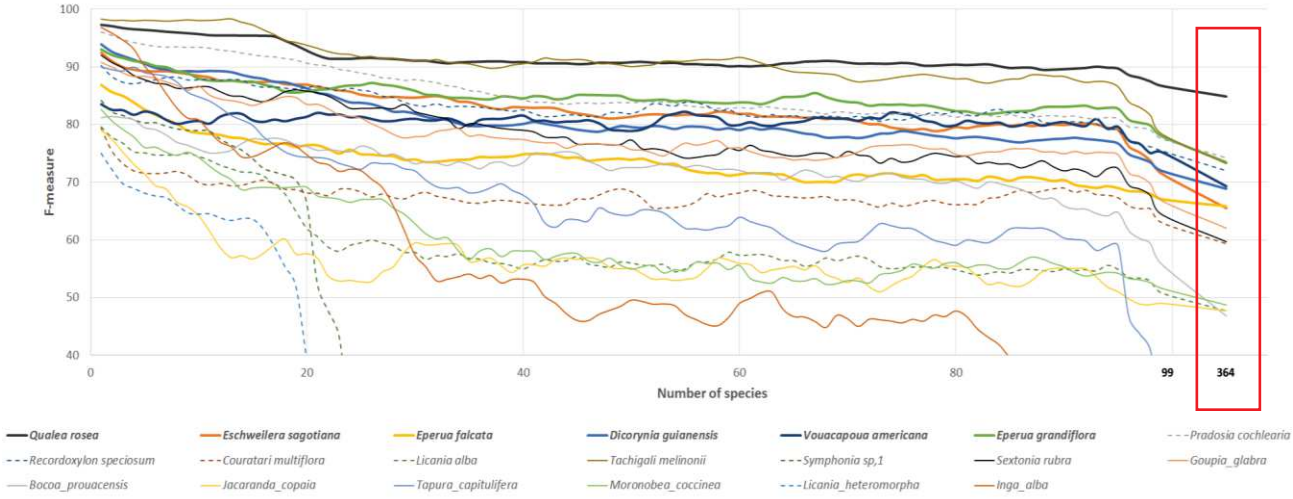

Figure 4. Evolution of species discrimination by successive addition of new species. F-measure value is smoothed using a five-step moving average, except after the 99th species when all remaining species were added at once (red box). Commercially valuable species are plotted in bold line.

The F-measure value for the species *Licania alba*, *Licania heteromorpha* and *Inga alba* declined sharply with the added species. In these cases, the focal species was well retrieved (the recall was higher than 70%), but the precision was very low (a lot of pixels were wrongly predicted to belong to the focal class).

1.9. Experiment 3 –Exploring the Impact of Training Set Impurity

To evaluate the impact on the classification performance of mislabeled pixels in the non-focal class, we applied different levels of contamination bias (1%, 2% and 5%).

1.9.1. Low Level of Impurity (All Species)

We first examined to what degree the different classification methods were affected by a low level of impurity of the non-focal class, and whether optimizing the pixel threshold based on the criterion proposed by [29] could mitigate this bias.

A noise of 1% had no significant effect on the identification of species, whether an optimization step was used or not (Table 3). When no optimisation was applied, LDA was outperformed by the other two methods for most of the species. For some species, LR was more efficient (e.g., *Pradosia cochlearia*) while for other species (e.g., *Sextonia rubra*) RDA performed best. Once optimized, all methods performed on a par in this experiment.

Table 3. Object level F-measure (and its standard deviation in brackets) of species binary classifiers trained with 1% mislabeling of the non-focal class. Plain and optimized indicate whether an optimization step (selecting a threshold probability value) was applied to maximize either the F-measure (0% contamination) or the Lee and Liu 2003 criterion (1% contamination). Standard deviation is given in Appendix B. The best prediction for each species and method is in bold.

Bias	0%						1%					
	Plain			Optimized			Plain			Optimized		
Species	LDA	RDA	LR	LDA	RDA	LR	LDA	RDA	LR	LDA	RDA	LR
B.p.	35.8 (±10.3)	**57.7** (±12.4)	37.5 (±15.5)	**65.4**	57.5	42.1	36.4	**58.5**	43.4	**64.3**	56.2	43.0
C.m.	**68.3** (±10.8)	64.7 (±10.1)	68.0 (±10.5)	69.8	68.3	**71.4**	**67.4**	65.1	66.5	**69.8**	68.0	69.4
D.g.	57.7 (±3.9)	72.3 (±6.0)	**78.2** (±6.1)	75.8	72.0	**78.5**	57.5	72.6	**78.2**	74.8	72.2	**78.3**
E.f.	50.8 (±4.1)	67.7 (±6.1)	**69.2** (±6.2)	**71.5**	68.9	70.5	51.5	67.9	**68.0**	**72.1**	69.2	70.5
E.g.	65.1 (±7.0)	82.0 (±5.3)	**84.4** (±4.6)	82.4	82.7	**84.8**	64.9	81.9	**82.5**	82.0	**82.5**	84.7
E.s.	68.1 (±4.2)	**74.2** (±4.1)	71.0 (±5.2)	74.7	74.1	**75.9**	68.5	**74.7**	69.2	74.1	73.4	**74.9**
G.g.	62.5 (±12.6)	73.4 (±11.5)	**76.4** (±12.5)	74.5	73.5	**77.3**	62.8	72.6	**74.8**	74.5	74.4	**75.0**
I.a.	38.3 (±11.5)	**42.3** (±12.7)	38.5 (±11.6)	**43.0**	41.0	41.4	38.9	**42.7**	36.6	**43.0**	41.4	40.9
J.c.	50.7 (±19.7)	51.4 (±16.8)	**52.4** (±17.1)	50.9	49.3	**53.1**	52.2	49.0	**51.6**	51.5	50.4	**52.6**
L.a.	41.6 (±8.7)	**56.3** (±10.8)	53.9 (±10.5)	56.6	54.4	**62.3**	41.9	**56.8**	48.5	57.0	54.7	**59.7**
L.h.	25.6 (±8.9)	**43.6** (±11.0)	22.8 (±8.1)	**44.0**	34.2	25.5	26.5	**43.5**	23.5	**46.6**	33.7	26.0
M.c.	58.5 (±13.8)	55.5 (±16.9)	**60.0** (±14.5)	56.3	54.9	**62.7**	**58.6**	56.1	56.5	55.4	56.2	**61.7**
P.c.	69.5 (±4.5)	77.1 (±4.9)	**85.9** (±3.6)	77.6	77.0	**84.3**	69.6	78.0	**85.8**	78.2	77.4	**84.0**
Q.r.	86.6 (±2.5)	88.1 (±2.7)	**91.3** (±2.3)	86.4	85.8	**90.7**	86.9	88.1	**91.2**	86.6	86.2	**90.8**
R.s.	**82.5** (±3.8)	78.8 (±4.5)	81.2 (±5.2)	81.0	80.7	**84.9**	**82.4**	79.3	79.5	80.7	79.7	**83.4**
S.r.	54.3 (±9.2)	**70.8** (±10.7)	61.6 (±14.7)	**71.8**	66.9	68.0	53.8	**69.7**	52.3	**71.3**	66.4	61.1
S.s.	29.1 (±5.9)	**36.7** (±9.1)	21.3 (±7.1)	36.6	**36.7**	25.6	29.1	**35.8**	21.3	35.5	**35.8**	25.6
T.m.	64.7 (±6.4)	69.9 (±6.0)	**80.7** (±9.4)	76.3	75.8	**78.8**	64.9	70.4	**78.2**	76.0	75.6	**78.9**
T.c.	58.3 (±11.5)	65.4 (±12.5)	**72.2** (±14.8)	65.8	52.1	**72.5**	58.2	65.4	**69.6**	63.6	51.0	**70.6**
V.a.	68.2 (±8.7)	**72.8** (±12.2)	65.1 (±12.8)	**74.4**	**74.4**	66.9	66.5	**74.1**	63.9	**74.4**	73.6	64.6
Average	56.8	**65.0**	63.6	**66.7**	64.0	65.9	56.9	**65.1**	62.0	**66.6**	63.9	64.8

1.9.2. High Level of Impurity (Most Abundant Species)

This analysis was restricted to the five most abundant species in order to explore the e**ff**ect of high rates of contamination of the non-focal species, which required having su**ffi**cient focal pixels. Results are summarized in Figure 5.

A labeling error of 5% did not a**ff**ect LDA, with or without an optimization step. However, the optimization step was very beneficial in both cases. RDA was mildly a**ff**ected by labeling error and not improved by optimization. LR classification method was much more impacted by a 5% error level and the optimization only partially compensated for this. The average loss in F-measure was 11.8% (without optimization) and 6.7% (with optimization). More complete results on the average F-measure, with or without an optimization step are provided in Appendix B, including results obtained for 2% bias.

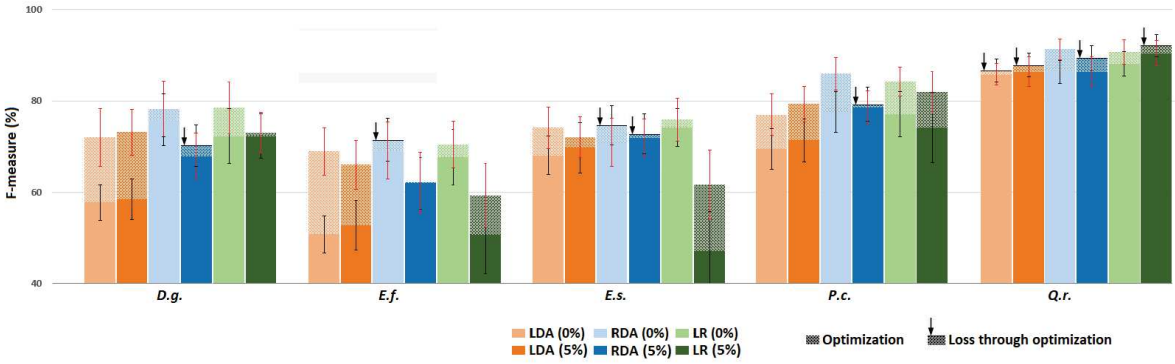

Figure 5. Object level F-measure (and its standard deviation) of species binary classifier trained with either 0% or 5% mislabeling of non-focal class. Paler color represents performance after applying an optimization step (see text). An arrow indicates cases when optimization decreased performance.

In summary, RDA appeared to be robust to high noise level even without optimization. LR was sensitive to noise even after including a modified optimization step criterion [29]. LDA performed very well, provided that an optimization step was included.

1.9.3. High Level of Impurity—Smaller Focal Training Class

To assess the impact of the size of the focal class on classification performance we reduced the number of crowns to 30 in the focal class while still considering 100,000 pixels randomly selected for the non-focal class. Only the five most abundant species were considered. Miss-labeled pixels were added to the non-focal class to achieve different levels of contamination (1%, 2% and 5%). Results were almost undistinguishable from those obtained with larger training sets for the same level of bias (see Appendix C).

1.9.4. Focal Class Purification

Mean crown Mahalanobis distance (computed across pixels within species) was used to try to detect outliers (Figure 6). Only *Licania alba* and *Bocoa prouacensis* had no outliers (outliers being defined as crowns showing a mean Mahalanobis distance greater than two standard deviations from the mean of all crowns). The most heavily impacted species were *Jacaranda copaia* and *Goupia glabra* (see table in Appendix D). Overall, for the 1297 representatives in the 20 focal classes, 55 crowns were potential outliers. There was thus a global potential bias of the focal classes of 4.2% or 2844 unreliable spectra.

We examined all those potential outliers on high-resolution RGB imagery. In a few cases, those outliers could be related to particular features (illustrated in Appendix D): low illumination leaving very few pixels per crown after filtering, higher than average contribution of wood caused by partial leaflessness (especially for *Eperua falcata*), presence of liana in the crown. However, in 90% of the cases, no peculiarity was detected on high-resolution imagery. One cannot exclude that crown mislabeling occurred, but this could not be ascertained.

To evaluate the improvement achieved after removing potential outlier crowns, a Student test by species was performed on each metric at object level for the three classifiers (without the optimization step). We compared those metrics before and after removal of the potential outliers. The most impacted species were the *Couratari multiflora* (from 68% to 71.6% using LR) and *Sextonia rubra* (from 61.6% from 64.7% using LR) (p-value <0.01). *Recordoxylon speciosum* (from 81.2% to 83.8% using LR) and *Jacaranda copaia* (from 50.7% to 59.2% using LDA) were less impacted (p-value **<**0.05). The RDA method was the method least sensitive to the introduction of outliers. For more detail, see the table in Appendix D.

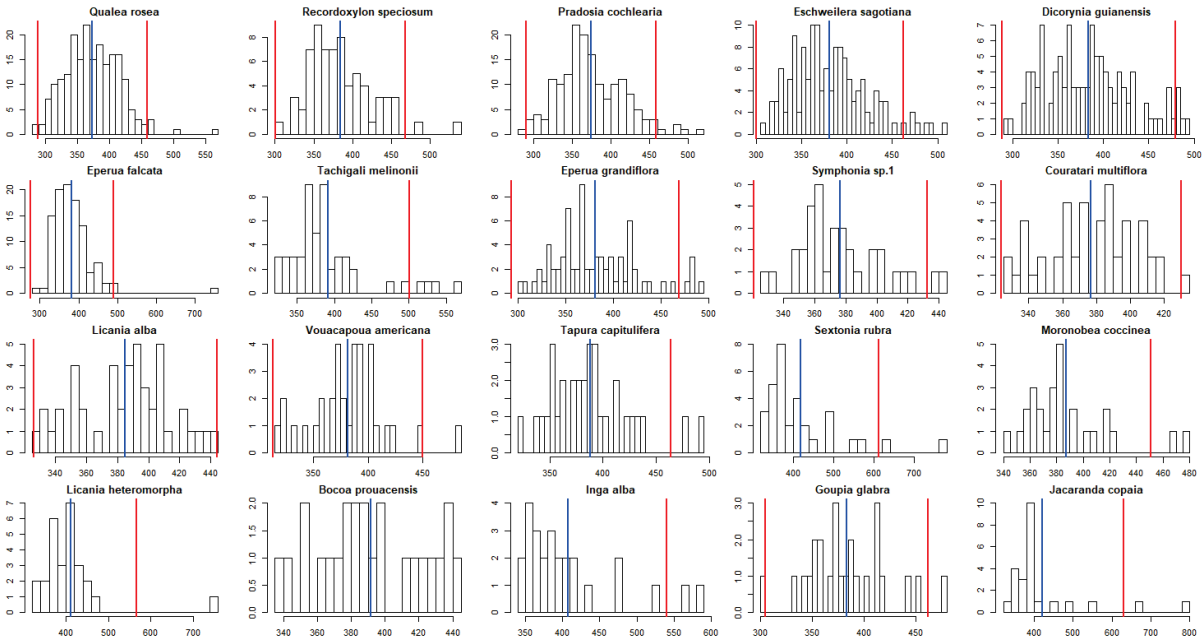

Figure 6. Within-species Mahalanobis distance (computed per crown). The blue line represents the mean of the Mahalanobis distance. The red lines represent the mean plus or minus two standard deviations.

1.10. Experiment 4

We related the recorded basal area per species per plot with the canopy area predicted to belong to the same species by the different classifiers. We used all the available crowns in our manual segments database to train the classifiers. For each focal species, we randomly selected 100,000 pixels to create the non-focal class for training the classifier. Results for LDA are reported in Figure 7 for the most abundant commercial species. Results for all other species are provided in Appendix E, Figure A3.

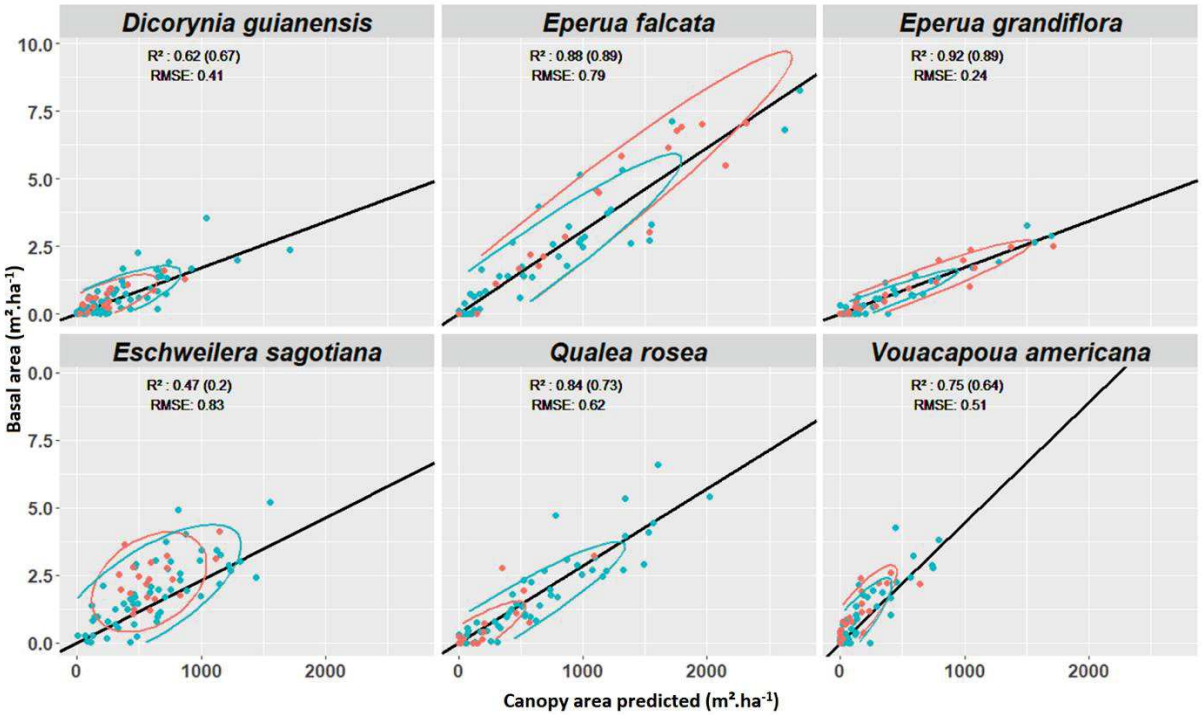

Figure 7. Relation between basal areal ($m^2 \cdot ha^{-1}$) and area predicted ($m^2 \cdot ha^{-1}$) using the optimized LDA. Plots not used for training are plotted in red and the R-squared for those observations given in brackets. The plot used for the training are in blue. Global R-squared values are given using all points (red and blue). In solid black line, linear regression based on all points passing through zero. The normal confidence ellipses are given, computed for each group points using the corresponding color (red or blue).

As expected, the effectiveness of the classifier conditions the correlation between the predicted canopy area and the basal area. The mean linear correlation coefficient across species was 0.70 (n = 20; Appendix F). Optimized LDA and RDA were the methods that provided the best average R^2 both at pixel and object level (not shown).

In addition, we compared the slope of the relations across species. We used a linear model without intercept. While species effect was significant overall ($p < 0.01$, e.g., *Eperua grandiflora* vs. *Eperua falcata*), some species shared a very similar slope (e.g., *Qualea rosea* vs. *Vouacapoua americana*). It might therefore be possible to use a common model between these species

Object-based prediction was better correlated to basal area than pixel-based prediction (4–8 point gain in R^2) for all classifiers. Less than 2% of the pixels were classified into more than one species for both LDA and RDA. Hence, the confusion was low, despite the fact that the species-specific classifiers were developed independently. After majority voting, no confusion remained.

4. Discussion

The first experiment unambiguously confirmed that addition of SWIR information vastly improves classification performance. Our results on the contribution of SWIR information to species identification are in line with those obtained in previous studies [23,26,44]. The authors of [23] found that the LDA classification accuracy of seven tropical forest tree species was improved from 67.3% to 83.6%. Similarly, we found an increase in accuracy of about 20%–25% on average across species depending on the method considered. We noted that the gain was more important for LDA and RDA than for LR. In addition, we assessed the quality of our database. In fact, our database may contain a few mislabeled cases, as suggested by the Mahalanobis distance analysis. However, these outliers had a limited negative effect on classification in general. It is unclear how many of these pseudo-outliers are actual outliers. Phenology or presence of lianas was not found to be significantly associated with potential outliers.

The comparison of the results obtained in this study (Table 3, without bias) with previously published results (Table 4 in Laybros et al., 2019) indicates that there is no significant difference in performance between the binary and multiclass LDA classifier at pixel level. However, at the object level, the binary classifiers are clearly outperformed, implying a strong influence of the majority vote. The relative majority can be effective in a multi-class problem, whereas the absolute majority is required when only two classes are considered.

An important characteristic of our study case is the unbalanced number of spectra (pixels) in each class. The minority classes are the most affected by this constraint. Unbalanced classes impact results obtained with SVM classifications [45]. In some studies, the authors used SVM without a specific method to control for this effect [20,21,24,26,46]. This method is, however, sensitive to unbalanced classes [47–50]. Biasing the classifier by adding weights and breaking down the penalization of the classifiers (parameter C) into two parameters was recommended by [19,51] in order to reduce the negative impact of the minority classes. Conversely, LDA is not sensitive to the number of elements per class [52] because it shares the variance–covariation matrix between the classes. In addition, training on the entire data set or a subset of only 30 crowns did not affect the performance of the classifier. The similar results between the LDA and the RDA methods seem to confirm that RDA is not sensitive to the disproportion between classes either.

It is preferable to optimize a specific metric in the biased cases and especially with a high bias rate. F-measure is not a good choice, because the precision is biased to an unknown extent. The use of the criterion proposed by [29] proved useful. In addition, when the bias is very severe, it is recommended to use a threshold that will optimize discrimination [49]. In the present study, LR was found to benefit much from this methodology. Conversely, the optimal probability threshold did not improve species discrimination for RDA.

The maximum level of bias tested here was 5%. How likely are larger biases to occur? In an operational case, the random selection of pixels on the study area will comprise a range of species

reflecting the local flora of the study site. Despite the extremely high local tree species diversity (alpha diversity) in tropical forests [53], some species are spatially aggregated as a result of either environmental preference or dispersal limitation. For example, the *Eperua falcata* measured in the Piste de St Elie region, is a species with an aggregated spatial distribution [54]. *Vouacapoua americana* and *Qualea rosea* species were also reported as being highly aggregated [55,56]. Extreme cases of aggregation may lead to monodominance [57], where more than 50% of the stems belong to a single species (e.g., *Spirotropis longifolia*). This, however, is a rare case and local frequency of the focal species will seldom reach 5% [58].

Most previous studies focused on discriminating a few well-known species against each other, which is of little relevance to an operational context. Here, we focused on assessing the potential of classification methods using hyperspectral information for retrieving particular species in a diverse canopy. This problem was previously addressed by [19], who highlighted the feasibility of identifying four tree species in a tropical forest canopy using biased SVM on the BCI site. As reported by [59], the taxonomic diversity in BCI is lower than in Paracou, whatever the diversity considered (alpha, beta or gamma diversity). On the BCI site, 225 tree species have been counted among more than 21,457 stems for an area of 50 hectares. On the Paracou site, 229 tree species were listed among 1124 stems for an area of 2 hectares. Therefore, the identification of focal tree species may be more challenging in the context of the present study than in the [19] study. As shown in Experiment 1, the performance of the classifiers tends to decrease with increasing diversity of the background.

The size of the training data set can be critical in an operational setting. The minimum number of crowns to be used for training in tropical canopies was investigated by [18,45]. Thirty crowns were found to be reasonable for species retrieval against 100,000 pixels. In the present study reducing the number of focal crowns from 74-144 to 30 crowns did not entail any loss in F-measure (comparison between Table 3 and Figure A1). However, the minimum training set size may depend on the species [18] and the diversity of the canopy. For the same number of crowns in the learning set (34 crowns, Table 1), the F-measure for *Vouacapoua americana* reached of 72.8% while it was only 36.7% for *Symphonia sp.1* (Table 3, RDA). In addition, because classification results were so variable across species (Figure 4), we examined correlations between the predictability of the species (F-measure), the size of the focal species training set, the intra-specific spectral variance—sensu [60]—and the dispersion of the Mahalanobis distance values. The only significant correlation detected was with the size of the training set (number of crowns or number pixels) highlighting that the detectability of a species is difficult to predict (Appendix F, Table A5).

Forest managers need to quantify the resource of the forest in order to orient the choices to be made. We found the relation between basal area and crown area in the canopy to vary across species. The authors of [61] mention that a global allometric model can predict the diameter of the crown from the diameter of the stem. These estimates are based on more than 108,753 stems with different geographic locations, consisting of 1,492 tree species. However, this general relation is only valid at coarse grain, and studies comparing species occurring in a restricted geographical range repeatedly found marked inter-species differences. This was the case for Indian monsoon forest [62] and Sumatran peneplains agroforests [63], for instance. In addition, it should be stressed that, albeit visible from above (and therefore segmented and labeled), a crown may be partially masked by its neighbors, so the area of the visible segment can differ from the total crown area. Finally, errors stemming from the automatic segmentation step may have further blurred the basal area–crown area relation.

The use of LiDAR data allowed us to cluster points into putative crowns using the AMS-3D algorithm. This segmentation step allowed improving the performance of classifications through majority voting by an average of 10%–18%. This suggests that improving the segmentation step might provide improvement in the classification results. Segmentation itself may take advantage of multisource data. Depending of the way data are merged, fine co-registration can be required but may be difficult to achieve [64,65].

The next major challenge to address in remote species classification is transferring the knowledge acquired on one site to another one [22]. The spectral signature may vary over both space and time. One reason is variation in the timing of leafing, fruiting and flowering, which is still not well characterized for the vast majority of species in evergreen forest [66]. Tropical rainforest trees produce leaves with widely variable phenology and longevity [67,68]. Leaf optical properties may differ significantly between young newly sprouted leaves and one- to two-month-old fully mature photosynthetically competent leaves [69]. In French Guiana where our study took place, a seminal work [70] monitored the phenology of 750 trees belonging to 35 species during two years. Various leafing patterns were reported to occur in that study: deciduous with synchronous or asynchronous leaf drop, or evergreen with continuous or seasonal leaf flushing. Local conditions (and notably soil water holding capacity) will contribute to desynchronize the phenological stage of deciduous trees submitted to similar weather conditions. Difference in tree water status may also alter leaf inclination through wilting which may affect the crown reflectance properties [71]. Additionally, differences in soil chemical fertility may induce differences in the spectral signatures of species. A long-known effect of an increase in nitrogen availability is an increased chlorophyll content and greenness [72]. The second major reason, and probably the most stringent issue, is the variability in imaging conditions. Several factors come into play to modify the apparent reflectance. On the one hand, during the acquisition, the sun–sensor angles change, as both the positions of the sun and the sensor are changing. The reflectance depends on those angles, and this dependency is idiosyncratic (i.e., depends on the characteristics of the object being imaged). Modelling the bidirectional reflectance distribution function (BDRF) to correct these spurious variations in spectral signature is extremely delicate. BRDF is likely to vary from tree to tree as it depends on characteristics such as foliage density, crown geometry (leaf size, leaf angle). Empirical or semi-empirical kernel-driven models [73] would probably require very large datasets to achieve a reasonable level of accuracy. The fact that reflectance in NIR is also significantly affected by neighbors identity [69] renders the hyper-diverse tropical forest particularly challenging in this respect.

5. Conclusions

We present a method that leverages the fusion of LiDAR data for ITC mapping and hyperspectral imagery for species identification. This method allowed us to accurately predict basal area of commercial species in a highly diverse tropical forest. As both types of sensors are likely to become more and more common, we hope that the proposed method will promote sustainable management of tropical forest.

Addition of the SWIR domain (1000nm to 2500nm) to the more commonly used VNIR spectrum was very useful and is strongly recommended whenever possible. In this study, we have examined different scenarios that forest managers may encounter to map specific species in tropical forests, including limited training set and labeling errors affecting both the focal and non-focal classes. We have successfully tested various methods to circumvent labeling error issues. We recommend using robust discriminant analysis, which seems to be the method least sensitive to bias while also offering very good results in the absence of bias in the training data. In addition, we have found that sampling can be limited to a few individuals (30 crowns) without reducing the accuracy of the classifications. Certain species were poorly identified with all classification methods. No indicator based on spectral characteristics allowed upstream identification of what made it possible to carry out an efficient classification of a given tree species.

Author Contributions: Conceptualization, A.L. and G.V.; methodology, A.L. and G.V.; validation, A.L.; formal analysis, A.L. and G.V.; investigation, A.L.; resources, G.V. and J.-B.F.; data curation, A.L.; writing—original draft preparation, A.L.; writing—review and editing, A.L., M.A.-K., J-B.F., C.B., O.B., G.D. and G.V.; visualization, A.L., G.V. and M.A.-K.; supervision, G.V.; project administration, C.B. and O.B.; All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the CartoDivDendroLiDAR project funded by the Fond stratégique de la forêt et du bois (FSFB) of the French Ministry of Agriculture (FSFB) and the Investissement d'Avenir grants from the Agence Nationale de la Recherche (CEBA, ref ANR-10-LABX-25-01 - CeMEB ANR-10-LABX-04-01).

Acknowledgments: A. Laybros acknowledges financial support from the French Ministry of Higher Education, Research and Innovation (MESRI) through CIFRE contract n° 2017/0405. Field support from Jean-Louis Smock (AMAP-IRD), Chantal Geniez (AMAP-IRD), Carla Baltzer (AMAP-IRD) and Raphaël Dutrieux (AMAP-IRD)
is gratefully acknowledged. Hyperspectral imagery acquisition was funded by the CNES (Centre National des Etudes Spatiales). J.-B. Féret acknowledges financial support from Agence Nationale de la Recherche (BioCop project—ANR-17-CE32-0001) and TOSCA program grant of the French Space Agency (CNES) (HyperTropik/HyperBIO project).

Conflicts of Interest: The authors declare no conflict of interest.

6. Appendix A. Experiment 1

Table A1. Pixel-level results from the one versus 19 other species method. Only the F-measure is given with the standard deviation. The best prediction for each species with either VNIR or VSWIR information is in bold.

	VNIR			VSWIR		
Species	**LDA**	**RDA**	**LR**	**LDA**	**RDA**	**LR**
B.p.	13.2(±6.5)	18.2(±10.4)	**24.1**(±12.8)	42.7(±7.7)	**65.2**(±6.5)	54.7(±7.8)
C.m.	66.8(±12.0)	65.0(±12.6)	**67.2**(±10.5)	75.3(±8.0)	72.2(±8.5)	**75.6**(±5.9)
D.g.	54.2(±4.3)	51.9(±5.8)	**60.1**(±4.8)	76.1(±3.3)	75.0(±3.6)	**78.9**(±3.0)
E.f.	33.6(±4.5)	14.1(±3.2)	**36.5**(±4.5)	59.3(±5.2)	66.6(±6.2)	**69.4**(±4.8)
E.g.	49.2(±6.5)	56.3(±4.9)	**69.6**(±4.2)	74.6(±3.5)	79.9(±3.4)	**81.5**(±2.7)
E.s.	59.3(±3.4)	**66.2**(±3.9)	65.7(±4.2)	75.8(±3.0)	**78.2**(±3.0)	75.6(±3.4)
G.g.	37.8(±9.7)	62.2(±9.1)	**67.4**(±9.3)	80.6(±6.2)	**81.3**(±8.2)	77.2(±10.2)
I.a.	38.1(±11.0)	45.9(±13.9)	**64.6**(±12.0)	63.6(±9.5)	**69.2**(±7.5)	64.6(±8.1)
J.c.	58.8(±19.4)	**59.5**(±22.7)	54.5(±20.4)	**56.8**(±16.9)	53.4(±20.5)	55.3(±12.5)
L.a.	34.0(±7.3)	52.1(±9.4)	**54.0**(±9.7)	58.0(±7.9)	**67.4**(±8.7)	64.2(±6.0)
L.h.	**9.4**(±3.9)	5.6(±4.8)	5.6(±4.8)	20.3(±8.3)	**46.5**(±12.3)	36.0(±9.4)
M.c.	45.5(±9.3)	50.0(±11.3)	**64.4**(±9.1)	68.8(±9.0)	70.0(±11.1)	**71.1**(±5.9)
P.c.	78.2(±3.2)	78.8(±1.5)	**81.3**(±1.6)	**89.1**(±1.6)	88.2(±1.4)	88.9(±1.6)
Q.r.	88.4(±2.0)	88.1(±1.6)	**90.8**(±0.8)	94.0(±0.8)	93.0(±1.0)	**94.4**(±0.8)
R.s.	76.6(±5.0)	75.1(±5.3)	**81.3**(±3.4)	**84.1**(±3.5)	82.4(±4.4)	81.7(±3.5)
S.r.	36.4(±7.8)	52.3(±10.3)	**55.3**(±10.0)	53.8(±10.4)	69.2(±6.6)	**71.7**(±4.6)
S.s.	**13.2**(±3.8)	0.4(±0.4)	10.4(±2.7)	33.9(±8.4)	**54.6**(±6.1)	52.4(±6.6)
T.m.	63.3(±7.8)	67.8(±7.6)	**68.5**(±5.8)	**87.7**(±3.4)	87.1(±3.5)	85.5(±3.4)
T.c.	7.4(±2.7)	8.8(±6.3)	**40.5**(±9.0)	30.3(±7.6)	**60.0**(±12.1)	59.2(±8.7)
V.a.	31.2(±8.6)	**48.6**(±10.0)	46.3(±12.5)	69.6(±5.8)	**72.4**(±7.8)	63.0(±7.1)
Mean F-measue	44.7	48.3	**55.4**	64.7	**71.6**	70.1

without (44.7% vs. 48.3%) an SWIR spectrum. At object level, LDA and RDA were the best methods.

7. Appendix B. Experiment 3 (Low Level of Impurity)

Table A2. Full table for the comparison between methods and bias applied, with and without optimisation step. Plain and optimized indicate whether an optimization step (selecting a threshold probability value) was applied to maximize either the F measure (0% contamination) or the Lee and Liu 2003 criterion (1% contamination). The F-measure is given at crown scale. The best prediction for each species, configurations and methods is in bold.

Bias	0%						1%					
	Plain			Optimized			Plain			Optimized		
Species	LDA	RDA	LR	LDA	RDA	LR	LDA	RDA	LR	LDA	RDA	LR
B.p.	35.8 (±10.3)	57.7 (±12.4)	37.5 (±15.5)	65.4 (±12.6)	57.5 (±17.4)	42.1 (±16.0)	36.4 (±9.8)	58.5 (±12.4)	43.4 (±13.6)	64.3 (±12.7)	56.2 (±17.3)	43.0 (±16.0)
C.m.	68.3 (±10.8)	64.7 (±10.1)	68.0 (±10.5)	69.8 (±11.2)	68.3 (±11.3)	71.4 (±7.9)	67.4 (±11.2)	65.1 (±10.3)	66.5 (±9.5)	69.8 (±11.6)	68.0 (±11.3)	69.4 (±8.4)
D.g.	57.7 (±3.9)	72.3 (±6.0)	78.2 (±6.1)	75.8 (±5.6)	72.0 (±6.3)	78.5 (±5.7)	57.5 (±3.8)	72.6 (±5.7)	78.2 (±5.5)	74.8 (±6.3)	72.2 (±5.4)	78.3 (±5.5)
E.f.	50.8 (±4.1)	67.7 (±6.1)	69.2 (±6.2)	71.5 (±4.6)	68.9 (±5.2)	70.5 (±5.0)	51.5 (±4.2)	67.9 (±5.2)	68.0 (±5.4)	72.1 (±4.8)	69.2 (±4.5)	70.5 (±4.9)
E.g.	65.1 (±7.0)	82.0 (±5.3)	84.4 (±4.6)	82.4 (±6.2)	82.7 (±6.2)	84.8 (±4.5)	64.9 (±6.9)	81.9 (±5.6)	82.5 (±6.0)	82.0 (±6.2)	82.5 (±6.0)	84.7 (±4.2)
E.s.	68.1 (±4.2)	74.2 (±4.1)	71.0 (±5.2)	74.7 (±4.3)	74.1 (±4.6)	75.9 (±4.7)	68.5 (±4.5)	74.7 (±3.9)	69.2 (±6.0)	74.1 (±4.3)	73.4 (±4.1)	74.9 (±5.0)
G.g.	62.5 (±12.6)	73.4 (±11.5)	76.4 (±12.5)	74.5 (±10.1)	73.5 (±10.9)	77.3 (±12.3)	62.8 (±14.2)	72.6 (±10.5)	74.8 (±11.6)	74.5 (±10.6)	74.4 (±11.2)	75.0 (±12.9)
I.a.	38.3 (±11.5)	42.3 (±12.7)	38.5 (±11.6)	43.0 (±12.9)	41.0 (±13.3)	41.4 (±13.1)	38.9 (±11.2)	42.7 (±12.5)	36.6 (±10.7)	43.0 (±14.4)	41.4 (±13.5)	40.9 (±14.2)
J.c.	50.7 (±19.7)	51.4 (±16.8)	52.4 (±17.1)	50.9 (±16.9)	49.3 (±17.5)	53.1 (±17.2)	52.2 (±17.2)	49.0 (±16.9)	51.6 (±16.6)	51.5 (±17.6)	50.4 (±16.5)	52.6 (±18.7)
L.a.	41.6 (±8.7)	56.3 (±10.8)	53.9 (±10.5)	56.6 (±11.1)	54.4 (±10.6)	62.3 (±10.3)	41.9 (±9.9)	56.8 (±10.2)	48.5 (±12.1)	57.0 (±10.5)	54.7 (±10.1)	59.7 (±8.8)
L.h.	25.6 (±8.9)	43.6 (±11.0)	22.8 (±8.1)	44.0 (±12.3)	34.2 (±13.1)	25.5 (±7.6)	26.5 (±8.9)	43.5 (±13.6)	23.5 (±8.3)	46.6 (±9.4)	33.7 (±11.5)	26.0 (±10.6)
M.c.	58.5 (±13.8)	55.5 (±16.9)	60.0 (±14.5)	56.3 (±15.9)	54.9 (±16.0)	62.7 (±12.0)	58.6 (±12.8)	56.1 (±16.5)	56.5 (±16.9)	55.4 (±16.5)	56.2 (±15.6)	61.7 (±14.8)
P.c.	69.5 (±4.5)	77.1 (±4.9)	85.9 (±3.6)	77.6 (±4.5)	77.0 (±4.6)	84.3 (±3.2)	69.6 (±4.9)	78.0 (±4.9)	85.8 (±3.2)	78.2 (±4.7)	77.4 (±4.8)	84.0 (±3.8)
Q.r.	86.6 (±2.5)	88.1 (±2.7)	91.3 (±2.3)	86.4 (±2.6)	85.8 (±2.4)	90.7 (±2.6)	86.9 (±2.5)	88.1 (±2.5)	91.2 (±2.2)	86.6 (±2.6)	86.2 (±2.8)	90.8 (±2.4)
R.s.	82.5 (±3.8)	78.8 (±4.5)	81.2 (±5.2)	81.0 (±3.7)	80.7 (±3.5)	84.9 (±4.5)	82.4 (±4.2)	79.3 (±5.0)	79.5 (±5.6)	80.7 (±4.4)	79.7 (±4.1)	83.4 (±5.3)
S.r.	54.3 (±9.2)	70.8 (±10.7)	61.6 (±14.7)	71.8 (±13.7)	66.9 (±15.8)	68.0 (±13.8)	53.8 (±9.1)	69.7 (±12.3)	52.3 (±18.4)	71.3 (±13.5)	66.4 (±13.9)	61.1 (±17.4)
S.s.	29.1 (±5.9)	36.7 (±9.1)	21.3 (±7.1)	36.6 (±8.0)	36.7 (±8.8)	25.6 (±14.2)	29.1 (±6.5)	35.8 (±7.9)	21.3 (±7.1)	35.5 (±9.2)	35.8 (±9.2)	25.6 (±11.7)
T.m.	64.7 (±6.4)	69.9 (±6.0)	80.7 (±9.4)	76.3 (±7.5)	75.8 (±6.8)	78.8 (±7.2)	64.9 (±6.5)	70.4 (±5.6)	78.2 (±9.9)	76.0 (±6.8)	75.6 (±7.2)	78.9 (±9.7)
T.c.	58.3 (±11.5)	65.4 (±12.5)	72.2 (±14.8)	65.8 (±14.4)	52.1 (±13.6)	72.5 (±15.1)	58.2 (±10.9)	65.4 (±12.7)	69.6 (±13.4)	63.6 (±14.2)	51.0 (±13.5)	70.6 (±14.2)
V.a.	68.2 (±8.7)	72.8 (±12.2)	65.1 (±12.8)	74.4 (±12.0)	74.4 (±12.2)	66.9 (±13.3)	66.5 (±10.4)	74.1 (±12.0)	63.9 (±13.3)	74.4 (±12.1)	73.6 (±11.1)	64.6 (±12.5)
Average	56.8	65.0	63.6	66.7	64.0	65.9	56.9	65.1	62.0	66.6	63.9	64.8

Bias	2%						5%					
	Plain			Optimized			Plain			Optimized		
Species	LDA	RDA	LR	LDA	RDA	LR	LDA	RDA	LR	LDA	RDA	LR
B.p.	--	--	--	--	--	--	--	--	--	--	--	--
C.m.	--	--	--	--	--	--	--	--	--	--	--	--
D.g.	56.2 (±4.2)	69.6 (±4.7)	76.5 (±5.7)	72.5 (±5.2)	67.2 (±4.9)	73.2 (±5.0)	58.5 (±4.4)	70.3 (±4.6)	72.3 (±4.8)	73.1 (±5.0)	68.0 (±4.9)	73.1 (±4.4)
E.f.	51.4 (±4.1)	62.5 (±6.1)	61.9 (±6.5)	67.3 (±6.0)	63.4 (±5.9)	65.0 (±5.5)	52.8 (±5.4)	62.0 (±5.7)	50.7 (±8.5)	66.0 (±5.4)	62.2 (±6.6)	59.3 (±7.0)
E.g.	--	--	--	--	--	--	--	--	--	--	--	--
E.s.	67.4 (±5.0)	74.1 (±4.3)	69.0 (±6.2)	74.2 (±4.4)	74.2 (±4.0)	73.3 (±4.6)	69.8 (±5.5)	72.8 (±4.4)	47.1 (±8.7)	72.0 (±4.5)	72.0 (±4.0)	61.7 (±7.6)
G.g.	--	--	--	--	--	--	--	--	--	--	--	--
I.a.	--	--	--	--	--	--	--	--	--	--	--	--
J.c.	--	--	--	--	--	--	--	--	--	--	--	--
L.a.	--	--	--	--	--	--	--	--	--	--	--	--
L.h.	--	--	--	--	--	--	--	--	--	--	--	--
M.c.	--	--	--	--	--	--	--	--	--	--	--	--
P.c.	70.4 (±4.9)	80.4 (±3.4)	84.6 (±4.6)	79.9 (±3.6)	79.8 (±3.5)	84.2 (±3.4)	71.4 (±4.8)	79.3 (±3.8)	74.2 (±7.7)	79.3 (±3.8)	78.7 (±3.5)	81.9 (±4.5)
Q.r.	88.0 (±2.5)	89.0 (±2.5)	92.2 (±2.0)	85.6 (±3.1)	85.7 (±2.7)	90.6 (±1.8)	87.9 (±2.6)	89.4 (±2.7)	92.1 (±2.4)	86.6 (±3.2)	86.4 (±3.0)	90.4 (±2.6)
R.s.	--	--	--	--	--	--	--	--	--	--	--	--
S.r.	--	--	--	--	--	--	--	--	--	--	--	--
S.s.	--	--	--	--	--	--	--	--	--	--	--	--
T.m.	--	--	--	--	--	--	--	--	--	--	--	--
T.c.	--	--	--	--	--	--	--	--	--	--	--	--
V.a.	--	--	--	--	--	--	--	--	--	--	--	--
Average	66.7	75.1	76.9	75.9	74.1	77.3	68.0	74.7	67.3	75.4	73.4	73.3

8. Appendix C. Experiment 3 (High Level of Impurity—Smaller Focal Training Class)

The procedure was similar to Experiment 3. We set aside 30% of the data for validation. However, we selected only 30 crowns of the focal class to use in the training set. In addition, we selected 100,000 pixels from the non-focal species to make up the non-focal class. We simulated three level of bias in the training set as described below.

Table A3. Comparison of classification methods with and without bias. Plain and optimized indicate whether an optimization step (selecting a threshold probability value) was applied or not. This step aims at maximizing either the F measure (0% contamination) or the Lee and Liu 2003 criterion (1% contamination). F-measure is given at object level. The best predictions for each species, configuration and method are in bold.

		LDA		RDA		LR	
SPID	**Bias**	**Plain**	**Optimized**	**Plain**	**Optimized**	**Plain**	**Optimized**
D. g	0	58.9 (±8.0)	69.7 (±6.4)	67.1 (±6.5)	66.7 (±6.7)	**71.6** (±7.4)	**74.6** (±6.5)
E. f	0	53.4 (±7.2)	**68.3** (±6.9)	**66.0** (±4.9)	64.7 (±7.1)	61.2 (±9.1)	63.6 (±7.6)
E. s	0	68.6 (±5.7)	**71.7** (±6.4)	**73.0** (±5.6)	71.5 (±6.3)	61.2 (±9.0)	68.4 (±8.4)
P. c	0	73.6 (±4.4)	79.9 (±5.6)	79.5 (±5.7)	79.2 (±5.1)	**83.6** (±4.2)	**84.9** (±5.1)
Q. r	0	88.8 (±2.7)	89.5 (±2.6)	89.4 (±2.6)	89.4 (±2.7)	**90.2** (±2.9)	**90.9** (±3.0)
Average		68.7	75.8	75.0	74.3	73.5	76.5
D. g	1%	58.8 (±8.3)	71.0 (±7.9)	67.8 (±7.7)	67.1 (±8.5)	**70.5** (±8.7)	**75.3** (±7.0)
E. f	1%	54.1 (±6.6)	**66.8** (±7.8)	**65.2** (±5.7)	65.1 (±5.8)	59.0 (±10.0)	63.6 (±6.7)
E. s	1%	69.0 (±6.6)	**73.2** (±5.1)	**73.4** (±5.3)	72.5 (±5.5)	52.9 (±8.1)	62.9 (±7.0)
P. c	1%	71.9 (±5.7)	76.5 (±5.8)	75.8 (±6.7)	76.7 (±5.9)	**77.1** (±5.0)	**79.5** (±4.7)
Q. r	1%	**89.2** (±2.9)	89.2 (±2.3)	**89.2** (±2.9)	**88.9** (±2.7)	87.2 (±3.6)	88.5 (±3.6)
Average		68.6	75.3	74.3	74.0	69.4	74.0
D. g	2%	57.0 (±8.2)	71.5 (±5.8)	68.3 (±6.2)	67.4 (±6.7)	**68.5** (±9.5)	**72.9** (±7.0)
E. f	2%	53.7 (±6.1)	**66.4** (±7.7)	**64.4** (±6.7)	64.6 (±7.3)	54.7 (±10.5)	62.2 (±9.6)
E. s	2%	68.2 (±6.1)	**73.3** (±5.1)	**73.6** (±5.3)	71.9 (±6.7)	48.6 (±11.3)	62.0 (±10.0)
P. c	2%	74.1 (±5.2)	79.6 (±4.0)	78.5 (±4.2)	78.6 (±5.0)	**78.6** (±3.5)	**82.3** (±3.4)
Q. r	2%	88.7 (±2.9)	**89.3** (±2.9)	**89.3** (±2.7)	**89.3** (±2.9)	84.2 (±3.0)	87.5 (±2.7)
Average		68.3	76.0	74.8	74.3	66.9	73.4
D. g	5%	58.8 (±6.8)	**71.6** (±5.2)	**68.2** (±6.2)	68.1 (±6.3)	61.3 (±5.8)	70.9 (±5.4)
E. f	5%	52.2 (±5.9)	**67.5** (±6.2)	**66.1** (±6.4)	66.2 (±5.9)	45.8 (±13.2)	59.0 (±8.6)
E. s	5%	69.2 (±5.9)	**73.2** (±5.5)	**73.7** (±4.8)	72.0 (±5.8)	41.6 (±11.2)	58.7 (±10.0)
P. c	5%	74.0 (±5.7)	**80.5** (±5.3)	**80.2** (±6.1)	79.9 (±5.8)	71.4 (±5.7)	78.4 (±4.9)
Q. r	5%	88.6 (±2.9)	**89.4** (±2.9)	**89.9** (±2.6)	89.2 (±3.1)	74.5 (±5.5)	83.4 (±6.0)
Average		68.6	76.4	75.6	75.1	58.9	70.1

The use of 30 crowns compared to the use of 70% of focal class crowns had no negative effect on the LDA and RDA method. Nevertheless, the LR method seemed to be more sensitive to the use of a smaller sample (from 79.1% to 73.5% or from 80.0% to 76.5% with optimisation step). The optimization step made up for part of the loss in performance of the F-measure.

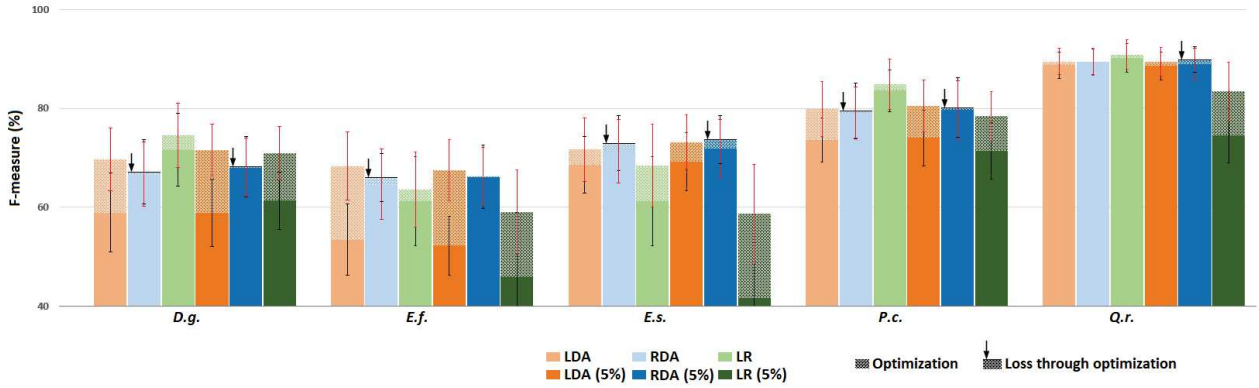

Figure A1. Comparison of classification methods with and without bias. F-measure and its standard deviation are given at object level. Only 30 crowns from the focus class and 100,000 pixels from the non-focus class were used to train the model. Bias consisted in adding 5% of the pixels of the focal species into the non-focal class. Optimization step is represented in the same paler colour. An arrow indicates when optimization decreased performance.

9. Appendix D. Experiment 3 (Focal Class Purification)

Removal of potential outliers' impact on F-measure of various classifiers.

Figure A2. From left to right: low illumination ratio from hyperspectral images, wood contribution, lianas contribution and ground contribution.

Illustration of tree crowns considered as outliers is proposed above. Low illumination led to a reduction in the number of pixels inside the crowns (given by hyperspectral imagery). Other possible causes of outlier crowns are illustrated by RGB imagery.

Table A4. Comparison between using all crowns and without some removed crowns based on the Mahalanobis distance. The F-measure is given at object level. Significance codes: * $p <$ 0.1, ** $p <$ 0.05, *** $p <$ 0.01. The best predictions for each species, configuration and method are in bold.

Removing	Without			With			Total crowns	Potential outliers	Proportion (%)
SPID	LDA	RDA	LR	LDA	RDA	LR			
B.p.	35.8 (±10.3)	**57.7** (±12.4)	37.5 (±15.5)	40.4 (±10.2)	**60.3** (±14.0)	35.0 (±13.3)	24	0	0.0
C.m.	**68.3** (±10.8)	64.7 (±10.1)	68.0 (±10.5)	68.4 (±11.6)**	65.5 (±9.8)**	**71.8** (±9.1)***	49	1	2.0
D.g.	57.7 (±3.9)	72.3 (±6.0)	**78.2** (±6.1)	55.9 (±5.2)	69.6 (±5.9)	**76.6** (±5.7)	108	5	4.6
E.f.	50.8 (±4.1)	67.7 (±6.1)	**69.2** (±6.2)	50.3 (±6.1)	66.4 (±6.1)	**69.3** (±5.9)*	106	2	1.9
E.g.	65.1 (±7.0)	82.0 (±5.3)	**84.4** (±4.6)	64.2 (±5.9)	81.9 (±5.2)	**84.6** (±5.4)	74	5	6.8
E.s.	68.1 (±4.2)	**74.2** (±4.1)	71.0 (±5.2)	66.7 (±4.7)	**74.6** (±3.9)	73.9 (±4.8)*	139	7	5.0
G.g.	62.5 (±12.6)	73.4 (±11.5)	**76.4** (±12.5)	58.4 (±8.3)	68.3 (±11.6)	**73.7** (±14.8)	25	2	8.0
I.a.	38.3 (±11.5)	**42.3** (±12.7)	38.5 (±11.6)	38.4 (±9.9)	**43.0** (±8.1)	33.4 (±13.2)	26	2	7.7
J.c.	50.7 (±19.7)	51.4 (±16.8)	**52.4** (±17.1)	**59.2** (±10.6)**	58.5 (±10.2)	57.4 (±12.6)	24	2	8.3
L.a.	41.6 (±8.7)	**56.3** (±10.8)	53.9 (±10.5)	43.3 (±7.0)	**53.3** (±8.3)	50.1 (±13.7)	46	0	0.0
L.h.	25.6 (±8.9)	**43.6** (±11.0)	22.8 (±8.1)	27.9 (±10.9)	**44.0** (±16.7)	26.9 (±8.6)	27	1	3.7
M.c.	58.5 (±13.8)	55.5 (±16.9)	**60.0** (±14.5)	60.1 (±11.1)	59.9 (±14.9)	**66.7** (±12.1)	27	2	7.4
P.c.	69.5 (±4.5)	77.1 (±4.9)	**85.9** (±3.6)	69.2 (±4.2)	79.0 (±3.7)	**85.7** (±3.7)	164	6	3.7
Q.r.	86.6 (±2.5)	88.1 (±2.7)	**91.3** (±2.3)	88.5 (±3.0)	89.9 (±2.4)	**91.9** (±2.2)	206	7	3.4
R.s.	**82.5** (±3.8)	78.8 (±4.5)	81.2 (±5.2)	83.1 (±7.6)	80.5 (±7.7)	**83.8** (±6.3)**	69	2	2.9
S.r.	54.3 (±9.2)	**70.8** (±10.7)	61.6 (±14.7)	55.6 (±8.3)***	**71.3** (±7.6)	64.7 (±12.9)***	32	2	6.3
S.s.	29.1 (±5.9)	**36.7** (±9.1)	21.3 (±7.1)	29.8 (±7.6)	**36.1** (±11.5)	22.4 (±6.9)	34	2	5.9
T.m.	64.7 (±6.4)	69.9 (±6.0)	**80.7** (±9.4)	64.6 (±11.5)	69.6 (±11.8)	**82.0** (±6.4)	51	4	7.8
T.c.	58.3 (±11.5)	65.4 (±12.5)	**72.2** (±14.8)	55.8 (±9.0)	68.0 (±12.8)*	**73.3** (±13.9)	32	2	6.3
V.a.	68.2 (±8.7)	**72.8** (±12.2)	65.1 (±12.8)	70.0 (±8.6)	**78.4** (±7.0)	67.8 (±10.7)	34	1	2.9
Average	56.8	**65.0**	63.6	57.5	**65.9**	64.5	1297.0	55	4.2

10. Appendix E. Experiment 4

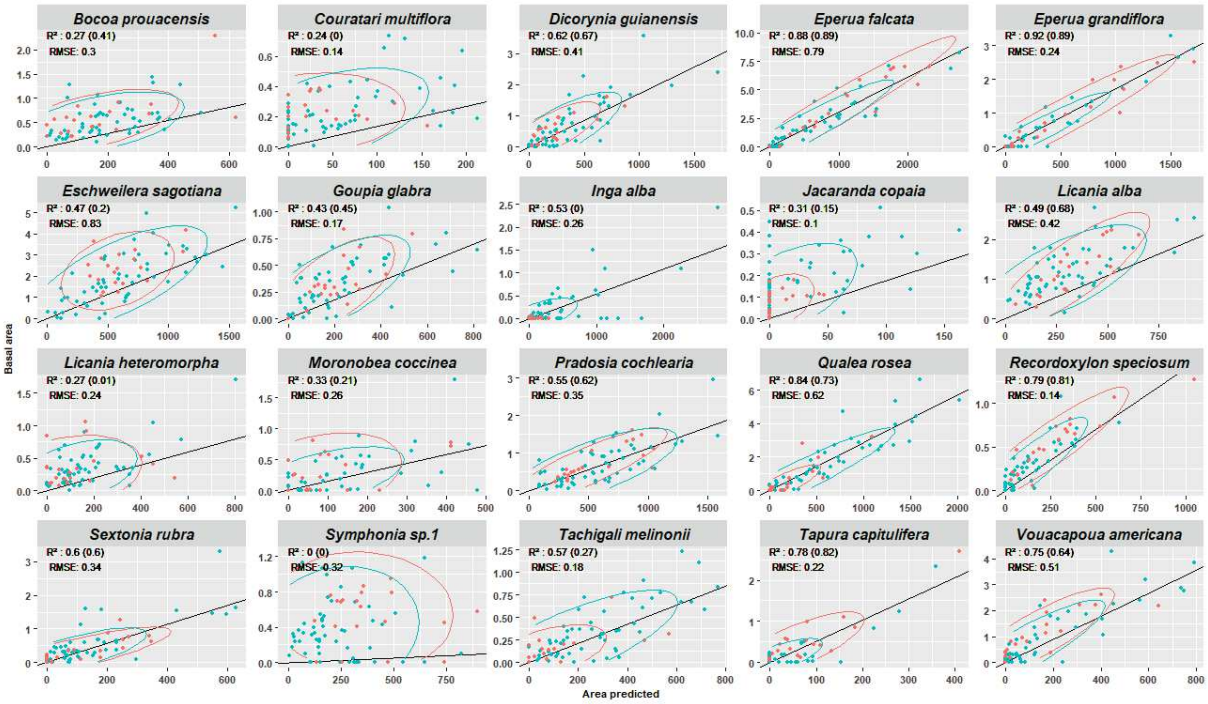

Figure A3. Relation between basal areal ($m^2 \cdot ha^{-1}$) and area predicted ($m^2 \cdot ha^{-1}$) using the optimized LDA. Plots not used for training are plotted in red and the R-squared for those observations are given in brackets. The plots used for the training are in blue. Global R-squares are given using all data points (red and blue). In solid black line, linear regression based on all points passing through zero. The normal confidence ellipses are given, computed for each group of points using the corresponding color (red or blue).

Figure A3 (basal area vs. prediction) is a fine representation of the behaviour of the classifier and makes it possible to understand the constraints. For example, one might think that the predictive performance of *Inga alba* is poor. However, on the plots never used for training (red point) we see that this species is not present. The classifier predicts few pixels of this species. The *Jacaranda copaia* is a species that has a low basal area; therefore, a prediction error is potentially more impacting than for the *Eperua grandiflora*. Note that the relationship between the basal area and the predicted area is different depending on the species considered, even for the species that have a high R^2.

11. Appendix F. Test of Correlation

Data used to test the relationship between species object level F measure (obtained using and optimized LDA) and different metrics are given in Table A5. AB~CA r2 is the linear regression coefficient between the predicted canopy coverage area for a particular species and its recorded basal area. Number of pixels is given per species. Intra-group variance was computed as proposed by [60]. Species mean Mahalanobis distance was computed as described in Section 2.6.3. Pseudo-outlier refers to number of outliers detected if Mahalanobis distance value is higher or lower than two standard deviations.

Table A5. F-measure and various metrics tested for correlation.

Species	F-measure (object level)	BA~CA r^2	Number of pixels	Number of crowns	Intra group variance	Mean species Mahalanobis distance	Pseudo outlier (%)
B.p.	65.4	0.27	2375	24	60716044	398	0
C.m.	69.8	0.24	4850	49	42710171	382	2
D.g.	75.8	0.62	18589	108	88937469	383	4.6
E.f.	71.5	0.84	15355	106	41768295	383	1.9
E.g.	82.4	0.92	10859	74	37064303	378	6.8
E.s.	74.7	0.47	12559	139	42879098	382	5
G.g.	74.5	0.43	4998	25	45436876	387	8
I.a.	43.0	0.53	3846	26	73845080	394	7.7
J.c.	50.9	0.31	1705	24	50504397	417	8.3
L.a.	56.6	0.49	3894	46	49323222	386	0
L.h.	44.0	0.27	1437	27	35529833	412	3.7
M.c.	56.3	0.33	3355	27	53471810	381	7.4
P.c.	77.6	0.55	38349	164	46398429	374	3.7
Q.r.	86.4	0.84	27828	206	51460549	374	3.4
R.s.	81.0	0.79	7944	69	120765823	382	2.9
S.r.	71.8	0.60	4070	32	31609536	417	6.3
S.s.	36.6	0.00	3355	34	38010280	378	5.9
T.m.	76.3	0.57	6745	51	80924196	392	7.8
T.c.	65.8	0.78	1224	32	36278962	388	6.3
V.a.	74.4	0.75	3218	34	137806130	387	2.9
Pearson Cor. Coeff	--	0.71**	0.54*	0.59***	0.28	0.08	0.48
Spearman Cor. Coeff.	--	0.69**	0.77**	0.71***	0.22	-0.44	-0.09

12. References

1. Corlett, R.T. Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Divers.* **2016**, *38*, 10–16. [CrossRef]
2. Peres, C.A.; Gardner, T.A.; Barlow, J.; Zuanon, J.; Michalski, F.; Lees, A.C.; Vieira, I.C.G.; Moreira, F.M.S.; Feeley, K.J. Biodiversity conservation in human-modified Amazonian forest landscapes. *Biol. Conserv.* **2010**, *143*, 2314–2327. [CrossRef]
3. Leisher, C.; Touval, J.; Hess, S.M.; Boucher, T.M.; Reymondin, L. Land and Forest Degradation inside Protected Areas in Latin America. *Diversity* **2013**, *5*, 779–795. [CrossRef]
4. Sala, O.E. Global Biodiversity Scenarios for the Year 2100. *Science* **2000**, *287*, 1770–1774. [CrossRef] [PubMed]
5. Burivalova, Z.; Şekercioğlu, Ç.H.; Koh, L.P. Thresholds of Logging Intensity to Maintain Tropical Forest Biodiversity. *Curr. Biol.* **2014**, *24*, 1893–1898. [CrossRef] [PubMed]
6. Putz, F.E.; Zuidema, P.A.; Pinard, M.A.; Boot, R.G.A.; Sayer, J.A.; Sheil, D.; Sist, P.; Vanclay, J.K. Improved Tropical Forest Management for Carbon Retention. *PLoS Biol.* **2008**, *6*, e166. [CrossRef]
7. Sist, P.; Mazzei, L.; Blanc, L.; Rutishauser, E. Large trees as key elements of carbon storage and dynamics after selective logging in the Eastern Amazon. *For. Ecol. Manag.* **2014**, *318*, 103–109. [CrossRef]
8. Putz, F.E. Woody vines and forest management in Malaysia. *Commonw. For. Rev.* **1985**, *64*, 359–365.
9. *The Biology of Vines*; Putz, F.E.; Mooney, H.A. (Eds.) Cambridge University Press: Cambridge, UK; New York, NY, USA, 1991; ISBN 978-0-521-39250-1.
10. Ouédraogo, D.-Y.; Beina, D.; Picard, N.; Mortier, F.; Baya, F.; Gourlet-Fleury, S. Thinning after selective logging facilitates floristic composition recovery in a tropical rain forest of Central Africa. *For. Ecol. Manag.* **2011**, *262*, 2176–2186. [CrossRef]
11. Alroy, J. E**ff**ects of habitat disturbance on tropical forest biodiversity. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6056. [CrossRef]
12. Kennard, D.K.; Gould, K.; Putz, F.E.; Fredericksen, T.S.; Morales, F. E**ff**ect of disturbance intensity on regeneration mechanisms in a tropical dry forest. *For. Ecol. Manag.* **2002**, *162*, 197–208. [CrossRef]
13. Chaudhary, A.; Burivalova, Z.; Koh, L.P.; Hellweg, S. Impact of Forest Management on Species Richness: Global Meta-Analysis and Economic Trade-O**ff**s. *Sci. Rep.* **2016**, *6*, 23954. [CrossRef] [PubMed]
14. Huang, C.; Asner, G.P. Applications of Remote Sensing to Alien Invasive Plant Studies. *Sensors* **2009**, *9*, 4869–4889. [CrossRef] [PubMed]
15. Ustin, S.L.; DiPietro, D.; Olmstead, K.; Underwood, E.; Scheer, G.J. Hyperspectral remote sensing for invasive species detection and mapping. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; Volume 3, pp. 1658–1660.
16. Waite, C.E.; van der Heijden, G.M.F.; Field, R.; Boyd, D.S. A view from above: Unmanned aerial vehicles (UAVs) provide a new tool for assessing liana infestation in tropical forest canopies. *J. Appl. Ecol.* **2019**. [CrossRef]
17. Sist, P. Reduced-impact logging in the tropics: Objectives, principles and impacts. *Int. For. Rev.* **2000**, *2*, 3–10.
18. Baldeck, C.; Asner, G. Improving Remote Species Identification through E**ffi**cient Training Data Collection. *Remote Sens.* **2014**, *6*, 2682–2698. [CrossRef]
19. Baldeck, C.A.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Knapp, D.E.; Kellner, J.R.; Wright, S.J. Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS ONE* **2015**, *10*, e0118403. [CrossRef]
20. Feret, J.-B.; Asner, G.P. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 73–84. [CrossRef]
21. Harrison, D.; Rivard, B.; Sánchez-Azofeifa, A. Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *66*, 93–105. [CrossRef]
22. Laybros, A.; Schläpfer, D.; Féret, J.-B.; Descroix, L.; Bedeau, C.; Lefevre, M.-J.; Vincent, G. Across Date Species Detection Using Airborne Imaging Spectroscopy. *Remote Sens.* **2019**, *11*, 789. [CrossRef]
23. Clark, M.; Roberts, D.; Clark, D. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* **2005**, *96*, 375–398. [CrossRef]
24. Ferreira, M.P.; Zortea, M.; Zanotta, D.C.; Feret, J.B.; Shimabukuro, Y.E.; Filho, C.R. On the use of shortwave infrared for tree species discrimination in tropical semideciduous forest. In Proceedings of the ISPRS Geospatial Week 2015, La Grande Motte, France, 28 September–3 October 2015; Volume XL-3/W3, pp. 473–476.
25. Clark, M.L.; Roberts, D.A. Species-Level Di**ff**erences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sens.* **2012**, *4*, 1820–1855. [CrossRef]
26. Dalponte, M.; Orka, H.O.; Gobakken, T.; Gianelle, D.; Naesset, E. Tree Species Classification in Boreal Forests with

Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2632–2645. [CrossRef]
27. Aubry-Kientz, M.; Dutrieux, R.; Ferraz, A.; Saatchi, S.; Hamraz, H.; Williams, J.; Coomes, D.; Piboule, A.; Vincent, G. A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests. *Remote Sens.* **2019**, *11*, 1086. [CrossRef]
28. Tochon, G.; Féret, J.B.; Valero, S.; Martin, R.E.; Knapp, D.E.; Salembier, P.; Chanussot, J.; Asner, G.P. On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images. *Remote Sens. Environ.* **2015**, *159*, 318–331. [CrossRef]
29. Lee, W.S.; Liu, B. Learning with Positive and Unlabeled Examples Using Weighted Logistic Regression. *ICML* **2003**, *3*, 448–455.
30. *Ecology and Management of a Neotropical Rainforest: Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*; Gourlet-Fleury, S.; Guehl, J.-M.; Laroussinie, O.; ECOFOR (Group). Elsevier: Paris, France, 2004; ISBN 978-2-84299-455-6.
31. Richter, R.; Schlapfer, D. *Atmospheric|Topographic Correction for Airborne Imagery (ATCOR-4 User Guide, Version 7.2.0)*; ReSe Applications LLC: Wil, Switzerland, 2018; p. 279.
32. Schlapfer, D. *PARametric Geocoding, Orthorectification for Airborne Scanner Data, User Manual Version 2.3*; ReSe Applications Schlaepfer and Remote Sensing Laboratories (RSL) of the University of Zurich: Zurich, Switzerland, 2006.
33. Galvão, L.S.; Ponzoni, F.J.; Epiphanio, J.C.N.; Rudorff, B.F.T.; Formaggio, A.R. Sun and view angle effects on NDVI determination of land cover types in the Brazilian Amazon region with hyperspectral data. *Int. J. Remote Sens.* **2004**, *25*, 1861–1879. [CrossRef]
34. Ferraz, A.; Saatchi, S.; Mallet, C.; Meyer, V. LiDAR detection of individual tree size in tropical forests. *Remote Sens. Environ.* **2016**, *183*, 318–333. [CrossRef]
35. Xiao, W.; Zaforemska, A.; Smigaj, M.; Wang, Y.; Gaulton, R. Mean Shift Segmentation Assessment for Individual Forest Tree Delineation from Airborne LiDAR Data. *Remote Sens.* **2019**, *11*, 1263. [CrossRef]
36. Roussel, J.-R.; Auty, D.; De Boissieu, F.; Meador, A. *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*; R Core Team: Vienna, Austria, 2018.
37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *Mach. Learn. PYTHON* **2011**, *12*, 2825–2830.
38. Kurtzer, G.M.; Sochat, V.; Bauer, M.W. Singularity: Scientific containers for mobility of compute. *PLoS ONE* **2017**, *12*, e0177459. [CrossRef]
39. Yu, H.; Yang, J. A direct LDA algorithm for high-dimensional data—With application to face recognition. *Pattern Recognit.* **2001**, *34*, 2067–2070. [CrossRef]
40. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2012; ISBN 978-1-118-58600-6.
41. Friedman, J.H. Regularized Discriminant Analysis. *J. Am. Stat. Assoc.* **1989**, *84*, 165–175. [CrossRef]
42. Guo, Y.; Hastie, T.; Tibshirani, R. *Regularized Discriminant Analysis and Its Application in Microarrays*; Dept. of Statistics, Stanford University: Stanford, CA, USA, 2005; p. 18.
43. Pohar, M.; Blas, M.; Turk, S. Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. *Metodol. Zv.* **2004**, *1*, 143.
44. Ferreira, M.P.; Zortea, M.; Zanotta, D.C.; Shimabukuro, Y.E.; de Souza Filho, C.R. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sens. Environ.* **2016**, *179*, 66–78. [CrossRef]
45. Graves, S.J.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Colgan, M.S.; Kalantari, L.; Bohlman, S.A. Tree Species Abundance Predictions in a Tropical Agricultural Landscape with a Supervised Classification Model and Imbalanced Data. *Remote Sens.* **2016**, *8*, 161. [CrossRef]
46. Naidoo, L.; Cho, M.A.; Mathieu, R.; Asner, G. 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 J. Photogramm. Remote Sens.* **2012**, *69*, 167–179. [CrossRef]
47. Akbani, R.; Kwek, S.; Japkowicz, N. Applying Support Vector Machines to Imbalanced Datasets. In Proceedings of the Machine Learning: ECML 2004, Pisa, Italy, 20–24 September 2004; Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D., Eds.; Springer: Berlin, Heidelberg, 2004; pp. 39–50.
48. Imam, T.; Ting, K.M.; Kamruzzaman, J. z-SVM: An SVM for Improved Classification of Imbalanced Data. In Proceedings of the AI 2006: Advances in Artificial Intelligence, Hobart, Australia, 4–8 December 2006; Sattar, A., Kang, B., Eds.; Springer: Berlin, Heidelberg, 2006; pp. 264–273.
49. Lin, W.-J.; Chen, J.J. Class-imbalanced classifiers for high-dimensional data. *Brief. Bioinform.* **2013**, *14*, 13–26.

[CrossRef]
50. Wu, G.; Chang, E.Y. Adaptive Feature-Space Conformal Transformation for Imbalanced-Data Learning. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2013; p. 8.
51. Baldeck, C.A.; Asner, G.P. Single-Species Detection with Airborne Imaging Spectroscopy Data: A Comparison of Support Vector Techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2501–2512. [CrossRef]
52. Xue, J.-H.; Titterington, D.M. Do unbalanced data have a negative e**ff**ect on LDA? *Pattern Recognit.* **2008**, *41*, 1558–1571. [CrossRef]
53. Ter Steege, H.; Pitman, N.C.A.; Sabatier, D.; Baraloto, C.; Salomao, R.P.; Guevara, J.E.; Phillips, O.L.; Castilho, C.V.; Magnusson, W.E.; Molino, J.-F.; et al. Hyperdominance in the Amazonian Tree Flora. *Science* **2013**, *342*, 1243092. [CrossRef]
54. Sabatier, D.; Grimaldi, M.; Prévost, M.-F.; Guillaume, J.; Godron, M.; Dosso, M.; Curmi, P. The influence of soil cover organization on the floristic and structural heterogeneity of a Guianan rain forest. *Plant Ecol.* **1997**, *131*, 81–108. [CrossRef]
55. Traissac, S. *Dynamique Spatiale de Vouacapoua Americana, Arbre de Foret Tropicale Humide a Repartition Agregee*; Université Claude Bernard Lyon 1: Villeurbanne, France, 2003.
56. Traissac, S.; Pascal, J.-P. Birth and life of tree aggregates in tropical forest: Hypotheses on population dynamics of an aggregated shade-tolerant species. *J. Veg. Sci.* **2014**, *25*, 491–502. [CrossRef]
57. Fonty, É.; Molino, J.-F.; Prévost, M.-F.; Sabatier, D. A new case of neotropical monodominant forest: Spirotropis longifolia (Leguminosae-Papilionoideae) in French Guiana. *J. Trop. Ecol.* **2011**, *27*, 641–644. [CrossRef]
58. Pitman, N.C.A.; Terborgh, J.; Silman, M.R.; Nuñez, V.P. Tree Species Distributions in an Upper Amazonian Forest. *Ecology* **1999**, *80*, 2651–2661. [CrossRef]
59. Marcon, E.; Scotti, I.; Hérault, B.; Rossi, V.; Lang, G. Generalization of the Partitioning of Shannon Diversity. *PLoS ONE* **2014**, *9*. [CrossRef]
60. Laliberté, E.; Schweiger, A.K.; Legendre, P. Partitioning plant spectral diversity into alpha and beta components. *Ecol. Lett.* **2020**, *23*, 370–380. [CrossRef]
61. Jucker, T.; Caspersen, J.; Chave, J.; Antin, C.; Barbier, N.; Bongers, F.; Dalponte, M.; van Ewijk, K.Y.; Forrester, D.I.; Haeni, M.; et al. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Glob. Chang. Biol.* **2017**, *23*, 177–190. [CrossRef]
62. Antin, C.; Pélissier, R.; Vincent, G.; Couteron, P. Crown allometries are less responsive than stem allometry to tree size and habitat variations in an Indian monsoon forest. *Trees* **2013**, *27*, 1485–1495. [CrossRef]
63. Harja, D.; Vincent, G.; Mulia, R.; Noordwijk, M. Tree shape plasticity in relation to crown exposure. *Trees* **2012**, *26*, 1275–1285. [CrossRef]
64. Brell, M.; Rogass, C.; Segl, K.; Bookhagen, B.; Guanter, L. Improving Sensor Fusion: A Parametric Method for the Geometric Coalignment of Airborne Hyperspectral and LiDAR Data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3460–3474. [CrossRef]
65. Tusa, E.; Laybros, A.; Monnet, J.-M.; Dalla Mura, M.; Barré, J.-B.; Vincent, G.; Dalponte, M.; Féret, J.-B.; Chanussot, J. Fusion of hyperspectral imaging and LiDAR for forest monitoring. In *Data Handling in Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 32, pp. 281–303. ISBN 978-0-444-63977-6.
66. Reich, P.B. Phenology of tropical forests: Patterns, causes, and consequences. *Can. J. Bot.* **1995**, *73*, 164–174. [CrossRef]
67. Laurans, M.; Martin, O.; Nicolini, E.; Vincent, G. Functional traits and their plasticity predict tropical trees regeneration niche even among species with intermediate light requirements. *J. Ecol.* **2012**, *100*, 1440–1452. [CrossRef]
68. Reich, P.B.; Uhl, C.; Walters, M.B.; Prugh, L.; Ellsworth, D.S. Leaf demography and phenology in Amazonian rain forest: A census of 40 000 leaves of 23 tree species. *Ecol. Monogr.* **2004**, *74*, 3–23. [CrossRef]
69. Korpela, I.; Heikkinen, V.; Honkavaara, E.; Rohrbach, F.; Tokola, T. Variation and directional anisotropy of reflectance at the crown scale—Implications for tree species classification in digital aerial images. *Remote Sens. Environ.* **2011**, *115*, 2062–2074. [CrossRef]
70. Loubry, D. *Déterminisme du Comportement Phénologique des Arbres en Forêt Tropicale Humide de Guyane Française (5° lat. n.)*; Université de Paris 6: Paris, France, 1994.
71. Saini, M.; Christian, B.; Joshi, N.; Vyas, D.; Marpu, P.; Krishnayya, N.S.R. Hyperspectral Data Dimensionality Reduction and the Impact of Multi-Seasonal Hyperion EO-1 Imagery on Classification Accuracies of Tropical Forest Species. Available online: https://www.ingentaconnect.com/content/asprs/pers/2014/00000080/ 00000008/art00005 (accessed on 2 May 2020).

72. Yadava, U.L. A Rapid and Non-destructive Method to Determine Chlorophyll in Intact Leaves. *HortScience* **1986**, *21*, 1449–1450.
73. Schlapfer, D.; Richter, R. Evaluation of brefcor BRDF e**ff**ects correction for HYSPEX, CASI, and APEX imaging spectroscopy data. In Proceedings of the 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 24–27 June 2014; IEEE: Lausanne, Switzerland, 2014; pp. 1–4.

C. Conclusion of the chapter

This chapter highlights the possibility of identifying tree species in tropical forests despite multiple constraints, the first of which is the diversity of species in the background. A random selection makes it possible to reduce the non-focal class to a small set of representative pixels, provided that the proportion of pixels from the focal class is not greater than 5%. In addition, species were well identified if they had at least 30 represntative crowns (focal class). The field inventory campaings can therefore be limited to 30 crowns per species of interest. Although not shown in the article, it is likely that the size of the crowns indirectly contributes to the correct identification. Species with small crowns may require more crowns.

The complementarity of LiDAR and hyperspectral data has a high potential and allows the quantification of the resource, providing an estimated cumulated basal area per species consistent with the field inventory. Improvements in point cloud segmentation methods from LiDAR may further improve predictions.

While these results are encouraging as they suggest that a limited training set size may be sufficient to conduct reliable inventories from the air, the next question is whether we could do without any site specific ground data collection. This poses the question of transferability of hyperspectral classifiers. It is expected that the identification of a species at the same site on a new date will offer lower results. In a simplified configuration as discussed in Chapter 1, using a second date with a time interval of one day reduces performance by 10%. In a more difficult case presented in this chapter performance may decrease by more than 10%. Thus, the next chapter addresses the key issue of consistency of spectral signature across acquisitions.

IX. Chapter 3: Between date sources of variation in hyperspectral airborne imagery over tropical forest canopy

Abstract :

Classifiers trained with airborne hyperspectral imagery can identify tree species in hyper-diverse tropical rainforest. However, the spectral signature varies with the atmospheric conditions and other acquisition characteristics leading to a drop in the performance of classifiers when applied to new imagery (different site or different date). Here we take advantage of repeat overflights of two tropical forest sites in French Guiana to investigate factors affecting spectral similarity across dates. Atmospheric corrections improved spectral consistency between flights in the spectral region from 400 nm to 2000 nm but degraded spectral consistency in the 2000-2500 nm region. This spectral region also showed the lowest SNR. Predictability of SAM divergence between successive overflights varied with the spatial resolution and the spectral region considered but was consistently low (<30%). Between dates SAM divergence in canopy reflectance was primarily explained by the difference in solar position (contributing to about 50-80% of explained variance depending on the spectral region considered) and secondarily by view angle. Local topography only contributed marginally. However application of a RossThick-LiSparse BRDF correction did not improve spectral consistency between dates and only marginally reduced the proportion of SAM divergence explained by the model.

Using the same predictors as those used to predict SAM divergence, the consistency of selected wavelengths was very much improved by a statistical correction. Both a Generalized Additive Model and Neural Network model trained on one part of the image and applied on another part reduced Root Mean Square Difference of reflectance by 10-90% (depending on the wavelength considered).

This study highlights the current limits to transferability of classifiers and the potential importance of integrating more efficient BRDF corrections to extract robust species-specific spectral signatures.

A. Introduction

A major benefit of remote sensing is to provide consistent mapping over large areas. Thanks to the numerous earth observation missions, it is possible to capture many images on many different sites at different times and with different sensors. For operationality and task automation, it is necessary to control for the side effects associated with the variations in acquisition parameters and also inter-calibrate sensors in case multiple sensors are used over time (Baraldi, 2009).

One of the first sources of variations may be the correction related to atmospheric variations in hyperspectral images. (Laybros et al., 2019), showed that single date discrimination of tropical species can be achieved using radiance images (as opposed to reflectance images which are corrected for atmospheric effects). However, when applied across dates species discrimination was improved when based on reflectance data rather than radiance data. Radiative transfer models, thanks to the decomposition of the contributions relative to the various physical processes disturbing the signal, make it possible to estimate the reflectance of the imaged object. Three components contributing to the at-sensor-radiance need to be considered: direct and diffuse solar radiation, radiation reflected by the object and radiation reflected from the object neighbourhood (Richter and Schlapfer, 2018). Each of these components is taken into account in the process of correcting atmospheric effects, which usually assumes Lambertian property of the reflecting surfaces. The atmospheric correction process is, however, imperfect. The review conducted by (Thompson et al., 2018) highlights the sensitivity of atmospheric corrections to errors in estimating the state of the atmosphere for space hyperspectral sensors. They identified several sources of error, the two most important of which are the estimation of water vapour content, which has the effect of absorbing the radiation and the estimation of the optical thickness at 550 nm, which has the effect of scattering the radiation, errors that can be cumulative (Bhatia et al., 2018). More specifically, the uncertainty in aerosol optical thickness (AOT) is more impactful than the uncertainty in water vapor (Bhatia et al., 2015). However, these two parameters are strongly positively correlated as shown on the city of Beijing (Che et al., 2016) and on a rural tropical area in India

(Srivastava et al. 2008). In the present study, two models, one for water vapour and one for optical thickness were used defining the absorption and scattering properties of the considered atmosphere. These models were chosen according to the geographical location. We followed recommendation made in (Richter and Schlapfer, 2018) and used "rural" aerosol type in our sites

A poor estimation of the water vapour column leads to an increase in the absorption of the wavelength at 940 nm and 1140 nm and changes in slope (derivative) especially in the spectral region between 2000 nm and 2500 nm (Thompson et al., 2018). APDA technique can remove the water vapour absorption consequences. Comparing the APDA technique (Schläpfer et al., 1998) applied on Sentinel-2 Imagery and AERONET measurements, (Makarau et al., 2017) showed a high correlation with a low root-mean-square error of about 0.1 cm for water vapour values from 0.2 to 5 cm. In addition, this error had a very low impact on the NDVI index. According to a simulation study, the aerosol concentration estimation errors would affect the slope of the spectrum but especially so for short wavelengths (Thompson et al., 2018). However, as noted by (Bhatia et al., 2015), vegetation spectra may or may not be influenced depending on the characteristic wavelengths of the object in the visible and the amplitude of the signal in the near infrared. From a theoretical point of view for an AOT variation from 0.04 to 0.1, the NDVI error could range from 2% to 5% (Thompson et al., 2018). (Pflug et al., 2015) estimated the AOT for 45 study sites imaged all around the globe by Landsat and Rapid Eye sensors. AOT estimated by ATCOR-4 based on the imagers was compared to data measured on the ground locally (AERONET). The authors found a mean AOT difference of 0.04 with a standard deviation of 0.02, despite the inclusion of cloudy and foggy areas. However, the tropics have a very high and very variable optical thickness, which can be modified by fires, leading to an increase in carbon dioxide content in the atmosphere (Reddington et al., 2016). In addition, in shaded areas, such as shadows cast by trees, light intensity reaching the sensor is close to zero, making the estimate of reflectance particularly complex for atmospheric inversion methods.

As natural surfaces are generally not Lambertian, the acquisition geometry and the illumination (sun-sensor geometry) have an effect on the observed reflected electromagnetic spectrum (Rajan et al., 2006). These variations will therefore have an effect on spectral indices. (Galvão et al., 2013) using the Hyperion sensor (EO-1) analysed the sensitivity of these images to the effects of viewpoint and direction of illumination. They highlighted that the spectral indices were strongly impacted, especially those using information from the NIR region such as the EVI. For a VZA (vertical zenith angle) of + or - 20 °, the EVI spectral index could vary from + 7% to -5%. Conversely, they found that certain spectral indices such as NDVI (Normalized Difference Vegetation Index), SIPI (Structure Insensitive Pigment Index) and VOG (Vogelmann Red Edge Index) were less sensitive to variations in lighting (1% variation for a VZA of + or - 20 °). (Verrelst et al., 2008) performed a similar analysis on a forest and a meadow, using the CHRIS sensor (PROBA), to highlight the effect of reflectance anisotropy on spectral indices. The SRI (Simple Ratio Index), NDVI, SIPI (Structure Insensitive Pigment Index), PRI (Photochemical Reflectance Index) and ARI (Anthocyanin Reflectance Index) had an angular response, which depended on the type of cover.

Airborne hyperspectral remote sensing further suffers from the fact that lighting conditions change from one flight line to the next which is acquired a few minutes later (Bréon and Vermote, 2012; Colgan et al., 2012; Gerard and North, 1997; Jia et al., 2020). Multiple view angles, solar positions and topographic configurations need to be jointly considered to determine the bidirectional reflectance distribution function (BRDF). In addition, BRDF varies with land cover type (Schlapfer et al., 2015). Knowing the local BRDF function allows to normalize the directional reflectance from the non-nadir view to the reflectance at the nadir. BRDF is also resolution dependent. For instance, (Roman et al. 2011) analyzed the surface BRDF at 3 m and 500 m resolution over a mixed agricultural landscape. They argued that coarser spatial resolution was preferable to finer spatial resolution using a RossThick-LiSparse Reciprocal (RTLSR) model, due to assumption of linearity in kernel-driven BRDF model, which depended on the weighted sum of isotropic-scattering, volume-scattering and geometric-optical scattering kernels.

(Pacifici et al., 2014) studied a time series of 21 images acquired between 2002 and 2009 by QuickBird over the city of Denver, Colorado. They found that the atmospheric and geometric properties of the acquisitions strongly affected the calculated reflectance values of the pixels. A Kappa improvement of 0.374 was achieved over raw pixels when the surface reflectance values were combined with the angular decomposition classifying 22 urban classes. In addition, the RPV model (Rahman – Pinty – Verstraete, (Rahman et al., 1993)) which decomposes the angular surface reflectance observed into three independent components representing the amplitude, anisotropy and the asymmetry factor (which controls the relative amount of back and front scattering) was beneficial in identifying vegetation classes. The anisotropy coefficient will depend on the structure of the object and of its surrounding. In case of trees many parameters may be influential : shape of the tree crown, branching pattern, tree foliage density, foliage elements size and orientation distributions, soil contribution, adjacency effect, etc… (Korpela et al., 2011).

Variation in signal intensity associated with lighting conditions can be controlled by normalizing spectra to a common baseline. One such option is the continuum removal, which considers the convex hull fit over the top of a spectrum as the base line (Clark and Roush, 1984). It was argued that this correction minimizes the residual effects of atmospheric correction and BRDF for a better estimate of the water concentration in the leaves (Kokaly and Clark, 1999).

An additional complicating parameter is the local slope of the relief with respect to the incident light (Richter and Schlapfer, 2018; Santini and Palombo, 2019). The topography is generally given by LiDAR measurements providing an accurate DEM (Digital Elevation Model) from which the slope and aspect can be mapped. The common point to all topographic correction methods is the correction of variable illuminance as a function of the slope. With slope values of 10 ° to 20 °, topographic effects are moderate (Richter et al., 2009). (Oliveira et al., 2019) applied a C-correction as proposed by (Teillet et al., 1982). From non-corrected to corrected data, a reduction in reflectance variability was observed. The standard deviation was reduced in the near-infrared (6.45%), the SWIR-1 (3.3%), the "red-edge" (2.47%) and the visible (0.37%). Cosine correction only modifies the direct part of the incoming irradiance while diffuse irradiance is not taken into account, as described by the correction equation (Meyer et al., 1993). In the same way, the red and near infrared reflectance increases when the rate of overlap between crowns is important and that whatever the direction of view (Gerard and North, 1997).

Spectral distance functions and metrics are useful tools to quantify spectral variation over an entire spectral range. (Deborah et al., 2015) performed an evaluation of function and spectral distance metrics for hyperspectral imagery taking care to consider the underlying assumptions and mathematical construction (triangular equality, identity of indistinguishable, monotonous increase). As noted by the authors, the RMSE (Root Mean Square Error) distance function differs from the SAM function because the latter does not comply with the assumption of the identity of the indistinguishable, due to the inverse cosine function. However, the advantage of the SAM is its insensitivity to the variation in intensity. Many other distance measurements considering a spectrum as a distribution can be used to quantify these differences and we refer the interested reader to (Deborah et al., 2015) .

Adaptation domain methods (Gross et al., 2019; Matasci et al., 2012) may help addressing the issue of transferring a species classifier trained under particular conditions to new conditions (different time or area for instance). It is however of interest to first identify the source and quantify the amplitude of the variations between acquisition dates. Indeed, identifying the causes of the differences could guide the development of generic methods. The aim of this study was therefore to characterize the sources of variations in hyperspectral data (at sensor radiance and reflectance). The ultimate goal is to develop robust methods to identify particular tree species over large areas of forest without the need to retrain a classifier with ground inventory for every new site. We first assess the noise level variation across spectral regions in order to assess signal quality for each spectral range. Then, we evaluate the gain in signal stability brought about by atmospheric corrections by comparing repeated flights over successive days. We do so both for individual crowns and patches of pixels.

We further take advantage of the particular flight plan on one site to explore the impact of the acquisition geometry (sun and the sensor positions) on spectral divergence between dates. Finally, we

examine the potential of wave length specific statistical correction to improve spectral consistency across dates.

B. Material & Method

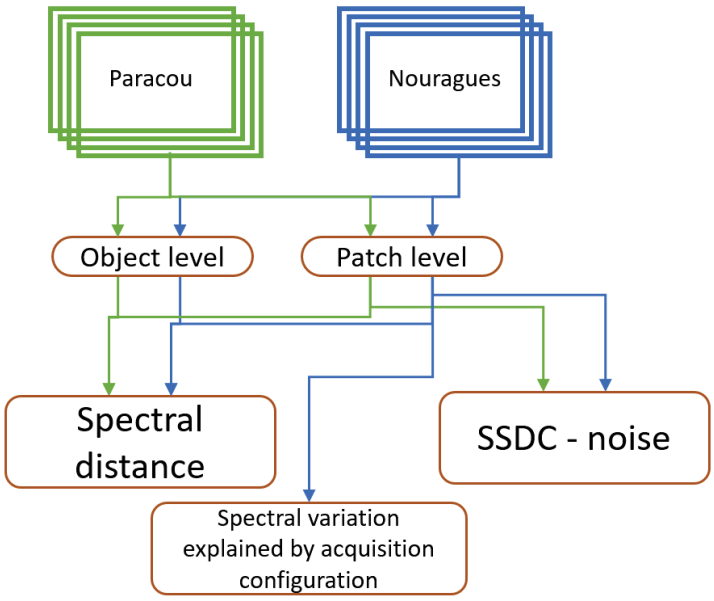

Figure 1: Workflow of the study. The blue boxes refer to hyperspectral imageries. The oranges boxes refer to the methods used in the study.

1. Study Sites

Two forest sites in French Guiana were considered for this study: Paracou and Nouragues.

The Paracou research station is located in the north of French Guiana (5°16' N, 52°55' W), about 15 km from the coast (Gourlet-Fleury et al., 2004). The annual average temperature is 26 ° C with a variation of + -1 ° C. Precipitation is highest in May and lowest in October-November. The rainfall is around 2875 + - 510 mm per year. The Paracou site is a hilly area. The altitude varies from 5 m to about 45 m above sea level.

The Nouragues station is located in the centre-east of French Guiana, 100 km from the coast (4°05' N, 52°40' W) (Bongers et al., 2013). The average temperature is similar to the Paracou site with a high average relative humidity, between 80 and 90%, depending on the season. The rainfall is 2990 mm / year. The relief is made of hills and plateaus. The elevation of the site varies from 60 m to 420 m at the top of the inselberg.

2. Hyperspectral data

For each site, hyperspectral data was acquired combining two sensors: a Hyspex VNIR-1600 (Hyspex NEO, Skedsmokorset, Norway) sensor mounted alongside a SWIR-384 (Hyspex NEO, Skedsmokorset, Norway) sensor. At the same time, a LiDAR laser scanner Riegl LMSQ780 acquired data point clouds. The VNIR sensor covers the range from 414 nm to 1 000 nm, discretized into 160 spectral bands. The SWIR-384 sensor covers the range from 1000 nm to 2500 nm discretized into 288 spectral bands. The spectral sampling was 3.7 nm and 5.45 for the VNIR and SWIR sensor respectively.

a) Date and time acquisitions

Overflights of the Paracou site took place on September 19, 2016 and September 20, 2016. The first acquisition started at 1:12 PM and finished at 3:38 PM local time, starting on the western side of the site. The second acquisition started at 2:14 PM and finished at 3:00 PM, covering a smaller area, about one third of the area imaged on the first date. Overflights of the Nouragues site took place on September 21, 2016 and September 22, 2016. The first acquisition started at 2:39 PM and finished at 4:38 PM. A low proportion of cloud shadows was present (Tab.1). The second acquisition started at 2:38 PM and

finished at 4:38 PM. A greater proportion of shadows was present on the second day (Tab.1). The flight plans were similar but the first overflight of this area started in the west and ended in the east, while the second flight started in the east and ended in the west.

Table 2: Overflights characteristics. The proportion of cloud shadow is based on visual delineation of shadows on the hyperspectral imagery.

Date	Site	Start time (UT -3, PM)	End time (UT -3, PM)	Altitude above the sea level (m)	Cloud shadow (%)	Poorly illuminated pixel (%)
09/19/2016	Paracou	1:12	3:38	920	4.8	9.3
09/20/2016	Paracou	2:14	3:00	920	1.2	25.1
09/21/2016	Nouragues	2:39	4:38	1100	23.0	82.4
09/22/2016	Nouragues	2:38	4:38	1100	63.9	86.6

b) Corrections applied

The hyperspectral images were radiometrically corrected using static calibration parameters of the manufacturer taking into account the dark signal, the bad pixels and the spectral response functions (SRF) depending of the FOV. Then, hyperspectral images were orthorectified and georeferenced at 1m spatial resolution for VNIR sensor and at 2m spatial resolution for SWIR sensor with PARGE software (Richter and Schlapfer, 2018). A 1m-DSM (Digital Surface Model) produced from the LiDAR point could was used for a fine orthorectification. The DSM was constructed using the data point cloud by selecting point of maximum height on a 1m resolution grid. VNIR and SWIR images were merged based on geographic positions using PARGE software at 2m spatial resolution.

Each flight line was corrected using ATCOR-4 software to remove atmospheric perturbations (Richter and Schlapfer, 2018). The same atmospheric look-up table file with a rural aerosol and a water vapour column of 0.4 g.cm^{-2} was used for both sites. As the aircraft flew between 920 and 1100 m altitude, the water vapour column was estimated for an altitude of 1000 m. Atmospheric corrections between Nouragues and Paracou can be compared to field measurement taken during the campaign and accessible using the following link: https://aeronet.gsfc.nasa.gov/cgi-bin/draw_map_display_aod_v3. According to (Bhatia et al., 2018), the amount of aerosol optical thickness can be considered low to moderate (less than 0.25) at 500 nm on both days at both study sites. A single AOT estimate per date was considered (rather than a spatially explicit version) (Schläpfer et al., 2018). The water vapour column, on the other hand, was high (higher than 3.5 cm). Some spectral ranges were not recoverable despite the use of atmospheric corrections due to excessive noise. We removed wavelengths between 1318 nm and 1503 nm, between 1742 nm and 1964 nm and beyond 2450 nm. Therefore, we used 355 spectral bands. Different sources of shadows can interfere with the spectral value measured by the sensor, the main ones being the shadows cast by trees and clouds. Shaded pixels negatively affect the results of classifications. We used a specific method to remove pixels that were considered poorly illuminated. We also masked non-forest areas such as rivers, roads and tracks. We did not mask out savannahs and rocky outcrops (Inselberg) which however were classified as non-forest areas. We calculated the illumination index using ATCOR-4 software based on the radiance signature and topographic information for the Paracou and Nouragues sites.

We applied a BRDF correction using BREFCOR model, available in ACTOR-4 software. BREFCOR method (Schlapfer et al., 2015) characterizes surface type using a fuzzy classification and then computes anisotropy factor for correction using a semi-empirical BRDF model. In this study, the BRDF model was computed for five classes defined by the BCI index as proposed by default. Before this step, we masked all non-forest land cover: bamboo patches, water bodies, and a rocky outcrop. We compared spectral consistency across dates either before or after applying a BRDF correction. We expected to see a reduction in the spectral signature difference between dates. For this study, three spectral regions were considered: VNIR (500 - 1318 nm) SWIR-1 (1508 - 1740 nm) and SWIR-2 (1970 – 2450 nm).

c) *Statistical preprocessing*

Radiance (also named "Geo") and reflectance (also named "Atmo") raster stacks were smoothed by applying a 3x3 mean filter. In addition, we normalized spectrum of each pixel by dividing by the mean VNIR spectrum value. Indeed spectral normalization was found to significantly improve tree species discrimination in a previous study (Laybros et al., 2019). This normalization was applied even though we finally selected the SAM index as a measure of spectral similarity. Normalization has no effect on the SAM values.

d) *Mosaicking*

For each site and each date, hyperspectral flight lines were mosaicked using the center-cropped option in PARGE, which reduced the variation in scan angle by preferentially selecting the most central part of each flight line. However, some flight lines were marred by cloud shadows. Cloud shadow masking may have resulted in a decrease in the rate of overlap between flight lines and increase scan angle variability.

We overlaid the hyperspectral mosaics at both dates to estimate the spatial shift and potential distortions and found that alignment was consistent within +/- 2 m.

In each site, the flight lines of the common area covered on successive days were separated into 3 groups to build three non-overlapping mosaics: west, center and east of the area. Width of the area repeatedly imaged differed significantly between sites (Paracou ~1.1km, Nouragues ~2.6km).

Spectral consistency of atmospheric-corrected data (reflectance) and uncorrected data (radiance) was systematically tested.

3. SNR evaluation

Noise affects the quality of the hyperspectral signal and denoising is delicate because of the need to preserve spatial and spectral information. The noise can come from different sources (Rasti et al. 2018). Several ways have been proposed to estimate the noise contribution relative to signal. As highlighted in (Gao et al., 2013) working with AVIRIS data, the different noise estimation methods generally give similar results. We evaluated the quality of the hyperspectral signal using the SSDC (spectral and spatial de-correlation) method (Roger and Arnold, 1996). The noise estimation is performed on different non-superimposed blocks of a fixed size and preferably homogeneous. It uses spectral correlation and spatial correlation to identify the noise component by linear regression. A pixel value is predicted using spectral and spatial neighbours value. The residue is considered noise.

The Signal-To-Noise ratio (SNR) was computed using a 13x13 pixels window (26x26m neighbourhoods). An illumination filter was used to exclude individual pixels with an illumination index below 0.6. For a robust linear regression and to correctly estimate the noise, the sampling must be sufficient. Among the 169 pixels, if more than 40% of pixels were removed then the calculation was not performed for the window.

4. Manually delineated crowns

In the Paracou and Nouragues sites, we manually delineated individual crowns well-illuminated on both dates (visual control). We used RGB images at 10 cm resolution as well as the canopy height model derived from LiDAR point clouds to delineate those crowns. No ground truth data was available to validate the quality of the segmentation.

Table 3: Number of crowns used for each study site.

Study site	Number of crowns	Area (m2)
Paracou	106	26260

Nouragues	145	44619

5. Patch level

We partitioned the hyperspectral images into 5x5 pixels patches. We used patches rather than single pixels to mitigate misalignment effects between dates. As the shadows have a strong impact on the measured reflectance, we filtered the shadows as explained in the noise evaluation paragraph (permissive configuration). Another configuration, the most restrictive (restrictive configuration), used a patch only if 100% of its pixels had an illumination ratio above 0.6. Analyses were conducted for both levels of illumination quality.

6. Dissimilarity index

The spectral angle between dates was computed between the spectral mean over all well-illuminated pixels of each patch/object. The SAM (Spectral Angle Mapper) metric determines the spectral similarity between two spectra by calculating the angle between the spectra treating them as vectors in a space with dimensionality equal to the number of bands. Its main advantage is it insensitivity to the illumination variation.

$$d_{SAM}(S_1, S_2) = cos^{-1}\left(\frac{\sum_\lambda S_{1,\lambda} \cdot S_{2,\lambda}}{\sqrt{\sum_\lambda {S_{1,\lambda}}^2}\ \sqrt{\sum_\lambda {S_{2,\lambda}}^2}}\right)$$

NRMSD (normalized Root Mean Square Deviation and SID (Spectral Information Divergence) were also computed. The responses were very similar whatever the spectral similarity measure used. Results are presented in appendix.

7. Evaluation of parameter acquisitions

To explore the contribution of the different predictors to the spectral divergence, we first built coarser resolution mosaics by aggregating radiance/reflectance values by 3x3, 5x5 or 9x9 pixels. We then computed between dates SAM values for each cell. We further selected the cells to which apply the statistical analyses by considering two illumination thresholds. In the most restrictive case only patches in which all the 2m pixels had an illumination ratio higher than 0.6 were kept. In a less stringent scenario, similarly to the SNR analysis, cells were considered if at least 60% of the pixels were well illuminated. In which case only those well-illuminated pixels were used to compute the cell mean spectrum. In addition, we also tested the application of a Savitzky-Golay spectral smoother by second order polynomial fit over five consecutive bands (equal to 18.5 nm for the VNIR sensor and 27.3 nm for the SWIR sensor).

We explored how between dates spectral divergence as measured by SAM was related to various predictors. We conducted this analysis on the Nouragues site only since it showed the largest solar variation between dates (see M&M section). To evaluate the contribution of different predictors we opted for Generalized Additive Models (GAM) thereby avoiding making parametric assumptions with regard to the shape of the response. In GAM the response variable depends linearly on unknown smooth functions of some predictor variables (Hastie and Tibshirani, 1999).

Because of the strong spatial correlation expected to occur in spectral response divergence and to avoid overfitting we used block cross-validation (Valavi et al., 2019) . We subdivided the area between north and south blocks and trained the model using one-half of the data and evaluated its prediction on the other half. We report the R^2 of the linear regression of predicted versus observed values as the goodness of fit statistic. We evaluated the contribution of the different predictors by recording the drop in explained variance when that predictor was omitted from the model (using the same cross validation method by blocks to assess the explained variance with or without the predictor examined). The

reflectance of many cover types depends on the viewing and solar illumination geometry. Added to this angular dependence between the sun and the sensor are the relief effects, which can be captured using topographic information. The Bidirectional Reflectance Distribution Function (BRDF) uses this information to restore the spectral signature so as to be independent of the acquisition geometry. In this study we explore how much the spectral consistency between dates is affected by viewing and solar angle, in conjunction with local topography.

Nouragues site is mostly covered in forest. The river in the south and the large barren rocky outcrop in the north were identified using RGB imagery for the analysis. In addition, in Nouragues bamboo thickets (locally referred as "cambrouse") create openings of a few hectares in the forest canopy. A mask was created for each of the non-forest categories.

We computed the elevation and azimuth angles of the sun for each pixel based on the time acquisition and geographic position. We also considered absolute scan azimuth (horizontal angle measured clock wise from true north), and the scan zenith angle. We also computed the slope and the aspect angles based on the canopy surface model re-sampled at 10 m spatial resolution. The predictors description and range of variation are reported in Tab.3.

Depending on whether the configuration was restrictive or permissive, and the size of the patches considered, the number of observations varied widely (from 790 000,000 to 202; see appendix 3). We divided the dataset in two by the median Y coordinate: the north part and the south part. We randomly selected 10,000 pixels from the training set when more than 10,000 observations were available to speed up calculation.

Table 4: Pixel level characteristic. Range is given in degree.

Origin	Predictor	Description	Range (degree)
Sun	Sun Elevation	Absolute difference in sun elevation between successive flights	26.4, 55.6 26.2, 55.8
	Sun Azimuth	Absolute difference in sun azimuth between successive flights	264.7, 268.7 264.0, 267.9
Sensor	Absolute Azimuth	Absolute difference in sensor azimuth viewing angle	0, 360
	Scan Zenith	Absolute difference in sensor zenith viewing angle	0.0, 14.6 0.0, 12.8
Topography	Slope	Slope of Canopy Surface Model	0.5, 53.1
	Aspect	Aspect of Canopy Surface Model	0, 360
Combination	Sun – sensor angle	Absolute difference in the angle between the pixel-sun vector and the pixel-sensor vector	29.7, 74.1 25.9, 72.3

8. Spectral correction

In this paragraph, we propose a method to correct the between date differences in five spectral bands (556 nm, 760 nm, 1058 nm, 1639 nm and 2246 nm) positioned in each of the spectral ranges (VNIR, SWIR-1, SWIR-2). A GAM model and a neural network model (NN) were used and compared. The NN model was made up of 17 layers, consisting of 32 to 102 neurons per layer. Only "Relu" functions were used, except for the last layer (linear) (more details are given in appendix 6).

The data acquired in the north of the Nouragues site are used for the training of classifiers. Data in the south is used for prediction. We then reversed the procedure using the southern half for the training and the northern half for the prediction. We pooled the prediction of the two steps and we analyzed them jointly. Only pixels with an illumination ratio higher than 0.6 were considered. The models aimed to predict the difference in reflectance of specific wavelengths observed between the two acquisition dates. No specific transformation was applied to topography related predictors (slope and aspect).

C. Results

1. SNR evaluation

We evaluated the noise per spectral region, both before and after atmospheric correction for three subzones west, center and east in each site.

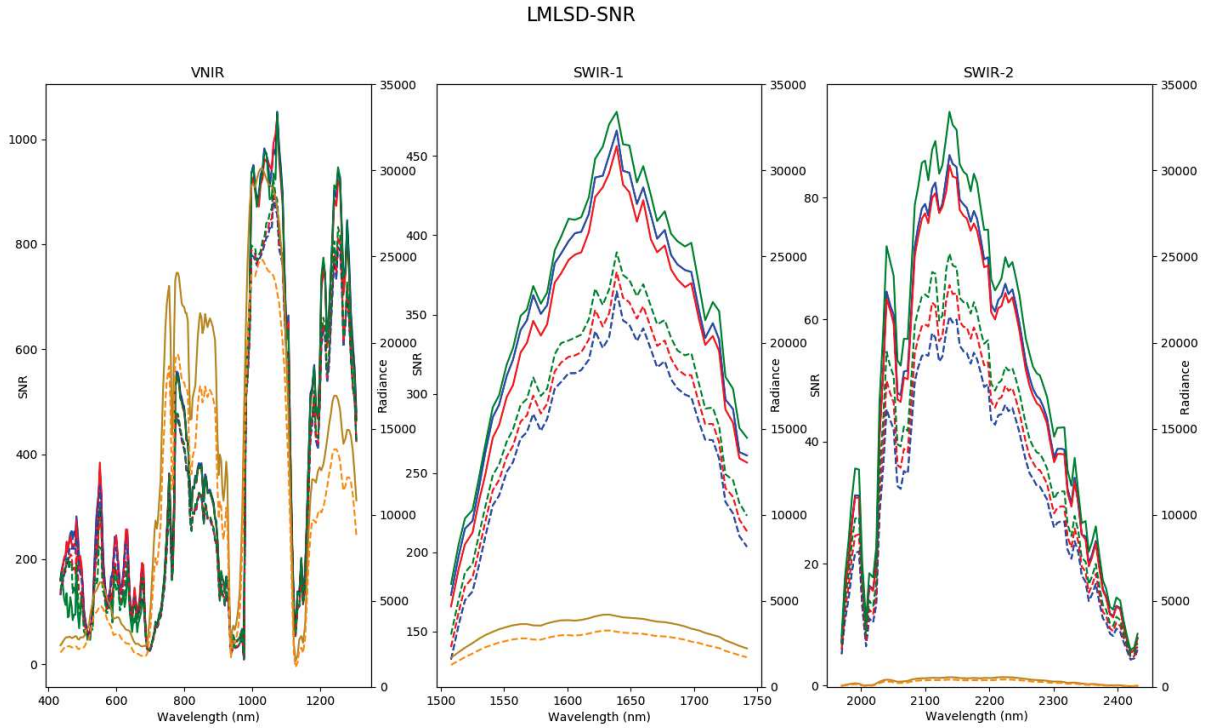

Figure 2: Paracou radiance. SNR in the west (green lines), the middle (red lines) and the east (blue lines) of the overflight prior to applying atmospheric corrections, per spectral region. The brown and orange spectra are the spectral means of the site. The first date in solid lines. The second date in dashed lines

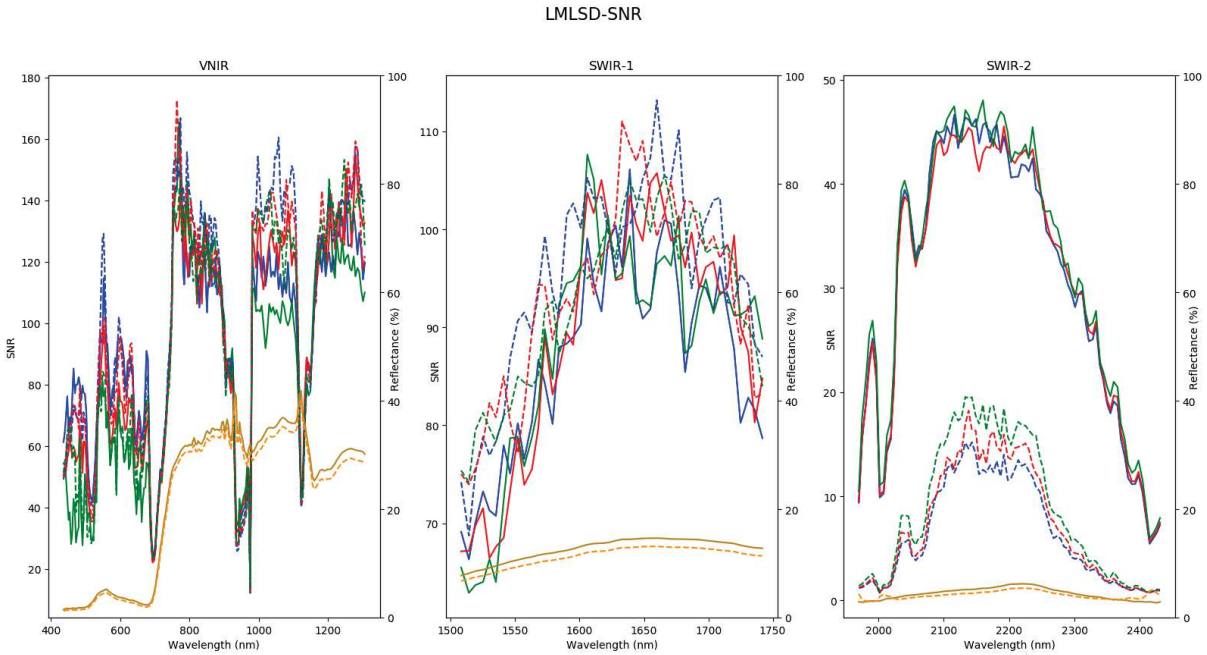

Figure 3: Paracou reflectance. SNR in the west (green lines), the middle (red lines) and the east (blue lines) of the overflight after atmospheric corrections were applied per spectral region. The brown and orange spectra are the spectral means of the site. The first date in solid lines. The second date in dashed lines.

The SNR decreased from VNIR to SWIR-1 and even more so to SWIR-2. The hyperspectral image acquired on Paracou on the first date was less affected by noise than on the second date (Fig.2 and Fig.3), in particular over the spectral range from 1900 nm to 2450 nm. Atmospheric corrections tended to reduce the SNR. There was no clear strong geographic (E-W) pattern.

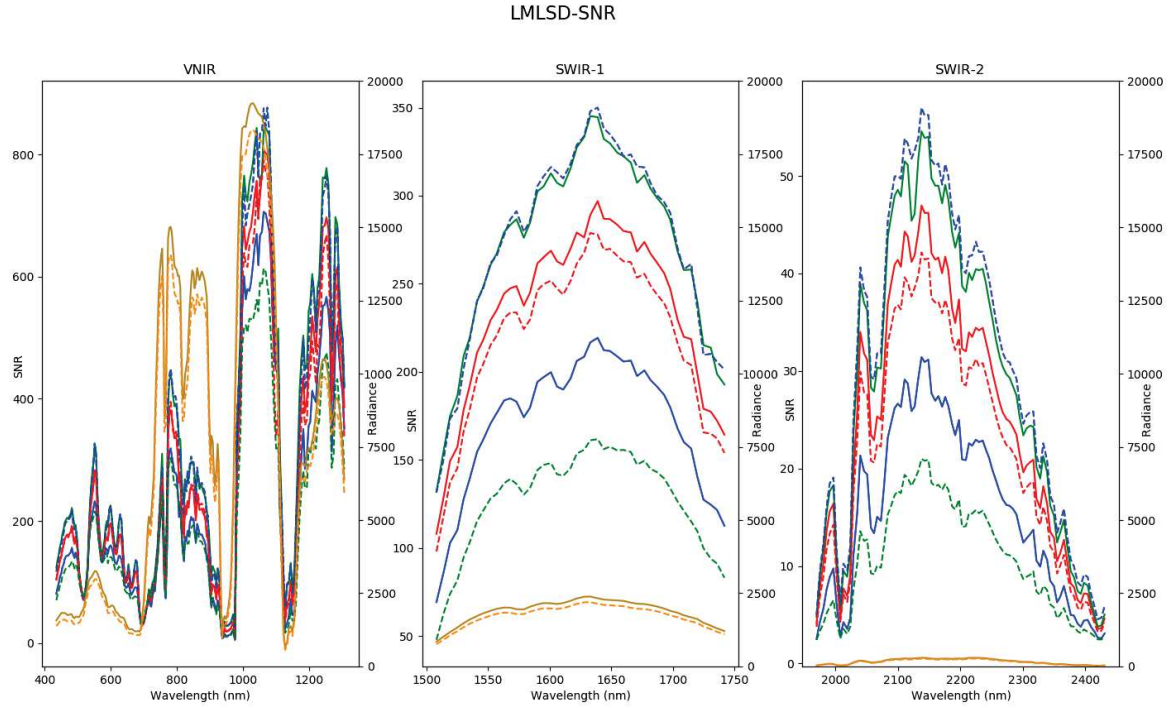

Figure 4: Nouragues radiance. In each of three spectral regions, we compared the SNR in the west (green lines), the middle (red lines) and the east (blue lines) of the overflight prior to applying atmospheric corrections. The brown and orange spectra are the spectral means of the site. The first date in solid lines. The second date in dashed lines.

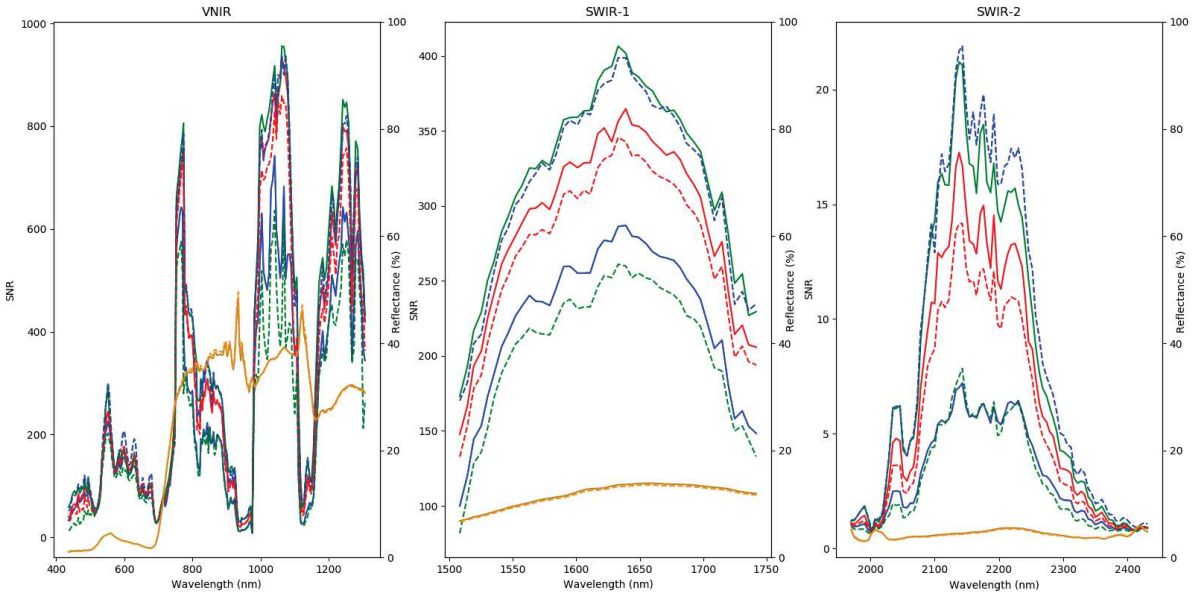

Figure 5: Nouragues reflectance - In each of three spectral regions, we compared the SNR in the west (green lines), the middle (red lines) and the east (blue lines) of the overflight after applying atmospheric corrections. The brown and orange spectra are the spectral means of the site. The first date in solid lines. The second date in dashed lines.

Reflectance-SNR (Fig.5) was also significantly lower than radiance-SNR (Fig.4) in the Nouragues site. In addition, the SNR showed a clear spatial pattern of variation. The central area had similar SNR on both dates, while data acquired later in the afternoon on either date had a lower SNR.

2. Spectral discrepancies

To begin, we present the analysis at object level, using manually segmented crowns for both sites. We then present results obtained using 5x5 pixels windows on a much larger data set.

a) Object level:

To ensure the best object matching between dates, crowns were delineated independently on each mosaic (flight date). We computed the spectral mean per object before comparing spectral distance in terms of R-square and SAM.

i. R-square

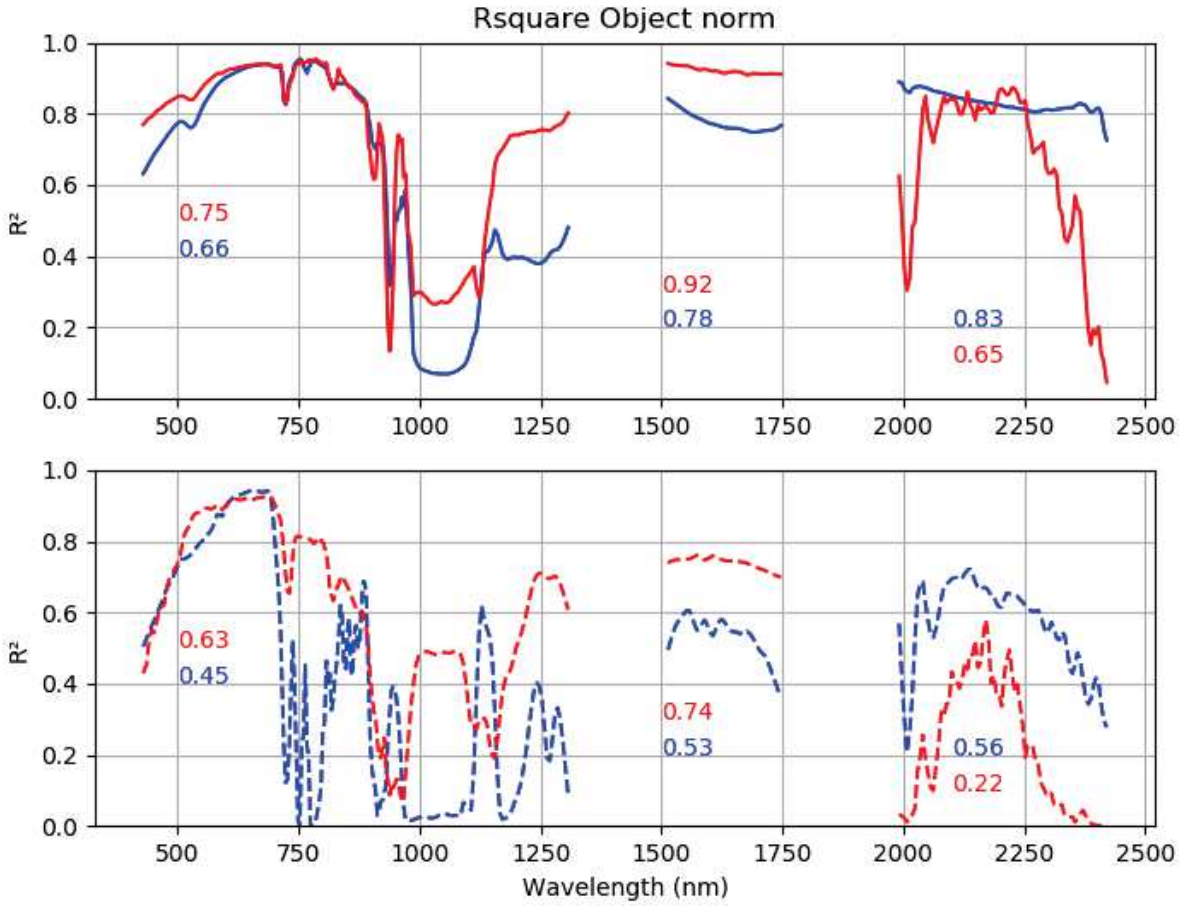

Figure 6: Spectral correlation between same crowns viewed on two dates in Paracou (full lines) and Nouragues (dashed lines) site. The red and blue lines represent the dataset atmospherically corrected or not respectively. The figures are the average R^2 value per spectral region.

Fig.6 compares the spectral consistency for the same objects seen at a time interval of 1 day using the squared Pearson correlation coefficient. Atmospheric correction mostly increased the consistency of the spectral signature from 500 nm to 1750 nm. Beyond this threshold, the atmospheric correction considerably degraded the consistency between the dates.

ii. SAM metric

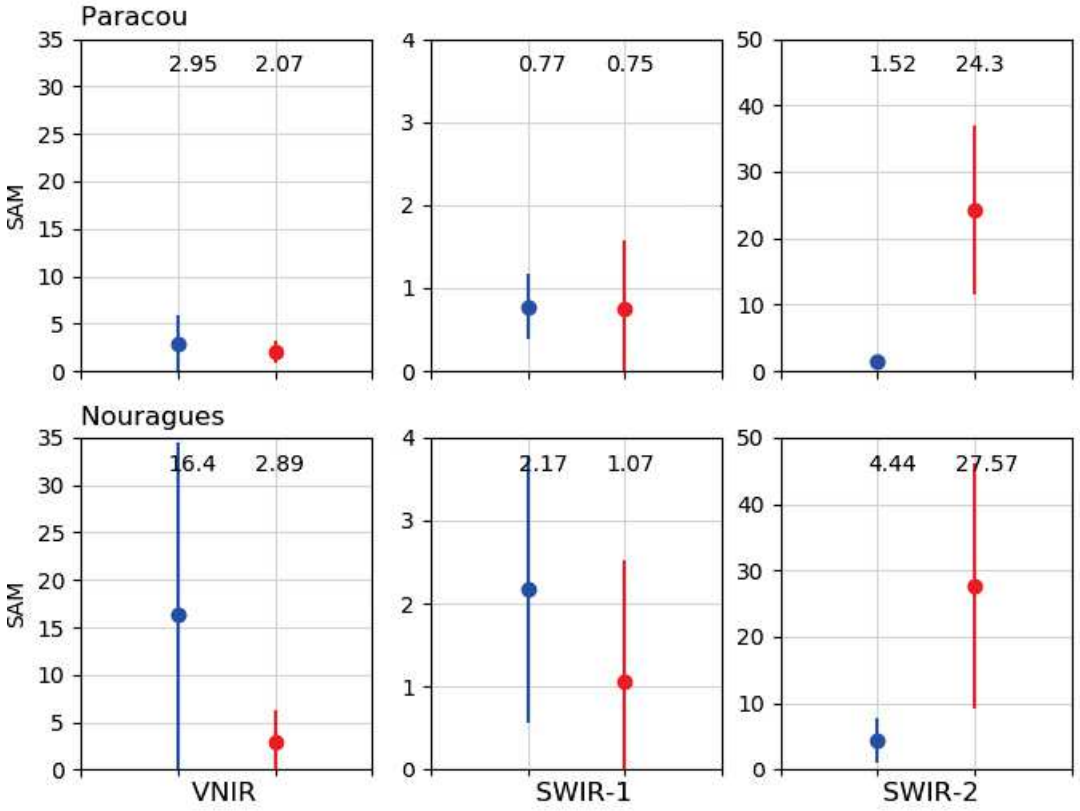

Figure 7: Paracou and Nouragues sites – For each site, comparison of individual crowns SAM metric according to the spectral range considered. Geo (blue) uses at sensor radiance values. Atmo (red) uses reflectance spectra (after atmospheric correction). Median value written above each boxplot.

Using SAM the average reflectance and radiance spectrum of each object was compared to the average reflectance and radiance spectrum of the same object seen on the second date (Fig.7). Data corrected for atmospheric effects were on average more consistent than uncorrected data (lower median SAM value) except in the SWIR-2 range. We observed the same behaviour for the RMSD and SID metrics. The correlation matrices between the spectral distances (calculated over the entire spectrum) SID, SAM and RMSD are given in the appendix (Fig.15 and Fig.16).

SAM values were always lower in Paracou indicating higher discrepancy between dates in Nouragues.

b) Patch analysis:

In this part, we considered a window size of 5x5 pixels to evaluate different metrics. This allowed to consider many more observations compared to the object based approach (Tab. 6).

i. R-square

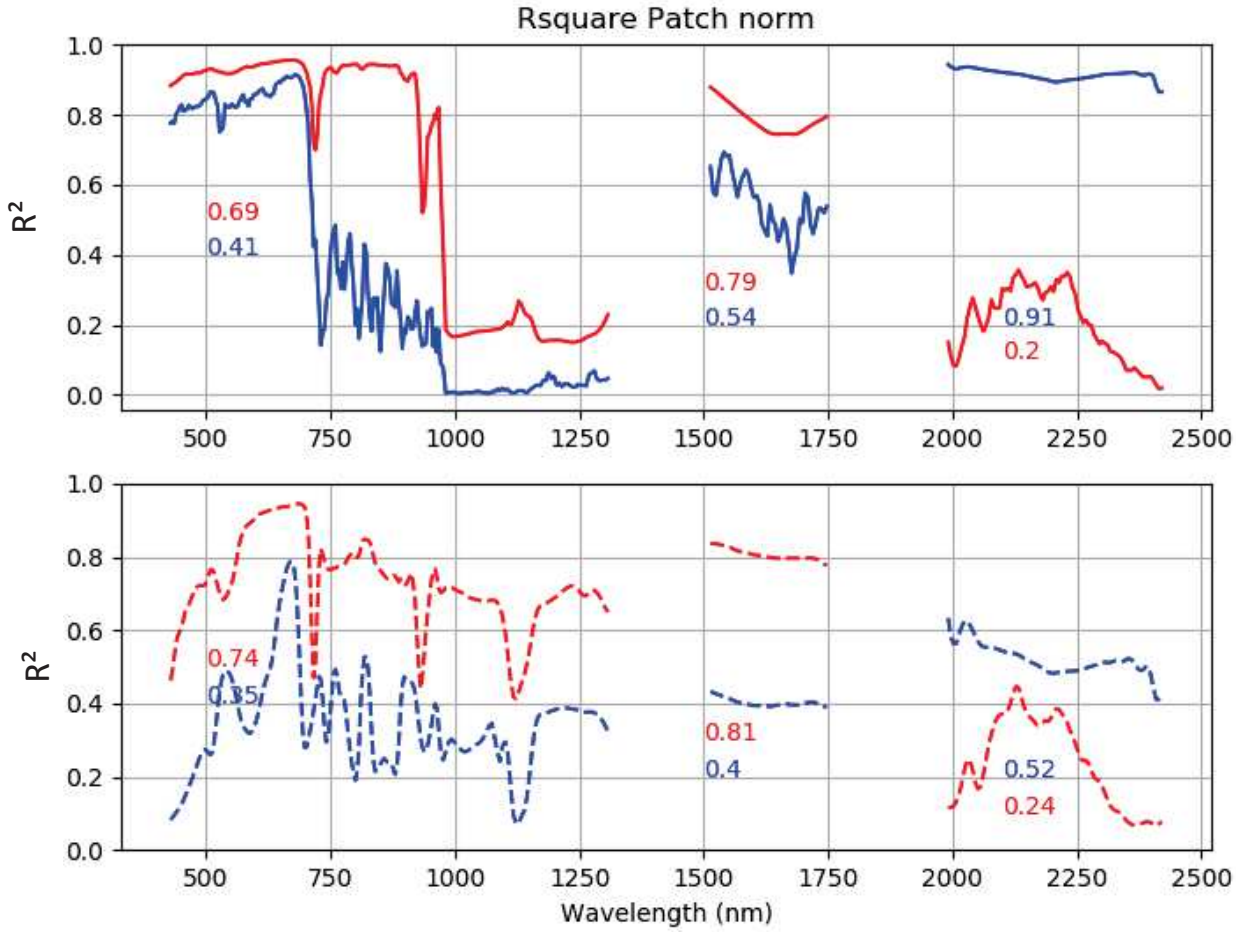

Figure 8: Relation between same patch (5x5 pixels) viewed at the two dates in Paracou (full lines) and Nouragues (dashed lines) site. The blue and red lines represent the dataset no corrected and corrected by atmospheric correction respectively.

The pattern found was similar to the one observed using selected crowns. Reflectance consistency was higher in Paracou than in Nouragues for patches like for individual crowns. A large fluctuation in R^2 was observable for crowns in Paracou in the SWIR-2 part (Fig.6). These fluctuations were much less when considering patches instead of crowns (Fig.9) but the overall consistency was also much lower.

Whether at the object or patch level, sudden changes in consistency appeared between two contiguous wavelengths. Inconsistencies around 950 nm may be due to the transition between the two sensors (Lenhard et al., 2015). The observed dip in R^2 around 1150 nm in Nouragues possibly reflected poorly corrected atmospheric disturbances.

ii. SAM metric

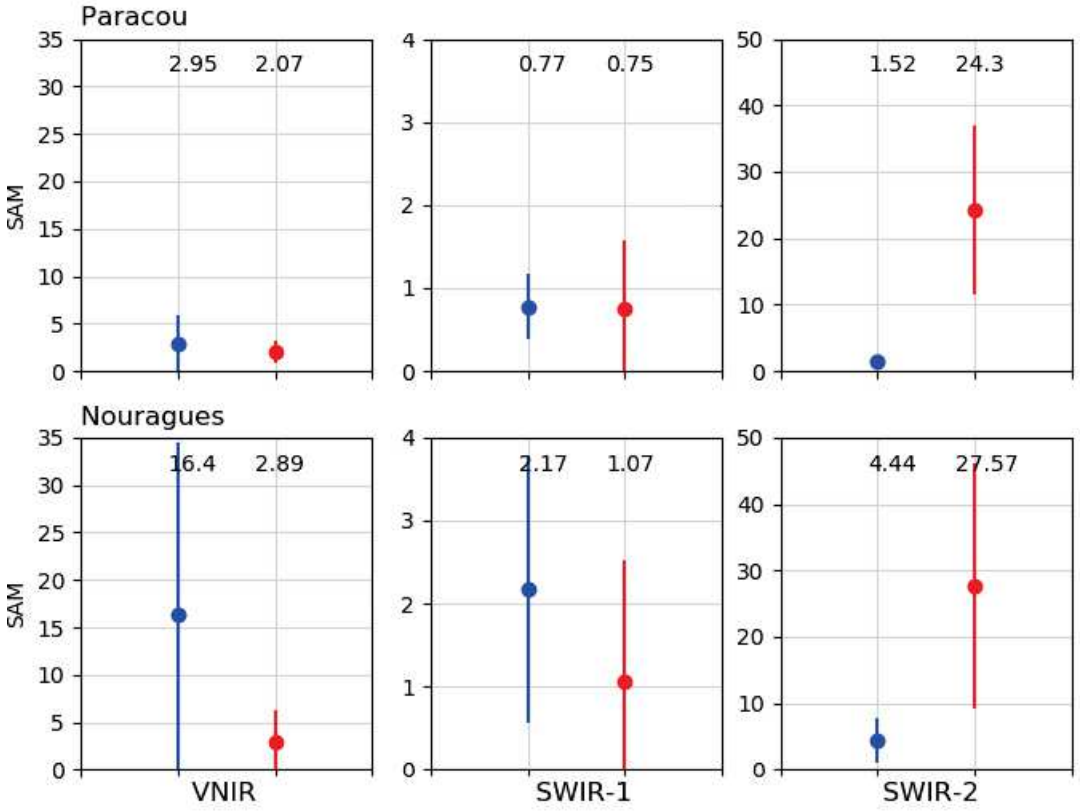

Figure 9: Paracou and Nouragues sites – Between date SAM for two sites and three spectral ranges. SAM are computed based on mean spectrum of 5x5 pixels patch. Geo (blue) uses at sensor radiance values. Atmo (red) uses reflectance spectra (after atmospheric correction). Median value written above each symbol.

The results were very similar for different similarity indices tested (see correlation matrix between indices in appendix, Fig.15 and Fig.16). The only slight discrepancy occurred for SWIR-1 in Paracou where R^2 was significantly increased following atmospheric corrections (Fig.8), but SAM was not (Fig. 9). Atmospheric corrections reduced SAM vales much more in Nouragues than in Paracou.

The results at the crown scale (Fig.8) or patch scale (Fig.11) were very similar. The different metrics were all consistent.

3. Sources of between date spectral divergence

We now explore the spatial variation in spectral divergence along the east-west direction. We consider 5x5 pixel windows to compute SAM dissimilarity index between dates and report those values for the three sub mosaics: on the western side, the middle, and the eastern side. Both radiance (blue color) and reflectance (red color) dissimilarity are computed.

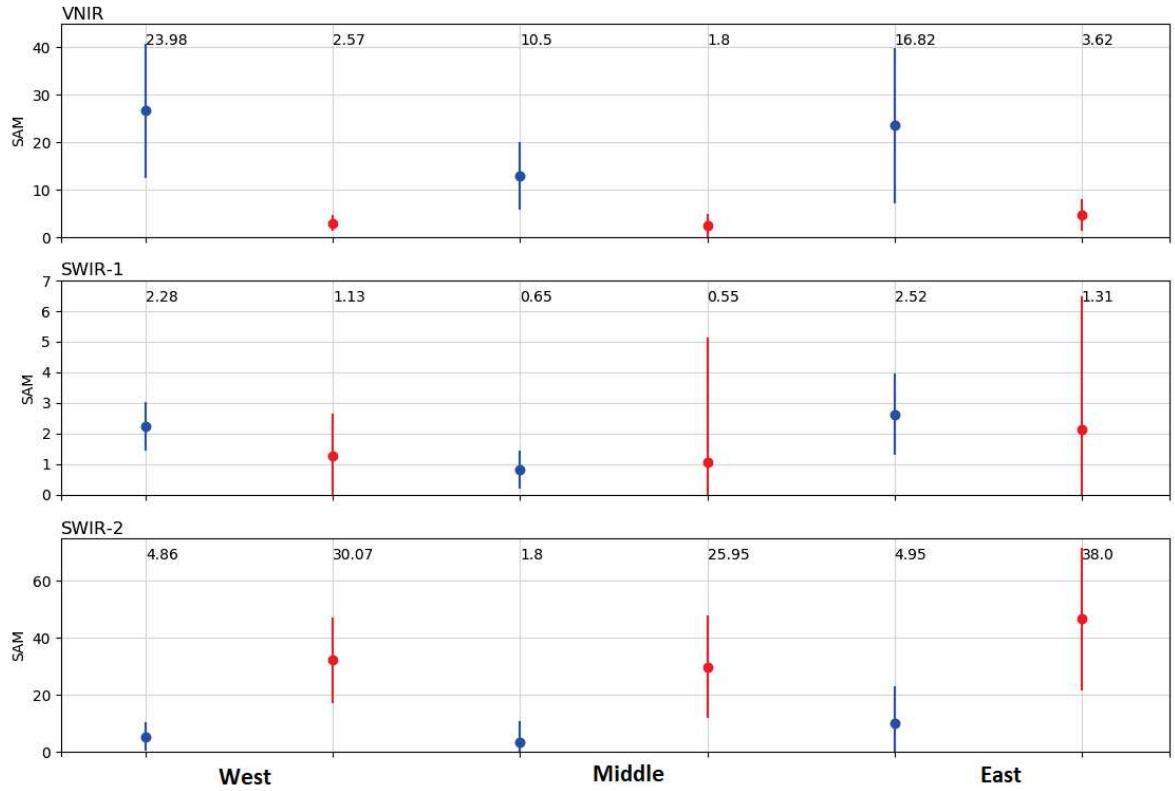

Figure 10: Characterization of spectral dissimilarity by the SAM index (5x5 patches) at the Nouragues site in three sub-zones (left to right) and three spectral ranges (top to bottom). Blue and red symbols stand for radiance and reflectance data respectively.

The overflights, which took place on successive days began on opposite sides of the area (Fig.10). The first one started in the east and the second one in the west. For each flight strip, the same trajectory (flight path and flight direction) were maintained. We compared the average and the median SAM value. The middle part of the mosaic with the least difference in solar position between dates (Table 5) was the most consistent across dates (see median value, Fig.10). In addition, data in the east were significantly less consistent that in the west.

Table 5: Proportion of well-illuminated pixels for each configuration. An illumination ratio higher than 0.6 was considered as good.

		Good illumination (%)	Solar acquisition time
Western side	T1	20.7	4:29PM to 4:38 PM
	T2	37.5	2:39PM to 2:48PM
Middle	T1	25.0	3:34PM to 3:49 PM
	T2	25.9	3:27PM to 3:39PM
Eastern side	T1	40.3	2:40PM to 2:49 PM
	T2	21.2	4:30PM to 4:39 PM

There was an opposite variation in the quantity of well-illuminated pixels between the two dates along the east-west axis. (Tab.4).

4. Sun-Sensor variation

In this analysis, we evaluated the contribution of the variation in acquisition parameters (sun position and sensor orientation, and topographic variations) to the spectral dissimilarity computed

using SAM index. We considered two levels of illumination for selecting patches. Only reflectance spectral signatures were used.

Table 6: SAM statistics (median and standard deviation) for each spectral ranges and considering two type of landcover. With and without SG indicates whether Savitzky-Golay spectral smoothing was applied or not. Permissive illumination configuration was used.

		Cambrouse					Forest				
	Size	Full	VNIR	SWIR-1	SWIR-2	Pixels	Full	VNIR	SWIR-1	SWIR-2	Pixels
Without SG	1	17.6 (±13.2)	4.0 (±4.4)	1.0 (±4.7)	26.8 (±42.4)	40113	18.5 (±11.7)	4.1 (±2.8)	1.3 (±2.2)	59.2 (±39.8)	716057
	3	15.5 (±12.5)	3.4 (±2.7)	0.9 (±2.6)	20.9 (±26.6)	4107	15.6 (±9.6)	3.6 (±2.0)	1.2 (±1.3)	44.9 (±21.4)	54710
	5	14.3 (±12.5)	2.8 (±2.3)	0.8 (±1.0)	21.0 (±21.4)	1352	12.7 (±8.7)	2.8 (±1.8)	1.0 (±0.8)	37.9 (±14.9)	12896
	9	13.1 (±12.5)	2.4 (±2.1)	0.8 (±0.7)	15.5 (±18.6)	355	11.1 (±7.9)	2.5 (±1.7)	0.9 (±0.6)	27.6 (±10.5)	1485
With SG	1	13.7 (±13.1)	2.6 (±4.3)	0.8 (±4.9)	15.2 (±37.1)		11.3 (±11.6)	2.3 (±2.5)	0.9 (±2.6)	27.2 (±34.6)	
	3	12.6 (±12.5)	2.1 (±2.6)	0.7 (±2.1)	11.9 (±22.8)		9.9 (±9.7)	2.0 (±1.7)	0.8 (±1.2)	20.0 (±18.2)	
	5	13.3 (±12.5)	2.3 (±2.1)	0.7 (±0.9)	11.4 (±19.3)		10.4 (±8.8)	2.0 (±1.5)	0.8 (±0.8)	18.1 (±13.1)	
	9	12.3 (±12.4)	2.0 (±1.8)	0.7 (±0.7)	9.3 (±16.0)		9.3 (±7.9)	1.7 (±1.4)	0.8 (±0.6)	13.7 (±9.9)	

Coarser spatial resolution moderately increased spectral consistency. The SAM reflectance median and the standard deviation decreased systematically with increasing spatial aggregation. The use of a Savitzky - Golay smoother reduced the median SAM value by an average 10-20% for clearings and 20-40% for forest. This reduction in SAM following spectral smoothing decreased when considering larger patches. The dissimilarity values were lower for the bamboo thickets compared to the forest canopy for the SWIR-2 spectral range.

Predictability of SAM divergence was analysed for windows of 1x1 and 3x3 pixels, based on a north-south cross validation (Figure 1111). Coarser spatial resolution significantly increased SAM divergence predictability for forest only and again only for VNIR and SWIR-1 (Fig.11).

In addition, considering only patches with all pixels illuminated (restrictive configuration), did not increase the explained variance (Fig.11). Although the spectral smoothing reduced the variations in SAM, it had no impact on the level of variance explained by the GAM models regardless of the spectral range considered (not shown).

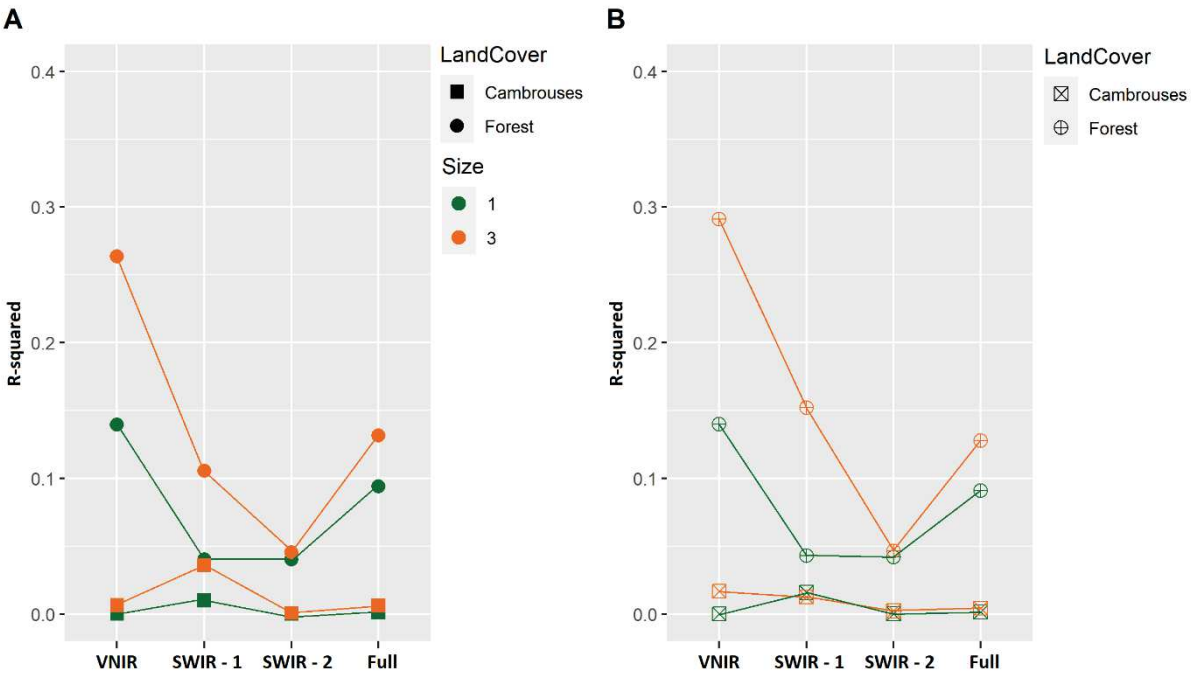

Figure 11: R-squared given by the prediction models using specific spectral range. A) Permissive illumination configuration was used, B) Restrictive illumination configuration was used.

Increasing of the aggregation factor decreased the number of available observations (Tab.6, in appendix). At 9x9 pixels window size, the number of observations became very low (permissive:1385, restrictive: 42). There was no clear trend of increasing R^2 in SAM predictability with increasing windows size beyond 3x3 pixels for forest canopy except for SWIR-1 (Fig 14 left and appendix, Fig.17).

Filtering by an illumination ratio greater than 0.6 (as computed by ATCOR-4 module) was quite effective in removing any type of shadow (Tab.4). By considering only fully well-lit areas (restrictive configuration) we expected that we would be observing areas more consistent across dates and that this would globally improve spectral consistency and increase the predictability of the SAM divergence. None of this was observed (Appendix, Tab.7).

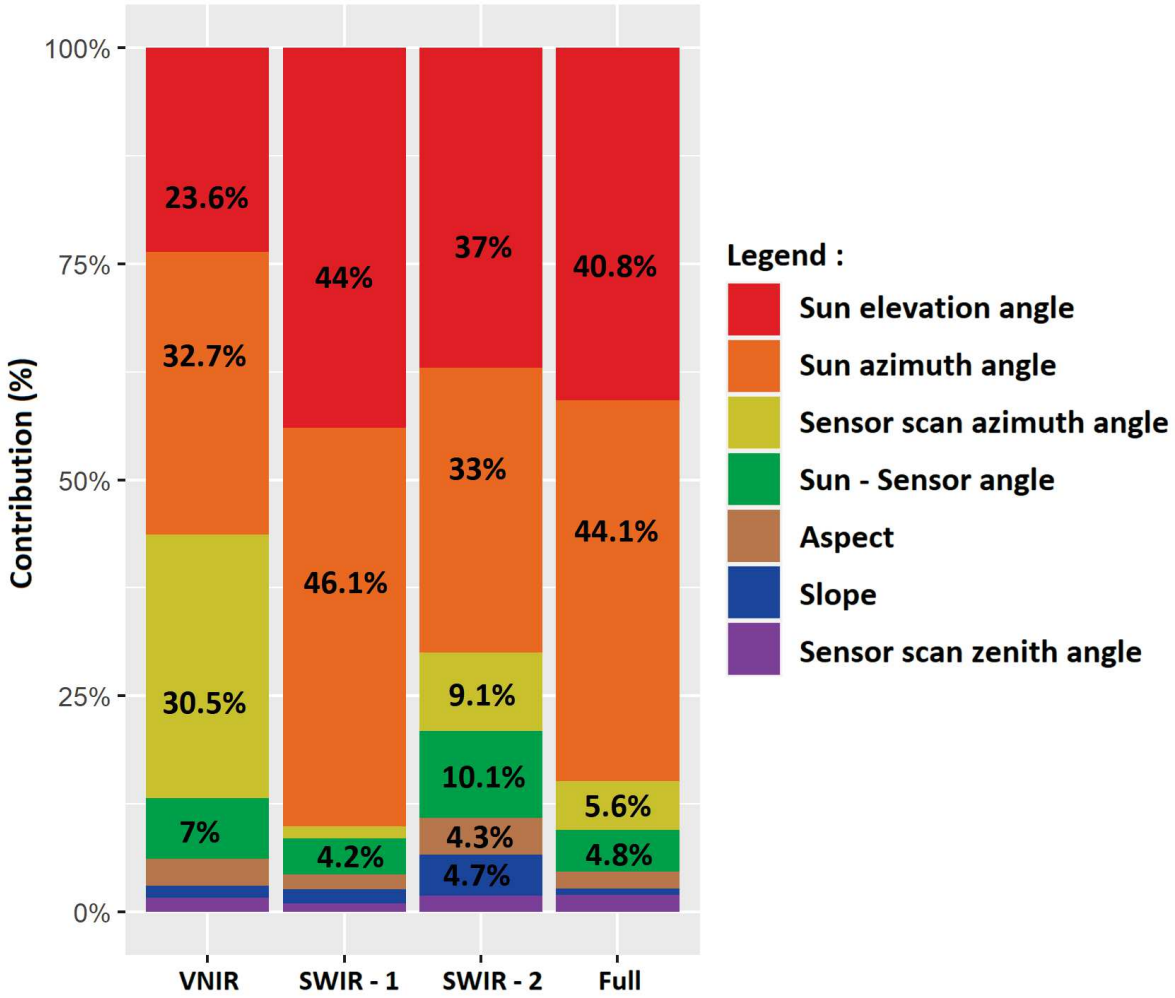

Figure 12: Relative importance of each component of the model estimated by block cross-validation. Data from the permissive configuration and 5x5 pixels resolution were used.

Differences in solar position between the two dates contributed most to the explained variance (Fig.12). The individual contribution of a given predictor was defined as the increase in explained variance achieved when adding this factor to the model once all other predictors were already included. We report the individual contribution of each factor divided by the sum of individual contributions of all factors. Solar position made the largest contribution to the explained variance in SAM. Sensor viewing angle and sun-sensor angle came second. Topography contributed the least in all spectral regions. Its largest contribution was for SWIR-2.

5. BRDF applied

A BRDF model correction was then applied on data atmospherically corrected using the BREFCOR model. We computed the SAM index for both configurations and we compared the mean, median and standard SAM values between configurations. A decrease in the mean SAM value and/or its standard deviation would indicate an improvement in consistency between dates. Then, we compared the variance explained between both configurations.

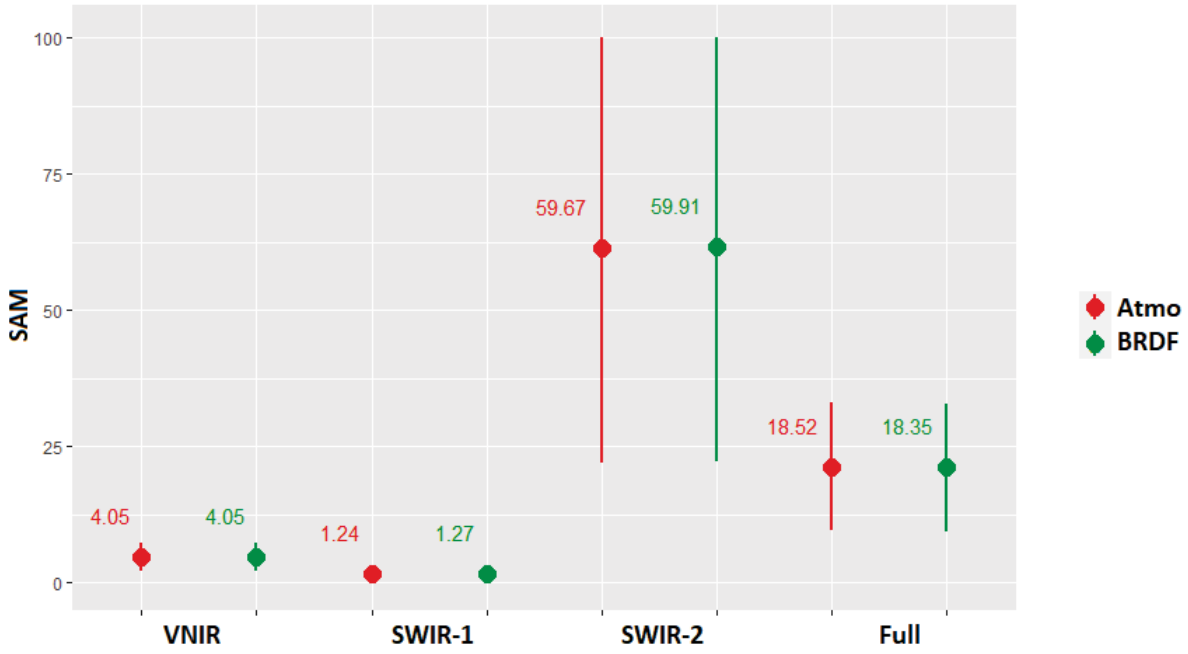

Figure 13: Distribution of SAM values between data with and without BRDF correction. In red and green are given the data atmospherically corrected and the data atmospherically corrected with a BRDF correction respectively. The 1x1 pixel size was considered. The graphic point represents the mean. The value given represents the median.

SAM values where unaffected by BRDF correction (Fig.13). This remained the case regardless of the size of the window considered (not shown).

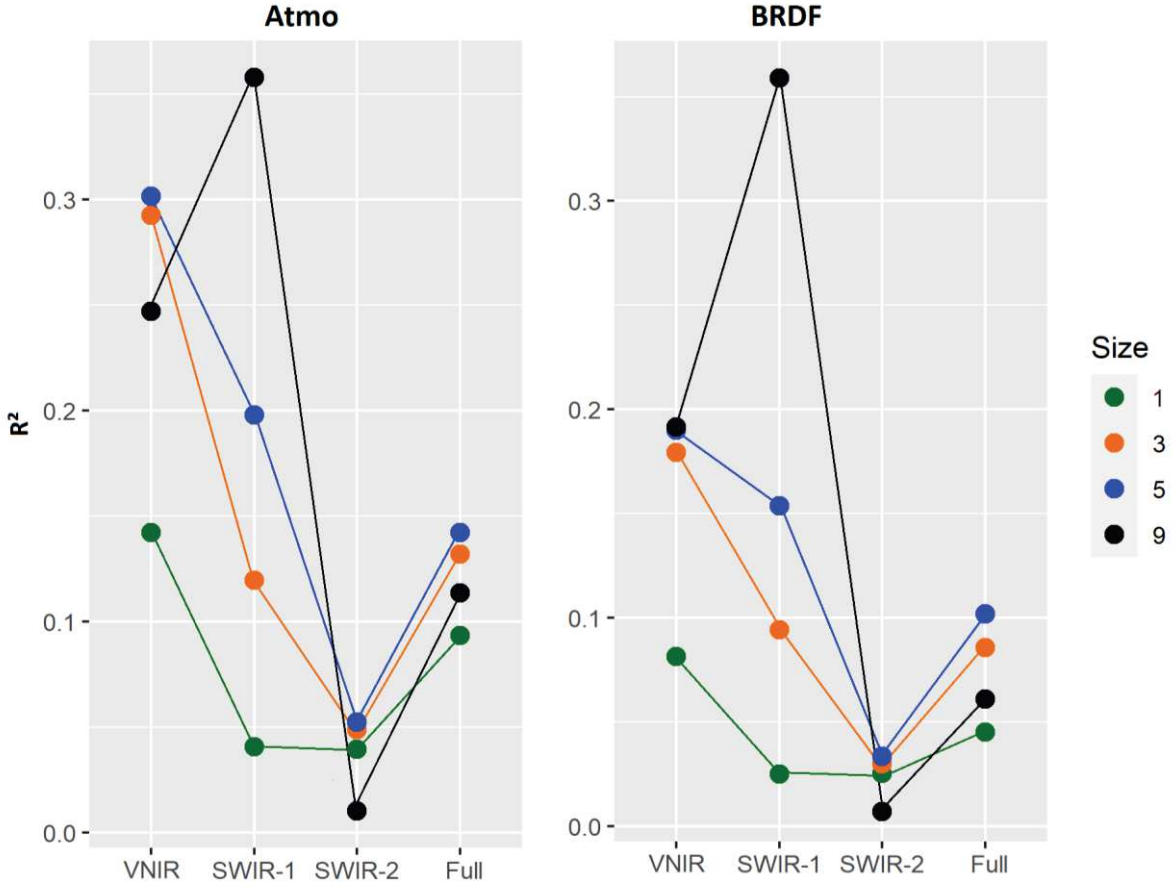

Figure 14: Comparison of variance explained using atmospheric correction data prior (left) or after (right) applying BRDF correction. Permissive illumination was used in both cases.

Applying a BRDF correction decreased the explained variance regardless of the spectral range considered (Fig.14). The largest patch size (9x9 pixels) showed an outlying behavior with an unexpectedly high variance explained in the SWIR-1 region only.

6. Spectral correction

We used acquisition parameters and reflectance values on date 1 and 2 to model the spectral distorsion between dates. We tested two model types (GAM and NN method).

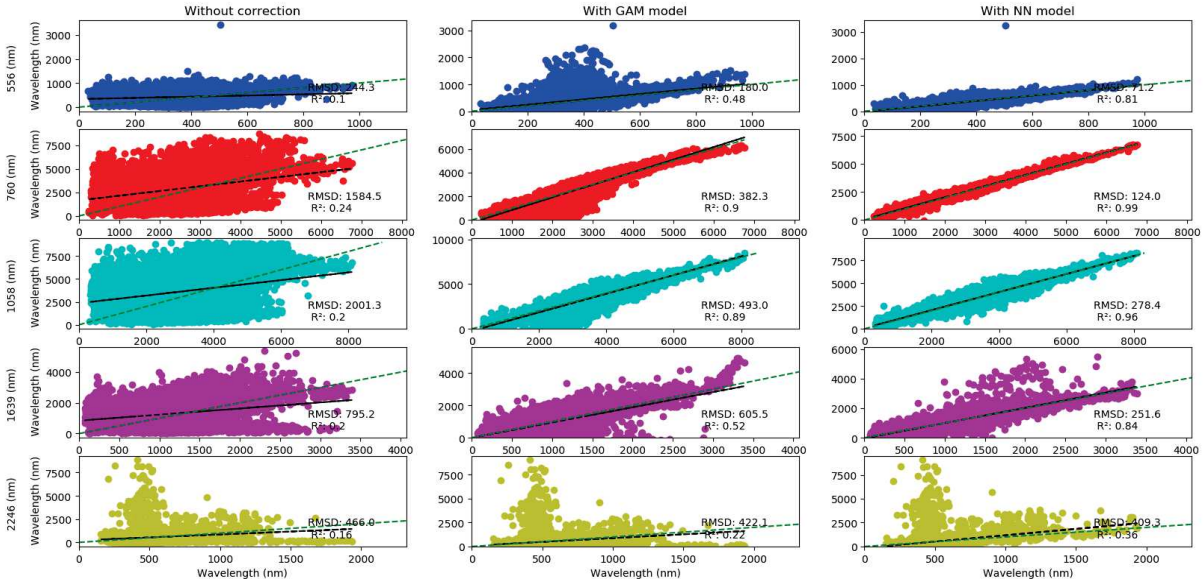

Figure 15: Five spectral bands retained and compared without correction and after corrected using a GAM or a NN model. The black dotted line is the regression line. The green dotted line is the line 1:1. The RMSD (Root Mean Squared Deviation) between dates either before (left column) or after correction (center and right columns) is given in each graph respectively. The first date is given on the X-axis. The second date (observed or predicted) is given on the Y-axis.

Initially, a strong dispersion and significant bias (difference with 1:1 line) were observed between the reflectance values at the two dates. The two methods (GAM and NN) achieved good results by eliminating bias and strongly reducing dispersion: RMSD was decreased by 10.3% to 93.5% (GAM) and by 16.2% to 93.9% (NN) (details in appendix 6). GAM seemed less capable than NN to correct the extreme values (see graphs at the 556 nm and 1639 nm). Considering data without atmospheric correction, using the same models, the results were also in favor of NN method (appendix 7, Tab.9). Importantly, the reflectance data were much more predictable than radiance data, except for the 2246 nm wavelength.

D. Discussion

SNR analysis indicated that SWIR-2 spectral range was extremely noisy and that atmospheric corrections tended to inflate that noise to the point that it reduced spectral consistency between dates in that spectral range (see Fig.2 and Fig.3). Two things may have contributed to the observed pattern: the characteristics of the sensor and the atmospheric correction. As reported by (Lenhard et al., 2015), characterizing the radiometric, spectral and geometric performances, the uncertainties of the radiometric responses evaluated for the center pixels of the sensor was higher in SWIR-2 compared to the rest of the spectral range. The fact that SAM dissimilarity index increased after atmospheric corrections indicated these corrections were less than satisfactory.

Even without considering the SWIR-2 region, the predictability of spectral divergence between dates was found to be low : the cross validation R^2 barely reached 0.3 in the best case scenario (i.e VNIR, forest only, well-illuminated 3x3 pixels window size in Figure 14 right panel).

The main contributing source to spectral difference between dates was the difference in solar position. This held true for the three spectral regions considered. However, the relative contribution of the various predictors included in the model varied with the spectral region (Fig.9). This would suggest that conducting the analysis on more narrow spectral regions or even per wavelength might help refine the interpretation.

(Doxani et al., 2018) made an inter comparison of four atmospheric correction methods (ATCOR, FORCE, iCOR, and LaSRC) applied to spatial data (Landsat-8 and Sentinel-2), positioned all around the globe. They reported that AOT at 550 nm was underestimated by ATCOR compared to AERONET data. AERONET data are based on very local measurements made by solar photometer allowing to assess the direct solar radiation which provides information on AOT. The calculation of AOT by ATCOR-4 based on shadows and dense vegetation uses the Dark Dense Vegetation (DDV) method (Richter et al., 2006). However, AOT is affected by noise of the limited sensor accuracy. The calculation of the SNR for the Paracou and Nouragues data indicated that certain spectral areas were very noisy, which could have led to a poor estimate of the AOT. Although field measurement using a sun photometer were conducted during the overflights analysed here those we measured at a single spatial position during the entire acquisition. A spatially variable AOT maybe estimated (Schläpfer et al., 2018) and should preferably have been applied here.

In our configuration and in view of the proportion of clouds present in the Nouragues study area, it is clear that the diffusive component of incoming light was locally high. The high concentration of water vapour induces an increase in the reflectance variation for certain wavelengths found mainly in the SWIR-2 and SWIR-1 spectral region. Although we removed the shadows cast by clouds and trees by filtering pixels based on the local illumination, it is possible that the second date had more variable water vapour concentrations given the higher cloud cover present. However, the SWIR-1 was the least impacted spectral zone, between dates.

The terrain slope has a significant effect on the surface albedo. (Wu et al., 2018) using DART (Discrete Anisotropic Radiative Transfer) simulations and in situ observations, showed that there is an absolute bias between the horizontal sloping surface albedo and the sloping surface albedo. The slopes and orientations were simulated for values from 0 ° to 60 ° (incremented by steps of 10 °) and from 0 ° to 360 ° (incremented by steps of 30°). The absolute biases found, averaging around 62.2%, greatly surpass the effect of the optical and structural parameters. In our study, the inclination / orientation of the slope did not explain much variance in SAM. This may be related to the relatively flat relief on the study site: in Nouragues (excluding the rocky outcrop) less than 2% of the slopes (based on the 10m resolution DTM) were larger than 20 degrees

BRDF correction was tested using BREFCOR model. The of BRDF corrections can be very strong as mentioned in ATCOR-4 manual (up to 30% changes in reflectance). Here, we found no improvement in spectral consistency across dates following BRDF correction, whatever the resolution (patch size) considered. Forest canopies have a strong anisotropic factor due to the clumping on the scale of the leaf and the crown (Liang et al., 2000) and BRDF models depend on the spatial scale of the measurements (Román et al., 2011). (Koukal et al., 2014) report significant improvement in forest type separability after modelling BRDF effects per forest type and computing a forest specific bidirectional reflectance factor. As shown on selected wavelengths, sun-sensor-topography geometry seem to affect consistency between dates in a predictable way. So, the reason why the applied BRDF correction did not help reduce spectral divergence measured using SAM deserves further scrutiny. GAM or NN method trained on one part of the image effectively reduced discrepancy between dates. NN models slightly outperformed GAM and may prove to have better generalization (extrapolation) capability.

However, any empirical or semi-empirical approach will be based on the estimated data and although it may be effective in certain cases, it will not easily be generalizable. A fully physically based approach, using ray tracing may serve to conduct sensitivity analyses to identify in which range of parameters (canopy roughness, solar position, etc) BRDF is likely to significantly affect spectral signature. For example, the DART (Gastellu-Etchegorry et al., 2017) model or LESS model (Qi et al., 2017) can create 3D scenes simulating the position of trees and all parameters useful for estimating the

BRDF. The reflectance of tree leaves is informed by means of a database or field measurements (Spectroradiometer). The DART model can use 3D scenes generated from LiDAR data providing realistic configuration and fine description of the vegetation canopy cover. LiDAR data are used to describe a scene by documenting individual voxels with specific turbidity characteristic. Ray-tracing allows to take into account the scattering within the canopy and to better correct the HDRF value over a large spectral range. (Fawcett et al., 2018) used the DART model to simulate the irradiance-scaling factor derived from top of canopy, which is more relevant than a coarse DEM. Nevertheless, the authors mentioned limitations such as computation time, the voxel size that should be higher than image resolution and a large number of scattering angle and iterations and over and underestimation of irradiance partially occurred for gaps between trees. This experimentation was conducted on temperate mixed forest. It would probably be beneficial to conduct a similar study on tropical dense rainforest.

In the previous analysis, we used the SAM dissimilarity index computed over large spectral regions (multiple wavelengths). However, the different wavelengths do not respond in the same way to the various disturbances (atmospheric or geometric). One follow-up study would then be to test an entirely data driven approach to BRDF corrections. One could build a statistical correction for each wavelength based on the same predictors used here to model the change between dates. Such models applied to atmospherically corrected data may allow improving spectral consistency between dates more than the semi-empirical models such as attempted here. As repeat overflights become available date, robust statistical models for correction of anisotropic reflectance in canopies may become effective.

E. Conclusion

The mapping of forest species over large areas using hyperspectral imagery requires to develop robust classification methods that can be generalized and can be adapted to a large range of acquisition conditions. It is therefore important to understand the sources of variation in spectral signal when only viewing conditions (and atmospheric conditions) are altered while the object remains essentially unchanged. In this study, which took advantage of repeated overflights of hyper diverse tropical canopies. Unsatisfactory atmospheric corrections in certain spectral ranges seem to be a major source of discrepancy in spectral signature between dates. While atmospheric corrections did significantly reduce the dissimilarity between dates for VNIR and SWIR-1 the opposite occurred beyond 2000 nm. We found that only a minor fraction of the difference in spectral signature between two dates acquired in a very short time interval could be attributed to sun-sensor-topography characteristics.

More advanced atmospheric correction such as applying a spatially variable AOT estimate (Schläpfer et al., 2018) and better calibrated BRDF corrections (Richter and Schlapfer, 2018; Thompson et al., 2019) might improve spectral consistency but probably not dramatically so. An alternative complementary strategy applicable for sun-sensor-relief effects standardisation might be a purely data driven statistical approach. Statistical models seem promising indeed and could be further improved by adding relevant canopy roughness, LAD density information which is readily available from co-occurring LiDAR data (if this the case like in the present study).

F. Appendix

1. Other metrics and their correlations

a) SID (Kullback-Leibler, symétrie):

To compare two datasets coming from same multidimensional distribution, it is possible to use the Kullback-Leibler (KL) (Kullback and Leibler 1951) divergence. KL does not assume any parametric form of the probability distributions being compared. KL is a measure based on the relative entropy of two probability density function and it is compute as follow:

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

P and Q are the discrete probability distribution.

KL divergence is not a distance because it does not verify the triangular inequality. The spectral mean of the two datasets are separately normalised and considered as distributions. KL divergence is not symmetric. Symmetry is ensured by computing the mean of $D_{KL}(P/Q)$ and $D_{KL}(Q/P)$ also known as Jeffrey divergence (Nguyen, Morell, and De Baets 2017) or SID measure (Chein-I Chang 2000).

$$SID = \frac{D_{KL}(P||Q) + D_{KL}(Q||P)}{2}$$

b) NRMSD

The normalized Root-Mean-Square Deviation is a Euclidian distance. We compute the spectral mean of the each objects and patches. Then, this metric was computed.

$$RMSD = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}$$

2. Correlation matrix

We evaluated the contribution of previous metrics and compared those with SAM. The correlation is therefore proposed.

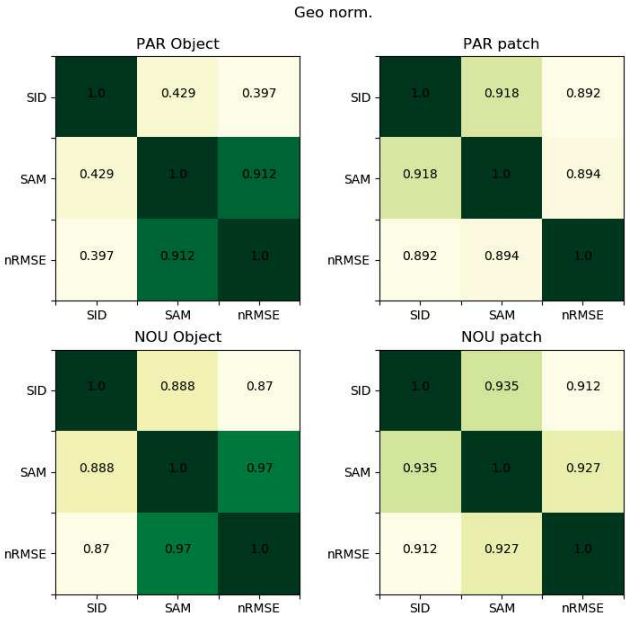

Figure 16: Correlation matrices between metrics, computed using all radiance information

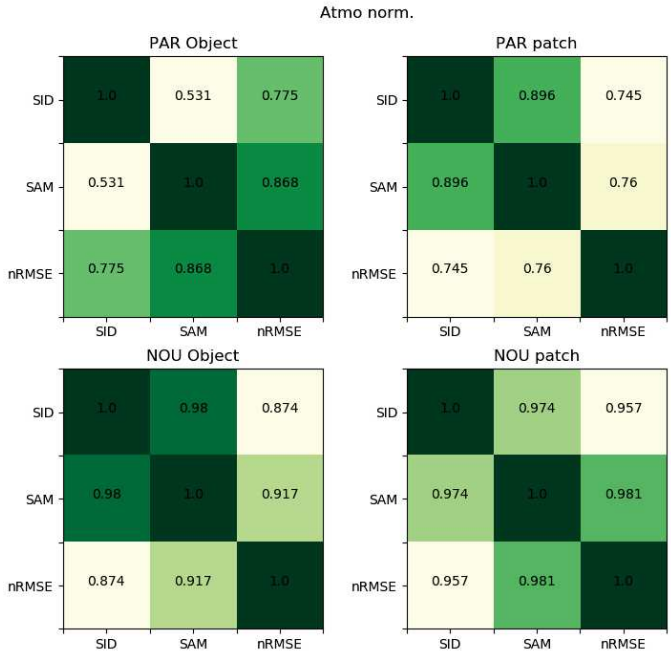

Figure 17: Correlation matrices between metrics, computed using all reflectance information

Whether before or after the atmospheric correction, the correlation between the different dissimilarity measurements is strong at the patch scale. This is similar between the two study sites. However, the scale of objects, the correlation is low, mainly in respect of objects delineated on the site Paracou

3. Proportion of pixels for each configuration

The details of pixels number per class is given in Tab.6.

Table 7: Proportion of pixels in each class.

Size		1	3	5	9
Permissive	Forest	716057	54710	12896	1385
	Cambrouse	40113	4107	1332	355
	Others	36577	2923	725	100
	Total	792747	61740	14953	1840
Restrictive	Forest	716057	14931	1125	42
	Cambrouse	40113	2707	735	154
	Others	36577	1110	128	6
	Total	792747	18748	1988	202

The proportion of forest pixels decreases more quickly than the pixels of cambrouse. This is explained by the surface roughness of two types of land use. Because of its smoother surface, the shadows cast on a cambrouse surface are much less than on a forest canopy.

4. SAM statistics

Table 8: Median and standard deviation are given following different configuration. This table must be related to the number of elements per configuration (see Tab.6).

			Permissive				Restrictive			
	Size	LandCover	Full range	VNIR	SWIR-1	SWIR-2	Full range	VNIR	SWIR-1	SWIR-2
Sav	1	Forest	15.57 (±11.57)	3.60 (±2.22)	1.17 (±2.10)	45.33 (±34.36)	15.58 (±11.54)	3.60 (±2.22)	1.17 (±2.10)	45.54 (±34.37)
		Cambrouse	15.34 (±14.28)	3.33 (±2.44)	0.92 (±1.11)	20.82 (±34.70)	15.33 (±14.23)	3.33 (±2.43)	0.92 (±1.11)	20.89 (±34.63)

	3	Forest	11.18 (±9.67)	2.44 (±1.58)	0.91 (±1.01)	27.74 (±18.07)	10.86 (±10.22)	2.27 (±1.55)	0.87 (±1.01)	22.25 (±17.65)
		Cambrouse	12.93 (±14.22)	2.37 (±2.00)	0.77 (±0.78)	15.33 (±24.10)	13.06 (±15.04)	2.36 (±1.92)	0.73 (±0.68)	13.48 (±25.17)
	5	Forest	10.02 (±8.79)	2.02 (±1.42)	0.83 (±0.77)	20.17 (±13.00)	10.13 (±9.35)	1.90 (±1.31)	0.81 (±0.70)	15.83 (±12.91)
		Cambrouse	12.31 (±14.08)	2.05 (±1.80)	0.73 (±0.70)	11.87 (±21.52)	12.90 (±14.63)	2.10 (±1.81)	0.67 (±0.66)	10.04 (±22.56)
	9	Forest	9.48 (±8.02)	1.72 (±1.28)	0.79 (±0.59)	13.73 (±9.88)	7.24 (±4.30)	1.67 (±0.91)	0.76 (±0.35)	9.53 (±3.81)
		Cambrouse	12.24 (±13.47)	1.85 (±1.70)	0.67 (±0.69)	9.21 (±19.71)	12.34 (±13.90)	1.82 (±1.72)	0.63 (±0.63)	7.93 (±20.94)
NoSav	1	Forest	18.52 (±11.65)	4.06 (±2.55)	1.24 (±1.61)	59.75 (±39.45)	18.53 (±11.63)	4.06 (±2.55)	1.24 (±1.61)	60.02 (±39.46)
		Cambrouse	17.39 (±14.44)	3.82 (±2.68)	0.99 (±0.92)	26.75 (±39.13)	17.38 (±14.39)	3.82 (±2.67)	0.99 (±0.92)	26.85 (±39.08)
	3	Forest	12.83 (±9.58)	2.75 (±1.90)	0.95 (±0.97)	38.06 (±21.12)	12.34 (±10.14)	2.58 (±1.88)	0.91 (±0.85)	30.55 (±20.53)
		Cambrouse	14.18 (±14.24)	2.77 (±2.24)	0.82 (±0.76)	20.83 (±26.56)	14.23 (±15.07)	2.76 (±2.19)	0.77 (±0.66)	18.10 (±27.45)
	5	Forest	11.43 (±8.72)	2.31 (±1.75)	0.86 (±0.74)	27.37 (±14.67)	11.35 (±9.29)	2.21 (±1.64)	0.84 (±0.64)	21.70 (±14.25)
		Cambrouse	13.55 (±14.12)	2.46 (±2.07)	0.76 (±0.67)	15.08 (±22.82)	13.91 (±14.67)	2.54 (±2.09)	0.71 (±0.64)	12.91 (±23.79)
	9	Forest	10.65 (±7.98)	2.03 (±1.61)	0.82 (±0.57)	18.08 (±10.37)	8.39 (±4.21)	2.06 (±1.25)	0.78 (±0.35)	13.35 (±4.19)
		Cambrouse	13.20 (±13.51)	2.30 (±1.98)	0.71 (±0.67)	11.26 (±20.07)	13.27 (±13.96)	2.32 (±1.99)	0.66 (±0.61)	9.75 (±21.33)

5. BCI distribution

Table 9: Distribution of BCI values estimated by BREFCOR model using standard calibration granularity

BCI interval	[-1500:-899]	[-900:-199]	[-200:399]	[400:899]	[900:1500]
Number (%)	0.3	0.0	0.0	2.4	97.3

The BCI values ranged widely from 900 to 1500.

6. Wavelength correction

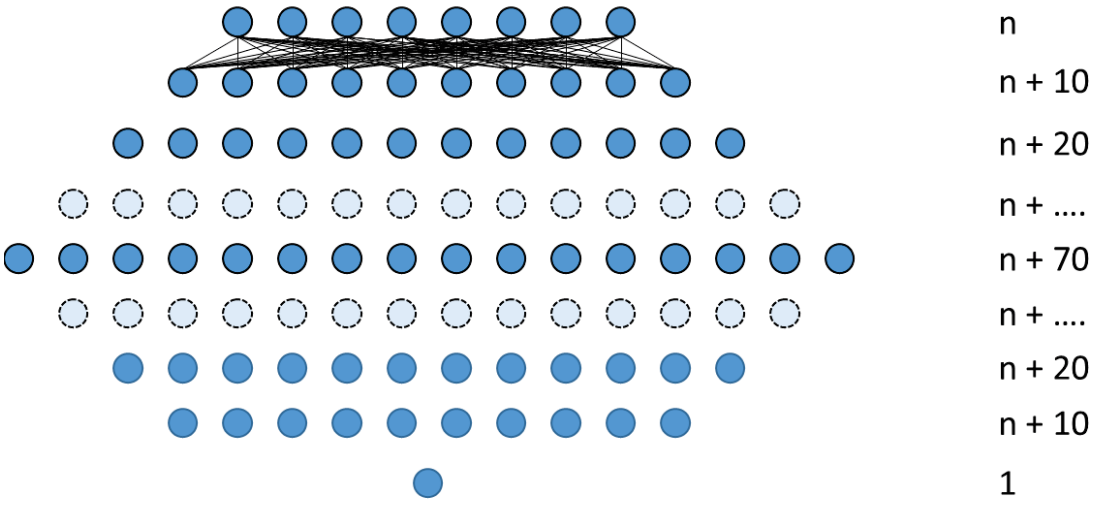

Figure 18: Neural network structure. 17 layers were used. For each layer, 10 neurons are added compared to the initial feature numbers. At n + 70, the number of neurons decreases with each layer by 10 to arrive at one. The activation function for all neurons are "Relu" functions, except for the last one which is a linear function.

The neural network was implemented using TensorFlow library on Python. The RMSProp optimizer was used.

Table 10: RMSE computed before and after correction using a GAM or a NN model for each wavelength. Same evaluation have been done using data with and without atmospheric compensation.

	Wavelength (nm)	Without correction	GAM model		NN model	
		RMSE	RMSE	Percentage	RMSE	Percentage
With	556	240	77.6	67.7	70.4	70.7
	760	1541	100.7	93.5	93.5	93.9
	1058	1925	271.4	85.9	303.9	84.2
	1639	770	256.8	66.6	195.3	74.6
	2246	432	387.6	10.3	362.2	16.2
Without	556	2330	993	57.3	964	58.6
	760	8005	3435	57.1	2819	64.8
	1058	19277	14763	23.4	13996	27.4
	1639	2061	973	52.8	917	55.5
	2246	257	148	42.4	113	55.8

We compared the GAM model and the NN model with and without atmospheric correction. The data without atmospheric correction were less predictable.

G. References

Baraldi, A., 2009. Impact of Radiometric Calibration and Specifications of Spaceborne Optical Imaging Sensors on the Development of Operational Automatic Remote Sensing Image Understanding Systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2, 104–134. https://doi.org/10.1109/JSTARS.2009.2023801

Bhatia, N., Iordache, M.-D., Stein, A., Reusen, I., Tolpekin, V.A., 2018. Propagation of uncertainty in atmospheric parameters to hyperspectral unmixing. Remote Sensing of Environment 204, 472–484. https://doi.org/10.1016/j.rse.2017.10.008

Bhatia, N., Tolpekin, V.A., Reusen, I., Sterckx, S., Biesemans, J., Stein, A., 2015. Sensitivity of Reflectance to Water Vapor and Aerosol Optical Thickness. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 8, 3199–3208. https://doi.org/10.1109/JSTARS.2015.2425954

Bongers, F., Charles-Dominique, P., Forget, P.-M., Théry, M., 2013. Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest. Springer Science & Business Media.

Bréon, F.-M., Vermote, E., 2012. Correction of MODIS surface reflectance time series for BRDF effects. Remote Sensing of Environment 125, 1–9. https://doi.org/10.1016/j.rse.2012.06.025

Che, H., Gui, K., Chen, Q., Zheng, Y., Yu, J., Sun, T., Zhang, X., Shi, G., 2016. Calibration of the 936 nm water-vapor channel for the China aerosol remote sensing NETwork (CARSNET) and the effect of the retrieval water-vapor on aerosol optical property over Beijing, China. Atmospheric Pollution Research 7, 743–753. https://doi.org/10.1016/j.apr.2016.04.003

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. Journal of Geophysical Research: Solid Earth 89, 6329–6340. https://doi.org/10.1029/JB089iB07p06329

Colgan, M., Baldeck, C., Féret, J.-B., Asner, G., 2012. Mapping Savanna Tree Species at Ecosystem Scales Using Support Vector Machine Classification and BRDF Correction on Airborne Hyperspectral and LiDAR Data. Remote Sensing 4, 3462–3480. https://doi.org/10.3390/rs4113462

Deborah, H., Richard, N., Hardeberg, J.Y., 2015. A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8, 3224–3234. https://doi.org/10.1109/JSTARS.2015.2403257

Doxani, G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., Vanhellemont, Q., 2018. Atmospheric Correction Inter-Comparison Exercise. Remote Sensing 10, 352. https://doi.org/10.3390/rs10020352

Fawcett, D., Verhoef, W., Schläpfer, D., Schneider, F.D., Schaepman, M.E., Damm, A., 2018. Advancing retrievals of surface reflectance and vegetation indices over forest ecosystems by combining imaging spectroscopy, digital object models, and 3D canopy modelling. Remote Sensing of Environment 204, 583–595. https://doi.org/10.1016/j.rse.2017.09.040

Galvão, L.S., Breunig, F.M., Santos, J.R. dos, Moura, Y.M. de, 2013. View-illumination effects on hyperspectral vegetation indices in the Amazonian tropical forest. International Journal of Applied Earth Observation and Geoinformation 21, 291–300. https://doi.org/10.1016/j.jag.2012.07.005

Gao, L., Du, Q., Zhang, B., Yang, W., Wu, Y., 2013. A Comparative Study on Linear Regression-Based Noise Estimation for Hyperspectral Imagery. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 6, 488–498. https://doi.org/10.1109/JSTARS.2012.2227245

Gastellu-Etchegorry, J.-P., Lauret, N., Yin, T., Landier, L., Kallel, A., Malenovsky, Z., Bitar, A.A., Aval, J., Benhmida, S., Qi, J., Medjdoub, G., Guilleux, J., Chavanon, E., Cook, B., Morton, D., Chrysoulakis, N., Mitraka, Z., 2017. DART: Recent Advances in Remote Sensing Data Modeling With Atmosphere, Polarization, and Chlorophyll Fluorescence. IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING 10.

Gerard, F.F., North, P.R.J., 1997. Analyzing the effect of structural variability and canopy gaps on forest BRDF using a geometric-optical model. Remote Sensing of Environment 62, 46–62. https://doi.org/10.1016/S0034-4257(97)00070-9

Gourlet-Fleury, S., Guehl, J.-M., Laroussinie, O., ECOFOR (Group) (Eds.), 2004. Ecology and management of a neotropical rainforest: lessons drawn from Paracou, a long-term experimental research site in French Guiana. Elsevier, Paris.

Gross, W., Tuia, D., Soergel, U., Middelmann, W., 2019. Nonlinear Feature Normalization for Hyperspectral Domain Adaptation and Mitigation of Nonlinear Effects. IEEE Transactions on Geoscience and Remote Sensing 57, 5975–5990. https://doi.org/10.1109/TGRS.2019.2903719

Hastie, T., Tibshirani, R., 1999. Generalized additive models. Chapman & Hall/CRC, Boca Raton, Fla.

Jia, W., Pang, Y., Tortini, R., Schläpfer, D., Li, Z., Roujean, J.-L., 2020. A Kernel-Driven BRDF Approach to Correct Airborne Hyperspectral Imagery over Forested Areas with Rugged Topography. Remote Sensing 12, 432. https://doi.org/10.3390/rs12030432

Kokaly, R.F., Clark, R.N., 1999. Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. Remote Sensing of Environment 67, 267–287. https://doi.org/10.1016/S0034-4257(98)00084-4

Korpela, I., Heikkinen, V., Honkavaara, E., Rohrbach, F., Tokola, T., 2011. Variation and directional anisotropy of reflectance at the crown scale — Implications for tree species classification in digital aerial images. Remote Sensing of Environment 115, 2062–2074. https://doi.org/10.1016/j.rse.2011.04.008

Koukal, T., Atzberger, C., Schneider, W., 2014. Evaluation of semi-empirical BRDF models inverted against multi-angle data from a digital airborne frame camera for enhancing forest type classification. Remote Sensing of Environment, Special Issue on 2012 ForestSAT 151, 27–43. https://doi.org/10.1016/j.rse.2013.12.014

Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J., Vincent, G., 2019. Across Date Species Detection Using Airborne Imaging Spectroscopy. Remote Sensing 11, 789. https://doi.org/10.3390/rs11070789

Lenhard, K., Baumgartner, A., Schwarzmaier, T., 2015. Independent Laboratory Characterization of NEO HySpex Imaging Spectrometers VNIR-1600 and SWIR-320m-e. IEEE Transactions on Geoscience and Remote Sensing 53, 1828–1841. https://doi.org/10.1109/TGRS.2014.2349737

Liang, S., Strahler, A.H., Barnsley, M.J., Borel, C.C., Gerstl, S.A.W., Diner, D.J., Prata, A.J., Walthall, C.L., 2000. Multiangle remote sensing: Past, present and future. Remote Sensing Reviews 18, 83–102. https://doi.org/10.1080/02757250009532386

Makarau, A., Richter, R., Schläpfer, D., Reinartz, P., 2017. APDA Water Vapor Retrieval Validation for Sentinel-2 Imagery. IEEE Geoscience and Remote Sensing Letters 14, 227–231. https://doi.org/10.1109/LGRS.2016.2635942

Matasci, G., Tuia, D., Kanevski, M., 2012. SVM-Based Boosting of Active Learning Strategies for Efficient Domain Adaptation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5, 1335–1343. https://doi.org/10.1109/JSTARS.2012.2202881

Meyer, P., Itten, K.I., Kellenberger, T., Sandmeier, S., Sandmeier, R., 1993. Radiometric corrections of topographically induced effects on Landsat TM data in an alpine environment. ISPRS Journal of Photogrammetry and Remote Sensing 48, 17–28. https://doi.org/10.1016/0924-2716(93)90028-L

Oliveira, L.M. de, Galvão, L.S., Ponzoni, F.J., 2019. Topographic effects on the determination of hyperspectral vegetation indices: a case study in southeastern Brazil. Geocarto International 0, 1–18. https://doi.org/10.1080/10106049.2019.1690055

Pacifici, F., Longbotham, N., Emery, W.J., 2014. The Importance of Physical Quantities for the Analysis of Multitemporal and Multiangular Optical Very High Spatial Resolution Images. IEEE Transactions on Geoscience and Remote Sensing 52, 6241–6256. https://doi.org/10.1109/TGRS.2013.2295819

Pflug, B., Main-Knorn, M., Makarau, A., Richter, R., 2015. Validation of aerosol estimation in atmospheric correction algorithm ATCOR. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-7/W3, 677–683. https://doi.org/10.5194/isprsarchives-XL-7-W3-677-2015

Qi, J., Xie, D., Guo, D., Yan, G., 2017. A Large-Scale Emulation System for Realistic Three-Dimensional (3-D) Forest Simulation. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 10, 4834–4843. https://doi.org/10.1109/JSTARS.2017.2714423

Rahman, H., Pinty, B., Verstraete, M.M., 1993. Coupled surface-atmosphere reflectance (CSAR) model: 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. J. Geophys. Res. 98, 20791. https://doi.org/10.1029/93JD02072

Rajan, S., Ghosh, J., Crawford, M.M., 2006. Exploiting Class Hierarchies for Knowledge Transfer in Hyperspectral Data. IEEE Transactions on Geoscience and Remote Sensing 44, 3408–3417. https://doi.org/10.1109/TGRS.2006.878442

Reddington, C.L., Spracklen, D.V., Artaxo, P., Ridley, D.A., Rizzo, L.V., Arana, A., 2016. Analysis of particulate emissions from tropical biomass burning using a global aerosol model and long-term surface observations. Atmospheric Chemistry and Physics 16, 11083–11106.

Richter, R., Kellenberger, T., Kaufmann, H., 2009. Comparison of Topographic Correction Methods. Remote Sensing 1, 184–196. https://doi.org/10.3390/rs1030184

Richter, R, Schlapfer, D., 2018. Atmospheric / Topographic Correction for Airborne Imagery (ATCOR-4 user Guide, Version 7.2.0) 279.

Richter, Rolf, Schlapfer, D., 2018. PARametric GEocoding : Orthorectification for airborne scanner data. User Manual, Version 3.4.

Richter, R., Schläpfer, D., Müller, A., 2006. An automatic atmospheric correction algorithm for visible/NIR imagery. International Journal of Remote Sensing 27, 2077–2085. https://doi.org/10.1080/01431160500486690

Roger, R.E., Arnold, J.F., 1996. Reliably estimating the noise in AVIRIS hyperspectral images. International Journal of Remote Sensing 17, 1951–1962. https://doi.org/10.1080/01431169608948750

Román, M.O., Gatebe, C.K., Schaaf, C.B., Poudyal, R., Wang, Z., King, M.D., 2011. Variability in surface BRDF at different spatial scales (30m–500m) over a mixed agricultural landscape as retrieved from airborne and satellite spectral measurements. Remote Sensing of Environment 115, 2184–2203. https://doi.org/10.1016/j.rse.2011.04.012

Santini, F., Palombo, A., 2019. Physically Based Approach for Combined Atmospheric and Topographic Corrections. Remote Sensing 11, 1218. https://doi.org/10.3390/rs11101218

Schläpfer, D., Borel, C.C., Keller, J., Itten, K.I., 1998. Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. Remote Sensing of Environment 65, 353–366. https://doi.org/10.1016/S0034-4257(98)00044-3

Schläpfer, D., Hueni, A., Richter, R., 2018. Cast Shadow Detection to Quantify the Aerosol Optical Thickness for Atmospheric Correction of High Spatial Resolution Optical Imagery. Remote Sensing 10, 200. https://doi.org/10.3390/rs10020200

Schlapfer, D., Richter, R., Feingersh, T., 2015. Operational BRDF Effects Correction for Wide-Field-of-View Optical Scanners (BREFCOR). IEEE Transactions on Geoscience and Remote Sensing 53, 1855–1864. https://doi.org/10.1109/TGRS.2014.2349946

Teillet, P.M., Guindon, B., Goodenough, D.G., 1982. On the Slope-Aspect Correction of Multispectral Scanner Data. Canadian Journal of Remote Sensing 8, 84–106. https://doi.org/10.1080/07038992.1982.10855028

Thompson, D.R., Guanter, L., Berk, A., Gao, B.-C., Richter, R., Schläpfer, D., Thome, K.J., 2019. Retrieval of Atmospheric Parameters and Surface Reflectance from Visible and Shortwave Infrared Imaging Spectroscopy Data. Surv Geophys 40, 333–360. https://doi.org/10.1007/s10712-018-9488-9

Thompson, D.R., Guanter, L., Berk, A., Gao, B.-C., Richter, R., Schläpfer, D., Thome, K.J., 2018. Retrieval of Atmospheric Parameters and Surface Reflectance from Visible and Shortwave Infrared Imaging Spectroscopy Data. Surveys in Geophysics. https://doi.org/10.1007/s10712-018-9488-9

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution 10, 225–232. https://doi.org/10.1111/2041-210X.13107

Verrelst, J., Schaepman, M.E., Koetz, B., Kneubühler, M., 2008. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. Remote Sensing of Environment, Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue 112, 2341–2353. https://doi.org/10.1016/j.rse.2007.11.001

Wu, S., Wen, J., You, D., Hao, D., Lin, X., Xiao, Q., Liu, Q., Gastellu-Etchegorry, J.-P., 2018. Characterization of Remote Sensing Albedo Over Sloped Surfaces Based on DART Simulations and In Situ Observations. Journal of Geophysical Research: Atmospheres 123, 8599–8622. https://doi.org/10.1029/2018JD028283

X. General discussion

The main objective of my work was to propose an operational method of airborne remote sensing to produce inventories of tree species by jointly using hyperspectral and LiDAR information. Throughout this work, new and complementary results have been provided. They will be summarized and discussed in the following paragraphs.

At the core of much of the work presented so far lies the issue of the quality and quantity of data used to training classifiers for automatic species recognition. This discussion will thefore be organized around this emergent nexus, as training data availability appears to be a most critical factor.

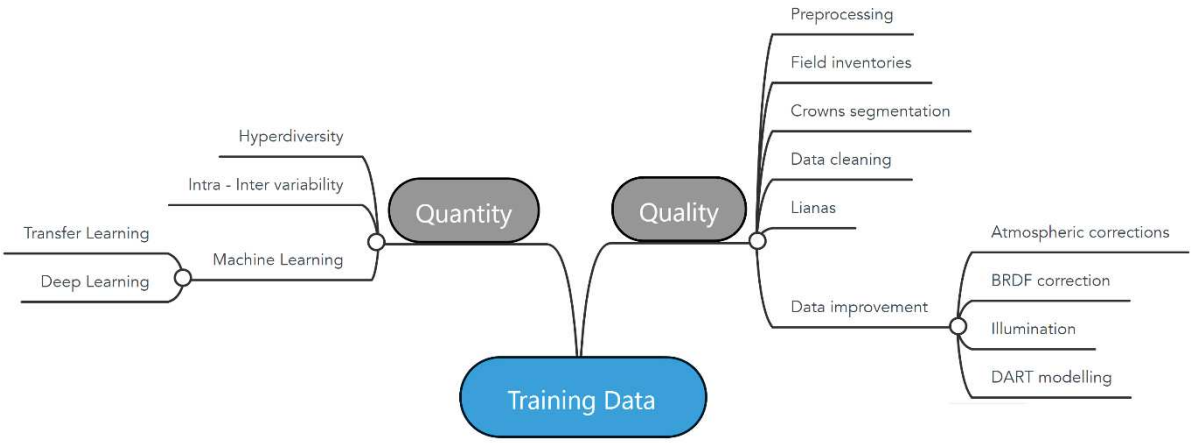

Figure 19: Map ideas from the thesis discussion related to training data uncertainty and possible improvements

Regardless of the imager, image acquisition is constantly improving. Driven by vectors such as drones more and more easily mobilized, sensors are improved. Increasingly efficient, faster and more robust detector placement through innovative cooling methods and improved optical path of light reduces sources of variation. However, inherent issues to the sensor should be addressed such as dead pixels (no signal) (Vidal and Amigo, 2012) or unexpected spectral values (Prats-Montalbán et al., 2011; Zhang et al., 2014). The purpose of preprocessing is to recover data free from prior contamination. However, it is necessary to identify the source of interference to use the most suitable pre-treatment. As reported by (Lenhard et al., 2015), the quality of the VNIR and SWIR imagers are satisfying. Both imagers are efficient radiometrically, spectrally and geometrically.

Other sources of noise caused by instrumentation can be reduced by spatial or spectral smoothing methods or both (Velasco-Forero and Manian, 2009). The noise often characterized as Gaussian and unstructured to simplify the working hypotheses, can be reduced by a spatial / spectral smoothing based on the average and those without altering the quality of the measurement. A commonly used technique is Savitzky-Golay spectral smoothing, using a moving average of a window size to be defined (King et al., 1999), because it does not reduce the identifying qualities of objects. Another method is to consider a spatial coherence and thus use a sliding window on each of the spectral bands. While the Savitzky-Golay was not helpful in discriminating between species, the spatial smoothing technique was profitable. In addition, the signal intensity normalization by the mean value contributed significantly to improving the distinction of the classes. This standardization, which is little used in remote sensing, although very simple to calculate, has offered better results than the use of radiance or raw reflectance. However, this normalization no longer makes it possible to consider the spectrum in reflectance value, known to be relatively invariant to the acquisition conditions. Normalizing the spectrum as expressed in the first chapter simply eliminates the effects of lighting, which is variable during acquisition and between dates. Therefore, it seems relevant to find a new methodology that takes into account this effect relating to lighting, integrated with atmospheric corrections. Note that the contribution of pretreatments is relative to the final objective. In our configuration, we assessed their contribution against the objective

of best identifying the 20 species of interest. Thus, their assessment will be relative to the defined criterion (eg: RMSE) or to the classification method considered.

In the perspective of developing species specific classifiers for a much larger range of species two additional difficulties may arise: taxonomic uncertainty and phylogenetic proximity. The identification of trees in the field is carried out primarily through their morphology. The presence of flower or fruit helps identification, but is not or only partially present during a year (Helmer et al., 2015). The tree's membership to a genus and a species can be difficult to ascertain (Duminil and Di Michele, 2009). Genomic information is now commonly used to delineate species (Heuertz et al., 2020). Most tropical forest inventories have many unidentified species, (Schweiger et al., 2018) found that dissimilarity of species' leaf spectra was positively related to functional dissimilarity and the evolutionary divergence time. In other words, species with strongly similar functional traits will tend to have more similar spectral signatures**.** The data set collected for this project over large plots with high quality botanical inventories could be further used to explore how spectral distance of tropical trees relates to phylogenetic and functional distance. This would open-up the way to mapping functional and phylogenetic diversity without the need for developing classifiers for every single canopy species.

The leaves orientation of trees plays the role of intercepting light (Muraoka et al., 1998) and may change with wilting. Change in leaf orientation and light scattering may affect the crown reflectance (Jonckheere et al., 2004). Even if the leaves are relatively small compared to the resolution of the sensor, change in leaf orientation may affect the anisotropic factor (Comar et al., 2014). More generally, a tree will adjust to its environment and the chemical composition will notably vary with nutrient availability, so the spectral signature of a particular species is expected to change with the local soil fertility. Indeed reflectance measurements can be used to infer the chemical and physical leaves' properties or the forest canopy. This makes it possible to estimate the nutrient content of tree leaves (Doughty et al., 2017; Feret et al., 2008b; Féret et al., 2017). The reflectance measured is therefore related to the availability of nutrients in the soil.

Be that as it may, remote sensing need information at crown level. The use of majority voting to increase accuracy has already been demonstrated. The errors' structuring within the crown explains the origin of the errors. Edges and differently lit parts tend to be less predicted than homogeneously lit central parts. However, majority voting can be improved in a binary classification. The principle of majority voting is to define the class of the object in relation to the most abundant class (greater than 50%) attributed to the pixels. However, the canopy structure can induce a reduction in illumination relative to the position of the sun and create a masking of crown areas. In addition, this estimate may depend on the species. Using the training information, we can identify the optimal fill rate of the object. By comparing the fill rate to that of the overall accuracy, we can find the threshold that defines the minimum fill for optimal prediction.

In a non-binary classification, this methodology is more complicated. However, in this type of configuration, the use of multiple classes allows the misattribution error to be distributed among the classes, which will have less effects on the final overall precision at the object scale. Thus, we recommend optimizing the filling rate of objects (crowns) and the use of several classes in order to improve the certainty of the classes.

On the other hand, the relationship between the area of the crown and the dbh of the trees is useful to make prediction of the dbh distribution within a given area. This is made possible by using models relating basal area and crown area or tree height, specific to each species. If most of this resource is in small stems, then this will have little commercial value. Based on the dbh distribution a better appraisal of costs and benefits becomes possible ex-ante.

As part of the same project, (Aubry-Kientz et al., 2019) first evaluated different individual crown segmentation strategies applied to dense tropical forest canopy. It was found that methods based on point clouds were more efficient than those based on digital surface models (DSM). In addition, the AMS-3D method (Ferraz et al., 2016b, 2012) was the one offering the best overall performance when segmentation quality of both small and large tree crowns was considered. The AMS-3D method uses three-dimensional information from LiDAR data only and considers the point cloud as a multimodal

3D distribution where each mode corresponds to the theoretical location of a tree crown. To find the modes, a Pollock function was used as bandwidth, the parameters changing with the height of the dots to allow taller crowns to be taller. (Aubry-Kientz et al., Submitted) further highlighted the benefits of fusion spectral and LiDAR fusion for the purpose of individual tree crown segmentation.

The use of CNN designed for object delineation applied to high spatial resolution multispectral data was shown to be effective for tree crown delineation (Weinstein et al., 2020, 2019). The spatial fusion of the two approaches, AMS-3D based on LiDAR corrected by the ImageNet neural network (Krizhevsky et al., 2017) built on RGB images, allowed an improvement in the characterization of the crowns (congruence: +4.7 %) (Aubry-Kientz et al., Submitted). (Dai et al., 2018) using multi-spectral aerial LiDAR showed that the use of multispectral LiDAR colorimetry allowed a 6% improvement over single LiDAR point clouds in delineating individual trees. Thus, the fusion of hyperspectral and LiDAR data (Dalponte et al., 2008; Tusa et al., 2020) using a very fine method (Brell et al., 2017) could offer improved results. In subtropical forest LiDAR has also been used in conjunction with spectral information for extracting additional features potentially useful for species identification, especially for differentiating conifers and deciduous trees (Shen and Cao, 2017). However, in the context of tropical forests, this information cannot help improving the discriminating power of hyperspectral due to complex ecosystems with overlapping species, nested crown structures and varying vertical distributions.

During the field campaign, there are many difficulties in associating trunk with crowns. Improving spectral stability involves detecting outliers. To do so, we calculated the Mahalanobis distance between the theoretical groups with respect to each of the elements in the group. This distance allowed us to exclude trees crowns which phenology were quite distinct (visual interpretation from RGB images), but leafless trees were a very small subset of the detected outliers. In addition, using multiple cross validation, we found out that some crowns were never attributed to their theoretical class. Based on this targeted information, a ground survey made it possible to correct a few labelling errors made in the original database. Therefore, hyperspectral technology has shown its usefulness to correct attribution errors.

Before considering the improvement of the data by knowing the origin of the disturbances, it is important to evaluate the confusions linked to the characterization of the measured surface. Lianas are an example of sources interference in the spectral signal.

An important and less well-known component of tropical forest diversity is the lianas (lianas sensu stricto and hemi-epiphytes) which include thousands of species (Dewalt et al., 2000). The current impact of climate change on the proliferation of lianas (Körner 2009) may be contributing to the reduction of carbon sinks (Brienen et al., 2014; Schnitzer, 2014; van der Heijden et al., 2015) by modifying the dynamics of the environment through different aspects. Lianas are effective competitors who play in multiple arenas. This is because the canopy part of the vine has the ability to move both horizontally and vertically with a relatively fast growth rate compared to the trees' (Van der Heijden and Phillips, 2009). They therefore do not have a well-defined structure (Rowe and Speck, 1996) which complicates their identification. In the presence of lianas, the diameter growth rate and stem sap flow are significantly reduced (Schnitzer, 2014) as is the height of the canopy. Carbon accumulation is also reduced by the root competitiveness of lianas. The presence of lianas decreases the recruitment of young tree seedlings by preempting the available space. In addition, tree mortality is also increasing because of the weight of the lianas that have climbed to the crown of their hosts. In the long term, the amount of carbon fixed by the rainforest may be reduced. Lianas development could be favored by logging (Laurance et al., 2001; Putz, 1985) or disturbances by increasingly frequent extreme climatic events (Schnitzer and Bongers, 2011). The proliferation of lianas can therefore have a profound impact on the functioning and the floristic composition of tropical forests. The lianas mapping in rich forest areas is therefore of direct interest for managers in tropical forest contexts for the characterization of current and future forest dynamics through the temporal monitoring of the development of this component of the vegetation.

An internship has been realized on this subject. Thanks to the visual interpretation, a delimitation of lianas and tree crowns classes have been realized. By two types of classifier (PLSDA and RandomForest), it has been shown that it is possible to distinguish the lianas and trees. For this work, hyperspectral preprocessed as proposed in the first chapter have been done. Textural information have been computed on RGB images. The results are satisfactory using RGB data (89.9%) or hyperspectral data (88.2%). The performance results were explained for the hyperspectral data by the difference of leaf characteristics between the lianas and the tree crown (photosynthetic activity higher for lianas). The texture information of the RGB images highlighted the low relief of the lianas on the surfaces which they cover. Both information are complementary and using both for discriminate trees and lianas, improves the accuracy by 3%. More details can be found in the appendix.

We have explored the contribution of hyperspectral signal processing such as the application of atmospheric corrections. When considering a single date, atmospheric corrections are not necessary. However, when considering multiple sites or dates, we observed a decrease in the ability (or even an inability) to identify species. Using a naive approach, we tried to predict the presence of certain species on the Montagne-Tortue and Nouragues sites. The training was based exclusively on data from Paracou. In addition, we found spectral discrepancies, which provenance we assessed. In this section, we discuss the stability of the spectral signature between different dates at the same site.

To assess the transferability of spectral signatures, it is first important to define the measured / estimated physical quantities. The term reflectance is often ambiguously used and does not strictly describe the product used (Schaepman-Strub et al., 2006). The radiances converted by the ATCOR processing chain are called "surface reflectance" and correspond to top of canopy hemispherical – directional reflectance factor (HDRF) i.e. the irradiance is considered hemispherically while only the vertically reflected radiance is considered.

The reflectance of the surface of a material is defined as its effectiveness in reflecting radiant energy. It is a function of the wavelength of the light and the angle of incidence. However what is directly measured is the "at sensor radiance" and the corresponding surface reflectance has to be inferred from the prevailing illumination conditions, atmospheric disturbance corrections and solar-sensor geometry. This is normally done by applying atmospheric correction methods such as ATCOR-4, which was used in the present study.

We evaluated the sources of variability between dates, considering the two same sites which were imaged on successive days. SNR calculation revealed a very weak signal for the SWIR-2 part (2000 nm to 2400 nm) of the spectral range. This weak signal reduced the reliability of the atmospheric correction, and this resulted in a greater inconsistency between dates. Using a GAM (Generalized additive model) model, we evaluated the contribution of factors relating to the acquisition geometry (sun, sensor and slope) in relation to the SAM dissimilarity index calculated between dates. The solar position (azimuth and solar elevation, in order of importance) was the most important one. These variations, linked to the time of acquisition, suggest making acquisitions during a restricted period of the day in order to limit this source of dissimilarity. While the acquisition time should be as close as possible to solar noon it should nevertheless avoid the passage of solar at zenith during the acquisition (hot spot effect).

The atmospheric corrections are also subject to errors due to the estimation of the water vapor concentration and the optical thickness of aerosols (Bhatia et al., 2018). The spatial variability of AOT has recently been considered (Schläpfer et al., 2018) but taking it into account did not improve our ingle-date classification. The errors made on WV and AOT may have compound effects further increasing the error in the estimate of reflectance. The approach recently proposed by (Thompson et al., 2016) using Bayesian inference to remove systematic errors in atmospheric correction is interesting. Indeed, this approach makes it possible to compensate for the effects of atmospheric perturbations by combining a physical approach using a radiative transfer model and a statistical approach. However, its contribution to improving the similarity between dates would need to be tested. (Duffy et al., 2019) compared a physical model and a deep learning approach to remove the effects of atmospheric scattering and absorption as well as effects due to acquisition geometry for a time series of satellite images. The use of a geostationary satellite (Himwari-8 AHI) guaranteed a large redundancy in spectral information and a

large range of acquisition geometries. In which case the deep learning method was able to emulate a physically based approach (itself not free from biases and errors of its own).

The BRDF correction model makes it possible to homogenize the successively acquired images. The solar position varying between two acquisitions, different visuals are perceptible. A BRDF model is composed in two kernel in general. The first one takes into account the geometric scattering and the second one takes into account the volume scattering. BRDF assumes a flat topographic. In addition, scattering effect depends on of the structure characterised, which it can be debatable for tropical forest context at a fine scale (Wen et al., 2018). Moreover, as showed by DART model, the anisotropy depends on leaf orientation and density distribution (Gastellu-Etchegorry et al., 2017). Soil contribution can also modified the albedo.

We have evaluated the contribution of acquisition parameters to understand the SAM dissimilarity, which showed a variation of contribution factors related to landcover type. The SAM dissimilarity for forest landcover was mostly affected by acquisition parameters. We tested this correction (using BREFCOR module, ATCOR-4 (R Richter and Schlapfer, 2018)) in our first study in order to assess its interest in species discrimination. This degrades performance. We will assess its contribution to the reduction of variations (using SAM dissimilarity) between dates, observed on the Nouragues site, considering exclusively forest surface (masking the other surfaces).

We found that spectral variations persisted between acquisitions over the same study site, despite atmospheric corrections aimed at normalizing the spectral signature. The passive character of this technology induces its dependence on the sun. Conversely, LiDAR, which is a so-called active technology, does not have this type of constraint. The fusion of data between the hyperspectral images and the LiDAR data by a back-projection method (Brell et al., 2016; Valbuena et al., 2011) would offer a finer method to correct the effects of variations in illumination compared to the intensity measured by LIDAR (Brell et al., 2017). We tried but failed to reproduce Valbuena's methodology for hyperspectral data but the acquisition method between RGB and hyperspectral (pushbroom) data are very different.

An aspect that was not covered in this thesis is the potential of radiative transfer model to better understand the spectral variations related to surface-solar-view geometry. DART (Discrete Anisotropic radiative transfer) is one of the most comprehensive radiative transfer models (Gastellu-Etchegorry et al., 2017). It can use a detailed 3D description of the scene of interest including the plant area density (m^2/m^3) of the vegetation cover which is accessible via LiDAR (Vincent et al., 2017) and leaf optical properties of leaves (Feret et al., 2008b). DART allows to specify the composition of the atmosphere, the solar geometry and the sensor specifications. Through simulations, it might be possible to better understand and rank the most factors contributing to the spectral variation both temporally and spatially. This kind of tool may also be used to generate many realistic acquisition configurations to train deep learning methods that would only need to be refined using real world data.

The quality of the hyperspectral was discussed with the aim of improving species identification and understanding the inconsistencies of spectral signatures. However, as it stands, we discuss the limits imposed by the object study and the possibilities to improve discrimination and spectral coherence.

The possibility of large-scale inventory of canopy trees in tropical forests was the main goal of this work. Beforehand, it is therefore necessary to make a detailed field inventory of the tree crowns to be characterized. We observed two major constraints for the identification of trees crowns by field measures in forest: hyper-diversity and tree density, limiting penetration into the forest. The number of data is limited by the fact that many tree species coexist. On the Paracou site for example, over the 118 hectares have counted more than 740 species distributed unfairly among more than 75,000 trees. A small proportion emerged in the canopy. Emerging crowns are also unevenly distributed among a large number of species. Among the crowns characterized on the Paracou site, only 20 species have a number of crowns greater than 24 crowns. For the purpose of exploiting the timber resource, it is preferable from a practical point of view to identify species which have a high added value timber (timber) and which are abundant. It would be necessary to think of grouping together the species with high added value in order to identify this procession of trees of interest.

The question of how to make a spectral signature unique was asked in the 1990s by (Price, 1994), arguing that a spectral signature for each species may not be possible. (Castro-Esau et al., 2006) examined the influence of intra- and inter-specific spectral variation at different scales and the direct consequence on the identification of tree species. This detailed study focused on tropical forests. They highlighted the difficulty of identifying the same species at different seasons and on different sites, considering more than 50 species in their study. There is therefore a factor linked to intra-species variability and a factor linked to seasonality.

Change in plant phenology over time will cause spectral variations to occur (Meerdink et al., 2019). Intra-crown variation in phenology was found to be lower during the rainy season by (Lopes et al., 2016) who monitored 267 individual tree crowns in an evergreen Central Amazon forest. (Lopes et al., 2016) calculated the EVI index of more than 200 tree crowns at different phenological stages using QuickBird images acquired at nadir. They identified that this index varied over time and that it was correlated with seasonality as well as with the EVI calculated from MODIS data, confirming that the EVI index could capture phenology and that it was not due to illumination geometry. However, (Dennison and Roberts, 2003) used the multiple endmember spectral mixture analysis (MESMA) algorithm to identify spectral variations due to seasonal variations. MESMA models mixed spectra as a linear combination of pure spectrum. They used the AVIRIS sensor to carry out this work, from which they were able to conclude that acquisition variations have a significant effect on the change in spectra. The study focused only on a wooded biome of chaparrals (a kind of maquisfound in California). (Ma et al., 2020) observed the phenology at sites of forests, woodlands and grasslands in South West Australia. Using a BRDF correction model based on the RossThick-LiSparse-Reciprocal kernel, they showed that indices like NDVI or EVI were very dependent on the acquisition geometry, noting that the EVI index was 50% less sensitive than the NDVI index. The EVI was more consistent over two successive phenological cycles.

Hyperspectral software exist to associate hyperspectral measurements with their metadata. It is possible to identify two main systems developed for close range (ASD) measurements: SPECCHIO (Bojinski et al., 2003) and SpectraProc (Hueni and Tuohy, 2006). SPECCHIO in particular offers data including detailed metadata describing the environment, sampling geometry, spatial position, target type, object type, measurement sensor and acquisition campaign. Sharing among a spectroscopy community labelled airborne images acquired over various vegetation types, might help understand the sources of spectral dissimilarity. However, there is no dedicated system for airborne imagery available. The latter would need to include radiance images and detailed acquisition characteristics to allow the database to be useful.

Faced with the high variability and low training sample rates, we evaluated the performance of classifiers in several training configurations.

Supervised classification methods such as the SVM (Support Vector Machine), Random Forest, LDA (Linear Discriminant Analysis), Logistic Regression methods have been widely tested by the scientific community. The use of hyperspectral data requires good management of large dimensionality (curse for dimensionality) (Qian et al., 2009). For this, dimensionality reduction methods such as Principal Component Analysis can be mobilized upstream. (Landgrebe, 2003) made the point that too many spectral bands can be a limit to the classification performance. (Fukunaga, 1990) demonstrated a relationship between dimensionality and the number of samples needed for classification. The number of training samples required is linearly related to the dimensionality and this for different types of classifiers. In our setup, the use of PCA followed by an SVM classification method was not found to be beneficial for species identification. In chapter 1, we justified the choice of the LDA method for its ability to manage the large dimensionality of the explanatory variables, as well as for its ease of calculation. As a reminder, this method maximizes inter-specific variability and minimizes intra-specific variability. However, this approach can be improved through the use of RDA (Regularized Discriminant Analysis) which estimates a co-variance matrix specific to each of the classes (used in Chapter 2).

We also confirmed that the increase in the number of classes induced a decrease in species classification accuracy. The efficiency limit of the classifier can come from its intrinsic properties. In

addition, the proportion of each class in the classification process is commonly unbalanced. This will be the case for species relative abundance in tropical forest.

Severals studies (Ghamisi et al., 2015; Ghamisi and Benediktsson, 2015; Melgani and Bruzzone, 2004; Pal and Mather, 2006) have shown that SVMs can efficiently classify the input data even if a limited number of samples learning is available. This is explained by the fact that only samples that are close to the boundary between classes are useful for delineating the boundary. (Feret and Asner, 2013), for instance, showed that for the tree species identification, SVM method (whatever the kernel) achieved accuracy comparable to linear classification methods (such as LDA and RDA). We found that RDA method was capable of identifying a particular species, using a small sample (30 crowns). In contrast the so-called non-focal class, consisted of pixels drawn randomly from the hyperspectral image (Chapter 2). A selection of 100,000 pixels was sufficient to clearly delineate the boundary between the two classes and represent the hyper-diversity contained in the hyperspectral data.

The random selection of pixels on the study site may be subject to an assignment error. This error is function of the representation of the species in the study space. In view of the great floristic diversity, the representation of a particular canopy species is often less than 5 % (except in specific cases: (Kamal and Phinn, 2011; Traissac and Pascal, 2014)). We assessed the sensitivity of 3 classifiers to assignment errors. The results were very sensitive to a pollution rate less than or equal to 5%. This is reassuring, especially since assignment errors in the field and consequently in the data used for the creation of the classification model certainly occur.

An aspect not evaluated in this thesis is the contribution of deep learning methods to the identification of species in tropical environments using hyperspectral data. This subject is also missing in the literature to my knowledge, although they are methods offering great potential for hyperspectral imagery (Zhang et al., 2006). The majority of studies based on hyperspectral data use well-known data sets such as “ROSIS-03 - Pavia University”, “AVIRIS - Indian Pines” or even “CASI - Houston”. The scientific community attributes to deep learning methods the ability to find invariant and abstract characteristics in hyperspectral data, regardless of the type of network SAE (Vincent et al. 2010), DBN (deep belief network) (Hinton, Osindero) , and Teh 2006) or CNN (convolutional neural network) (Krizhevsky et al., 2017). Their use may be limited due to the large amount of pixels required for training. For example, (Chen et al., 2014; Y. Chen et al., 2015; Pan et al., 2016), used half of the pixels in the annotated image. In our context, it is impossible to annotate so much information. On the other hand, neural networks use 2D convolutions to find the most relevant filters. The 2D convolutions which are morphological filters analyze the texture and shape of the objects to be identified. The shape of the crowns is very variable, constrained by its neighbors and the availability of light (Sterck and Bongers, 2001). These two types of information may not be very specific to species characteristics. Spatial patterns does not seem to be the most relevant characteristics to build upon to improve classification.

As described previously, we failed to identify the species on the other two study sites. We trained the classifier on the most complete data (Paracou site) in order to detect the species of interest on new sites (Montagne Tortue and Nouragues) which would limit the field inventories to targeted areas. We then turned to methodologies who allow limiting new field inventories.

Specific methods such as Transfer Learning method could help identifying tree species on a large scale. The main objective of Transfer Learning is to create effective classifications using two data sources, one of which is not or only partially labeled. In transfer learning, reference information, the most well-known and rich information, is considered to be part of the “source domain” to which we associate a probability distribution Ps (Xn, Yn). It is the conditional probability of a label being X given the observed features. The low density or zero information is part of the “target domain” to which we associate a probability distribution Pt (Xm, Ym). X is the vector corresponding to the spectrum. Y is the output variable (labeling) associated with X. The supervised methods of Transfer Learning assume that the number of labels of the source domain must be much higher than the labels of the target domain (in other words, Yn >> Ym).

There are several approaches to Transfer Learning: homogeneous and heterogeneous knowledge transfer. The one that concerns us is the homogeneous knowledge transfer that is to say that the input

data will be similar in the two domains. The number of explanatory variables at the input of the two domains must be strictly equal. In this branch of homogeneous Transfer Learning several approaches have been proposed: Inductive, transductive and unsupervised Transfer Learning. In the first case, we know the labeling of data in both fields. In the second case, we only know the labels for data in the source domain and we do not know the labels of the target domain. Unsupervised Transfer Learning is defined when we do not know the labels in the two fields. The choice of method therefore depends on the data available. Figure 2 of (Pan and Yang, 2010) describes in more detail the different strategies and their implications. For our purpose, we can consider that the source data is available but that the target data may be missing. However, we may still consider to blindly using the target data.

The search for an invariant, one of the knowledge transfer methods, would lead to descriptor reduction. (Persello and Bruzzone, 2016) used a method based on the optimization of two criteria which are the capacity of discrimination and the search for optimal invariants. They applied it to two hyperspectral images showing the ability to identify in both domains. However, the source domain discrimination capabilities have been reduced due to the reduction in the wavelength number. Indeed, the decrease in descriptor for identifying trees in a space as diverse as tropical forest will lead to a decrease in classification performance (Ferreira et al., 2016b). The invariant search could be replaced by an adaptation between the probability distributions of the two domains (Covariate Shift), which no reduce the dimentionality. The goal is to find a function that matches the two distribution probabilities (align the distribution) while minimizing the energy due to moving the data from an initial distribution to the most optimal. However, for very large dimensionality data this can be very penalizing (Polo and Vicente, 2020) and may not be computationally feasible (Pan et al., 2011, 2008). Another way is offered to us. Boosting Technology methods can contribute to transfer learning. These methods are iteratively improved by using poorly performing binary classifiers to form a better performing set of classifiers, where the weights of misclassified instances are adjusted so that the classifiers focus on the difficult cases. (Dai et al., 2007) proposed an algorithm named "TrAdaBoost" which is mainly based on the "AdaBoost" method. AdaBoost, which is also a meta-algorithm, separates data into several parts. By successive iteration and by the use of multiple classifiers, the predictive power is improved compared to the deemed weak algorithm. The idea of the TrAdaBoost method is that the distributions of the data in the two areas are different. However, some data from the source domain can be useful for learning the target domain. By automatically adjusting the weights of training instances, increasing it for well-assigned training instances, the meta-classifier is improved. However, this method, which is the basis of many publications, has some limitations such as the problem related to class imbalance and that of negative transfer, which can be corrected by a new strategy of initializing the new weights and updating them.

To summarize, we are limited both by the number of examples available but also by the data structure used. To maximize species identification, we must use all descriptors. However, this large dimensionality leads to computational complications. In addition, the strong disproportion between classes is also an issue. However, the amount of information in the source domain necessary for the transfer of knowledge has not yet been investigated to our knowledge except for Deep Learning methods (Soekhoe et al., 2016).

Another interesting point to explore would be whether or not it is necessary to know the class of interest in the target domain and replace it with another piece of data common to both domains. Identifying specific species in tropical forests is tricky for the many reasons mentioned above. However, common elements can potentially be identifiable by visual interpretation, such as roads or tracks. Whether such limited information common to the different domains would suffice to develop transfer function applicable to vegetation spectra remains to be explored. Little work has been conducted so far on transfer learning of hyperspectral classifiers applied to species recognition in tropical forest, certainly due to the limited amount of terrain data and the lack of suitable hyperspectral images of high spatial and spectral resolution.

XI. Conclusion

The aim of this work was to assess the feasibility identication and mapping of trees via airborne remote sensing. We explored the capacity of hyperspectral imagery and LiDAR to meet this challenge.

During this work, we tested the capabilities and limitations of classifiers applied to the identification of trees in tropical forests. We showed that pre-treatments such as spatial smoothing or standardization were necessary. In addition, atmospheric corrections proved useful only when tree identification had to be performed on multiple sites. Thus, we have confirmed the ability of hyperspectral information to discriminate tree species.

We made a first step towards operational deployment of the method by showing that target trees species could be retrieved in a highly diverse canopy even with limited ground truthing effort. For this step, the coupling with LiDAR information was very useful to segment individual crowns. Segmentation not only improved identification but also quantification of the resource since the dbh could be estimated from the crown size.

The major remaining hurdle to overcome is the transposability of hyperspectral signatures. In the first analysis presented, we touched upon the issue by using an acquisition made on one date for prediction on a second date, on the same study site. We noted a significant loss in classification accuracy by doing so. Using data from another site, as described in the third chapter, we sought to identify the sources of the spectral distortions that affected spectral signatures consistency and thus the potential for discriminating of trees at large scale. The identification of these sources of disturbance should help pave the way towards a robust approach to deal with those issues.

In view of the current state of the literature and our own results, we believe that the joint use of hyperspectral and LiDAR information should be encouraged. The fusion of LiDAR and hyperspectral data by back-projection based on the work of (Valbuena, 2014; Valbuena et al., 2011) (LiDAR and RGB) and (Brell et al., 2016) should be democratized by proposing a turnkey fusion method. As described (Brell et al., 2017), the fusion of these sources of information allows to better characterize the reflectance and to correct it for effects related to the acquisition geometry and atmospheric perturbations. By comparing the reflectances (LiDAR and hyperspectral), it is possible to refine the atmospheric corrections. On the other hand, the fusion of these two pieces of information could allow the improvement of segmentation methods. (Dai et al., 2018) have proposed a segmentation method based on the Mean-Shift algorithm, using the topographic characteristics (X, Y, Z) but also the colorimetric characteristics of some channels.

Still with the aim of improving the identification of trees on a large scale, the use of Deep Learning methods may prove useful. Indeed, by coupling the physical approach and Machine Learning, it has been possible to reduce the impact of shadows (Windrim et al., 2018, 2016) and to improve atmospheric corrections (Duffy et al., 2019; Thompson et al., 2016). In this sense, new acquisitions will have to be carried out to capture or automatically construct invariants.

The use of UAVs and new hyperspectral imagers is particularly valuable in tropical forests where cloud cover is important. The rapid deployment of this vector is its strength. In addition, the new HysSpex Mjolnir VS-620 has all the advantages of the sensors we used in this thesis.

It therefore has all the means to advance the understanding of this very special biome while carrying out reasoned resource extraction.

XII. Appendices

Appendix -1 – Lianas as a complicating factor to species recognition from the air

As part of an internship under my co-supervision, Fiston Nininahazwe had to identify the lianas in tropical environment by joint use of multispectral data with high spatial resolution RGB and imaging spectroscopy. Thus different sub-objectives were identified: 1 / Processing of acquired images, orthorectification, mosaicking. 2 / Manual delimitation of crowns entangled by lianas. 3 / Evaluation of the detectability of lianas on hyperspectral images and quantification of the degree of liana coverage.

To carry out this study, the Nouragues area was selected. This area, known for a stronger presence of lianas than on the other study sites available to us, was therefore favored. A preliminary work was to identify the lianas by multiple aerial RGB images (drone). This work is very delicate because it is difficult to assess the presence / absence of lianas, even at a spatial resolution of a few cm. Thanks to these data, 230 delimitations, including 129 and 101 polygons lianas and trees respectively, were mobilized. Hyperspectral data in the spectral range of 414 nm to 994 nm, at 1m spatial resolution was also used. Hyperspectral data had been corrected for atmospheric effects. Spatial smoothing with a window of 3x3 pixels followed by normalization by the mean was applied as described in chapter 1. RGB images were mobilized to calculate seven texture indices (Haralick, Shanmugam, and Dinstein 1973) on a window of 11x11 pixels, averaged in 4 directions (0 °, 45 °, 90 ° and 135 °): mean - variance- correlation - homogeneity - dissimilarity -entropy - contrast.

The evaluation of the discrimination of lianas compared to trees was carried out by PLSDA (Discriminant Analysis by Partial Least Squares) and RandomForest. These methods also made it possible to identify the contribution of variables to the determination of the class of membership by the classifier. We found an equivalent identification rate between the two data used as well as a small increase due to their association. The hyperspectral, the texture indices and the combined use of these two sources of information allowed an identification of 88.2% (± 4.1%), 89.9% (± 2.6%) and 93.1% (± 2.7%) respectively. Remarkably texture was more informative than imaging spectroscopy for liana detecition in that study.

The average reflectance for the wavelength around 680 nm is lower for lianas compared to trees. The reverse is observed for the wavelength around 550 nm where the lianas have a greater reflectance. This is mainly explained by a greater photosynthetic activity, especially for chlorophyll pigments (Sims and Gamon, 2002). This is true at the leaf scale but also at the canopy scale (Castro-Esau, 2004). The lianas development strategy may explain this. The lianas are not self-supporting as they reach the canopy thanks to the trees present and can thus quickly capture a maximum of light in order to grow (principle of autotrophy). By analyzing the most contributing wavelengths from classification methods, a greater proportion of near-infrared bands emerge. The visible bands contribute, but to a lesser extent.(Kalacska et al., 2007; Marvin et al., 2016) confirmed the importance of these spectral regions at the scale of the canopy, while at the scale of the organ (the leaf) the visible was the most contributory area.

While the hyperspectral focuses on spectral bands relating to the water content between liana leaves and tree leaves (Castro-Esau, 2004; Sánchez-Azofeifa et al., 2009; Schnitzer, 2005), the texture from RGB images apprehends the spatial relationships between different objects (Haralick et al., 1973). In extreme cases (as here in Nourageus) lianas may literelly blanket the tree canopy and contribute to reducing the canopy roughness. The tangled areas therefore appear smoother (Waite et al., 2019).

The results of this internship could not be confirmed / invalidated due to the lack of suitable field data. This is partly explained by the fact that lianas are difficult to observe and identify, especially in the canopy. (Cox et al., 2019) found that there was no relationship between the number of lianas stems at ground level and the level of infestation of canopy crowns. Lianas may grow horizontally to reach the canopy and then develop on another crown or may not emerge in the crown where they started growing. Only an assessment in the canopy can be reliable as can be done using close-range drone imagery (Waite et al., 2019).

On the other hand, it is necessary not to consider lianas as a single category in the same way as trees. Indeed, we consider here the lianas as unique same species whereas each species has different growth strategies (Rowe et al., 2004; Rowe and Speck, 2005).

XIII. Références

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., Sousa, J., 2017. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. Remote Sens. 9, 1110. https://doi.org/10.3390/rs9111110

Alonzo, M., Bookhagen, B., Roberts, D.A., 2014. Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sens. Environ. 148, 70–83. https://doi.org/10.1016/j.rse.2014.03.018

Asner, G., 2008. Hyperspectral Remote Sensing of Canopy Chemistry, Physiology, and Biodiversity in Tropical Rainforests, in: Kalacska, M., Arturo Sanchez-Azofeifa (Eds.), Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. CRC Press, pp. 261–296. https://doi.org/10.1201/9781420053432.ch12

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., Coomes, D., Piboule, A., Vincent, G., 2019. A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests. Remote Sens. 11, 1086. https://doi.org/10.3390/rs11091086

Aubry-Kientz, M., Laybros, A., Weinstein, B.G., Ball, J., Jackson, T., Coomes, D., Vincent, G., Submitted. Multi-sensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. J. Sel. Top. Appl. Earth Obs. Remote Sens.

Baldeck, C., Asner, G., 2014. Improving Remote Species Identification through Efficient Training Data Collection. Remote Sens. 6, 2682–2698. https://doi.org/10.3390/rs6042682

Ballanti, L., Blesius, L., Hines, E., Kruse, B., 2016. Tree Species Classification Using Hyperspectral Imagery: A Comparison of Two Classifiers. Remote Sens. 8, 445. https://doi.org/10.3390/rs8060445

Baltzer, C., 2015. Fusion de données de télédétection et dendrométriques pour l'étude des variations allométriques des arbres tropicaux (rapport de stage d'année de césure). Montpellier.

Baraldi, A., 2009. Impact of Radiometric Calibration and Specifications of Spaceborne Optical Imaging Sensors on the Development of Operational Automatic Remote Sensing Image Understanding Systems. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2, 104–134. https://doi.org/10.1109/JSTARS.2009.2023801

Barlow, J., França, F., Gardner, T.A., Hicks, C.C., Lennox, G.D., Berenguer, E., Castello, L., Economo, E.P., Ferreira, J., Guénard, B., Gontijo Leal, C., Isaac, V., Lees, A.C., Parr, C.L., Wilson, S.K., Young, P.J., Graham, N.A.J., 2018. The future of hyperdiverse tropical ecosystems. Nature 559, 517–526. https://doi.org/10.1038/s41586-018-0301-1

Bazezew, M.N., Hussin, Y.A., Kloosterman, E.H., 2018. Integrating Airborne LiDAR and Terrestrial Laser Scanner forest parameters for accurate above-ground biomass/carbon estimation in Ayer Hitam tropical forest, Malaysia. Int. J. Appl. Earth Obs. Geoinformation 73, 638–652. https://doi.org/10.1016/j.jag.2018.07.026

Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L., Girard, S., 2009. Retrieval of Mars surface physical properties from OMEGA hyperspectral images using regularized sliced inverse regression. J. Geophys. Res. Planets 114. https://doi.org/10.1029/2008JE003171

Bhatia, N., Iordache, M.-D., Stein, A., Reusen, I., Tolpekin, V.A., 2018. Propagation of uncertainty in atmospheric parameters to hyperspectral unmixing. Remote Sens. Environ. 204, 472–484. https://doi.org/10.1016/j.rse.2017.10.008

Bhatia, N., Tolpekin, V.A., Reusen, I., Sterckx, S., Biesemans, J., Stein, A., 2015. Sensitivity of Reflectance to Water Vapor and Aerosol Optical Thickness. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 3199–3208. https://doi.org/10.1109/JSTARS.2015.2425954

Bojinski, S., Schaepman, M., Schläpfer, D., Itten, K., 2003. SPECCHIO: a spectrum database for remote sensing applications. Comput. Geosci. 29, 27–38. https://doi.org/10.1016/S0098-3004(02)00107-3

Boltz, F., Holmes, T.P., Carter, D.R., 2003. Economic and environmental impacts of conventional and reduced-impact logging in Tropical South America: a comparative review. For. Policy Econ. 5, 69–81. https://doi.org/10.1016/S1389-9341(01)00075-2

Bongers, F., Charles-Dominique, P., Forget, P.-M., Théry, M., 2013. Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest. Springer Science & Business Media.

Brell, M., Rogass, C., Segl, K., Bookhagen, B., Guanter, L., 2016. Improving Sensor Fusion: A Parametric Method for the Geometric Coalignment of Airborne Hyperspectral and Lidar Data. IEEE Trans. Geosci. Remote Sens. 54, 3460–3474. https://doi.org/10.1109/TGRS.2016.2518930

Brell, M., Segl, K., Guanter, L., Bookhagen, B., 2017. Hyperspectral and Lidar Intensity Data Fusion: A Framework for the Rigorous Correction of Illumination, Anisotropic Effects, and Cross Calibration. IEEE Trans. Geosci. Remote Sens. 55, 2799–2810. https://doi.org/10.1109/TGRS.2017.2654516

Bréon, F.-M., Vermote, E., 2012. Correction of MODIS surface reflectance time series for BRDF effects. Remote Sens. Environ. 125, 1–9. https://doi.org/10.1016/j.rse.2012.06.025

Brienen, R.J.W., Phillips, O.L., Feldpausch, T.R., Gloor, E., Baker, T.R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo Mendoza, A., Malhi, Y., Lewis, S.L., Vásquez Martinez, R., Alexiades, M., Álvarez Dávila, E.A., Alvarez-Loayza, P., Andrade, A., Aragão, L.E.O.C., Araujo Murakami, A., Arets, E.J.M.M., Arroyo, L., Aymard C., G.A., Bánki, O.S., Baraloto, C., Barroso, J.G., Bonal, D., Boot, R., Camargo, J.L., Castilho, C., Chama, V., Chao, K.-J., Chave, J., Comiskey, J.A., Cornejo, F., Da Costa, L., De Oliveira, E.A., Di Fiore, A., Erwin, T.L., Fauset, S., Forsthofer, M., Grahame, S.E., Groot, N.E., Herault, B., Higuchi, N., Honorio C., E., Keeling, H., Killeen, T., Laurance, W., Laurance, S., Licona, J.-C., Magnussen, W.E., Marimon, B.S., Marimon-Junior, B.H., Mendoza, C., Neill, D., Nogueira, E.M., Nunez, P., Pallqui Camacho, N.C., Parada, A., Pardo, G., Peacock, J., Pena-Claros, M., Pickavance, G.C., Pitman, N.C.A., Poorter, L., Prieto, A., Quesada, C.A., Ramírez, F., Ramírez-Angulo, H., Restrepo, Z., Roopsind, A., Rudas, A., Salomão, R.P., Schwarz, M., Silva, N., Silva-Espejo, J.E., Silveira, M., Stropp, J., Talbot, J., Ter Steege, H., Teran-Aguilar, J., Terborgh, J., Thomas-Caesar, R., Toledo, M., Torello-Raventos, M., Umetsu, R.K., Van Der Heijden, G.M.F., Van Der Hout, P., Guimaraes Vieira, I., Vieira, Simone A, Vilanova, E., Vos, V., Zagt, R.J., 2014. Long-term decline of the Amazon carbon sink. https://doi.org/10.5521/ForestPlots.net/2014_4

Brown, K.A., Gurevitch, J., 2004. Long-term impacts of logging on forest diversity in Madagascar. Proc. Natl. Acad. Sci. 101, 6045–6049. https://doi.org/10.1073/pnas.0401456101

Burivalova, Z., Hua, F., Koh, L.P., Garcia, C., Putz, F., 2017. A Critical Comparison of Conventional, Certified, and Community Management of Tropical Forests for Timber in Terms of Environmental, Economic, and Social Variables: Certified and community forest management. Conserv. Lett. 10, 4–14. https://doi.org/10.1111/conl.12244

Carreño-Rocabado, G., Peña-Claros, M., Bongers, F., Alarcón, A., Licona, J.-C., Poorter, L., 2012. Effects of disturbance intensity on species and functional diversity in a tropical forest. J. Ecol. 100, 1453–1463. https://doi.org/10.1111/j.1365-2745.2012.02015.x

Castro-Esau, K., 2004. Discrimination of lianas and trees with leaf-level hyperspectral data. Remote Sens. Environ. 90, 353–372. https://doi.org/10.1016/j.rse.2004.01.013

Castro-Esau, K.L., Sanchez-Azofeifa, G.A., Rivard, B., Wright, S.J., Quesada, M., 2006. Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. Am. J. Bot. 93, 517–530. https://doi.org/10.3732/ajb.93.4.517

Che, H., Gui, K., Chen, Q., Zheng, Y., Yu, J., Sun, T., Zhang, X., Shi, G., 2016. Calibration of the 936 nm water-vapor channel for the China aerosol remote sensing NETwork (CARSNET) and the effect of the retrieval water-vapor on aerosol optical property over Beijing, China. Atmospheric Pollut. Res. 7, 743–753. https://doi.org/10.1016/j.apr.2016.04.003

Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014. Deep Learning-Based Classification of Hyperspectral Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 2094–2107. https://doi.org/10.1109/JSTARS.2014.2329330

Chen, Y., Zhao, X., Jia, X., 2015. Spectral–Spatial Classification of Hyperspectral Data Based on Deep Belief Network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 2381–2392. https://doi.org/10.1109/JSTARS.2015.2388577

Chen, Y.-J., Cao, K.-F., Schnitzer, S.A., Fan, Z.-X., Zhang, J.-L., Bongers, F., 2015. Water-use advantage for lianas over trees in tropical seasonal forests. New Phytol. 205, 128–136. https://doi.org/10.1111/nph.13036

Clark, J.A., Covey, K.R., 2012. Tree species richness and the logging of natural forests: A meta-analysis. For. Ecol. Manag. 276, 146–153. https://doi.org/10.1016/j.foreco.2012.04.001

Clark, M., Roberts, D., Clark, D., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sens. Environ. 96, 375–398. https://doi.org/10.1016/j.rse.2005.03.009

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. J. Geophys. Res. Solid Earth 89, 6329–6340. https://doi.org/10.1029/JB089iB07p06329

Colgan, M., Baldeck, C., Féret, J.-B., Asner, G., 2012. Mapping Savanna Tree Species at Ecosystem Scales Using Support Vector Machine Classification and BRDF Correction on Airborne Hyperspectral and LiDAR Data. Remote Sens. 4, 3462–3480. https://doi.org/10.3390/rs4113462

Comar, A., Baret, F., Obein, G., Simonot, L., Meneveaux, D., Viénot, F., de Solan, B., 2014. ACT: A leaf BRDF model taking into account the azimuthal anisotropy of monocotyledonous leaf surface. Remote Sens. Environ. 143, 112–121. https://doi.org/10.1016/j.rse.2013.12.006

Cox, C.J., Edwards, W., Campbell, M.J., Laurance, W.F., Laurance, S.G.W., 2019. Liana cover in the canopies of rainforest trees is not predicted by local ground-based measures. Austral Ecol. 44, 759–767. https://doi.org/10.1111/aec.12746

Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. ISPRS J. Photogramm. Remote Sens. 144, 400–411. https://doi.org/10.1016/j.isprsjprs.2018.08.010

Dai, W., Yang, Q., Xue, G.-R., Yu, Y., 2007. Boosting for transfer learning, in: Proceedings of the 24th International Conference on Machine Learning - ICML '07. Presented at the the 24th international conference, ACM Press, Corvalis, Oregon, pp. 193–200. https://doi.org/10.1145/1273496.1273521

Dalponte, M., Bruzzone, L., Gianelle, D., 2008. Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. IEEE Trans. Geosci. Remote Sens. 46, 1416–1427. https://doi.org/10.1109/TGRS.2008.916480

Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. 7, 1236–1245. https://doi.org/10.1111/2041-210X.12575

de Avila, A.L., Ruschel, A.R., de Carvalho, J.O.P., Mazzei, L., Silva, J.N.M., Lopes, J. do C., Araujo, M.M., Dormann, C.F., Bauhus, J., 2015. Medium-term dynamics of tree species composition in response to silvicultural intervention intensities in a tropical rain forest. Biol. Conserv. 191, 577–586. https://doi.org/10.1016/j.biocon.2015.08.004

de Lima, R.A.F., Mori, D.P., Pitta, G., Melito, M.O., Bello, C., Magnago, L.F., Zwiener, V.P., Saraiva, D.D., Marques, M.C.M., de Oliveira, A.A., Prado, P.I., 2015. How much do we know about the endangered Atlantic Forest? Reviewing nearly 70 years of information on tree community surveys. Biodivers. Conserv. 24, 2135–2148. https://doi.org/10.1007/s10531-015-0953-1

Deborah, H., Richard, N., Hardeberg, J.Y., 2015. A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 3224–3234. https://doi.org/10.1109/JSTARS.2015.2403257

Dennison, P.E., Roberts, D.A., 2003. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sens. Environ. 87, 123–135. https://doi.org/10.1016/S0034-4257(03)00135-4

Dewalt, S.J., Schnitzer, S.A., Denslow, J.S., 2000. Density and diversity of lianas along a chronosequence in a central Panamanian lowland forest. J. Trop. Ecol. 16, 1–19.

Dieterle, G., 2009. Sustaining the World's Forests: Managing Competing Demands for a Vital Resource – The Role of the World Bank, in: Spathelf, P. (Ed.), Sustainable Forest Management in a

Changing World: A European Perspective, Managing Forest Ecosystems. Springer Netherlands, Dordrecht, pp. 9–32. https://doi.org/10.1007/978-90-481-3301-7_2

Doughty, C.E., Santos-Andrade, P.E., Goldsmith, G.R., Blonder, B., Shenkin, A., Bentley, L.P., Chavana-Bryant, C., Huaraca-Huasco, W., Díaz, S., Salinas, N., Enquist, B.J., Martin, R., Asner, G.P., Malhi, Y., 2017. Can Leaf Spectroscopy Predict Leaf and Forest Traits Along a Peruvian Tropical Forest Elevation Gradient? J. Geophys. Res. Biogeosciences 122, 2952–2965. https://doi.org/10.1002/2017JG003883

Doxani, G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., Vanhellemont, Q., 2018. Atmospheric Correction Inter-Comparison Exercise. Remote Sens. 10, 352. https://doi.org/10.3390/rs10020352

Duffy, K., Vandal, T., Wang, W., Nemani, R., Ganguly, A.R., 2019. Deep Learning Emulation of Multi-Angle Implementation of Atmospheric Correction (MAIAC). ArXiv191013408 Cs Eess Stat.

Duminil, J., Di Michele, M., 2009. Plant species delimitation: A comparison of morphological and molecular markers. Plant Biosyst. - Int. J. Deal. Asp. Plant Biol. 143, 528–542. https://doi.org/10.1080/11263500902722964

Dutrieux, R., 2018. A comparative assessment of the performance of Individual Tree Crowns delineation algorithms from ALS data in tropical forests (stage de césure Ingénieur). AgroParisTech, Nancy.

Ellis, E.A., Montero, S.A., Hernández Gómez, I.U., Romero Montero, J.A., Ellis, P.W., Rodríguez-Ward, D., Blanco Reyes, P., Putz, F.E., 2019. Reduced-impact logging practices reduce forest disturbance and carbon emissions in community managed forests on the Yucatán Peninsula, Mexico. For. Ecol. Manag. 437, 396–410. https://doi.org/10.1016/j.foreco.2019.01.040

Ellis, P.W., Gopalakrishna, T., Goodman, R.C., Putz, F.E., Roopsind, A., Umunay, P.M., Zalman, J., Ellis, E.A., Mo, K., Gregoire, T.G., Griscom, B.W., 2019. Reduced-impact logging for climate change mitigation (RIL-C) can halve selective logging emissions from tropical forests. For. Ecol. Manag. 438, 255–266. https://doi.org/10.1016/j.foreco.2019.02.004

Eshleman, K.N., 2004. Hydrological consequences of land use change: A review of the state-of-science. Wash. DC Am. Geophys. Union Geophys. Monogr. Ser. 153, 13–29. https://doi.org/10.1029/153GM03

Fawcett, D., Verhoef, W., Schläpfer, D., Schneider, F.D., Schaepman, M.E., Damm, A., 2018. Advancing retrievals of surface reflectance and vegetation indices over forest ecosystems by combining imaging spectroscopy, digital object models, and 3D canopy modelling. Remote Sens. Environ. 204, 583–595. https://doi.org/10.1016/j.rse.2017.09.040

Feret, J.-B., Asner, G.P., 2013. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. IEEE Trans. Geosci. Remote Sens. 51, 73–84. https://doi.org/10.1109/TGRS.2012.2199323

Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., 2008a. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012

Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., 2008b. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012

Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215. https://doi.org/10.1016/j.rse.2017.03.004

Ferraz, A., Bretar, F., Jacquemoud, S., Gonçalves, G., Pereira, L., Tomé, M., Soares, P., 2012. 3-D mapping of a multi-layered Mediterranean forest using ALS data. Remote Sens. Environ. 121, 210–223. https://doi.org/10.1016/j.rse.2012.01.020

Ferraz, A., Saatchi, S., Longo, M., Clark, D.B., 2020. Tropical tree size–frequency distributions from airborne lidar. Ecol. Appl. https://doi.org/10.1002/eap.2154

Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016a. Lidar detection of individual tree size in tropical forests. Remote Sens. Environ. 183, 318–333. https://doi.org/10.1016/j.rse.2016.05.028

Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016b. Lidar detection of individual tree size in tropical forests. Remote Sens. Environ. 183, 318–333. https://doi.org/10.1016/j.rse.2016.05.028

Ferreira, M.P., Zortea, M., Zanotta, D.C., Shimabukuro, Y.E., de Souza Filho, C.R., 2016a. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. Remote Sens. Environ. 179, 66–78. https://doi.org/10.1016/j.rse.2016.03.021

Ferreira, M.P., Zortea, M., Zanotta, D.C., Shimabukuro, Y.E., de Souza Filho, C.R., 2016b. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. Remote Sens. Environ. 179, 66–78. https://doi.org/10.1016/j.rse.2016.03.021

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global Consequences of Land Use. Science 309, 570–574. https://doi.org/10.1126/science.1111772

Fukunaga, K., 1990. Introduction to statistical pattern recognition. Acad. Press, San Diego.

Galvão, L.S., Breunig, F.M., Santos, J.R. dos, Moura, Y.M. de, 2013. View-illumination effects on hyperspectral vegetation indices in the Amazonian tropical forest. Int. J. Appl. Earth Obs. Geoinformation 21, 291–300. https://doi.org/10.1016/j.jag.2012.07.005

Gao, L., Du, Q., Zhang, B., Yang, W., Wu, Y., 2013. A Comparative Study on Linear Regression-Based Noise Estimation for Hyperspectral Imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 488–498. https://doi.org/10.1109/JSTARS.2012.2227245

Gastellu-Etchegorry, J.-P., Lauret, N., Yin, T., Landier, L., Kallel, A., Malenovsky, Z., Bitar, A.A., Aval, J., Benhmida, S., Qi, J., Medjdoub, G., Guilleux, J., Chavanon, E., Cook, B., Morton, D., Chrysoulakis, N., Mitraka, Z., 2017. DART: Recent Advances in Remote Sensing Data Modeling With Atmosphere, Polarization, and Chlorophyll Fluorescence. IEEE J. Sel. Top. Appl. EARTH Obs. REMOTE Sens. 10.

Gerard, F.F., North, P.R.J., 1997. Analyzing the effect of structural variability and canopy gaps on forest BRDF using a geometric-optical model. Remote Sens. Environ. 62, 46–62. https://doi.org/10.1016/S0034-4257(97)00070-9

Ghamisi, P., Benediktsson, J.A., 2015. Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization. IEEE Geosci. Remote Sens. Lett. 12, 309–313. https://doi.org/10.1109/LGRS.2014.2337320

Ghamisi, P., Couceiro, M.S., Benediktsson, J.A., 2015. A Novel Feature Selection Approach Based on FODPSO and SVM. IEEE Trans. Geosci. Remote Sens. 53, 2935–2947. https://doi.org/10.1109/TGRS.2014.2367010

Giambelluca, T.W., 2002. Hydrology of altered tropical forest. Hydrol. Process. 16, 1665–1669. https://doi.org/10.1002/hyp.5021

Gibbs, H.K., Ruesch, A.S., Achard, F., Clayton, M.K., Holmgren, P., Ramankutty, N., Foley, J.A., 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s 6.

Goetz, A.F.H., Vane, G., Solomon, J.E., Rock, B.N., 1985. Imaging Spectrometry for Earth Remote Sensing. Science 228, 1147–1153. https://doi.org/10.1126/science.228.4704.1147

Gourlet-Fleury, S., Guehl, J.-M., Laroussinie, O., ECOFOR (Group) (Eds.), 2004. Ecology and management of a neotropical rainforest: lessons drawn from Paracou, a long-term experimental research site in French Guiana. Elsevier, Paris.

Griscom, B.W., Ellis, P.W., Burivalova, Z., Halperin, J., Marthinus, D., Runting, R.K., Ruslandi, Shoch, D., Putz, F.E., 2019. Reduced-impact logging in Borneo to minimize carbon emissions and impacts on sensitive habitats while maintaining timber yields. For. Ecol. Manag. 438, 176–185. https://doi.org/10.1016/j.foreco.2019.02.025

Gross, W., Tuia, D., Soergel, U., Middelmann, W., 2019. Nonlinear Feature Normalization for Hyperspectral Domain Adaptation and Mitigation of Nonlinear Effects. IEEE Trans. Geosci. Remote Sens. 57, 5975–5990. https://doi.org/10.1109/TGRS.2019.2903719

Hamraz, H., Contreras, M.A., Zhang, J., 2017. Vertical stratification of forest canopy for segmentation of understory trees within small-footprint airborne LiDAR point clouds. ISPRS J. Photogramm. Remote Sens. 130, 385–392. https://doi.org/10.1016/j.isprsjprs.2017.07.001

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hari Poudyal, B., Maraseni, T., Cockfield, G., 2018. Evolutionary dynamics of selective logging in the tropics: A systematic review of impact studies and their effectiveness in sustainable forest management. For. Ecol. Manag. 430, 166–175. https://doi.org/10.1016/j.foreco.2018.08.006

Hastie, T., Tibshirani, R., 1999. Generalized additive models. Chapman & Hall/CRC, Boca Raton, Fla.

Helmer, E., Goodwin, N.R., Gond, V., Souza, C., Asner, G., 2015. Characterizing tropical forests with multispectral imagery. pp. 363–391.

Heuertz, M., Caron, H., Scotti-Saintagne, C., Pétronelli, P., Engel, J., Tysklind, N., Miloudi, S., Gaiotto, F.A., Chave, J., Molino, J.-F., Sabatier, D., Loureiro, J., Budde, K.B., 2020. The hyperdominant tropical tree Eschweilera coriacea (Lecythidaceae) shows higher genetic heterogeneity than sympatric Eschweilera species in French Guiana. Plant Ecol. Evol. 153, 67–81. https://doi.org/10.5091/plecevo.2020.1565

Hu, J., Herbohn, J., Chazdon, R.L., Baynes, J., Wills, J., Meadows, J., Sohel, Md.S.I., 2018. Recovery of species composition over 46 years in a logged Australian tropical forest following different intensity silvicultural treatments. For. Ecol. Manag. 409, 660–666. https://doi.org/10.1016/j.foreco.2017.11.061

Hueni, A., Tuohy, M., 2006. Spectroradiometer data structuring, pre-processing and analysis – an IT based approach. J. Spat. Sci. 51, 93–102. https://doi.org/10.1080/14498596.2006.9635084

Jackson, S.M., Fredericksen, T.S., Malcolm, J.R., 2002. Area disturbed and residual stand damage following logging in a Bolivian tropical forest. For. Ecol. Manag. 166, 271–283. https://doi.org/10.1016/S0378-1127(01)00681-8

Jia, W., Pang, Y., Tortini, R., Schläpfer, D., Li, Z., Roujean, J.-L., 2020. A Kernel-Driven BRDF Approach to Correct Airborne Hyperspectral Imagery over Forested Areas with Rugged Topography. Remote Sens. 12, 432. https://doi.org/10.3390/rs12030432

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. Agric. For. Meteorol. 121, 19–35. https://doi.org/10.1016/j.agrformet.2003.08.027

Kalacska, M., Bohlman, S., Sanchez-Azofeifa, G.A., Castro-Esau, K., Caelli, T., 2007. Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels. Remote Sens. Environ. 109, 406–415. https://doi.org/10.1016/j.rse.2007.01.012

Kamal, M., Phinn, S., 2011. Hyperspectral Data for Mangrove Species Mapping: A Comparison of Pixel-Based and Object-Based Approach. Remote Sens. 3, 2222–2242. https://doi.org/10.3390/rs3102222

King, R.L., Ruffin, C., LaMastus, F.E., Shaw, D.R., 1999. The analysis of hyperspectral data using Savitzky-Golay filtering-practical issues. 2, in: IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293). Presented at the IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99, IEEE, Hamburg, Germany, pp. 398–400. https://doi.org/10.1109/IGARSS.1999.773512

Kokaly, R.F., Clark, R.N., 1999. Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. Remote Sens. Environ. 67, 267–287. https://doi.org/10.1016/S0034-4257(98)00084-4

Koponen, P., Nygren, P., Sabatier, D., Rousteau, A., Saur, E., 2004. Tree species diversity and forest structure in relation to microtopography in a tropical freshwater swamp forest in French Guiana. Plant Ecol. Former. Veg. 173, 17–32. https://doi.org/10.1023/B:VEGE.0000026328.98628.b8

Korpela, I., Heikkinen, V., Honkavaara, E., Rohrbach, F., Tokola, T., 2011. Variation and directional anisotropy of reflectance at the crown scale — Implications for tree species classification in digital aerial images. Remote Sens. Environ. 115, 2062–2074. https://doi.org/10.1016/j.rse.2011.04.008

Kremen, C., Razafimahatratra, V., Guillery, R.P., Rakotomalala, J., Weiss, A., Ratsisompatrarivo, J.-S., 1999. Designing the Masoala National Park in Madagascar Based on Biological and Socioeconomic Data. Conserv. Biol. 13, 1055–1068.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386

Landgrebe, D.A., 2003. Signal Theory Methods in Multispectral Remote Sensing: Landgrebe/Multispectral Remote w/CD. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/0471723800

Laurance, W.F., Pérez-Salicrup, D., Delamônica, P., Fearnside, P.M., D'Angelo, S., Jerozolinski, A., Pohl, L., Lovejoy, T.E., 2001. RAIN FOREST FRAGMENTATION AND THE STRUCTURE OF AMAZONIAN LIANA COMMUNITIES. Ecology 82, 105–116. https://doi.org/10.1890/0012-9658(2001)082[0105:RFFATS]2.0.CO;2

Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J., Vincent, G., 2019. Across Date Species Detection Using Airborne Imaging Spectroscopy. Remote Sens. 11, 789. https://doi.org/10.3390/rs11070789

Lenhard, K., Baumgartner, A., Schwarzmaier, T., 2015. Independent Laboratory Characterization of NEO HySpex Imaging Spectrometers VNIR-1600 and SWIR-320m-e. IEEE Trans. Geosci. Remote Sens. 53, 1828–1841. https://doi.org/10.1109/TGRS.2014.2349737

Lewis, S.L., Edwards, D.P., Galbraith, D., 2015. Increasing human dominance of tropical forests. Science 349, 827–832. https://doi.org/10.1126/science.aaa9932

Likens, G.E., Bormann, F.H., Pierce, R.S., Reiners, W.A., 1978. Recovery of a Deforested Ecosystem. Science 199, 492–496. https://doi.org/10.1126/science.199.4328.492

Liu, C., Straif, C., Flügel-Paul, T., Zeitner, U.D., Gross, H., 2017. Comparison of hyperspectral imaging spectrometer designs and the improvement of system performance with freeform surfaces. Appl. Opt. 56, 6894–6901. https://doi.org/10.1364/AO.56.006894

Lopes, A.P., Nelson, B.W., Wu, J., Graça, P.M.L. de A., Tavares, J.V., Prohaska, N., Martins, G.A., Saleska, S.R., 2016. Leaf flush drives dry season green-up of the Central Amazon. Remote Sens. Environ. 182, 90–98. https://doi.org/10.1016/j.rse.2016.05.009

Lugo, A.E., Helmer, E., 2004. Emerging forests on abandoned land: Puerto Rico's new forests. For. Ecol. Manag. 190, 145–161. https://doi.org/10.1016/j.foreco.2003.09.012

Ma, X., Huete, A., Tran, N.N., Bi, J., Gao, S., Zeng, Y., 2020. Sun-Angle Effects on Remote-Sensing Phenology Observed and Modelled Using Himawari-8. Remote Sens. 12, 1339. https://doi.org/10.3390/rs12081339

Makarau, A., Richter, R., Schläpfer, D., Reinartz, P., 2017. APDA Water Vapor Retrieval Validation for Sentinel-2 Imagery. IEEE Geosci. Remote Sens. Lett. 14, 227–231. https://doi.org/10.1109/LGRS.2016.2635942

Marie, C.-V., Rallu, J.-L., n.d. Les tendances démographiques et migratoires dans les régions ultrapériphériques : quel impact sur leur cohésion économique, sociale et territoriale ? (Commandé par la Commission Européenne, Direction Générale des Politiques Régionale). Institue national d'études démographique.

Martínez-Izquierdo, L., García, M.M., Powers, J.S., Schnitzer, S.A., 2016. Lianas suppress seedling growth and survival of 14 tree species in a Panamanian tropical forest. Ecology 97, 215–224. https://doi.org/10.1890/14-2261.1

Marvin, D.C., Asner, G.P., Schnitzer, S.A., 2016. Liana canopy cover mapped throughout a tropical forest with high-fidelity imaging spectroscopy. Remote Sens. Environ. 176, 98–106. https://doi.org/10.1016/j.rse.2015.12.028

Matangaran, J.R., Putra, E.I., Diatin, I., Mujahid, M., Adlan, Q., 2019. Residual stand damage from selective logging of tropical forests: A comparative case study in central Kalimantan and West Sumatra, Indonesia. Glob. Ecol. Conserv. 19, e00688. https://doi.org/10.1016/j.gecco.2019.e00688

Matasci, G., Tuia, D., Kanevski, M., 2012. SVM-Based Boosting of Active Learning Strategies for Efficient Domain Adaptation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 1335–1343. https://doi.org/10.1109/JSTARS.2012.2202881

McMichael, A.J., Woodruff, R.E., Hales, S., 2006. Climate change and human health: present and future risks. Lancet Lond. Engl. 367, 859–869. https://doi.org/10.1016/S0140-6736(06)68079-3

Meerdink, S.K., Roberts, D.A., Roth, K.L., King, J.Y., Gader, P.D., Koltunov, A., 2019. Classifying California plant species temporally using airborne hyperspectral imagery. Remote Sens. Environ. 232, 111308. https://doi.org/10.1016/j.rse.2019.111308

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790. https://doi.org/10.1109/TGRS.2004.831865

Meyer, P., Itten, K.I., Kellenberger, T., Sandmeier, S., Sandmeier, R., 1993. Radiometric corrections of topographically induced effects on Landsat TM data in an alpine environment. ISPRS J. Photogramm. Remote Sens. 48, 17–28. https://doi.org/10.1016/0924-2716(93)90028-L

Miles, L., Grainger, A., Phillips, O., 2004. The impact of global climate change on tropical forest biodiversity in Amazonia. Glob. Ecol. Biogeogr. 13, 553–565. https://doi.org/10.1111/j.1466-822X.2004.00105.x

Mitchard, E.T.A., Feldpausch, T.R., Brienen, R.J.W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T.R., Lewis, S.L., Lloyd, J., Quesada, C.A., Gloor, M., ter Steege, H., Meir, P., Alvarez, E., Araujo-Murakami, A., Aragão, L.E.O.C., Arroyo, L., Aymard, G., Banki, O., Bonal, D., Brown, S., Brown, F.I., Cerón, C.E., Chama Moscoso, V., Chave, J., Comiskey, J.A., Cornejo, F., Corrales Medina, M., Da Costa, L., Costa, F.R.C., Di Fiore, A., Domingues, T.F., Erwin, T.L., Frederickson, T., Higuchi, N., Honorio Coronado, E.N., Killeen, T.J., Laurance, W.F., Levis, C., Magnusson, W.E., Marimon, B.S., Marimon Junior, B.H., Mendoza Polo, I., Mishra, P., Nascimento, M.T., Neill, D., Núñez Vargas, M.P., Palacios, W.A., Parada, A., Pardo Molina, G., Peña-Claros, M., Pitman, N., Peres, C.A., Poorter, L., Prieto, A., Ramirez-Angulo, H., Restrepo Correa, Z., Roopsind, A., Roucoux, K.H., Rudas, A., Salomão, R.P., Schietti, J., Silveira, M., de Souza, P.F., Steininger, M.K., Stropp, J., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., van Andel, T.R., van der Heijden, G.M.F., Vieira, I.C.G., Vieira, S., Vilanova-Torre, E., Vos, V.A., Wang, O., Zartman, C.E., Malhi, Y., Phillips, O.L., 2014. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites: Divergent forest carbon maps from plots & space. Glob. Ecol. Biogeogr. 23, 935–946. https://doi.org/10.1111/geb.12168

Molino, J.-F., Sabatier, D., Engel, J., Frame, D., Lucas, E.J., Delprete, P.G., Grenand, P., Odonne, G., Davy, D., Fleury, M., Martin, C., In prep. An annotated checklist of the tree species of French Guiana.

Muraoka, H., Takenaka, A., Tang, Y., Koizumi, H., Washitani, I., 1998. Flexible Leaf Orientations of Arisaema heterophyllum Maximize Light Capture in a Forest Understorey and Avoid Excess Irradiance at a Deforested Site. Ann. Bot. 82, 297–307. https://doi.org/10.1006/anbo.1998.0682

Oliveira, L.M. de, Galvão, L.S., Ponzoni, F.J., 2019. Topographic effects on the determination of hyperspectral vegetation indices: a case study in southeastern Brazil. Geocarto Int. 0, 1–18. https://doi.org/10.1080/10106049.2019.1690055

Olson, J.S., Watts, J.A., Allison, L.J., 1983. Carbon in Live Vegetation of Major World Ecosystems. Oak Ridge National Laboratory.

Pacifici, F., Longbotham, N., Emery, W.J., 2014. The Importance of Physical Quantities for the Analysis of Multitemporal and Multiangular Optical Very High Spatial Resolution Images. IEEE Trans. Geosci. Remote Sens. 52, 6241–6256. https://doi.org/10.1109/TGRS.2013.2295819

Pal, M., Mather, P.M., 2006. Some issues in the classification of DAIS hyperspectral data. Int. J. Remote Sens. 27, 2895–2916. https://doi.org/10.1080/01431160500185227

Pan, B., Shi, Z., Zhang, N., Xie, S., 2016. Hyperspectral Image Classification Based on Nonlinear Spectral–Spatial Network. IEEE Geosci. Remote Sens. Lett. 13, 1782–1786. https://doi.org/10.1109/LGRS.2016.2608963

Pan, S.J., Kwok, J.T., Yang, Q., 2008. Transfer Learning via Dimensionality Reduction 6.

Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q., 2011. Domain Adaptation via Transfer Component Analysis. IEEE Trans. Neural Netw. 22, 199–210. https://doi.org/10.1109/TNN.2010.2091281

Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Patz, J.A., Campbell-Lendrum, D., Holloway, T., Foley, J.A., 2005. Impact of regional climate change on human health. Nature 438, 310–317. https://doi.org/10.1038/nature04188

Patz, J.A., Norris, D., 2004. Land use change and human health. Ecosyst. Land Use Change 2004 159–167. https://doi.org/10.1029/153GM13

Pereira, R., Zweede, J., Asner, G.P., Keller, M., 2002. Forest canopy damage and recovery in reduced-impact and conventional selective logging in eastern Para, Brazil. For. Ecol. Manag. 168, 77–89. https://doi.org/10.1016/S0378-1127(01)00732-0

Persello, C., Bruzzone, L., 2016. Kernel-Based Domain-Invariant Feature Selection in Hyperspectral Images for Transfer Learning. IEEE Trans. Geosci. Remote Sens. 54, 2615–2626. https://doi.org/10.1109/TGRS.2015.2503885

Pflug, B., Main-Knorn, M., Makarau, A., Richter, R., 2015. Validation of aerosol estimation in atmospheric correction algorithm ATCOR. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XL-7/W3, 677–683. https://doi.org/10.5194/isprsarchives-XL-7-W3-677-2015

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., Liu, R., Mao, J., Myneni, R.B., Peng, S., Poulter, B., Shi, X., Xiao, Z., Zeng, N., Zeng, Z., Wang, Y., 2015. Detection and attribution of vegetation greening trend in China over the last 30 years. Glob. Change Biol. 21, 1601–1609. https://doi.org/10.1111/gcb.12795

Pickering, J., Stehman, S.V., Tyukavina, A., Potapov, P., Watt, P., Jantz, S.M., Bholanath, P., Hansen, M.C., 2019. Quantifying the trade-off between cost and precision in estimating area of forest loss and degradation using probability sampling in Guyana. Remote Sens. Environ. 221, 122–135. https://doi.org/10.1016/j.rse.2018.11.018

Polisar, J., de Thoisy, B., Rumiz, D.I., Santos, F.D., McNab, R.B., Garcia-Anleu, R., Ponce-Santizo, G., Arispe, R., Venegas, C., 2017. Using certified timber extraction to benefit jaguar and ecosystem conservation. Ambio 46, 588–603. https://doi.org/10.1007/s13280-016-0853-y

Polo, F.M., Vicente, R., 2020. Covariate Shift Adaptation in High-Dimensional and Divergent Distributions. ArXiv201001184 Cs Stat.

Powers, J.S., Marín-Spiotta, E., 2017. Ecosystem Processes and Biogeochemical Cycles in Secondary Tropical Forest Succession. Annu. Rev. Ecol. Evol. Syst. 48, 497–519. https://doi.org/10.1146/annurev-ecolsys-110316-022944

Prats-Montalbán, J.M., de Juan, A., Ferrer, A., 2011. Multivariate image analysis: A review with applications. Chemom. Intell. Lab. Syst. 107, 1–23. https://doi.org/10.1016/j.chemolab.2011.03.002

Price, J.C., 1994. How unique are spectral signatures? Remote Sens. Environ. 49, 181–186.

Putz, F.E., 1985. Woody Vines And Forest Management In Malaysia. Commonw. For. Rev. 64, 359–365.

Putz, F. E., Sist, P., Fredericksen, T., Dykstra, D., 2008. Reduced-impact logging: Challenges and opportunities. For. Ecol. Manag., Moving beyond reduced impact-logging towards a more holistic management of tropical forests 256, 1427–1433. https://doi.org/10.1016/j.foreco.2008.03.036

Putz, Francis E, Zuidema, P.A., Pinard, M.A., Boot, R.G.A., Sayer, J.A., Sheil, D., Sist, P., Elias, Vanclay, J.K., 2008. Improved Tropical Forest Management for Carbon Retention. PLoS Biol. 6, e166. https://doi.org/10.1371/journal.pbio.0060166

Qi, J., Xie, D., Guo, D., Yan, G., 2017. A Large-Scale Emulation System for Realistic Three-Dimensional (3-D) Forest Simulation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 4834–4843. https://doi.org/10.1109/JSTARS.2017.2714423

Qian, Y., Yao, F., Jia, S., 2009. Band selection for hyperspectral imagery using affinity propagation. IET Comput. Vis. 3, 213–222. https://doi.org/10.1049/iet-cvi.2009.0034

Rahman, H., Pinty, B., Verstraete, M.M., 1993. Coupled surface-atmosphere reflectance (CSAR) model: 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. J. Geophys. Res. 98, 20791. https://doi.org/10.1029/93JD02072

Rajan, S., Ghosh, J., Crawford, M.M., 2006. Exploiting Class Hierarchies for Knowledge Transfer in Hyperspectral Data. IEEE Trans. Geosci. Remote Sens. 44, 3408–3417. https://doi.org/10.1109/TGRS.2006.878442

Rametsteiner, E., Simula, M., 2003. Forest certification—an instrument to promote sustainable forest management? J. Environ. Manage. 67, 87–98. https://doi.org/10.1016/S0301-4797(02)00191-3

Reddington, C.L., Spracklen, D.V., Artaxo, P., Ridley, D.A., Rizzo, L.V., Arana, A., 2016. Analysis of particulate emissions from tropical biomass burning using a global aerosol model and long-term surface observations. Atmospheric Chem. Phys. 16, 11083–11106.

Réjou-Méchain, M., Tymen, B., Blanc, L., Fauset, S., Feldpausch, T.R., Monteagudo, A., Phillips, O.L., Richard, H., Chave, J., 2015. Using repeated small-footprint LiDAR acquisitions to infer spatial and temporal variations of a high-biomass Neotropical forest. Remote Sens. Environ. 169, 93–101. https://doi.org/10.1016/j.rse.2015.08.001

Richter, R., Kellenberger, T., Kaufmann, H., 2009. Comparison of Topographic Correction Methods. Remote Sens. 1, 184–196. https://doi.org/10.3390/rs1030184

Richter, R, Schlapfer, D., 2018. Atmospheric / Topographic Correction for Airborne Imagery (ATCOR-4 user Guide, Version 7.2.0) 279.

Richter, Rolf, Schlapfer, D., 2018. PARametric GEocoding : Orthorectification for airborne scanner data. User Manual, Version 3.4.

Richter, R., Schläpfer, D., Müller, A., 2006. An automatic atmospheric correction algorithm for visible/NIR imagery. Int. J. Remote Sens. 27, 2077–2085. https://doi.org/10.1080/01431160500486690

Rocha de Souza Pereira, F., Kampel, M., Gomes Soares, M., Estrada, G., Bentz, C., Vincent, G., Rocha de Souza Pereira, F., Kampel, M., Gomes Soares, M.L., Estrada, G.C.D., Bentz, C., Vincent, G., 2018. Reducing Uncertainty in Mapping of Mangrove Aboveground Biomass Using Airborne Discrete Return Lidar Data. Remote Sens. 10, 637. https://doi.org/10.3390/rs10040637

Roger, R.E., Arnold, J.F., 1996. Reliably estimating the noise in AVIRIS hyperspectral images. Int. J. Remote Sens. 17, 1951–1962. https://doi.org/10.1080/01431169608948750

Román, M.O., Gatebe, C.K., Schaaf, C.B., Poudyal, R., Wang, Z., King, M.D., 2011. Variability in surface BRDF at different spatial scales (30m–500m) over a mixed agricultural landscape as retrieved from airborne and satellite spectral measurements. Remote Sens. Environ. 115, 2184–2203. https://doi.org/10.1016/j.rse.2011.04.012

Rowe, N., Isnard, S., Speck, T., 2004. Diversity of Mechanical Architectures in Climbing Plants: An Evolutionary Perspective. J. Plant Growth Regul. 23, 108–128. https://doi.org/10.1007/s00344-004-0044-0

Rowe, N., Speck, T., 2005. Plant growth forms: an ecological and evolutionary perspective. New Phytol. 166, 61–72. https://doi.org/10.1111/j.1469-8137.2004.01309.x

Rowe, N.P., Speck, T., 1996. Biomechanical Characteristics of the Ontogeny and Growth Habit of the Tropical Liana Condylocarpon guianense (Apocynaceae). Int. J. Plant Sci. 157, 406–417. https://doi.org/10.1086/297357

Saatchi, S., Mascaro, J., Xu, L., Keller, M., Yang, Y., Duffy, P., Espírito-Santo, F., Baccini, A., Chambers, J., Schimel, D., 2015. Seeing the forest beyond the trees: Correspondence. Glob. Ecol. Biogeogr. 24, 606–610. https://doi.org/10.1111/geb.12256

Sabatier, D., PRÉVOST, M., 1990. Variations du peuplement forestier à l'échelle stationnelle. Cas Stn. Nouragues En Guyane Fr. Actes Atelier MAB-IUFRO Ecosystèmes For.

Sánchez-Azofeifa, G.A., Castro, K., Wright, S.J., Gamon, J., Kalacska, M., Rivard, B., Schnitzer, S.A., Feng, J.L., 2009. Differences in leaf traits, leaf internal structure, and spectral reflectance between two communities of lianas and trees: Implications for remote sensing in tropical environments. Remote Sens. Environ. 113, 2076–2088. https://doi.org/10.1016/j.rse.2009.05.013

Santini, F., Palombo, A., 2019. Physically Based Approach for Combined Atmospheric and Topographic Corrections. Remote Sens. 11, 1218. https://doi.org/10.3390/rs11101218

Schaepman-Strub, G., Schaepman, M.E., Painter, T.H., Dangel, S., Martonchik, J.V., 2006. Reflectance quantities in optical remote sensing—definitions and case studies. Remote Sens. Environ. 103, 27–42. https://doi.org/10.1016/j.rse.2006.03.002

Schläpfer, D., Borel, C.C., Keller, J., Itten, K.I., 1998. Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. Remote Sens. Environ. 65, 353–366. https://doi.org/10.1016/S0034-4257(98)00044-3

Schläpfer, D., Hueni, A., Richter, R., 2018. Cast Shadow Detection to Quantify the Aerosol Optical Thickness for Atmospheric Correction of High Spatial Resolution Optical Imagery. Remote Sens. 10, 200. https://doi.org/10.3390/rs10020200

Schlapfer, D., Richter, R., Feingersh, T., 2015. Operational BRDF Effects Correction for Wide-Field-of-View Optical Scanners (BREFCOR). IEEE Trans. Geosci. Remote Sens. 53, 1855–1864. https://doi.org/10.1109/TGRS.2014.2349946

Schnitzer, S.A., 2014. Lianas in gaps reduce carbon accumulation in a tropical forest 95, 10.

Schnitzer, S.A., 2005. A Mechanistic Explanation for Global Patterns of Liana Abundance and Distribution. Am. Nat. 166, 262–276.

Schnitzer, S.A., Bongers, F., 2011. Increasing liana abundance and biomass in tropical forests: emerging patterns and putative mechanisms: Increasing lianas in tropical forests. Ecol. Lett. 14, 397–406. https://doi.org/10.1111/j.1461-0248.2011.01590.x

Schweiger, A.K., Cavender-Bares, J., Townsend, P.A., Hobbie, S.E., Madritch, M.D., Wang, R., Tilman, D., Gamon, J.A., 2018. Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. Nat. Ecol. Evol. 2, 976–982. https://doi.org/10.1038/s41559-018-0551-1

Shen, X., Cao, L., 2017. Tree-Species Classification in Subtropical Forests Using Airborne Hyperspectral and LiDAR Data. Remote Sens. 9, 1180. https://doi.org/10.3390/rs9111180

Silva, L.N., Freer-Smith, P., Madsen, P., 2019. Production, restoration, mitigation: a new generation of plantations. New For. 50, 153–168. https://doi.org/10.1007/s11056-018-9644-6

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81, 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

Sist, P., Ferreira, F.N., 2007. Sustainability of reduced-impact logging in the Eastern Amazon. For. Ecol. Manag. 243, 199–209. https://doi.org/10.1016/j.foreco.2007.02.014

Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., Berg, E. van den, Bernacci, L., Bispo, P. da C., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Hockemba, M.B.-N., Bunyavejchewin, S., Matos, D.C.L., Castillo-Santiago, M., Catharino, E.L.M., Chai, S.-L., Chen, Y., Colwell, R.K., Chazdon, R.L., Clark, C., Clark, D.B., Clark, D.A., Culmsee, H., Damas, K., Dattaraja, H.S., Dauby, G., Davidar, P., DeWalt, S.J., Doucet, J.-L., Duque, A., Durigan, G., Eichhorn, K.A.O., Eisenlohr, P.V., Eler, E., Ewango, C., Farwig, N., Feeley, K.J., Ferreira, L., Field, R., Filho, A.T. de O., Fletcher, C., Forshed, O., Franco, G., Fredriksson, G., Gillespie, T., Gillet, J.-F., Amarnath, G., Griffith, D.M., Grogan, J., Gunatilleke, N., Harris, D., Harrison, R., Hector, A., Homeier, J., Imai, N., Itoh, A., Jansen, P.A., Joly, C.A., Jong, B.H.J. de, Kartawinata, K., Kearsley, E., Kelly, D.L., Kenfack, D., Kessler, M., Kitayama, K., Kooyman, R., Larney, E., Laumonier, Y., Laurance, S., Laurance, W.F., Lawes, M.J., Amaral, I.L. do, Letcher, S.G., Lindsell, J., Lu, X., Mansor, A., Marjokorpi, A., Martin, E.H., Meilby, H., Melo, F.P.L., Metcalfe, D.J., Medjibe, V.P., Metzger, J.P., Millet, J., Mohandass, D., Montero, J.C., Valeriano, M. de M., Mugerwa, B., Nagamasu, H., Nilus, R., Ochoa-Gaona, S., Onrizal, Page, N., Parolin, P., Parren, M.,

Parthasarathy, N., Paudel, E., Permana, A., Piedade, M.T.F., Pitman, N.C.A., Poorter, L., Poulsen, A.D., Poulsen, J., Powers, J., Prasad, R.C., Puyravaud, J.-P., Razafimahaimodison, J.-C., Reitsma, J., Santos, J.R. dos, Spironello, W.R., Romero-Saltos, H., Rovero, F., Rozak, A.H., Ruokolainen, K., Rutishauser, E., Saiter, F., Saner, P., Santos, B.A., Santos, F., Sarker, S.K., Satdichanh, M., Schmitt, C.B., Schöngart, J., Schulze, M., Suganuma, M.S., Sheil, D., Pinheiro, E. da S., Sist, P., Stevart, T., Sukumar, R., Sun, I.-F., Sunderland, T., Suresh, H.S., Suzuki, E., Tabarelli, M., Tang, J., Targhetta, N., Theilade, I., Thomas, D.W., Tchouto, P., Hurtado, J., Valencia, R., Valkenburg, J.L.C.H. van, Do, T.V., Vasquez, R., Verbeeck, H., Adekunle, V., Vieira, S.A., Webb, C.O., Whitfeld, T., Wich, S.A., Williams, J., Wittmann, F., Wöll, H., Yang, X., Yao, C.Y.A., Yap, S.L., Yoneda, T., Zahawi, R.A., Zakaria, R., Zang, R., Assis, R.L. de, Luize, B.G., Venticinque, E.M., 2015. An estimate of the number of tropical tree species. Proc. Natl. Acad. Sci. 112, 7472–7477. https://doi.org/10.1073/pnas.1423147112

Soekhoe, D., van der Putten, P., Plaat, A., 2016. On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks, in: Boström, H., Knobbe, A., Soares, C., Papapetrou, P. (Eds.), Advances in Intelligent Data Analysis XV, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 50–60. https://doi.org/10.1007/978-3-319-46349-0_5

Solomon, A.M., Prentice, I.C., Leemans, R., Cramer, W.P., 1993. The Interaction of Climate and Land Use in Future Terrestrial Carbon Storage and Release, in: Wisniewski, J., Sampson, R.N. (Eds.), Terrestrial Biospheric Carbon Fluxes: Springer Netherlands, Dordrecht, pp. 595–614. https://doi.org/10.1007/978-94-011-1982-5_40

Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. Nature 560, 639–643. https://doi.org/10.1038/s41586-018-0411-9

Steege, H. ter, Boot, R.G.A., Brouwer, L.C., Caesar, J.C., Ek, R.C., Hammond, D.S., Haripersaud, P.P., Hout, P. van der, Jetten, V.G., Kekem, A.J. van, Kellman, M.A., Khan, Z., Polak, A.M., Pons, T.L., Pulles, J., Raaimakers, D., Rose, S.A., Sanden, J.J. van der, Zagt, R.J., 1996. Ecology and logging in a tropical rain forest in Guyana: with recommendations for forest management. Ecol. Logging Trop. Rain For. Guyana Recomm. For. Manag.

Sterck, F.J., Bongers, F., 2001. Crown development in tropical rain forest trees: patterns with tree height and light availability. J. Ecol. 89, 1–13. https://doi.org/10.1046/j.1365-2745.2001.00525.x

St-Onge, B., Audet, F.-A., Bégin, J., 2015. Characterizing the Height Structure and Composition of a Boreal Forest Using an Individual Tree Crown Approach Applied to Photogrammetric Point Clouds. Forests 6, 3899–3922. https://doi.org/10.3390/f6113899

Teillet, P.M., Guindon, B., Goodenough, D.G., 1982. On the Slope-Aspect Correction of Multispectral Scanner Data. Can. J. Remote Sens. 8, 84–106. https://doi.org/10.1080/07038992.1982.10855028

ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomao, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.-F., Monteagudo, A., Nunez Vargas, P., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J., Wittmann, F., Andrade, A., Laurance, W.F., Laurance, S.G.W., Marimon, B.S., Marimon, B.-H., Guimaraes Vieira, I.C., Amaral, I.L., Brienen, R., Castellanos, H., Cardenas Lopez, D., Duivenvoorden, J.F., Mogollon, H.F., Matos, F.D. d. A., Davila, N., Garcia-Villacorta, R., Stevenson Diaz, P.R., Costa, F., Emilio, T., Levis, C., Schietti, J., Souza, P., Alonso, A., Dallmeier, F., Montoya, A.J.D., Fernandez Piedade, M.T., Araujo-Murakami, A., Arroyo, L., Gribel, R., Fine, P.V.A., Peres, C.A., Toledo, M., Aymard C., G.A., Baker, T.R., Ceron, C., Engel, J., Henkel, T.W., Maas, P., Petronelli, P., Stropp, J., Zartman, C.E., Daly, D., Neill, D., Silveira, M., Paredes, M.R., Chave, J., Lima Filho, D. d. A., Jorgensen, P.M., Fuentes, A., Schongart, J., Cornejo Valverde, F., Di Fiore, A., Jimenez, E.M., Penuela Mora, M.C., Phillips, J.F., Rivas, G., van Andel, T.R., von Hildebrand, P., Hoffman, B., Zent, E.L., Malhi, Y., Prieto, A., Rudas, A., Ruschell, A.R., Silva, N., Vos, V., Zent, S., Oliveira, A.A., Schutz, A.C., Gonzales, T., Trindade Nascimento, M., Ramirez-Angulo, H., Sierra, R., Tirado, M., Umana Medina, M.N., van der Heijden, G., Vela, C.I.A., Vilanova Torre, E., Vriesendorp, C., Wang, O., Young, K.R., Baider, C., Balslev, H., Ferreira, C., Mesones, I., Torres-Lezama, A., Urrego Giraldo, L.E., Zagt, R., Alexiades, M.N., Hernandez, L.,

Huamantupa-Chuquimaco, I., Milliken, W., Palacios Cuenca, W., Pauletto, D., Valderrama Sandoval, E., Valenzuela Gamarra, L., Dexter, K.G., Feeley, K., Lopez-Gonzalez, G., Silman, M.R., 2013. Hyperdominance in the Amazonian Tree Flora. Science 342, 1243092–1243092. https://doi.org/10.1126/science.1243092

Thompson, D.R., Guanter, L., Berk, A., Gao, B.-C., Richter, R., Schläpfer, D., Thome, K.J., 2019. Retrieval of Atmospheric Parameters and Surface Reflectance from Visible and Shortwave Infrared Imaging Spectroscopy Data. Surv. Geophys. 40, 333–360. https://doi.org/10.1007/s10712-018-9488-9

Thompson, D.R., Guanter, L., Berk, A., Gao, B.-C., Richter, R., Schläpfer, D., Thome, K.J., 2018. Retrieval of Atmospheric Parameters and Surface Reflectance from Visible and Shortwave Infrared Imaging Spectroscopy Data. Surv. Geophys. https://doi.org/10.1007/s10712-018-9488-9

Thompson, D.R., Roberts, D.A., Gao, B.C., Green, R.O., Guild, L., Hayashi, K., Kudela, R., Palacios, S., 2016. Atmospheric correction with the Bayesian empirical line. Opt. Express 24, 2134. https://doi.org/10.1364/OE.24.002134

Traissac, S., Pascal, J.-P., 2014. Birth and life of tree aggregates in tropical forest: hypotheses on population dynamics of an aggregated shade-tolerant species. J. Veg. Sci. 25, 491–502. https://doi.org/10.1111/jvs.12080

Trolliet, F., Forget, P.-M., Doucet, J.-L., Gillet, J.-F., Hambuckers, A., 2017. Frugivorous birds influence the spatial organization of tropical forests through the generation of seedling recruitment foci under zoochoric trees. Acta Oecologica 85, 69–76. https://doi.org/10.1016/j.actao.2017.09.010

Turner, I.M., 1996. Species Loss in Fragments of Tropical Rain Forest: A Review of the Evidence. J. Appl. Ecol. 33, 200–209. https://doi.org/10.2307/2404743

Tusa, E., Laybros, A., Monnet, J.-M., Dalla Mura, M., Barré, J.-B., Vincent, G., Dalponte, M., Féret, J.-B., Chanussot, J., 2020. Fusion of hyperspectral imaging and LiDAR for forest monitoring, in: Data Handling in Science and Technology. Elsevier, pp. 281–303. https://doi.org/10.1016/B978-0-444-63977-6.00013-4

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol. Evol. 10, 225–232. https://doi.org/10.1111/2041-210X.13107

Valbuena, R., 2014. Integrating Airborne Laser Scanning with Data from Global Navigation Satellite Systems and Optical Sensors, in: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), Forestry Applications of Airborne Laser Scanning. Springer Netherlands, Dordrecht, pp. 63–88. https://doi.org/10.1007/978-94-017-8663-8_4

Valbuena, R., Mauro, F., Arjonilla, F.J., Manzanera, J.A., 2011. Comparing airborne laser scanning-imagery fusion methods based on geometric accuracy in forested areas. Remote Sens. Environ. 115, 1942–1954. https://doi.org/10.1016/j.rse.2011.03.017

Van der Heijden, G.M.F., Phillips, O.L., 2009. Liana infestation impacts tree growth in a lowland tropical moist forest. Biogeosciences 6, 2217–2226.

van der Heijden, G.M.F., Powers, J.S., Schnitzer, S.A., 2015. Lianas reduce carbon accumulation and storage in tropical forests. Proc. Natl. Acad. Sci. 112, 13267–13271. https://doi.org/10.1073/pnas.1504869112

Vantomme, P., 2011. The Silviculture of Tropical Nonwood Forest Products, Between Farming and Forestry, in: Günter, S., Weber, M., Stimm, B., Mosandl, R. (Eds.), Silviculture in the Tropics, Tropical Forestry. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 119–127. https://doi.org/10.1007/978-3-642-19986-8_10

Velasco-Forero, S., Manian, V., 2009. Improving Hyperspectral Image Classification Using Spatial Preprocessing. IEEE Geosci. Remote Sens. Lett. 6, 297–301. https://doi.org/10.1109/LGRS.2009.2012443

Verrelst, J., Schaepman, M.E., Koetz, B., Kneubühler, M., 2008. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. Remote Sens. Environ., Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue 112, 2341–2353. https://doi.org/10.1016/j.rse.2007.11.001

Vidal, M., Amigo, J.M., 2012. Pre-processing of hyperspectral images. Essential steps before image analysis. Chemom. Intell. Lab. Syst. 117, 138–148. https://doi.org/10.1016/j.chemolab.2012.05.009

Vincent, G., Antin, C., Laurans, M., Heurtebize, J., Durrieu, S., Lavalley, C., Dauzat, J., 2017. Mapping plant area index of tropical evergreen forest by airborne laser scanning. A cross-validation study using LAI2200 optical sensor. Remote Sens. Environ. 198, 254–266. https://doi.org/10.1016/j.rse.2017.05.034

Vincent, G., Sabatier, D., Blanc, L., Chave, J., Weissenbacher, E., Pélissier, R., Fonty, E., Molino, J.F., Couteron, P., 2012. Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. Remote Sens. Environ. 125, 23–33. https://doi.org/10.1016/j.rse.2012.06.019

Waite, C.E., van der Heijden, G.M.F., Field, R., Boyd, D.S., 2019. A view from above: Unmanned aerial vehicles (UAVs) provide a new tool for assessing liana infestation in tropical forest canopies. J. Appl. Ecol. https://doi.org/10.1111/1365-2664.13318

Wang, J., Chagnon, F.J.F., Williams, E.R., Betts, A.K., Renno, N.O., Machado, L.A.T., Bisht, G., Knox, R., Bras, R.L., 2009. Impact of deforestation in the Amazon basin on cloud climatology. Proc. Natl. Acad. Sci. 106, 3670–3674. https://doi.org/10.1073/pnas.0810156106

Wang, T., Yan, G., Ren, H., Mu, X., 2010. Improved Methods for Spectral Calibration of On-Orbit Imaging Spectrometers. IEEE Trans. Geosci. Remote Sens. 48, 3924–3931. https://doi.org/10.1109/TGRS.2010.2067220

Watson, J.E.M., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., Thompson, I., Ray, J.C., Murray, K., Salazar, A., McAlpine, C., Potapov, P., Walston, J., Robinson, J.G., Painter, M., Wilkie, D., Filardi, C., Laurance, W.F., Houghton, R.A., Maxwell, S., Grantham, H., Samper, C., Wang, S., Laestadius, L., Runting, R.K., Silva-Chávez, G.A., Ervin, J., Lindenmayer, D., 2018. The exceptional value of intact forest ecosystems. Nat. Ecol. Evol. 2, 599–610. https://doi.org/10.1038/s41559-018-0490-x

Weinstein, Ben.G., Graves, S.J., Marconi, S., Singh, A., Zare, A., Stewart, D., Bohlman, S.A., White, E.P., 2020. A benchmark dataset for individual tree crown delineation in co-registered airborne RGB, LiDAR and hyperspectral imagery from the National Ecological Observation Network (preprint). Ecology. https://doi.org/10.1101/2020.11.16.385088

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sens. 11, 1309. https://doi.org/10.3390/rs11111309

Wen, J., Liu, Qiang, Xiao, Q., Liu, Qinhuo, You, D., Hao, D., Wu, S., Lin, X., 2018. Characterizing Land Surface Anisotropic Reflectance over Rugged Terrain: A Review of Concepts and Recent Developments. Remote Sens. 10, 370. https://doi.org/10.3390/rs10030370

Whitehead, P.G., Robinson, M., 1993. Experimental basin studies—an international and historical perspective of forest impacts. J. Hydrol., The Balquhidder Catchement and Process Studies 145, 217–230. https://doi.org/10.1016/0022-1694(93)90055-E

Williams, J., Schonlieb, C.-B., Swinfield, T., Lee, J., Cai, X., Qie, L., Coomes, D.A., 2020. 3D Segmentation of Trees Through a Flexible Multiclass Graph Cut Algorithm. IEEE Trans. Geosci. Remote Sens. 58, 754–776. https://doi.org/10.1109/TGRS.2019.2940146

Windrim, L., Melkumyan, A., Murphy, R., Chlingaryan, A., Nieto, J., 2016. Unsupervised feature learning for illumination robustness, in: 2016 IEEE International Conference on Image Processing (ICIP). Presented at the 2016 IEEE International Conference on Image Processing (ICIP), pp. 4453–4457. https://doi.org/10.1109/ICIP.2016.7533202

Windrim, L., Ramakrishnan, R., Melkumyan, A., Murphy, R.J., 2018. A Physics-Based Deep Learning Approach to Shadow Invariant Representations of Hyperspectral Images. IEEE Trans. Image Process. 27, 665–677. https://doi.org/10.1109/TIP.2017.2761542

Wright, S.J., 2010. The future of tropical forests. Ann. N. Y. Acad. Sci. 1195, 1–27. https://doi.org/10.1111/j.1749-6632.2010.05455.x

Wu, S., Wen, J., You, D., Hao, D., Lin, X., Xiao, Q., Liu, Q., Gastellu-Etchegorry, J.-P., 2018. Characterization of Remote Sensing Albedo Over Sloped Surfaces Based on DART Simulations and In Situ Observations. J. Geophys. Res. Atmospheres 123, 8599–8622. https://doi.org/10.1029/2018JD028283

Xu, H., Li, Y., Liu, S., Zang, R., He, F., Spence, J.R., 2015. Partial recovery of a tropical rain forest a half-century after clear-cut and selective logging. J. Appl. Ecol. 52, 1044–1052. https://doi.org/10.1111/1365-2664.12448

Zhang, H., He, W., Zhang, L., Shen, H., Yuan, Q., 2014. Hyperspectral Image Restoration Using Low-Rank Matrix Recovery. IEEE Trans. Geosci. Remote Sens. 52, 4729–4743. https://doi.org/10.1109/TGRS.2013.2284280

Zhang, J., Rivard, B., Sánchez-Azofeifa, A., Castro-Esau, K., 2006. Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. Remote Sens. Environ. 105, 129–141. https://doi.org/10.1016/j.rse.2006.06.010