

Université de Toulouse
MASTER 2 GEOMATIQUE

« **S**cience de l'**I**nformation **G**éoréférencée pour la **M**aîtrise de
l'environnement et l'**A**ménagement des territoires » (**SIGMA**)
<http://sigma.univ-toulouse.fr>

RAPPORT DE STAGE

**Cartographie des végétations particulières du Parc
amazonien de Guyane**



KARASIAK Nicolas



Maître de stage : Pauline Perbet
Tuteur-enseignant : Mathieu Fauvel (ENSAT)

Septembre 2016

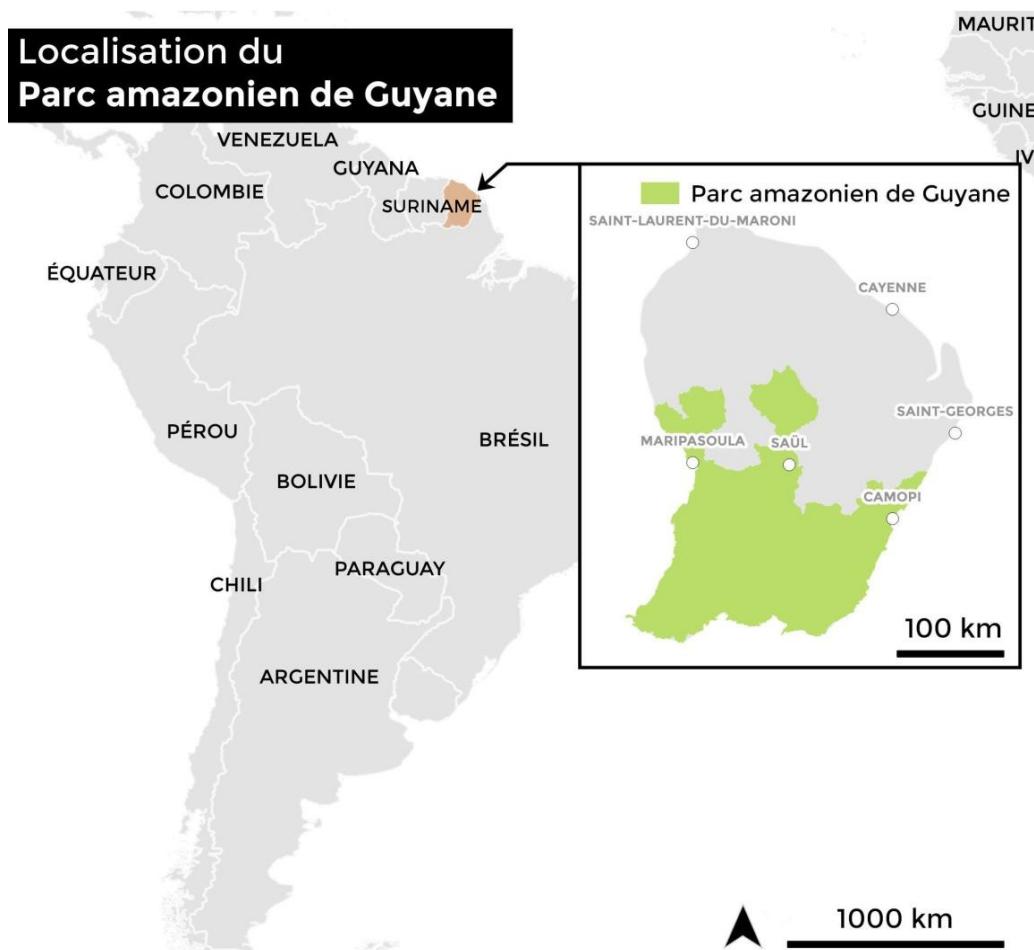
Résumé

Le Parc amazonien de Guyane est le plus grand parc national de France avec 3,4 millions d'hectares de forêt protégée. Plusieurs jours de transports sont nécessaires pour accéder aux lieux les plus au sud, ce qui fait de la télédétection un outil très utile. Ce stage a eu pour but de cartographier, avec des images SPOT-5 de 2007 à 2014, cinq types de végétation : palmier pinôt (*Euterpe oleracea*), palmier bâche (*Mauritia flexuosa*), parinari (*Parinari campestris*), végétation basse (*Guadua macrostachya*, *Lasiacis ligulata*, et *Heliconiaceae*), et savane-roche. Une analyse des données terrain a permis de réaliser un premier jeu de parcelles d'entraînement sur l'ensemble des 55 images nécessaires pour couvrir la totalité du territoire. Après une première classification, le faible nombre de données terrain n'a pas permis d'obtenir l'amplitude spectrale des essences pour chaque scène. Il a été donc nécessaire d'utiliser des végétations prédictes à la jonction entre deux scènes dans un second temps.

La cartographie finale a été réalisée en utilisant Random Forest (kappa : 97%), qui s'est montré plus performant que le Gaussian Mixture Model (kappa : 87%). Les indices de végétation de type NDVI ou EVI2 n'ont pas amélioré ces résultats.

Ce travail a permis notamment de cartographier une présence très importante de palmiers pinôts au sud-ouest du parc et de délimiter précisément les parinaris.

Les scripts (R, Python) et modèles Qgis utilisés pour ce travail sont disponibles à l'adresse github.com/lennepkade/ScriptsParcAmazonien/. Le plugin de classification *dzetsaka* créé pour ce stage est disponible dans le dépôt officiel de Qgis.



Abstract

With 37,840 sq mi, the French Guiana Amazonian Park is the largest French national park. Very difficult to get through, especially in the very south, it may took several days to accede to some point of interest by combining different way of travelling. Remote Sensing is here a great alternative for mapping tree species. SPOT-5 images from 2007 to 2014 were used to map five different species : *Euterpe oleracea*, *Mauritia flexuosa*, *Parinari campestris*, Savannah-rock (vegetation near to rock), and low vegetation which is a mix bamboo (*Guadua macrostachya*, *Lasiacis ligulata*) and heliconias (*Heliconiaceae*).

As there was no tool for classifying with Random Forest on Qgis, a plugin called *dzetsaka* was created and published on the official repository. Then, on each of the 55 images needed to cover the whole area, Regions Of Interest (ROI) have been drawn, using field data. Clouds have been eliminating by using Clouds Index (NDCI) for each image. After a first whole classification, lot of scenes didn't have enough ROI to train a model. To predict missing vegetation, vegetations predicted at the border of two scenes have been used as new ROI.

The final map has been generated using Random Forest (kappa : 97%) which was more efficient than the Gaussian Mixture Model (kappa : 87%). Thanks to this work, many *Euterpe oleracea* settlements in the southwest have been discovered and *Parinari Campestris* have been map precisely for the first time.

Qgis models and scripts (R, Python) used for this work are available on github.com/lennepkade/ScriptsParcAmazonien/.



Remerciements

Je tiens tout d'abord à remercier Pauline Perbet de m'avoir donné l'opportunité de travailler sur ce magnifique projet et de m'avoir fait confiance tout au long du stage. Un très grand merci aussi à Pierre Joubert, responsable SIG du Parc amazonien de Guyane, pour ses excellentes playlists tout au long du stage, et merci également à David Franco, le responsable informatique du parc, pour avoir été très réactif et pour m'avoir fourni un serveur virtuel afin d'effectuer tous les calculs nécessaires et un espace de stockage centralisé pour travailler sans peur de perdre mes données.

Je remercie aussi Daniel Sabatier (IRD) et Stéphane Guitet (ONF-IRD) pour leurs réponses précises et rapides sur les particularités des essences forestières tropicales.

Enfin je remercie mon tuteur enseignant, Mathieu Fauvel, pour son expertise python et son implémentation de l'algorithme Random Forest. Merci aussi à tous les enseignants du Master 2 SIGMA pour m'avoir apporté beaucoup durant cette année d'étude.

Sommaire

| | |
|---|----|
| Résumé..... | 2 |
| Abstract | 3 |
| Remerciements..... | 4 |
| Sommaire | 5 |
| Introduction..... | 7 |
| 1. Présentation du contexte de l'étude..... | 8 |
| 1.1. Présentation générale des Parcs nationaux | 8 |
| 1.1.1. Le Parc amazonien de Guyane | 8 |
| 1.1.2. Les activités du Parc | 9 |
| 1.1.2.1. Développement local | 10 |
| 1.1.2.2. Surveillance de l'orpaillage | 10 |
| 1.1.2.3. Connaissance du territoire | 10 |
| 1.1.3. Organisation du parc..... | 10 |
| 1.2. État de l'art | 11 |
| 2. Matériel et méthodes | 13 |
| 2.1. Zone d'étude | 13 |
| 2.2. Les végétations à cartographier | 13 |
| 2.3. Données satellitaires | 15 |
| 2.4. Données auxiliaires..... | 17 |
| 2.4.1. Les relevés terrain..... | 17 |
| 2.4.2. La récolte sur le terrain | 18 |
| 2.4.3. Les données agricoles et orpaillage | 19 |
| 2.5. Protocole de classification | 20 |
| 2.5.1. Nommage et organisation des fichiers..... | 20 |
| 2.5.2. Calcul des indices | 20 |
| 2.5.3. Création des parcelles d'entraînement | 21 |
| 2.5.3.1 Calcul des statistiques des ROI..... | 21 |
| 2.5.4. Traitement des images et des nuages..... | 22 |
| 2.5.5. Classification | 22 |
| 2.5.6. Synthèse des données | 23 |
| 2.6. Adaptation de la méthodologie à Qgis | 24 |
| 2.6.1. Création du plugin de classification dzetsaka..... | 24 |
| 2.6.1.1. Adaptation de l'algorithme Random Forest à Windows | 25 |
| 2.6.1.2. Modification de l'algorithme GMM | 26 |

| | |
|--|----|
| 2.6.1.3. Ergonomie et intuitivité | 26 |
| 2.6.2. Création des scripts Qgis | 27 |
| 3. Résultats et discussions | 28 |
| 3.1. Résultats de classification..... | 28 |
| 3.1.1. Zone Sud-ouest | 28 |
| 3.1.1.1. Indice Kappa et taux d'accord global | 29 |
| 3.1.1.2. Validation par photographie aérienne..... | 30 |
| 3.1.2. Zone de la pénéplaine de la Waki..... | 31 |
| 3.1.2.1. Indice kappa et taux d'accord | 31 |
| 3.1.2.2. Courbe ROC | 32 |
| 3.1.2.3. Validation par photographie aérienne..... | 33 |
| 3.2. Apport des indices | 33 |
| 3.3. Étude de l'effet biogéographique | 35 |
| 3.4. Étude de la phénologie | 36 |
| 3.5. Les différents problèmes rencontrés..... | 38 |
| 3.5.1. Détection des palmiers pinot | 38 |
| 3.5.2. Difficulté à seuiller l'intensité des parinaris | 38 |
| 3.5.3. Les variations d'éclairement..... | 39 |
| 3.6. Validation et comparaison avec le précédent travail | 39 |
| 3.7. Les limites de la cartographie des végétations particulières..... | 40 |
| 3.7.1. La juxtaposition des images..... | 40 |
| 3.7.2. La présence de palmiers bâche | 41 |
| 3.8. Carte d'identité des végétations et des capteurs | 41 |
| 4. Bilan et perspectives..... | 43 |
| 4.1. Conclusion | 43 |
| 4.2. Bilan personnel | 44 |
| Bibliographie..... | 45 |
| Table des illustrations | 47 |
| Table des tableaux..... | 48 |
| Annexes..... | 49 |

Introduction

La Guyane dispose d'un climat tropical stable à l'abri des cyclones et des tornades, la faune et la flore y connaissent donc peu de perturbations, d'autant plus que la population humaine est très faiblement présente (3 habitants/km²). Ainsi 96% du territoire est classé en forêt tropicale et le seul parc national présent sur le territoire est le plus grand de France avec 3,4 millions d'hectares de forêt amazonienne. Loin d'être simplement vert, vaste et uniforme, cet écosystème riche en biodiversité compte plus de 5200 espèces de plantes (Pavé, 2010) mais est assez mal cartographié tant les conditions d'accès y sont difficiles. Il est en effet impossible de s'y rendre par voie routière. Le moyen de locomotion privilégié est la pirogue, voire l'avion pour les villages de Saül ou de Maripasoula, et certains lieux au sud du parc ne sont accessibles qu'après plusieurs jours de transport.

Dans cet ensemble de forêt tropicale se trouvent des végétations inhabituelles et très différentes de la canopée habituelle : c'est ce qu'on appelle les végétations particulières. Il s'agit d'ensembles de palmiers pinôts, palmiers bâches, de parinaris, de bambous épineux ou encore de savane-roche (des végétations typiques d'inselbergs ou îlots rocheux).

L'objectif principal du stage est de cartographier de manière précise ces cinq types de végétations particulières présentes dans la forêt amazonienne. Selon le niveau de description des données terrain ou de la reconnaissance possible par photo-interprétation, la classification sera faite par essence – parinaris (*Parinari campestris*), palmiers bâches (*Mauritia flexuosa*) et pinotières (*Euterpe oleracea*) – ou par phisonomie – savane-roche et végétation basse qui regroupe les cambrousées de bambous (*Guadua macrostachya* et *Lasiacis ligulata*) et les héliconias (*Heliconiaceae*).

Un premier objectif est de savoir s'il est possible, et à quelle échelle, de cartographier les différents types de végétation sélectionnés sur l'ensemble du territoire. Il s'agit aussi de tester les différents indices de végétations (EVI2, NDVI), de comparer les algorithmes (Random Forest, Gaussian Mixture Model). L'hypothèse est faite que ces végétations se distinguent suffisamment bien à une date unique en utilisant un algorithme robuste comme Random Forest, et que la résolution spatiale peut compenser le manque de résolution temporelle (Brondizio et al. 1994).

Le deuxième objectif est de mettre en place à partir de Qgis une méthodologie semi-supervisée facilement reproductible, afin que le Parc amazonien de Guyane puisse renouveler cette cartographie dans quelques années avec des capteurs différents (Sentinel-2, SPOT-6/7, Pléiades...).

Cet objectif pose plusieurs questions en télédétection : l'usage d'images multi-temporelles est-il indispensable pour reconnaître les végétations sélectionnées ? L'utilisation d'indices spectraux améliore-t-il la qualité des résultats comme ce fut le cas sur des résolutions spatiales plus faibles ? Quelle chaîne de traitement mettre en place pour traiter les images et avoir un résultat unique à l'échelle du parc ? Quelles sont les bandes spectrales les plus importantes et les particularités nécessaires des capteurs pour poursuivre ce travail ?

1. Présentation du contexte de l'étude

1.1. Présentation générale des Parcs nationaux

Le Parc amazonien de Guyane est l'un des dix Parcs nationaux présents sur le territoire français. Reconnus au niveau international comme des patrimoines d'exception, ces territoires mélagent des espaces remarquables et un mode de gouvernance et de gestion qui permettent d'en préserver les richesses. Les parcs nationaux couvrent des domaines terrestres mais aussi maritimes et représentent près de 9,5% du territoire français (soit 60 728 km²).

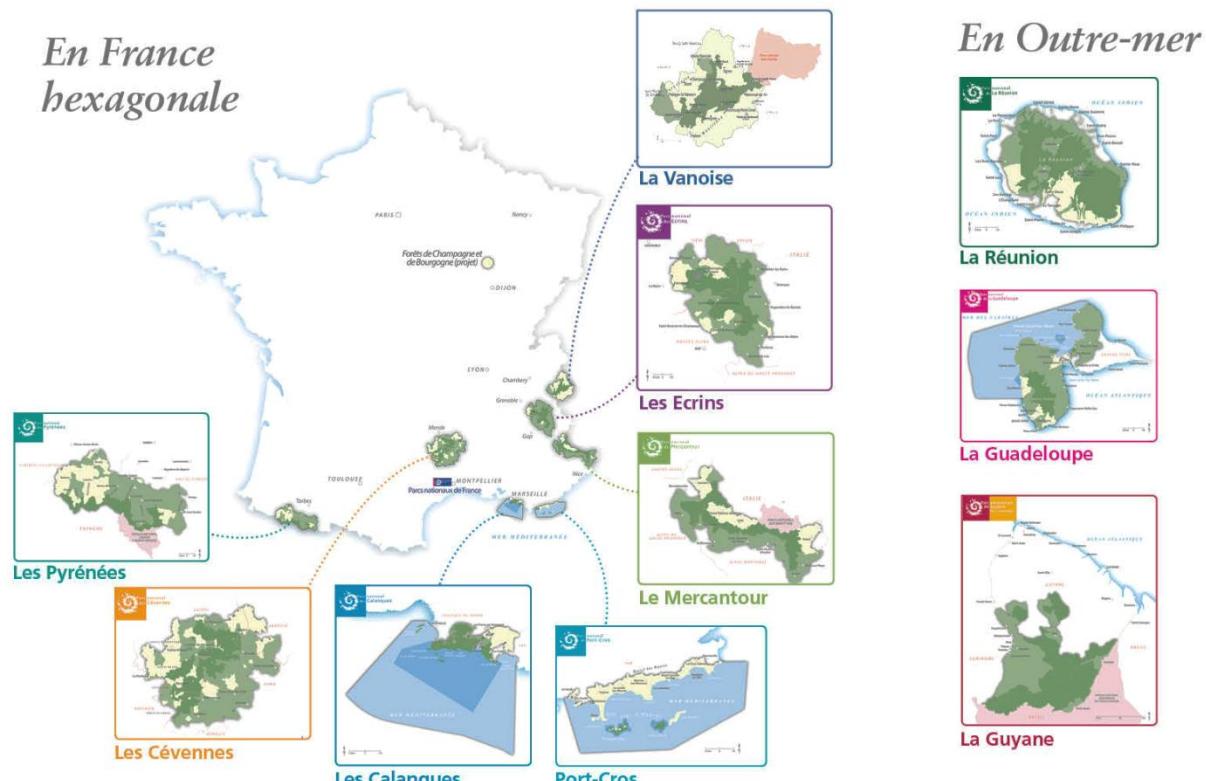


Figure 1 : Les différents parcs nationaux. Source : Parc Nationaux de France

Les parcs nationaux attirent chaque année plus de 8,5 millions de visiteurs. Le Parc amazonien de Guyane a été créé la même année que celui de La Réunion, en 2007.

1.1.1. Le Parc amazonien de Guyane

Le Parc amazonien de Guyane est de loin le plus grand des dix parcs nationaux avec une superficie totale représentant près du double de celle des neufs autres parcs nationaux cumulés. Créé en février 2007, sa superficie est de 3,4 millions d'hectares dans le centre et le sud de la Guyane française. Il est constitué d'une zone de cœur, un espace réglementé dans le but de protéger l'environnement, et d'une zone d'adhésion, où la priorité est donnée au développement local et adapté, dans le respect des modes de vie. Selon Demaze (2008), la zone de libre adhésion est “*destinée à la gestion durable des ressources naturelles pour le développement des collectivités territoriales*”.

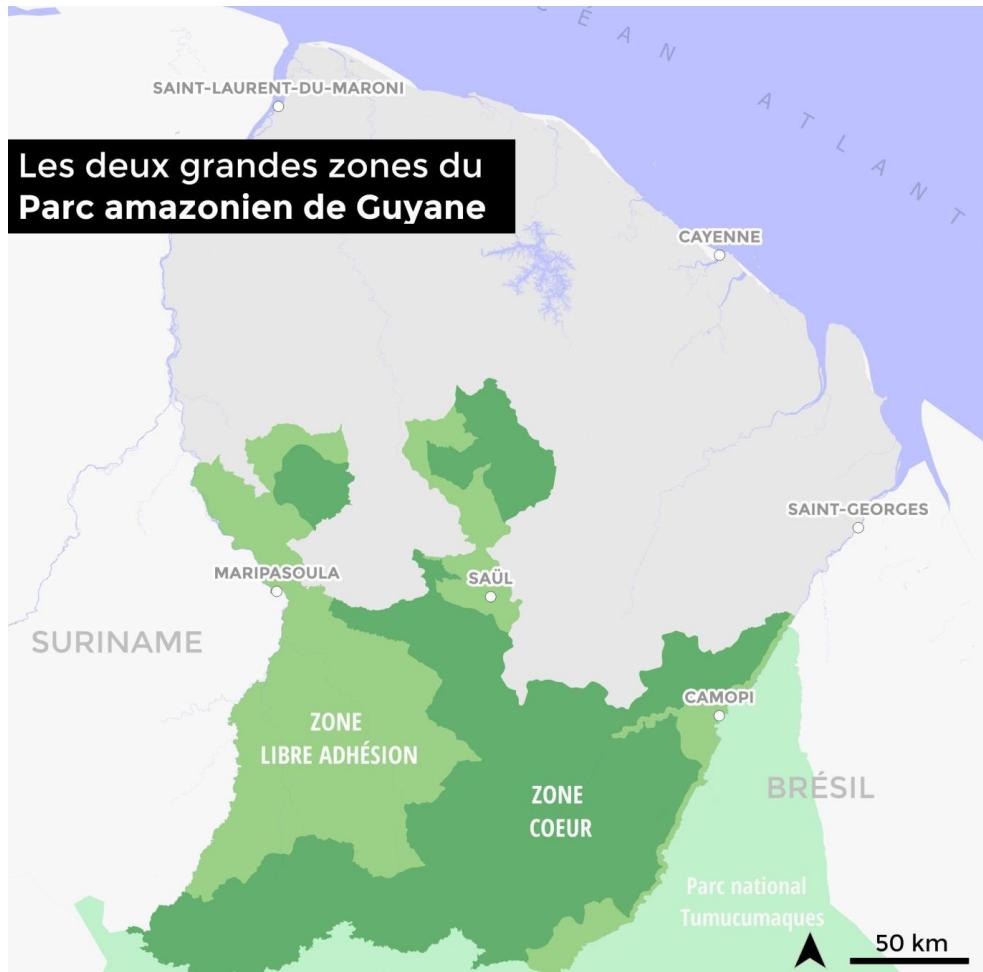


Figure 2 : Les deux grandes zones du Parc amazonien de Guyane

Cinq communes se trouvent sur le territoire où vivent plus de 18 000 habitants : des communautés amérindiennes (Wayana, Wayãpi, Téko), des Noirs Marrons (Aluku) et des Créoles. L'action du Parc amazonien de Guyane est orientée par la charte des territoires, un document élaboré pendant cinq ans par l'ensemble des acteurs concernés : élus locaux, autorités coutumières, associations, socio-professionnels, habitants, services de l'État. La charte vise à apporter des réponses sur deux thématiques centrales : le développement durable des territoires et la valorisation des patrimoines naturels et culturels.

1.1.2. Les activités du Parc

Le sud de la Guyane française, du Suriname et le nord-est du Brésil (État de l'Amapá et nord du Pará) forment le plus vaste massif forestier tropical non fragmenté, d'une superficie de près de 30 millions d'hectares. Avec le Parc National du Tumucumaques au Brésil, il s'agit du plus grand espace protégé de forêt tropical au monde.

A l'initiative des gestionnaires d'aires protégées, un dialogue technique s'est engagé à l'échelle régionale. Ainsi le Parc amazonien soutient différentes actions de coopération telles que des rencontres transfrontalières communautaires, des partages de données faune et flore, ou encore des expériences en matière de gouvernance des aires protégées. À une échelle plus large, le Parc amazonien est également membre du réseau REDPARQUES, qui regroupe les aires protégées d'Amérique du Sud et de la Caraïbe.

1.1.2.1. Développement local

Un des enjeux du parc national est de favoriser le développement d'activités économiques locales, adaptées et durables, en concertation avec la population et dans le respect des modes de vie. Il s'agit de contribuer à la structure de micro-filières, d'accompagner les porteurs de projets dans les domaines de l'agriculture, de l'artisanat ou du tourisme équitable, mais aussi de soutenir les collectivités locales dans leurs projets alliant développement local et protection de l'environnement. Les actions se focalisent autour de quatre grands enjeux :

- Amélioration du cadre de vie (équipements, déchets, etc.)
- Accompagnement des porteurs de projet
- Appui à la réalisation de projets éco-touristiques
- Éducation à l'environnement et au développement durable

1.1.2.2. Surveillance de l'orpaillage

Une autre mission importante du parc est la surveillance du territoire car ce dernier subit les impacts destructeurs de l'orpaillage clandestin. Les équipes du Parc national effectuent des missions de diagnostic environnemental (aérien, fluvial et terrestre) afin d'alerter les forces de l'ordre en charge de la lutte contre ce fléau. Elles participent également à des opérations conjointes sur le terrain avec les forces armées de Guyane et la gendarmerie.

Pour assurer un suivi des zones orpaillées illégalement, l'usage de la télédétection se fait en collaboration avec l'ONF et à partir des images obtenues à travers le projet SEAS (Surveillance de l'Environnement Amazonien par Satellite).

1.1.2.3. Connaissance du territoire

L'acquisition de connaissances en vue d'une bonne gestion des ressources est une des missions majeures du Parc national. L'Établissement public et ses partenaires développent des programmes participatifs, en lien avec les populations, afin de répondre à des questions primordiales sur la pérennité des ressources en gibier et en poisson. Les habitants sont informés des résultats des études par le biais de restitutions publiques.

Le Parc amazonien organise ou participe à la réalisation d'inventaires naturalistes dans l'optique d'améliorer la connaissance de la biodiversité. Il contribue à alimenter les bases de données nationales en transmettant de nombreuses données collectées sur son territoire. Les équipes de l'Établissement public mènent également des missions de reconnaissance dans des zones peu ou pas explorées. La reconnaissance et la valorisation de la diversité des cultures du territoire constituent ici un enjeu fondamental. C'est dans le cadre de cette connaissance de la biodiversité que ce stage s'inscrit.

1.1.3. Organisation du parc

Le parc est organisé autour du siège qui se situe à Remire-Montjoly (commune de Cayenne) et qui compte 34 employés. Trois délégations territoriales se répartissent le reste des effectifs avec tout d'abord la Délégation Territoriale du Maroni (DTM) qui se situe à Maripasoula et compte 30 agents ; la Délégation Territoriale de l'Oyapock (DTO) à Camopi avec 16 agents sur le terrain ; enfin 7 agents présents à la Délégation Territoriale du Centre (DTC) à Saül.

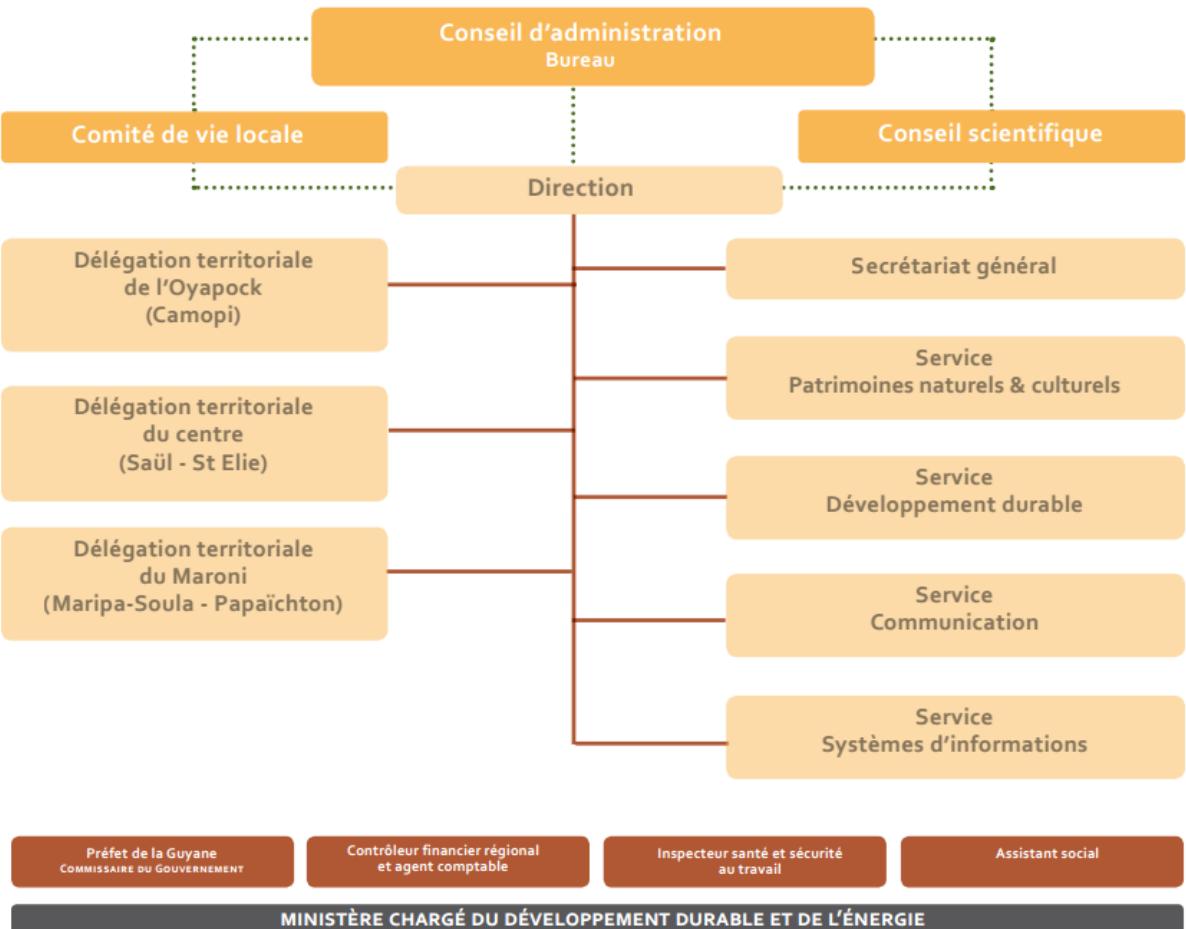


Figure 3 : Organigramme du Parc. Source : PAG

Le service Systèmes d'informations qui compte trois agents est dirigé par David Franco, informaticien et également membre du comité de direction (CODIR).

La tutrice du stage, Pauline Perbet, est géomaticienne sous la direction de Pierre Joubert, responsable SIG.

Les différentes missions du service SIG sont entre autre de créer les cartographies pour les sentiers du Parc amazonien de Guyane, de fournir des outils de suivi des surfaces agricoles, d'établir des cartes de toponymie (et leur récolte à travers de nombreuses missions terrain) et de les restituer auprès des territoires,

1.2. État de l'art

L'utilisation de la télédétection pour recenser les formations végétales se fait depuis plusieurs décennies à partir de capteurs avec des résolutions spatiales assez faibles (Landsat, SPOT-4...). Les résultats varient, mais des taux d'accord de plus de 90% selon les classes peuvent être observés sur les forêts tropicales (Foody et Hill, 1996). D'une manière générale, ces travaux ont pour but de cartographier les habitats forestiers et non les essences.

La forêt guyanaise est très méconnue car elle est non seulement très vaste mais aussi très difficile d'accès. La plupart des travaux en télédétection se concentre donc sur les forêts côtières où la présence de mangroves est importante (Fromar et al., 2004). Des travaux à l'échelle de la Guyane ont mis en avant des différences phénologiques selon les grandes zones de végétations (Pennec et al. 2010). Ces résultats mettent en avant que le Parc amazonien de Guyane est surtout composé de forêt à canopée ouverte et de mélange de forêt

haute et de canopée ouverte. Cependant l'usage du satellite MODIS (250m minimum de résolution spatiale, mais 500m dans la majorité des bandes spectrales) ne permet pas une grande précision de la donnée et la télédétection précise des essences forestières. En 2011 Valéry Gond et al. ont publié un article sur les différents types de forêts dans le bouclier guyanais (la formation géologique d'Amérique du Sud sur laquelle se trouve la Guyane) en utilisant le satellite SPOT-4.

Une mosaïque faite par l'IGN en 2008 avait servi de base pour réaliser une carte d'occupation du sol, cependant, elle ne recensait pas les végétations voulues par le parc (Stach et al., 2009).

Le seul recensement plus précis des forêts a été calculé à partir de la géomorphologie et de la texture en utilisant le modèle altimétrique (SRTM 30m), et localise les grands habitats forestiers (Guitet et al., 2013). Un livre *Catalogue des habitats forestiers de Guyane* (2015) est d'ailleurs paru suite à la thèse et aux recherches de Stéphane Guitet.

A plus petite échelle, Saatchi et al. (2008) ont combiné plusieurs capteurs, dont MODIS, pour développer plusieurs variables dont une spécialement sur la végétation qui regroupe cinq essences différentes *Calophyllum brasiliense* (*Clusiaceae*), *Carapa guianensis* (*Meliaceae*), *Hura crepitans* (*Euphorbiaceae*), *Manilkara bidentata* (*Sapotaceae*), et *Virola surinamensis* (*Myristicaceae*).

Pour trouver des travaux plus proches de celui de ce stage, il faut aller au Brésil, dans la région du Mato Grosso. En 2006 Carreiras et ses pairs ont utilisé le satellite SPOT-4 afin de cartographier différents habitats, et ce en utilisant des images multi-temporelles sur douze mois avec un taux d'accord global de 88%. Avec Landsat™, Brondizio et al. (1994) ont cartographié cette fois-ci les essences comme le palmier pinôt, avec un taux d'accord global de 94%.

Au parc, Pauline Perbet, géomaticienne, a réalisé un premier essai de cartographie des savanes-roches et des végétations basses sur l'ensemble du territoire. La méthode consistait à définir manuellement un arbre de décision, et il fallait à chaque image retrouver les bons paramètres. Cependant, cette méthode ne permettait pas de détecter les végétations spectralement proches de la forêt comme les palmiers pinôt, bâche ou les parinaris.

Afin d'avoir des données précises sur les habitats des essences et leur répartition potentielle en Guyane, Jean-Jacques de Granville et Marc Gayot ont publié en 2014 le *Guide des Palmiers de Guyane*. Cet ouvrage, fortement utilisée dans ce stage, recense l'ensemble des palmiers présents sur le territoire, leurs caractéristiques (stipe, feuilles, pennes, distribution spatiale, type habitat), et leur saison de floraison.

Comme on peut le voir, la télédétection des essences forestières n'est pas la plus courante en milieu tropical, il s'agit plutôt de localiser des habitats (forêt haute de feuillus, forêt à canopée ouverte...), car les grands ensembles d'une essence unique sont en fait assez peu fréquents.

2. Matériel et méthodes

2.1. Zone d'étude

La zone d'étude correspond à la totalité du territoire du Parc amazonien de Guyane, zone de libre adhésion et zone de cœur, soit la totalité des 3,4 millions d'hectares. Majoritairement recouvert de forêt, cet espace comporte néanmoins quelques villages et plusieurs zones d'orpaillage qu'il conviendra de ne pas traiter dans la cartographie des végétations particulières.

2.2. Les végétations à cartographier

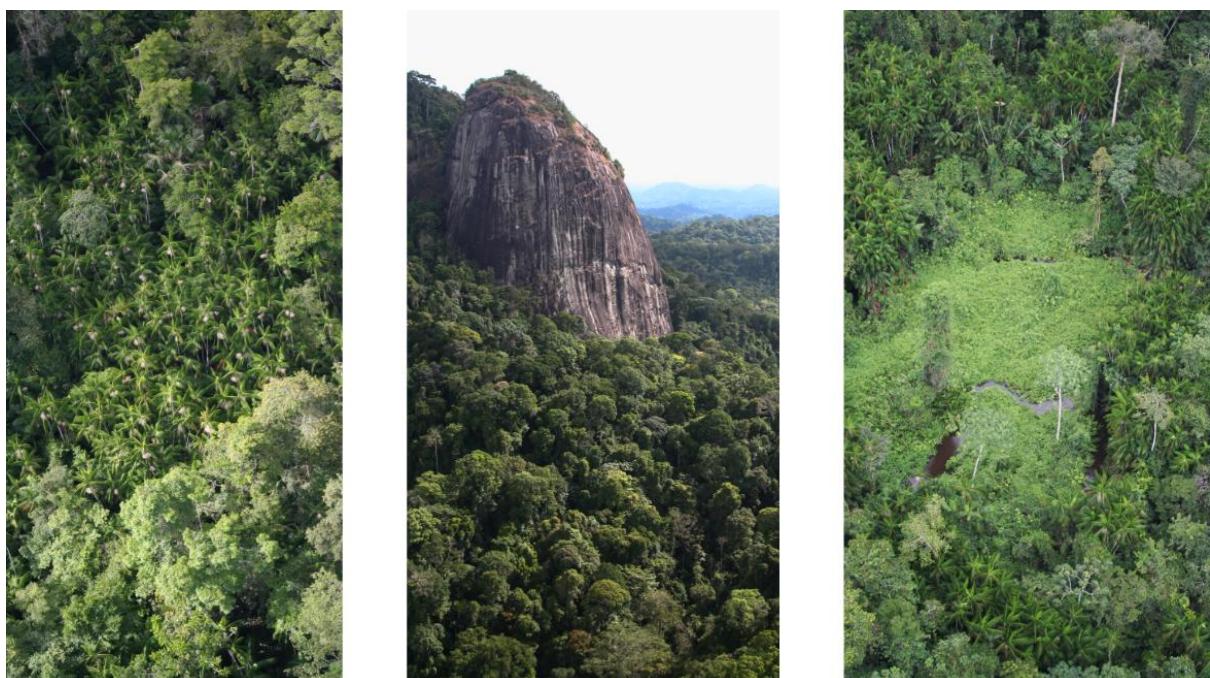


Figure 4 : De gauche à droite : palmiers pinôts (*Euterpe oleracea*), savane-roche, végétation basse (cambrouse de bambous *Lasiacis ligulata*). Crédits : Daniel Sabatier (IRD)

Selon le niveau de description des données terrain ou de la reconnaissance possible par photo-interprétation, il sera réalisé une classification soit par essence, soit par phystionomie. La savane-roche par exemple représente un mélange de végétations présentes sur les inselbergs (littéralement les îlots rocheux) et leur cartographie se fait en recensant les roches de Guyane.

Ces végétations sont appelées particulières car elles se démarquent de l'ensemble de la forêt amazonienne, qui peut être considérée comme uniforme (on dit couramment que sa texture ressemble à celle d'un chou-fleur). Ces essences étant différentes de la majorité de la forêt, elles constituent donc un indicateur important de la biodiversité. Ce sont aussi des groupes de végétations homogènes, et donc cartographiables car elles forment un ensemble assez uni qui peut être détecté en utilisant les images satellites.

Nul ne sait aujourd'hui comment des cambrousse de bambous se sont retrouvées dans certains endroits de la forêt guyanaise, mais en faire un inventaire permettra de suivre leur évolution et peut-être de mieux comprendre leur origine.



Figure 5 : Canopée de la forêt amazonienne. Photo : Daniel Sabatier (IRD)

Dans certaines d'entre elles des traces de poteries ont été découvertes, ce qui laisse supposer que ce lieu était habité auparavant. Certaines hypothèses parlent d'ouvertures créées par des chablis (chute d'arbres), d'autres par une action anthropique car il est extrêmement difficile de parcourir une cambrouse de bambous épineux. Cela a donc pu constituer un moyen de défense. Les cambrousse peuvent donc avoir plusieurs origines, et afin d'en savoir plus à l'avenir il est important d'établir un inventaire, qui constitue le préalable pour effectuer leur suivi par la suite.

Tableau 1 : Aperçu des végétations particulières sous SPOT-5

| Végétation | Vue par satellite (SPOT-5 : Composition 4-1-2) |
|---|--|
| Végétation basse, (<i>Héliconiaceae</i> , <i>Lasiacis ligulata</i> et <i>Guadua macrostachya</i>) (vert clair) | |
| Parinari, (<i>Parinari campestris</i>) (rouge/marron foncé) | |
| Palmier bâche (<i>Mauritia flexuosa</i>) (vert clair, plus lisse) | |
| Palmier pinôt (<i>Euterpe oleracea</i>) (vert lisse) | |
| Savane roche (rouge) | |

Concernant le palmier pinôt, il est très utilisé en Guyane car il entre, selon Granville & Gayot (2014), dans la confection de toitures, planchers, vannerie (balais, récipients). Il a aussi un usage médicinal et alimentaire (cœur de palmier).

L'usage du palmier bâche est aussi important : utilisé aussi comme complément alimentaire (fruits et sève pour faire du jus et du vin, consommation du cœur du palmier), mais aussi pour la fabrication de hamac ou de natte. Toujours selon Granville & Gayot, au niveau médical il sert comme diurétique, antidiarrhéique, ou encore antalgique.

2.3. Données satellitaires

Pour réaliser la cartographie des végétations particulières à l'échelle du Parc amazonien de Guyane, des images SPOT-5 ont été acquises dans le cadre de la phase I du projet SEAS (Surveillance de l'Environnement Amazonien par Satellite) présent depuis 2006 en Guyane. Ce projet mené par l'IRD et financé par plusieurs institutions, notamment la région, a permis d'obtenir des images satellites grâce à l'installation d'une antenne réceptrice en Guyane. L'amplitude temporelle des 55 images utilisées pour la cartographie va de 2007 à 2014, essentiellement en saison sèche (début août à mi novembre), seule époque de l'année où la nébulosité est acceptable, et où l'intensité spectrale des végétations est la plus élevée (Cherrington et al., 2016). Le nombre limité de données annuelles s'explique par la forte nébulosité en Guyane. Sur la plupart des images, y compris en saison sèche, il est très fréquent d'avoir des taux de couverture nuageuse supérieurs à 50%.

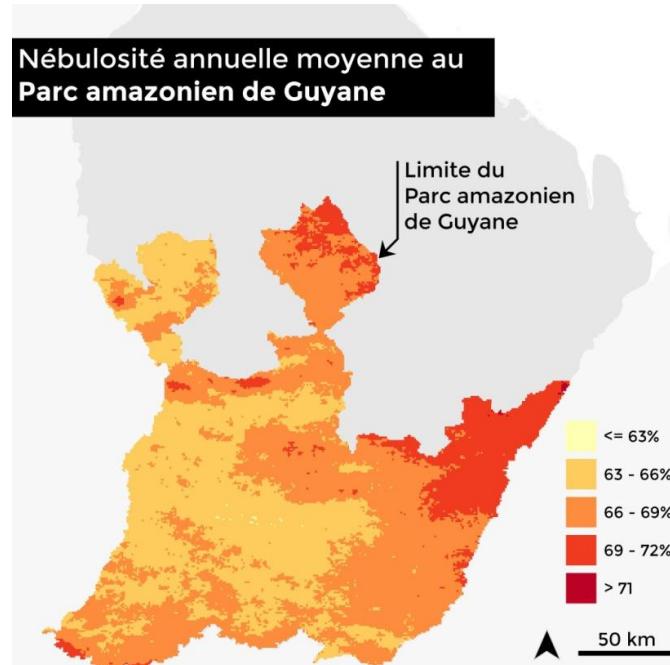


Figure 6 : La nébulosité moyenne au Parc. Donnée source de EarthEnv

Comme le montre le graphique ci-dessus, la couverture nuageuse moyenne annuelle est de l'ordre de 63 à 70% sur l'ensemble du territoire. La télédétection a donc ses limites et la faible résolution temporelle de SPOT-5 limite l'opportunité de saisir de larges étendues avec pas ou peu de nuages durant la saison sèche.



Figure 7 : Date d'acquisition des images SPOT-5

Les images acquises sont fournies en traitement 2A¹ et contiennent deux canaux du visible et deux dans l'infrarouge, sur une emprise totale de 3 600 km² (60x60km).

Tableau 2 : Détail des bandes spectrales de SPOT-5

| N° Canal | Couleur | Bande spectrale | Résolution (m) |
|----------|-------------------------|-----------------|----------------|
| B1 | Vert (V) | 0,50–0,59µm | 10x10 |
| B2 | Rouge (R) | 0,61–0,68µm | 10x10 |
| B3 | Proche infrarouge (PIR) | 0,78–0,89µm | 10x10 |
| B4 | Moyen infrarouge (MIR) | 1,58–1,75µm | 20x20 |

Pour couvrir la totalité du parc, 23 scènes SPOT-5 sont nécessaires. L'activité de ce satellite, commencée en 2002, s'est terminée en 2015. Il n'est donc plus possible d'obtenir des images avec les mêmes particularités. Par exemple son successeur SPOT-6 ne possède plus la bande du moyen infrarouge (MIR).

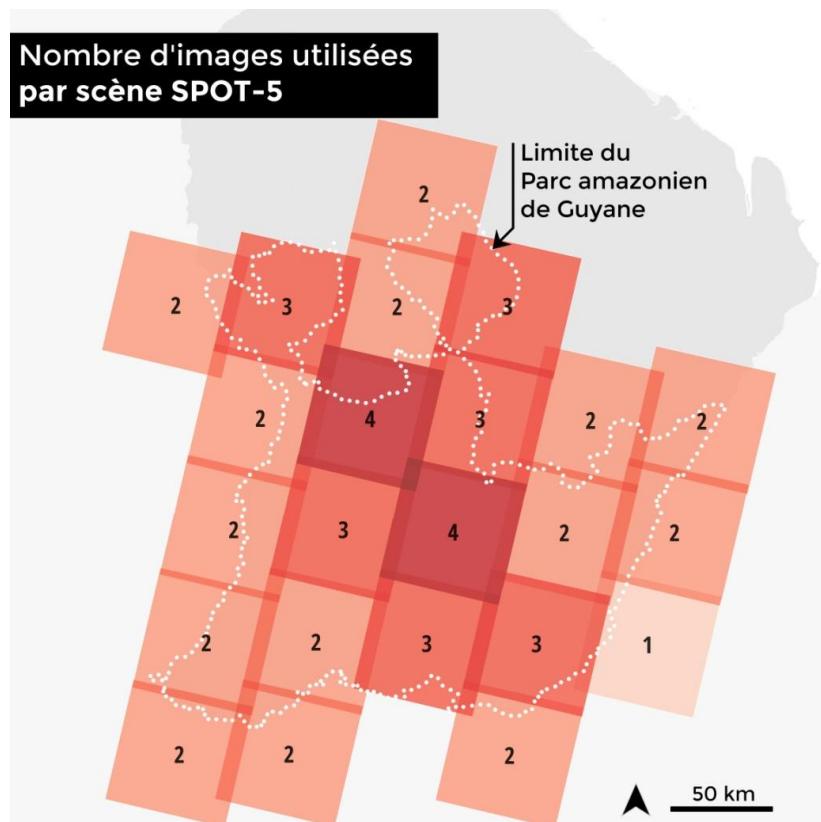


Figure 8 : Nombre d'images utilisées pour chaque scène SPOT-5

Autre particularité : la plupart des images de 2009 étaient comme floutées, elles étaient donc presque inexploitables. Cependant, elles se sont révélées d'une grande utilité pour la détection de la savane-roche car leur moyen infrarouge était très fort et faisait mieux ressortir les zones d'intérêt. Nous n'avons pas d'explication quant à la mauvaise qualité de

¹ Correction géométrique des effets systématiques (effet panoramique, courbure et rotation de la Terre). Correction géométrique effectuée dans la projection cartographique standard (UTM WGS84 par défaut) sans prise de points d'appui.

ces images (peut-être une mauvaise mise au point du capteur) mais cela a permis d'améliorer sensiblement la cartographie des savane-roche.

De manière générale, les différences de localisation entre plusieurs images d'une même scène ont nécessité un nouveau géoréférencement à partir de la mosaïque de la Guyane faite par l'IGN en 2008. Parfois, un écart de plus de 100m était constaté entre deux images. Comme ce travail implique la synthèse des scènes et des images entre elles, il faut donc qu'elles se juxtaposent du mieux possible.

2.4. Données auxiliaires

2.4.1. Les relevés terrain

Les données terrain sont indispensables en télédétection car elles servent de base de travail et permettent de bien repérer à quelle zone correspond telle espèce. Le territoire du Parc amazonien de Guyane s'étendant sur 3,4 millions d'hectares, toutes les zones ne sont donc pas égales en qualité et quantité de données. Certaines sont beaucoup mieux connues que d'autres, c'est le cas du Mont Itoupé qui abrite toutes les espèces végétales particulières recensées à l'échelle du parc et demandées pour ce stage. Le service SIG du parc recense aussi les points GPS de relevés des différentes expéditions scientifiques qui ont eu lieu dans le but de recenser les essences forestières. Les données proviennent majoritairement du CIRAD (3470 relevés soit 77% du total) et le reste de différents relevés du parc (photos aériennes, points GPS...).

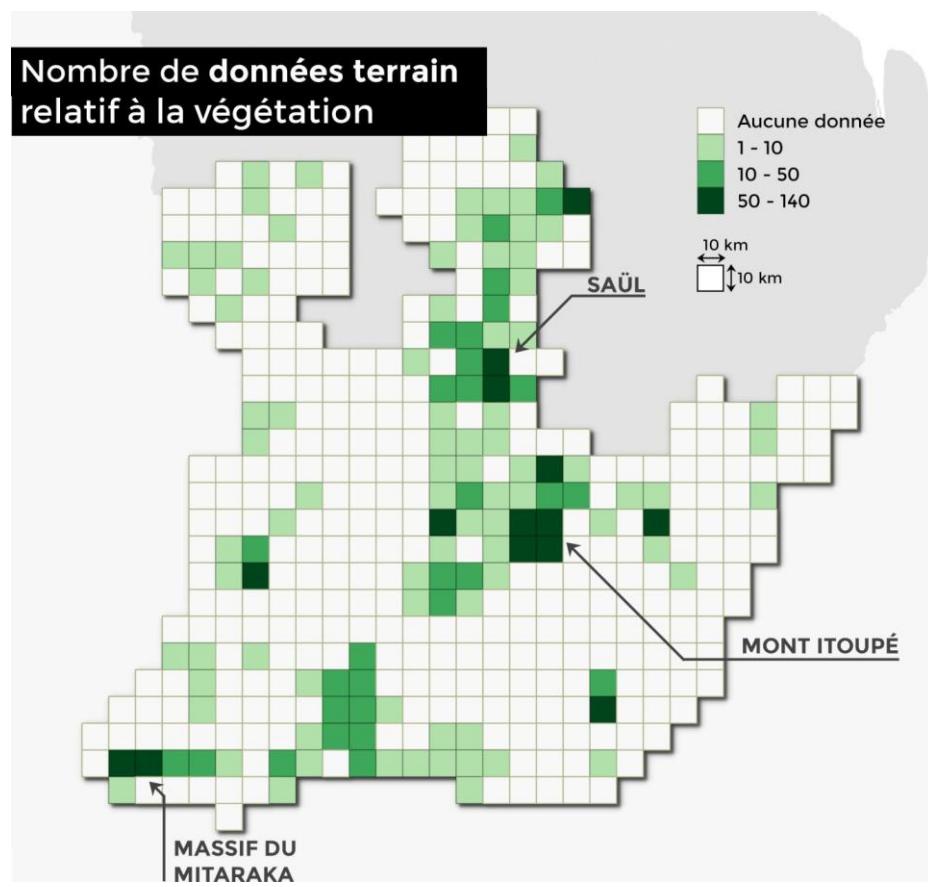


Figure 9 : Nombre de données terrain (pour 100km²)

Cependant, les informations du CIRAD n'indiquent pas le type de végétation présente. Ces données indiquent seulement la présence du parinari ou du tachigali s'il y en a, et à quel pourcentage. Ainsi sur les 3470 relevés 34D, seuls 169 relevés terrains indiquent la présence de parinaris, le reste précise juste l'absence de parinaris. De plus les relevés sont faits sur des parcelles de 100m², ce qui signifie que le point en question vue par satellite ne peut très bien ne pas correspondre à du parinari et pourtant il sera mentionné comme tel dans la base. Il faut donc créer manuellement toutes les parcelles d'entraînement qui seront utilisées pour la classification.

L'un des principaux problèmes reste le manque d'information à l'échelle du parc. Il n'est pas rare que selon les scènes SPOT-5 il n'y ait aucun point de référence terrain, cela va donc compliquer la recherche et il faudra être d'autant plus vigilant sur les choix des parcelles d'entraînement. Toutefois, la présence importante de relevés terrain comme sur le Mont Itoupé permettra d'entraîner un modèle de classification et de l'exporter sur une autre scène afin de vérifier en photo-interprétation si la classification est correcte et de valider ou non de nouvelles zones d'apprentissage. Cependant, les différences temporelles entre les images risquent de fausser de manière importante l'entraînement et la classification d'une scène à l'autre.

Tableau 3 : Nombre de données terrain par type de végétation

| Végétation | Nombre de relevés |
|------------------|-------------------|
| Végétation basse | 319 |
| Parinari | 208 |
| Palmier pinôt | 99 |
| Savane-roche | 33 |
| Palmier bâche | 16 |

2.4.2. La récolte sur le terrain

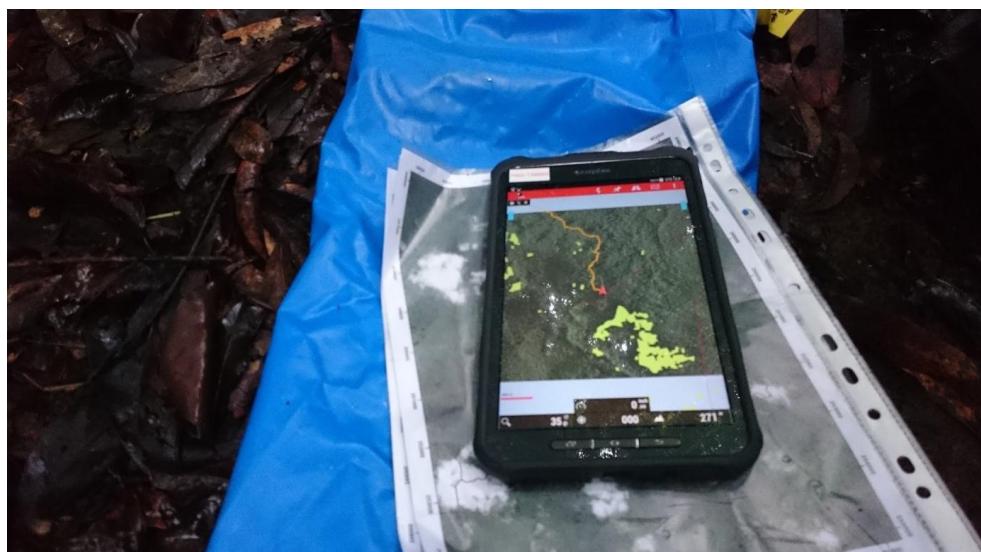


Figure 10 : Utilisation d'une tablette/GPS pour vérifier les végétations détectées

Au début du stage, une expédition de 2 jours (cf carte en annexe 2) s'est déroulée à proximité de Saül en présence de Pauline Perbet (géomaticienne), Sébastien Sant (naturaliste à Saül) dans le but d'aller repérer sur le terrain quelques végétations basses recensées à partir d'images satellites. Après le vol vers Saül et plusieurs heures de marche sous la pluie et la chaleur, nous avons posé notre carbet bâche (un hamac posé entre deux arbres et abrité par une bâche) sous des trombes d'eau. Le lendemain nous avons pu alors commencer à nous rendre sur les lieux identifiés comme des cambrousse. Nous avons constaté la grande hétérogénéité de ces milieux. Alors que nous nous attendions à trouver un ensemble assez homogène, il est apparu alors que ces endroits étaient non seulement composés de bambous, mais aussi d'héliconias, voire de lianes. Le plus souvent ce type de végétation était très mélangé. C'est à ce moment qu'il a été décidé d'essayer de cartographier les végétations basses sans inclure de différenciation entre cambrousse ou héliconias.



Figure 11 : De grandes difficultés d'accès aux végétations

2.4.3. Les données agricoles et orpaillage

Afin de ne pas recenser des données en végétations ou en savane-roche lorsqu'il s'agit d'agriculture ou d'orpailage, un masque des données à éviter a été créé. Celui-ci repose tout d'abord sur la couche *AgriPag* qui contient la localisation de l'ensemble des parcelles agricoles présentes sur le territoire. Pour l'orpailage, la couche se base sur l'OAM (Observatoire des Activités Minières), c'est une donnée confidentielle qui recense l'ensemble des zones orpailées, qu'elles soient légales ou illégales.

Par la suite une zone tampon de 100m autour du masque a été générée afin de corriger les différences de localisation et des évolutions potentielles légères agricoles ou d'orpailage.

Les zones à déclasser représentent 128 000 hectares, soit 3,76% du territoire du parc.

2.5. Protocole de classification

Une illustration facilitant la compréhension du protocole de classification (du traitement de l'image jusqu'au rendu final de la cartographie) est consultable (annexe 1).

2.5.1. Nommage et organisation des fichiers

Afin de faciliter l'exécution en lot des scripts de classification, une organisation et un nommage des fichiers a été choisi. A chaque dossier correspond une scène SPOT-5, dans lequel se trouve la totalité des images, des indices, et des vecteurs associés.

Les fichiers ont été nommés tout d'abord par leur date de prise de vue, puis leur numéro de scène, et enfin leur indice si tel était le cas. Exemple :

```
date-scene.tif (exemple : 20120505-685-344.tif)
date-scene.shp (exemple : 20120505-685-344.shp)
date-scene_nomIndice.tif (exemple : 20120505-685-344_ndvi.tif)
```

Cela permet pour les statistiques qui seront réalisées par la suite de pouvoir enregistrer la date de la prise de vue mais aussi le numéro de la scène et le nom de l'indice si c'en est un.

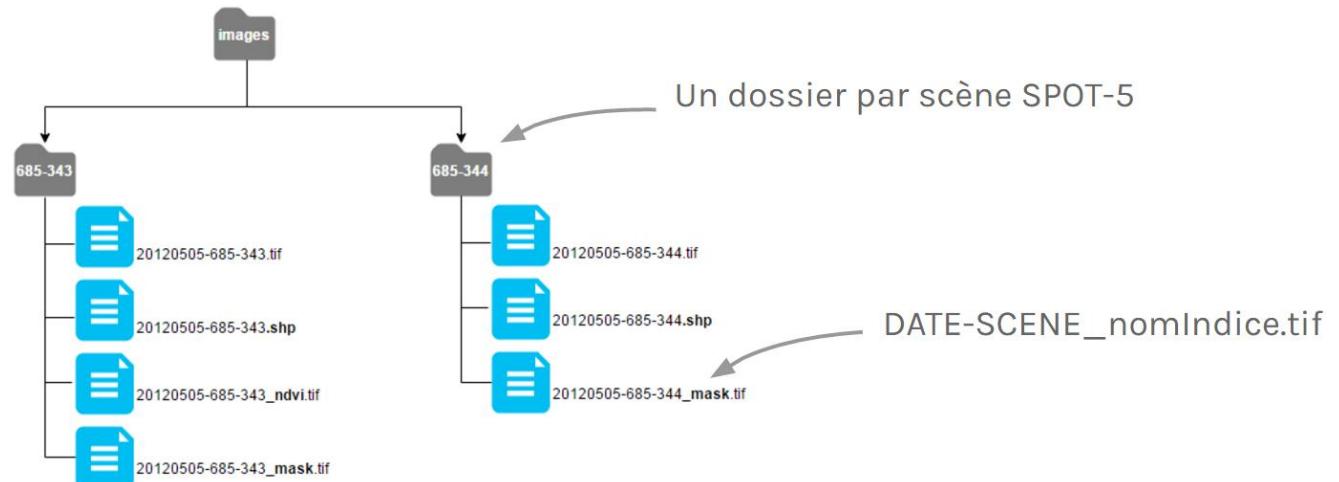


Figure 12 : Schéma d'organisation des images

2.5.2. Calcul des indices

Les indices de végétation (NDVI, EVI2), d'eau (NDWI) ou de nébulosité (NDCI) étant nécessaires afin d'évaluer leur impact dans la classification des images, leur calcul a donc été automatisé. Afin de faciliter le travail et de permettre un traitement optimal des données, un script Qgis écrit en python a été développé afin de pouvoir traiter en lot les images. La totalité des indices a été généré en quelques clics et en quelques minutes depuis Qgis. Le nommage des fichiers se fait selon la norme mise au point pour ce travail, c'est-à-dire que le nom de l'indice sera justaxposé par un underscore au nom du fichier.

Le script compute_indices est disponible comme tous les autres scripts et modèles Qgis sur la page github.com/lennepkade/ScriptsParcAmazonien/.

2.5.3. Création des parcelles d'entraînement

La création des parcelles d'entraînement (ROI) s'est faite par photo-interprétation en s'aidant de la précédente cartographie des végétations basses et des savanes-roche réalisée par Pauline Perbet mais aussi des quelques relevés terrains qui ont grandement facilité la détection des végétations.

Un fichier vectoriel contenant les ROI (Regions Of Interest) a été créé et chaque polygone correspondant au type de végétation dessiné. Les nuages, ombres et franges nuageuses ont aussi été détournés. Étant donnée les différences de géolocalisation des images et de la couverture nuageuse très changeante, il a été décidé de créer un fichier vectoriel par image car il y avait trop de différences entre les dates d'une même scène.

La forêt étant très variée, elle a été divisée en deux classes, forêt et forêt déstructurée. La seconde regroupe les parcelles contenant beaucoup de lianes par exemple et est plus claire sur les images.

Au total ce sont 1608 polygones qui ont été créés sur les 55 différentes images, soit en moyenne une trentaine par image.

La plus grande difficulté a été de trouver les parcelles d'entraînement pour les palmiers bâches et palmiers pinôts. En effet, très peu de données terrains étaient disponibles, et ils sont spectralement assez proches de la forêt. La solution choisie a été de classifier les images avec des parcelles d'entraînement validées sur le terrain, puis de voir s'il y avait des végétations intéressantes détectées à la marge des images (et donc potentiellement à la marge aussi des autres scènes SPOT-5).

La différence spectrale entre deux scènes était trop importante car cela pouvait varier de quelques mois à une année ou plus. Il n'était donc pas possible d'entraîner un modèle d'une image et de l'utiliser sur une image voisine. Sur les scènes où aucune donnée aux marges des images n'était disponible, elles ont donc été analysées finement afin de détecter des ensembles de palmiers. Quelques spécificités des végétations ont permis de réaliser ce travail. Par exemple les palmiers pinôts se situent selon Granville & Gayot (2014) en "*bord de crique et [sur] sol hydromorphe*". On les trouve donc dans les bas-fonds et, chose importante pour la télédétection, ils ont une texture légèrement plus lisse que la forêt sur les images SPOT-5.

2.5.3.1 Calcul des statistiques des ROI

Une fois les parcelles d'entraînement créées et les indices générés, il a pu être calculé pour l'ensemble des images les statistiques sur la valeur spectrale des végétations.

Afin d'extraire des parcelles d'entraînements la valeur moyenne des pixels et de les sauvegarder dans un fichier unique contenant la totalité des 55 images, le script *runStatsBatch.R* (annexe 3) a été développé. Il parcourt chaque dossier, extraits les valeurs de chaque polygone, et enregistre dans une colonne cette valeur, mais aussi la date de la prise de vue, le numéro de la scène, le type de végétation.

Pour traiter en premier l'image et non les indices, car un premier fichier qui contient les données est créé pour chaque image, il a simplement été ordonné sous R la liste des images selon la longueur de leur nom, ainsi les images des indices ont forcément un nom plus long :

```
# Get all tif from folder
inRaster <- list.files(folder, pattern='*.tif$')
# Order list to have main image first
inRaster[order(nchar(inRaster))]
```

Tableau 4 : Extrait des valeurs des ROI des 55 images

| type | classe | imageName | date | scene | Vert | Rouge | NIR | ... | evi2 |
|----------|--------|------------------|----------|---------|------|-------|-----|-----|------|
| vegbasse | 2 | 20101005-687-342 | 20101005 | 687-342 | 292 | 35 | 54 | ... | 118 |
| foret | 10 | 20101005-687-342 | 20101005 | 687-342 | 215 | 95 | 43 | ... | 88 |
| foret | 10 | 20101005-687-342 | 20101005 | 687-342 | 219 | 47 | 46 | .. | 95 |
| vegbasse | 2 | 20101005-687-342 | 20101005 | 687-342 | 356 | 37 | 61 | ... | 164 |
| foret | 10 | 20101005-687-342 | 20101005 | 687-342 | 205 | 47 | 43 | ... | 93 |
| nuage | 6 | 20101005-687-342 | 20101005 | 687-342 | 240 | 129 | 54 | ... | 150 |
| roche | 4 | 20110912-688-342 | 20110912 | 688-342 | 185 | 91 | 73 | ... | 198 |
| vegbasse | 2 | 20110912-688-342 | 20110912 | 688-342 | 397 | 53 | 69 | ... | 196 |
| ombre | 7 | 20110912-688-342 | 20110912 | 688-342 | 150 | 43 | 55 | ... | 68 |
| Pinot | 11 | 20110912-688-342 | 20110912 | 688-342 | 263 | 50 | 66 | ... | 112 |
| vegbasse | 2 | 20120915-688-342 | 20120915 | 688-342 | 317 | 40 | 57 | ... | 140 |
| foret | 10 | 20120915-688-342 | 20120915 | 688-342 | 194 | 48 | 45 | ... | 87 |

2.5.4. Traitement des images et des nuages

Pour chaque image SPOT-5 il a été appliqué le traitement TOA (Top Of Atmosphere) sous Monteverdi. Ensuite, les images ont été géoréférencées selon une mosaïque de la Guyane réalisée en 2008 par l'IGN car leur localisation était trop aléatoire (parfois plus de 100m d'écart avec les données de terrain).

Pour faciliter la classification et éviter les erreurs aux alentours des nuages et des fleuves, il a été choisi d'utiliser l'indice NDCI (Normalized Difference Cloud Index) afin d'avoir un masque efficace et unique pour chaque image (Marshak, 2000). Avec l'indice NDCI calculé automatiquement, il a été défini pour chaque image le seuil optimal (aux alentours de 0.40 ~ 0.45) jusqu'au delà duquel les nuages n'étaient plus filtrés. Pour faciliter ce traitement, un script Qgis a été développé et permet d'attribuer automatiquement selon un seuil défini par l'utilisateur la partie qui servira à la classification et celle qui sera masquée et d'enregistrer le masque.

2.5.5. Classification

Pour traiter les 55 images SPOT-5, un script a parcouru automatiquement toutes les scènes à classifier et a rattaché les images avec leurs parcelles d'entraînement. Plus de trois jours complets de calculs ont été nécessaires pour réaliser une première classification avec l'algorithme Random Forest, et environ une heure en utilisant le modèle de mélange gaussien.

Après ces premiers résultats, il a été nécessaire de vérifier les images une à une, en utilisant les taux de confusion et indices kappa mais aussi par validation visuelle.

A chaque image où des végétations particulières étaient détectées sur les bordures et donc étaient aussi visibles depuis d'autres scènes SPOT-5, ces peuplements étaient vérifiés en photo-interprétation et si cela correspondait bien, l'endroit détecté était alors saisi comme parcelle d'entrainement pour la scène où les données étaient manquantes. C'était surtout le cas pour les palmiers pinôts et les palmiers bâches, qui sont plus difficiles à repérer.

Certaines végétations particulières (comme les savanes roches, ou végétations basses) sont en fait des zones d'orpaillage ou d'agriculture. Le masque généré à partir de la donnée de l'OAM et d'AgriPag a donc servi pour enlever les végétations détectées dans ces lieux anthropisés.

2.5.6. Synthèse des données

Une fois chaque image classifiée, un modèle Qgis a permis de traiter en lot les données afin d'effectuer un tamis puis un reclassement des pixels pour classer en *no data* (sans donnée) les nuages et l'eau, et ce pour pouvoir choisir de combler seulement les parties manquantes d'une image.

Puis, pour chaque scène une synthèse a été générée en gardant les spécificités de chaque prise de vue. Comme certaines images n'étaient exploitables que pour la savane-roche, il a alors été spécifié, par des calculs d'images avec l'outil *mapcalc* de l'outil Grass (disponible dans Qgis), quelle classe garder ou non.

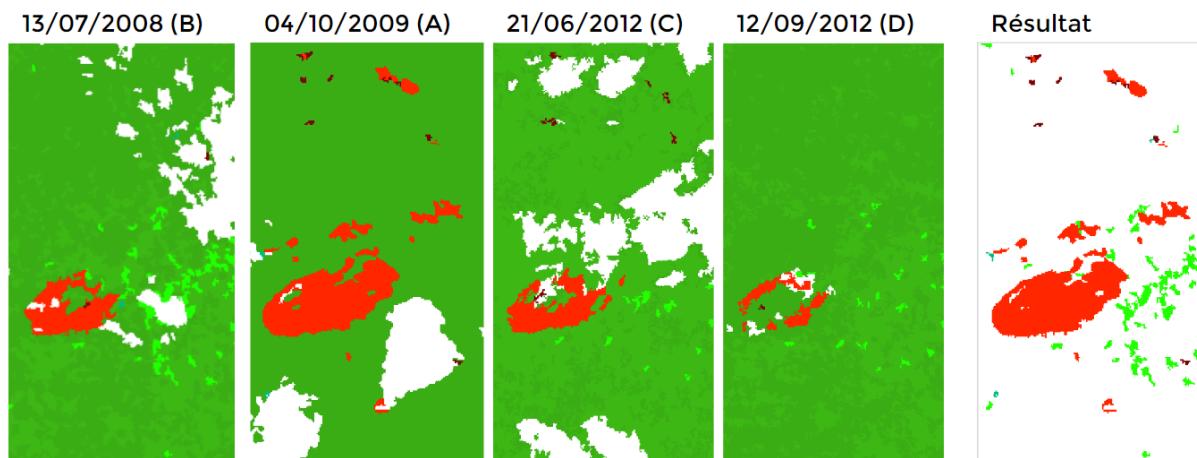


Figure 13 : Exemple de synthèse de 4 images d'une même scène

Sur l'extrait de la scène 690-344 ci-dessus, la formule *mapcalc* utilisée à consister à supprimer systématique les forêts, et pour l'image A de ne pas prendre en compte en plus la végétation basse (classe numéro 2) :

```

if((isnull(A) ||| A==2 ||| A==10 ||| A==12),
  if( (isnull(B) ||| B==10 ||| B==12 ||| B==1),
    if((isnull(C) ||| C==10 ||| C==12 ||| C==1),
      D,
      C),
    B),
  A)
  
```

Un autre exemple plus simple avec deux images : quand les deux images étaient à garder, la règle sous mapcalc consistait à rendre les classes 10 et 12 (forêt et forêt déstructurée) transparente :

```
if((isnull(A) ||| A==10 ||| A==12),B,A)
```

L'usage des `|||` pour signifier OU dans Grass s'explique car il est le seul à supporter les valeurs nulles. Dans cet autre exemple, la classe 3 et 11 (Palmier bâche et pinôt) sont à garder uniquement sur l'image A, et il ne faut pas garder la classe 1 (Parinari) de l'image B :

```
if((isnull(A) ||| B==3 ||| B==11),if(B==1,A,B),A)
```

La plupart des règles de concaténation ont seulement pour but de garder le maximum des classifications des images (comme montré dans le premier exemple).

Une fois qu'une image a été obtenue pour chaque scène, une fusion a été faite en utilisant Qgis. Pour que cette méthode se fasse parfaitement, les images finales de chaque scène ne contiennent plus les données de forêt, elles sont classées en *no data* (sans valeur), afin de pouvoir superposer les différentes scènes. Puis un découpage des données avec le masque du parc et celui des données minières et agricoles a été effectué. Afin d'optimiser la taille du fichier masque, seules les données à retirer avaient été classées, et le reste de l'image était classé en *no data*. Si le pixel du masque est de valeur nulle, alors on le remplace par le pixel classé, sinon il est classé comme nul :

```
if(isnull(B), A, null())
```

2.6. Adaptation de la méthodologie à Qgis

Lors des premières classifications il avait été utilisé le langage R avec l'algorithme Random Forest. Cependant les temps d'entraînement et de classification étaient importants (au minimum deux heures sur l'ordinateur de travail), et la méthodologie difficilement intégrable directement dans un plugin Qgis.

Pour cela le Gaussian Mixture Model (GMM), développé par Mathieu Fauvel et implémenté par mes soins dans Qgis avec Antoine Lomellini lors de l'atelier de fin d'année du master 2 SIGMA, était une solution adaptée. Cependant, les autres algorithmes, comme Random Forest qui a largement fait ses preuves sur la classification des forêts, n'étaient pas exploitables avec Windows suite à un bug non identifié au début de ce stage. L'objectif était alors de créer un nouveau plugin plus généraliste, permettant de faire une simple classification à partir de différents algorithmes, et compatible entièrement avec Windows.

2.6.1. Création du plugin de classification dzetsaka

Comme le but de ce travail est d'être répété ultérieurement par le personnel du parc et aussi de servir à l'ensemble de la communauté utilisant Qgis et la télédétection, il fallait absolument rendre le plugin le plus ergonomique possible, et compatible avec Random Forest sous Windows. Le plugin a été créé à partir de Qt Designer pour son interface graphique, et le reste a été programmé en python à partir du logiciel Spyder.

Le nom *dzetsaka* représente en Teko (une langue amérindienne parlée sur le fleuve Oyapock, à l'est de la Guyane) les différents objets utilisés pour voir le monde, comme un télescope, un appareil photo, ou encore une loupe. C'est d'ailleurs ce dernier outil qui est l'icône du plugin, et qui symbolise ici la télédétection (avec les pixels de fleuve en bleu et le reste en vert).

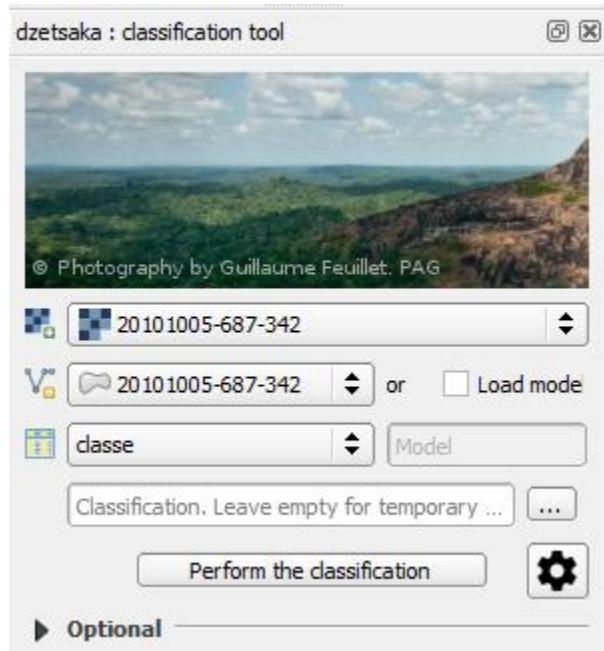


Figure 14 : Logo de dzetsaka

Figure 15 Fenêtre principale de dzetsaka

Ce travail s'inscrit dans la lignée du plugin Historical Map, créé durant l'atelier de fin d'année du master Sigma, et qui permet d'extraire les forêts des cartes anciennes de l'État-Major (Herrault et al. 2015). En effet, pour réaliser les différents tests de classification au début du stage, il a été utilisé une version modifiée d'Historical Map afin de bénéficier d'une interface graphique pratique. Suite à cela, il paraissait indispensable de publier dans Qgis un plugin plus généraliste de classification, qui serait également destiné au personnel du Parc amazonien pour poursuivre le travail réalisé sur la cartographie des végétations particulières.

2.6.1.1. Adaptation de l'algorithme Random Forest à Windows

Parmi les modifications à effectuer, la plus importante était d'implémenter l'algorithme Random Forest sous Windows. En effet, le classifieur fonctionne parfaitement bien sous linux mais un bug empêchait son utilisation avec Windows dans Qgis. A noter que le script python fonctionne très bien sans Qgis.

Après plusieurs essais et tentatives de débogage, il s'avère que lorsque le script est exécuté depuis Qgis, il est impossible d'utiliser plusieurs cœurs du processeur. Pour palier à ce bug de Qgis, il a simplement été indiqué que l'usage du multicœur est autorisé depuis linux, sinon le script force l'utilisation d'un seul cœur :

```
# AS Qgis in Windows doesn't manage multiprocessing, force to use 1 thread
for not linux system
if os.name == 'posix':
n_jobs=-1
else:
```

```
n_jobs=1
```

Par la suite la variable *n_jobs* sera insérée comme paramètre pour la classification.

2.6.1.2. Modification de l'algorithme GMM

Initialement, cet algorithme développé par Mathieu Fauvel ne gérait pas les classes non continues, il fallait systématiquement avoir des classes qui se suivaient (1 : forêt, 2 : palmier bâche, 3 : palmier pinôt...). Or comme le présent travail nécessite de garder la même nomenclature des végétations et que sur certaines scènes par exemple les palmiers bâches n'étaient pas présents, il manquait alors une valeur (1 : forêt, 3 : palmier pinôt...).

Il a donc fallu adapter le script afin de pouvoir gérer une classification sur n'importe quelle image du Parc amazonien de Guyane sans avoir à classer systématiquement en numérotation continue.

Auparavant le script se basait sur la valeur maximale des classes (ex : 3 pour palmier pinôt) pour compter le nombre de classes présentes :

```
C = int(y.max(0))
```

Comme cela ne convenait plus, il a été compté le nombre unique de classes (*y* correspondant au numéro de chaque classe) puis une nouvelle variable (*self.classnum*) a été créée pour les sauvegarder :

```
C = sp.unique(y).shape[0]

# ajout d'une nouvelle variable pour sauvegarder les numéros de classe

self.classnum = sp.empty(C).astype('uint8')

for c,cR in enumerate(sp.unique(y)):
    self.classnum[c] = cR      # Sauvegarde le numéro de la classe du pixel
```

Lors de la classification, le script assigne la classe de chaque pixel :

```
# Assigné la valeur de la classe à chaque pixel
yp = self.classnum[sp.argmin(K,1)]
```

2.6.1.3. Ergonomie et intuitivité

Afin de rendre le script le plus simple, ergonomique et intuitif possible, plusieurs adaptations ont été faites pour faciliter son utilisation.

Tout d'abord, contrairement à Historical Map, la fenêtre du plugin dzetsaka est directement intégrée à l'espace de travail de Qgis, par défaut sous la liste des couches (raster et vecteur). Cette fenêtre a été optimisée pour prendre le moins de place possible et présenter seulement les éléments les plus essentiels. Cependant, il fallait prendre en compte la possibilité de sauvegarder un modèle d'entraînement, d'indiquer la présence d'un masque, d'enregistrer la matrice de confusion, etc...

Pour cela toutes les données optionnelles ont été mises dans le menu “Optional” qui est replié par défaut et ne prend donc pas de place. Puis pour les paramètres constants, la roue dentée permet de choisir l'algorithme de classification, mais aussi d'indiquer le suffixe par

défaut du masque, ainsi l'utilisateur n'aura pas à le saisir à chaque fois, et si par défaut son masque a le même nom que son fichier et termine par “*_mask*” alors dzetsaka le détectera automatiquement. Afin de garder en mémoire les différents paramètres la bibliothèque configParser a été utilisée et le fichier *config.txt* contient tous les paramètres des utilisateurs (suffixe du masque, algorithme de classification...).

Ensuite, il n'est plus question de plusieurs étapes de traitement (entraînement puis classification), tout cela se fait discrètement sans que l'utilisateur ait à intervenir. De plus pour effectuer différents tests, l'utilisateur n'a plus besoin de saisir le fichier à enregistrer. Par défaut le script enregistre les résultats dans un fichier temporaire si l'utilisateur ne saisit aucun fichier de destination.

Grâce au script de Mathieu Fauvel qui gère l'entraînement et la classification à partir de la librairie Scikit-learn, le plugin dzetsaka a été le premier à gérer sous Windows l'algorithme très réputé Random Forest, mais aussi le Support Vecteur Machine (SVM).

Le plugin est disponible dans le dépôt officiel de Qgis et a déjà atteint les 2 000 téléchargements au 18 août 2016.

2.6.2. Création des scripts Qgis

Comme évoqué précédemment pour le calcul des indices, les scripts Qgis sont des fichiers python (.py) exécutables depuis la boîte à outil traitement du logiciel SIG. Afin que des variables soient disponibles et donc saisissables par l'utilisateur lors de l'exécution du script, il faut au début du script utiliser les ‘##’ suivis du nom de la variable, et de son type (raster, number, folder, vector...) :

```
##inRaster = raster
##outFolder = folder
```

L'intérêt de cette méthode est de fournir une interface graphique très utile et très rapide à mettre en place. Chose non négligeable, la boîte à outil traitement de Qgis permet d'exécuter en lot les scripts : on peut lui dire en deux clics de traiter tous les fichiers d'un dossier par exemple.

Ci-dessous un exemple pour appeler les variables afin de créer un masque à partir du seuil choisi par l'utilisateur sur l'indice NDCI :

```
##Raster=group
##Mask from NDCI=name
##raster=raster
##thresold=number 0.45
##mask=output raster

outputs_GDALOGRRASTERCALCULATOR_1=processing.runalg('gdalogr:rastercalculator',
raster,'1',None,'1',None,'1',None,'1',None,'1',None,'1','numpy.where(A<='+str(thresold)+',0,1)',-1,0,None,mask)
```

Ce script (*maskFromNDCI.py*) a été créé en exportant le modèle Qgis en python puis en ajoutant la variable de seuil (thresold) et en l'intégrant dans la formule dans la calculatrice raster “*numpy.where(A<='+str(thresold)+',0,1)*”.

3. Résultats et discussions

A chaque classification la matrice de confusion était sauvegardée dans un dossier séparé. Puis pour calculer la moyenne des indices kappa et des taux d'accord global (OA), un script R (*BatchOaKappaFromConfusionMatrix.R*) a été développé pour générer un tableau contenant l'ensemble des données (nom de l'image, OA, Kappa) de chaque image.

3.1. Résultats de classification

Une cartographie à l'échelle du Parc amazonien de Guyane des végétations particulières a été produite. Sur l'ensemble des images le taux moyen d'accord global est de 99% et l'indice kappa de 0.97 en utilisant la moitié des pixels pour l'entraînement et l'autre pour la validation et avec l'algorithme Random Forest. Cela montre une très grande qualité de classification et très peu d'erreurs de manière générale. Ces très bons chiffres s'expliquent en partie par la présence dans certaines images de la végétation basse uniquement, un type de végétation très facile à détecter, et par le nombre de pixels de forêt plus important que dans les autres classes.

Cependant, sur certaines scènes, notamment celles avec la totalité des végétations particulières à cartographier, les statistiques peuvent descendre jusqu'à 0.87 (kappa) ce qui montre que des confusions ont eu lieu entre la forêt et les végétations particulières.

3.1.1. Zone Sud-ouest

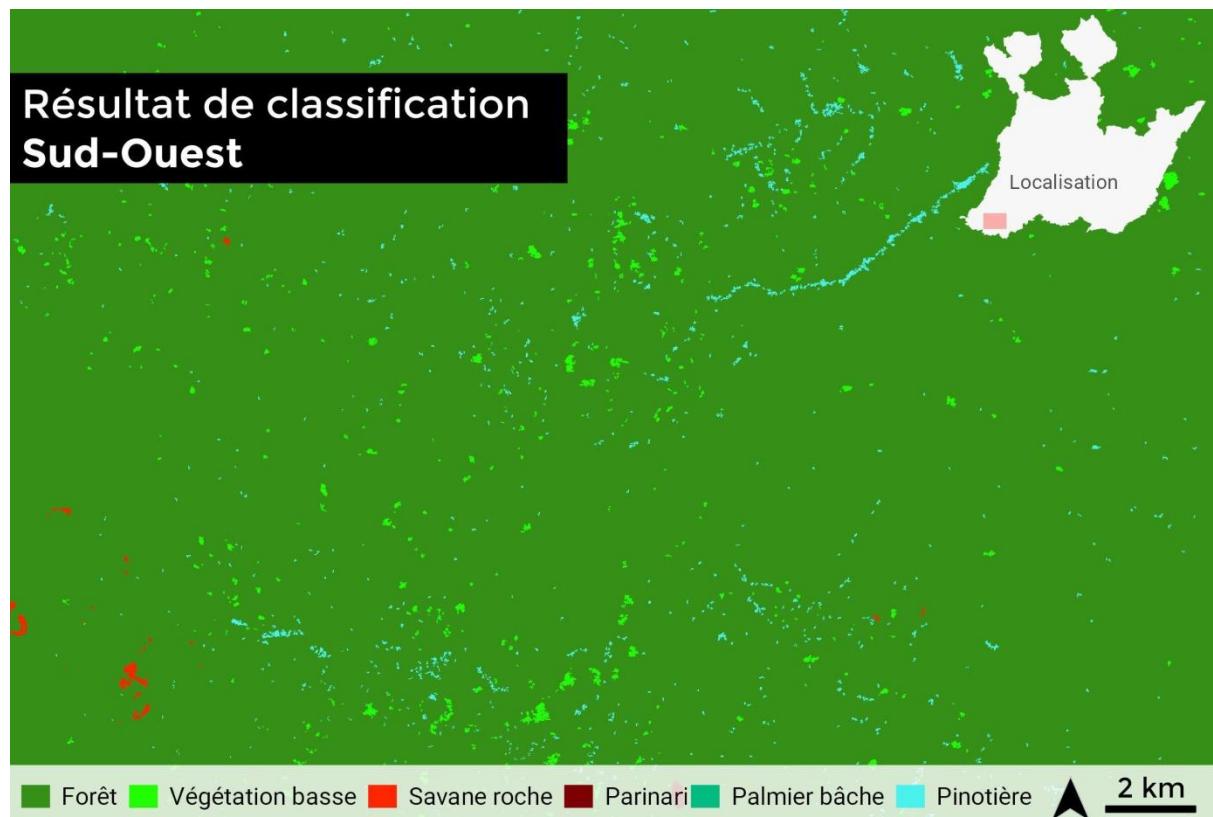


Figure 16 : Exemple de résultat de classification au Sud-ouest

3.1.1.1. Indice Kappa et taux d'accord global

Les taux d'accord global et les indices kappa permettent de vérifier les différences entre les parcelles qui ont servi d'entraînement et le résultat final.

Dans cette classification on observe des pinotières le long des cours d'eau. Il n'y avait qu'une parcelle d'entraînement dans le sud-ouest guyanais. Selon le Guide des palmiers de Guyane (Granville et Gayot, 2014), l'habitat des palmiers pinôts se trouve en forêt et plus précisément sur "*le bord de crique et sol hydromorphe*", ce qui correspond bien à la télédétection trouvée. Afin de vérifier ces données, des photos avaient été prises lors d'un survol hélicoptère sur une mission de 2006. Cela a permis de bien attester, quand une photographie prise par un agent depuis l'hélicoptère existait, de la présence réelle de ce palmier. A chaque point photo qui se trouvait au dessus d'une pinoterie détectée, la photo attestait de la présence réelle de cette dernière.

Cependant, parmi les zones d'entraînement classées comme pinotières, 90,96% des pixels qui étaient donnés de cette classe ont été classifiés comme tel. Ainsi 117 pixels de forêt se sont retrouvés par erreur comme palmiers pinôts lors de la classification, et 495 pixels donnés comme palmiers pinôts à l'entraînement ont été classées comme la forêt.

Ces erreurs de classification ont pu être corrigées à l'aide d'un tamis afin d'homogénéiser les résultats. La forêt qui étaient classée en pinôt ou inversement consistait en pixels souvent isolés ou inférieurs à une taille de 10 pixels.

Tableau 5 : Matrice de confusion de la scène 688-345

| | | CLASSIFICATION | | | | |
|---------------------------------|-----------------------|------------------|--------------|--------|---------------|-----------------------|
| | | Végétation basse | Savane roche | Forêt | Palmier pinôt | Précision réalisateur |
| T E R R A I N | Végétation basse | 444 | 0 | 1 | 0 | 99,78% |
| | Savane roche | 0 | 2067 | 0 | 0 | 100% |
| | Forêt | 1 | 0 | 138420 | 495 | 99,64% |
| | Palmier pinôt | 0 | 0 | 117 | 1177 | 90,96% |
| | Précision utilisateur | 99,33% | 100% | 99,91% | 70,39% | |

Moyenne des statistiques des 2 images classifiées sur la scène 688-345. Kappa = 0.97

L'indice Kappa, qui reflète la qualité globale de la classification, permet d'enlever l'effet du hasard. Ainsi, une valeur kappa de 0,90 indique que 90% des erreurs sont évitées par rapport à une méthode de classification travaillant complètement aléatoirement (Girard et Girard, 1999).

A noter que le nombre important de pixels de forêt sur l'ensemble des classifications a tendance à améliorer les taux de réussite (accord global et indice kappa). Cela s'explique car il y a très peu de parcelles d'entraînement des végétations particulières et la variété de la forêt étant importante, il faut donc bien faire attention à recenser le plus possible la large variabilité spectrale de cette classe.

3.1.1.2. Validation par photographie aérienne

Plusieurs campagnes aériennes ont été menées par le Parc amazonien de Guyane en collaboration avec l'IRD. Ces données sont récoltées à partir d'un hélicoptère où un utilisateur photographie les formations végétales particulières. Ainsi, le survol avec l'IRD en 2008 qui avait pour but d'observer les savanes-roche a aussi été utile : de nombreuses photos de pinotières ou de parinaris ont été réalisées lors de cette expédition.

Un prochain survol devrait avoir lieu vers octobre 2016 dans le cadre d'une expédition menée en partenariat avec l'IRD et le parc. Il permettra de vérifier certaines des observations faites avec la présente cartographie des végétations particulières. Cela permettra aussi d'affiner dans le futur les parcelles d'entraînement.

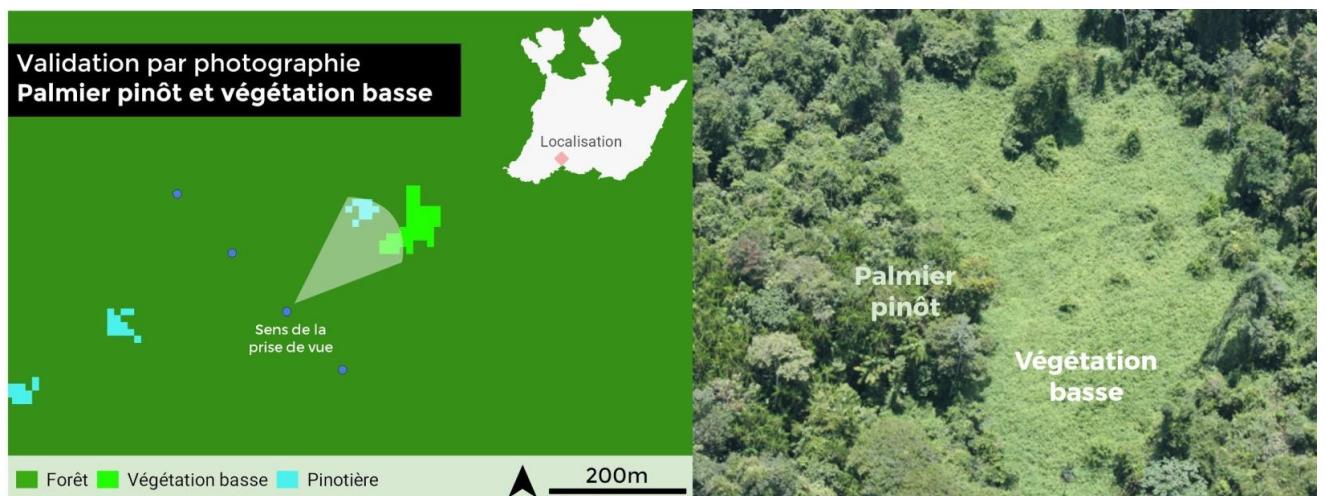


Figure 17 : Validation par photographie héliportée

3.1.2. Zone de la pénéplaine de la Waki

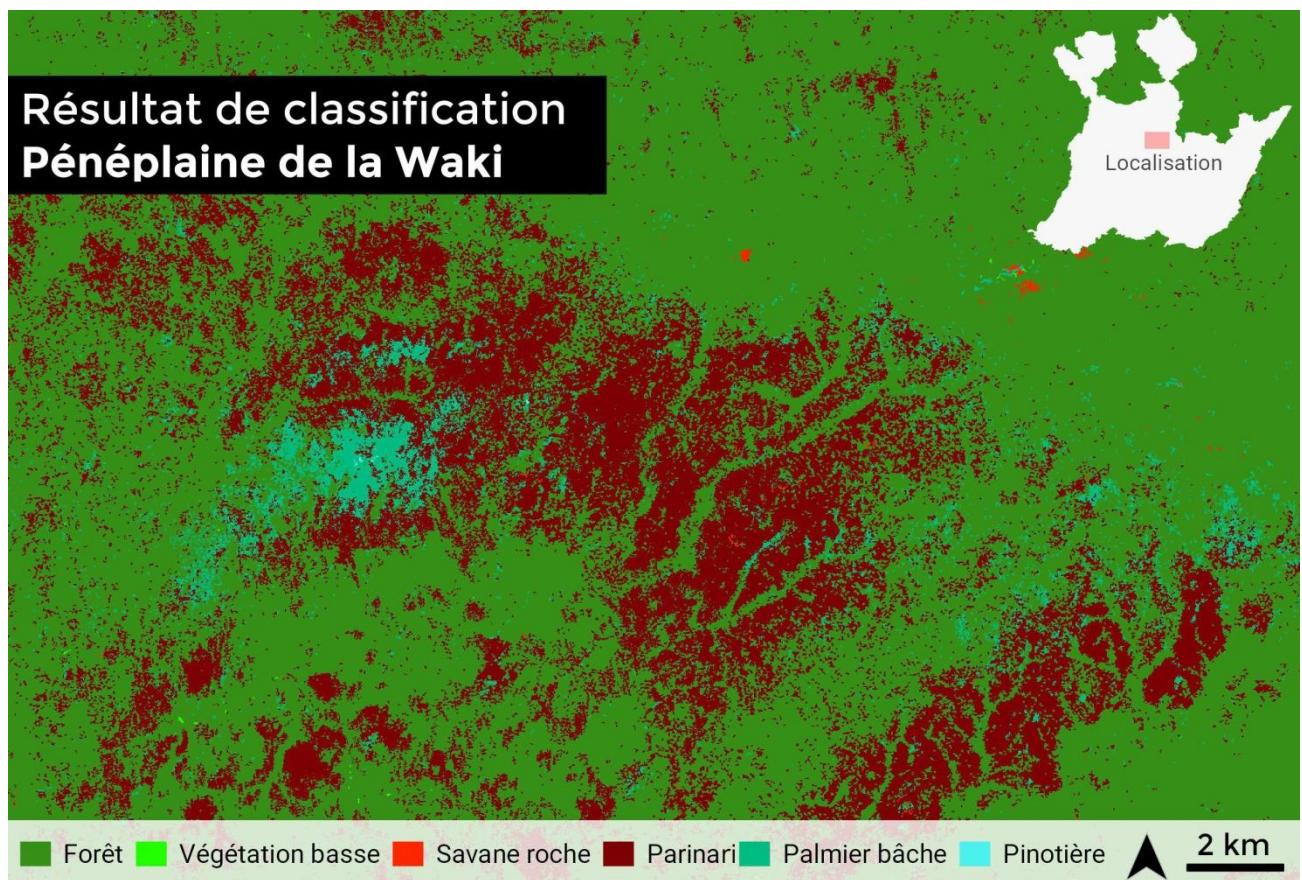


Figure 18 : Exemple de résultat de classification dans la pénéplaine de la Waki

3.1.2.1. Indice kappa et taux d'accord

Tableau 6 : Matrice de confusion de la scène 690-344

CLASSIFICATION

| | Parinari | Palmier bâche | Végétation basse | Savane roche | Forêt | Précision réalisateur | |
|---------------------------------|-----------------------|---------------|------------------|--------------|--------|-----------------------|--------|
| T E R R A I N | Parinari | 3588 | 3 | 24 | 14 | 16 | 98,44% |
| | Palmier bâche | 1 | 518 | 0 | 0 | 1 | 99,62% |
| | Végétation basse | 21 | 11 | 957 | 0 | 6 | 96,18% |
| | Savane roche | 0 | 0 | 0 | 1166 | 0 | 100% |
| | Forêt | 14 | 27 | 7 | 1 | 39078 | 99,87% |
| | Précision utilisateur | 99,01% | 92,67% | 96,86% | 98,73% | 99,94% | |

Moyenne des statistiques des 3 images classifiées sur la scène 690-344. Kappa = 0.99

On remarque que le palmier bâche est la classe où le taux d'erreurs est le plus important. En précision utilisateur il y a 92,67% de réussite, le plus faible après la végétation basse à 96,86%. La classe du palmier bâche se confond tout d'abord avec la forêt, puis la végétation basse. Cela signifie que des pixels ont été assignés comme palmier bâche alors qu'ils n'en étaient pas. Par contre ce qui était catégorisé comme palmier bâche a bien été reconnu comme tel, à l'exception de deux pixels.

3.1.2.2. Courbe ROC

Pour vérifier la présence des parinaris sur la pénéplaine de la Waki, un test statistique a été calculé entre la cartographie et les données terrain du CIRAD qui contiennent le pourcentage de parinaris observés sur une parcelle de 100m². La courbe ROC dessinée représente alors l'optimum de lien entre les données de vérification et la cartographie finale.

Pour générer ce graphique sous R, une carte de proximité a été créée sous Grass (*Distance Raster*) à partir de la classe des parinaris et indique la distance la plus proche de chaque pixel n'étant pas de ce type de végétation. Le résultat nous montre qu'à une distance de 29,5m du point de référence, on peut considérer que la cartographie représente très bien la réalité du terrain (88,2%).

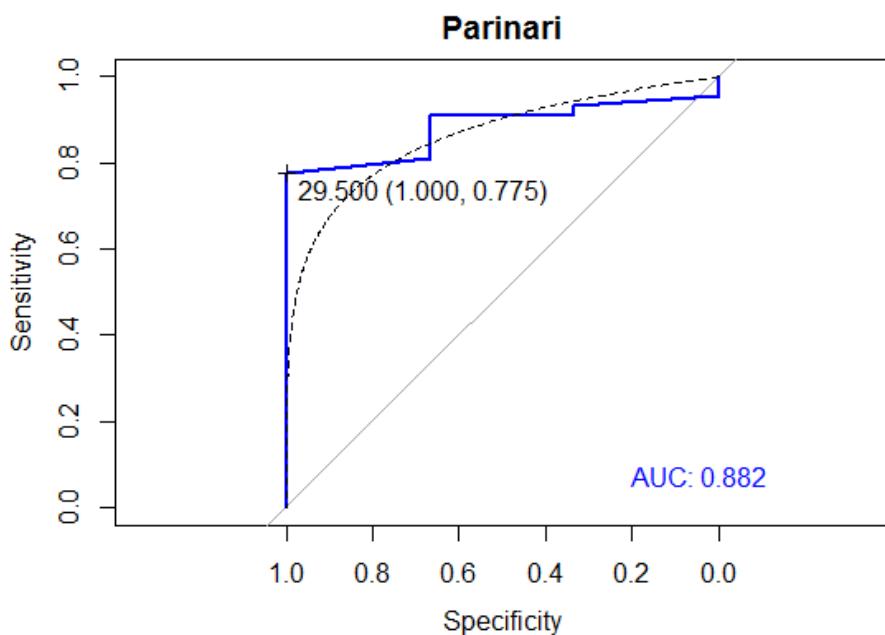


Figure 19 : Courbe ROC des Parinaris

Les données du CIRAD contiennent aussi des informations sur le pourcentage de présence d'une autre essence forestière : le tachigali. Le résultat d'aire sous la courbe (AUC = Area Under the Curve) pour le tachigali est de 0,652, donc sans lien significatif avec les données fournies. Ce n'était pas l'essence à cartographier mais les tachigalis sont aussi être présents dans la forêt de la pénéplaine de Waki. Cela confirme que la cartographie représente bien les parinaris sur les points sélectionnés.

3.1.2.3. Validation par photographie aérienne

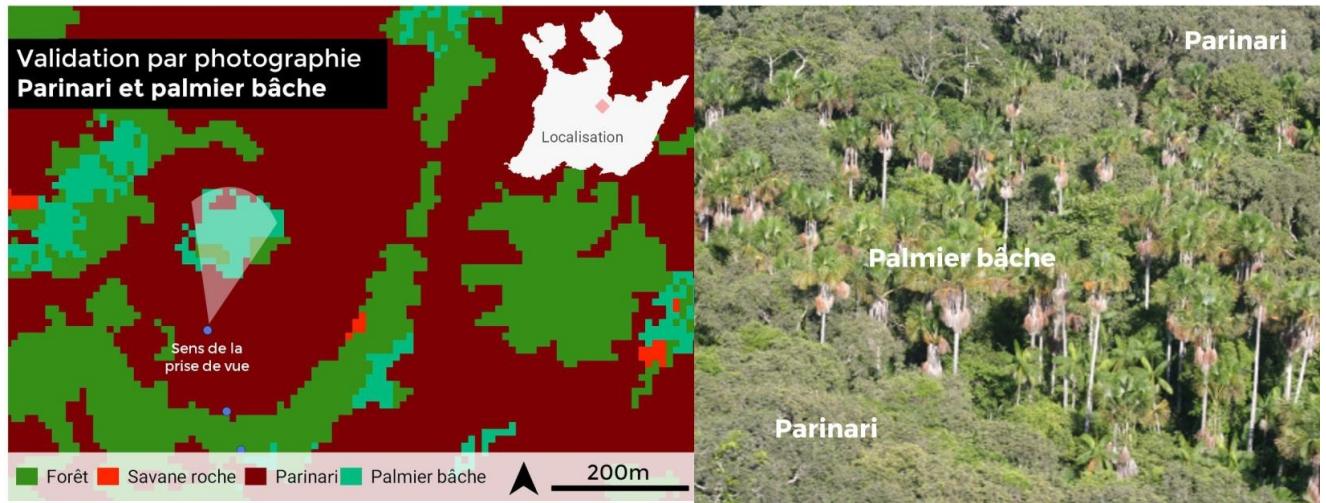


Figure 20 : Validation par photographie héliportée

Grace aux photos aériennes, la majorité des résultats a pu être validée comme c'est le cas sur cette photo où l'on distingue bien les palmiers bâches entre les parinaris.

3.2. Apport des indices

Des travaux ont montré la corruption des données spectrales et l'importance d'utiliser des indices de végétation comme le NDVI ou le EVI. Samanta et al. (2012) observent que plus de 60% des données spectrales sont corrompues à cause des aérosols. De leur côté, Morton et al (2014) ont montré que le verdissement de la canopée en saison sèche était un artefact, mais sur des données sans traitement atmosphérique (TOA). Dans l'article “*BRDF-corrected vegetation indices confirm seasonal pattern in greening of French Guiana's forests*”, Emil Cherrington et al (2016) a conclu à l'aide de données traitées en TOA que le verdissement en Guyane était bien réel en comparant trois forêts tropicales aux mêmes latitudes. L'apport des indices dans la classification des végétations particulières est donc important à évaluer.

Les différents tests ont été réalisés sur la scène la scène 690-344.

Afin de vérifier leur utilité pour le type de végétation étudié et dans la forêt amazonienne, des classifications ont été effectuées en gardant 50% des pixels d'entraînement pour la classification, le reste pour la validation. Les deux algorithmes, GMM et Random Forest ont été comparés.

Tableau 7 : Accord Global (OA) et indice Kappa par algorithme en combinant les bandes spectrales (SB) avec les indices, Image du 13/07/2008

| | SB | | SB + NDVI | | SB + NDWI | | SB + EVI2 | |
|-----|-------|-------|-----------|-------|-----------|-------|-----------|-------|
| | OA(%) | Kappa | OA(%) | Kappa | OA(%) | Kappa | OA(%) | Kappa |
| GMM | 81 | 0.67 | 81 | 0.67 | 80 | 0.66 | 77 | 0.62 |
| RF | 93 | 0.87 | 91 | 0.85 | 91 | 0.85 | 91 | 0.85 |

En étudiant les données du 13 juillet 2008 on observe qu'aucun indice, et ce peu importe l'algorithme, n'a permis d'améliorer la classification. Au contraire leur ajout fait même diminuer l'indice Kappa pour l'ensemble des classifications, à l'exception du NDVI avec GMM où aucun changement n'a été repéré.

Tableau 8 : Accord Global (OA) et indice Kappa par algorithme en combinant les bandes spectrales (SB) avec les indices, Image du 21/06/2012

| | SB | | SB + NDVI | | SB + NDWI | | SB + EVI2 | |
|-----|-------|-------|-----------|-------|-----------|-------|-----------|-------|
| | OA(%) | Kappa | OA(%) | Kappa | OA(%) | Kappa | OA(%) | Kappa |
| GMM | 88 | 0.84 | 78 | 0.71 | 88 | 0.83 | 87 | 0.83 |
| RF | 96 | 0.95 | 93 | 0.89 | 87 | 0.78 | 87 | 0.78 |

L'analyse des statistiques de l'image du 21 juin 2012 confirme les résultats précédemment obtenus. Aucun ajout d'indice n'a permis d'améliorer les résultats, que ce soit avec l'algorithme GMM ou Random Forest.

Tableau 9 : Accord Global (OA) et indice Kappa par algorithme en combinant les bandes spectrales (SB) avec les indices, Image du 12/09/2012

| | SB | | SB + NDVI | | SB + NDWI | | SB + EVI2 | |
|-----|-------|-------|-----------|-------|-----------|-------|-----------|-------|
| | OA(%) | Kappa | OA(%) | Kappa | OA(%) | Kappa | OA(%) | Kappa |
| GMM | 87 | 0.79 | 88 | 0.81 | 87 | 0.79 | 89 | 0.82 |
| RF | 96 | 0.94 | 96 | 0.94 | 96 | 0.94 | 96 | 0.94 |

Comme constaté précédemment, l'algorithme Random Forest obtient de meilleurs résultats (kappa 0.87 à 0.95) que GMM (kappa 0.67 à 0.84). L'ajout d'indices avec l'algorithme Random Forest est au mieux négligeable (même taux de confusion et même indice kappa), voire diminue sensiblement la qualité de la classification (passant de 0.95 à 0.78 pour l'indice EVI2 et à 0.89 pour le NDVI). Les résultats ne sont guères meilleurs avec l'algorithme GMM, même s'il est constaté une légère amélioration avec l'indice EVI2 sur la dernière image.

3.3. Étude de l'effet biogéographique

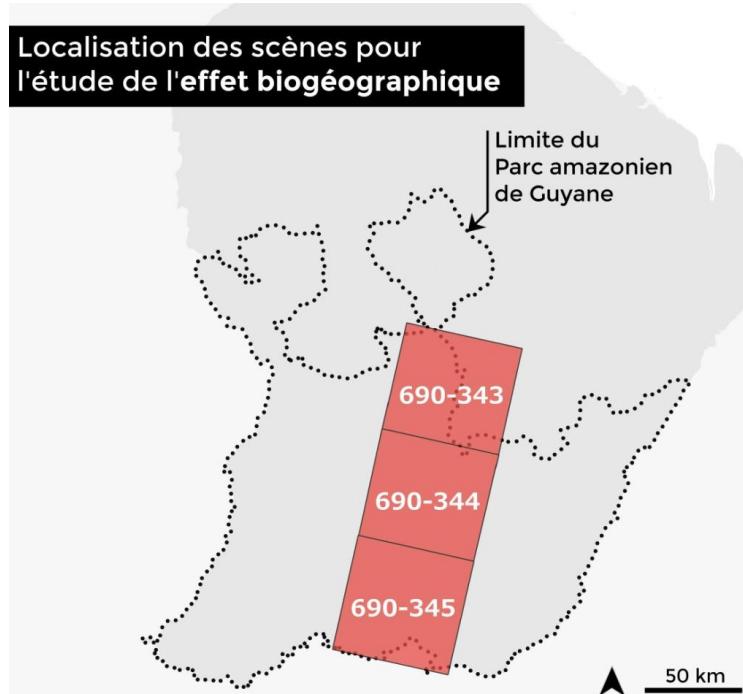


Figure 21 : Les scènes utilisées pour étudier l'effet biogéographique

En utilisant le langage de programmation R, des analyses ont permis d'observer les variabilités spectrales des végétations suivant leur emplacement géographique. Afin de bien isoler les différences, trois scènes ont été étudiées à la date unique du 12 septembre 2012. L'ensemble des végétations étudiées montre un verdissement de la canopée vers le sud. La différence est cependant assez faible et il faudrait d'autres données à d'autres dates pour confirmer ou infirmer cette hypothèse.

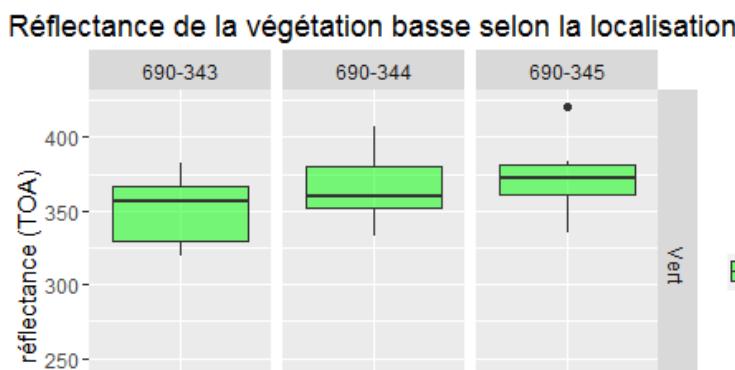


Figure 22 : Évolution de la bande verte selon la latitude pour la végétation basse (SPOT-5)

En comparant avec les images SPOT-6 prises le 12/09/2015 avec l'image SPOT-5 du 12/09/2012 soit trois ans jour pour jour plus tôt, on observe que le phénomène est inverse. Plus la végétation se trouve au Sud (Trois-Sauts), moins sa réflectance dans les verts est élevée.

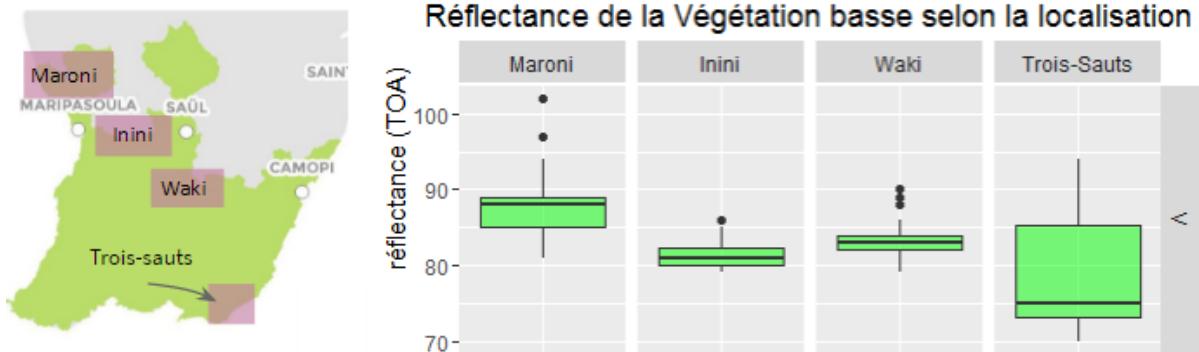


Figure 23 : Évolution de la bande verte selon la latitude pour la végétation basse (SPOT-6)

D'autres données sont nécessaires afin de comprendre ce phénomène. Est-ce un retard de la saison sèche, des végétations basses avec des essences différentes ?

3.4. Étude de la phénologie

La phénologie est l'évolution de la végétation selon les saisons. Cela peut correspondre à la perte de feuilles, la sécheresse, au débourrement... Les essences forestières peuvent donc selon les saisons avoir une évolution différente qui permet de mieux les différencier (Sheeren et al., 2015)

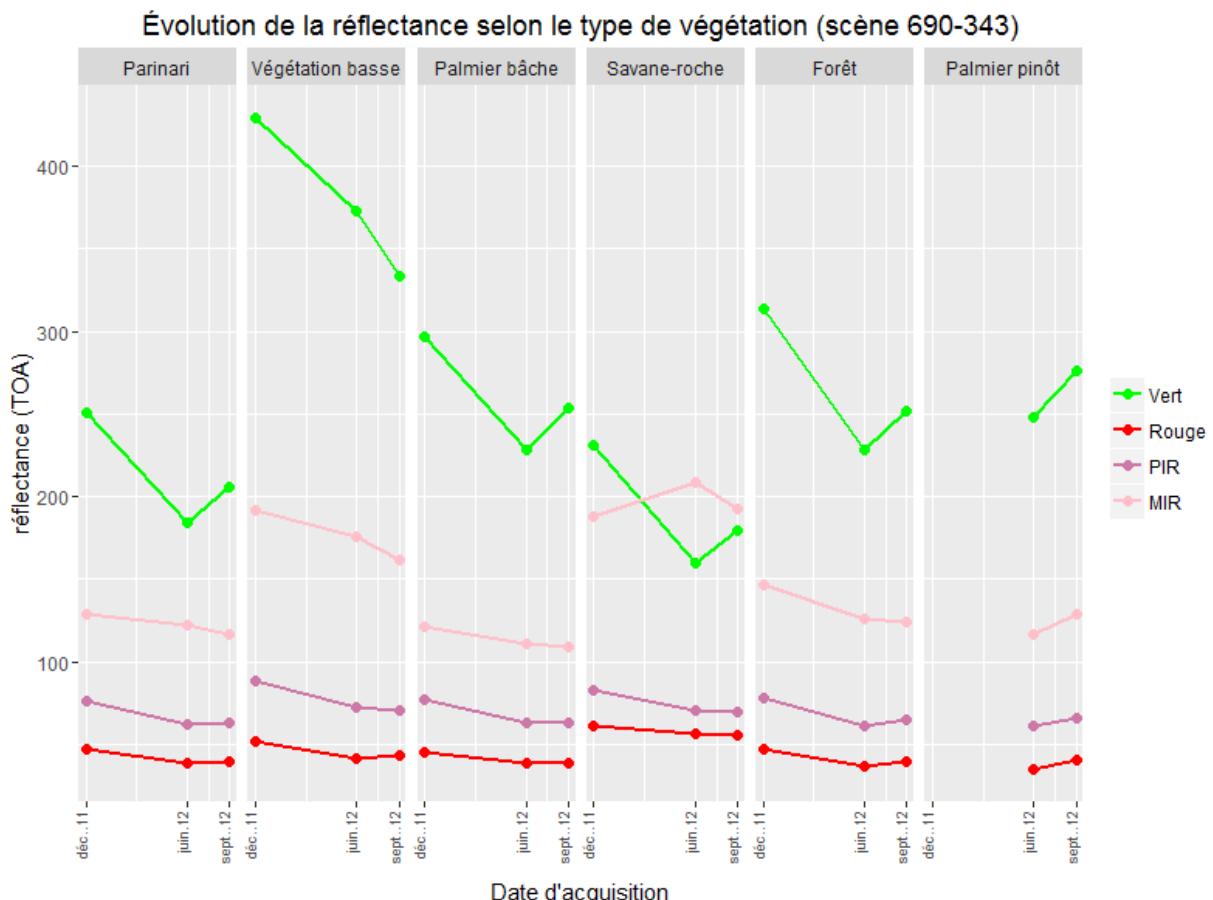


Figure 24 : Phénologie des végétations avec SPOT-5

En comparant l'évolution des végétations entre décembre 2011 et juin et septembre 2012, on observe que la végétation basse est la seule à avoir une baisse du verdissement entre juin et septembre. La savane-roche qui s'observe avec les dalles rocheuses apparentes est la seule à avoir le moyen infrarouge aussi fort, voire plus fort, que le vert. Les palmiers bâches et pinôts qui peuvent se confondre ont une phénologie différente. En effet entre juin et septembre, le palmier bâche a tendance à baisser, voire stagner, dans le moyen infrarouge.

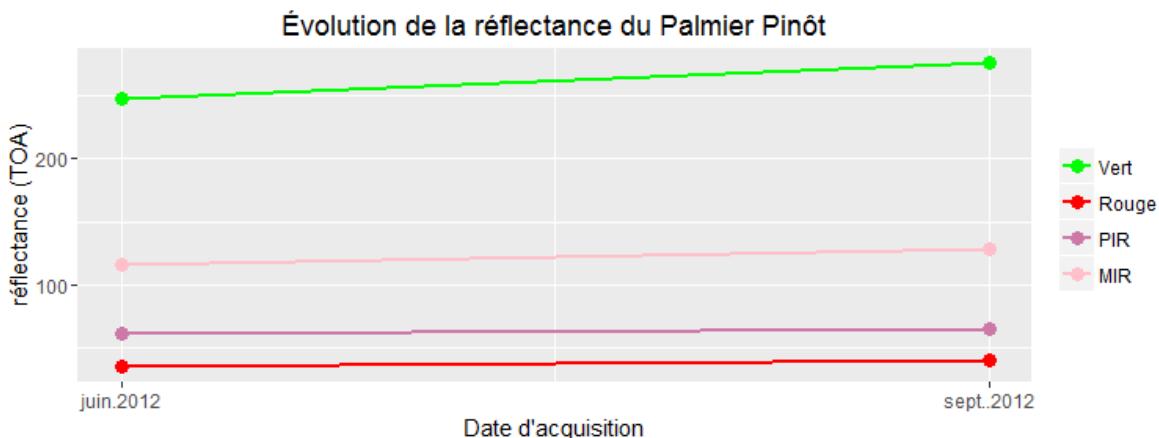


Figure 25 : Évolution de la réflectance du Palmier Pinôt

Le palmier pinôt est peu présent sur cette zone, il n'y a donc que deux dates pour suivre son évolution. On note toutefois un verdissement en saison sèche (+69 d'intensité) ainsi qu'une augmentation des infrarouges par rapport à juin (+12 d'intensité en MIR et +14 en PIR). Dans le moyen infrarouge le maximum est de 116 pour juin contre 128 vers la fin de l'année.

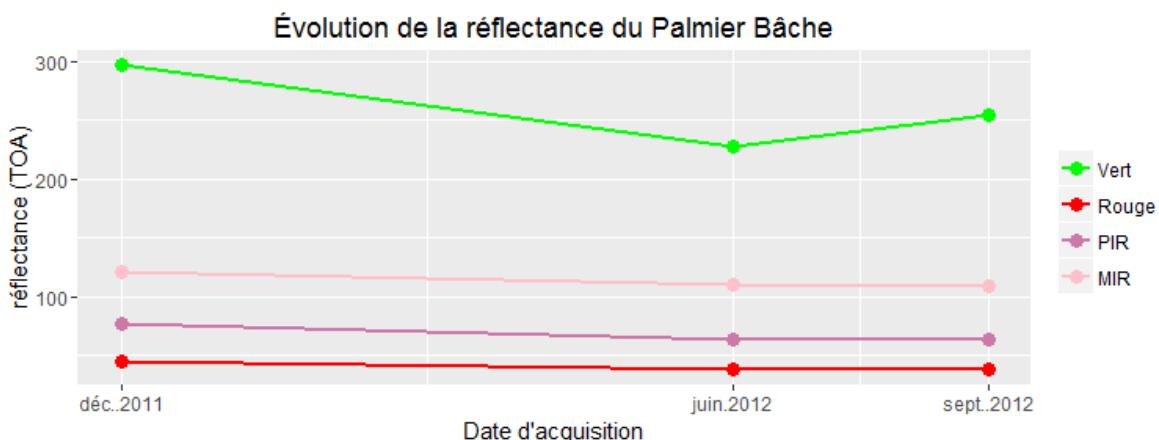


Figure 26 : Évolution de la réflectance du Palmier Bâche

Le palmier bâche est assez constant dans le rouge et les infrarouges à la différence du vert qui fluctue selon la saison avec un minimum de 228 en moyenne en juin, contre un maximum de 297 en moyenne en décembre. Dans le moyen infrarouge le maximum est de 109 pour juin contre 121 vers la fin de l'année.

Combiner des images à plusieurs dates afin de mieux séparer spectralement les végétations permettrait de mieux prendre en compte la phénologie des végétations et ainsi de bien les discerner grâce à leur variation spécifique dans le temps (Somers 2014).

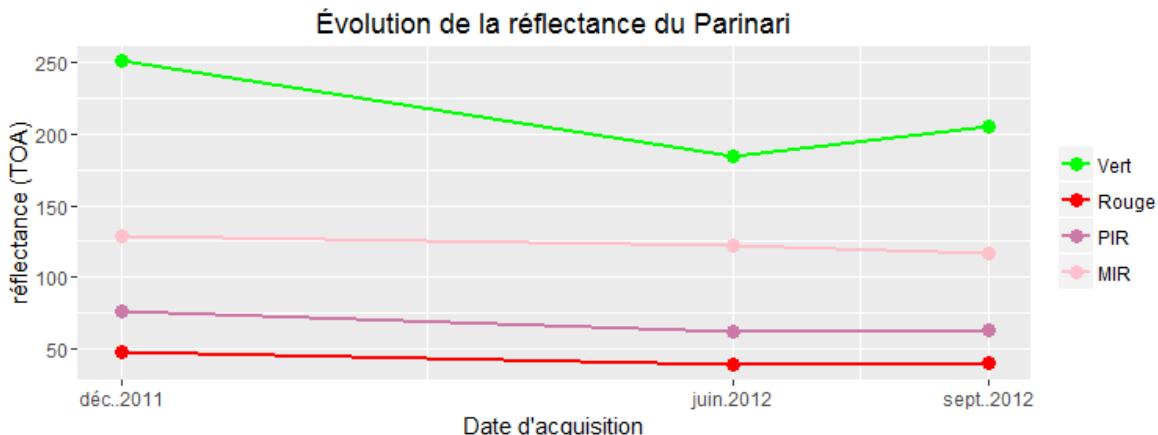


Figure 27 : Évolution de la réflectance du Parinari

Les parinaris sont des arbres présents sur un sol de type djougoun pété, c'est-à-dire de "micro-dépressions d'une mètre de diamètre sur 30 à 50 cm de profondeur, parfois remplies d'eau en saison des pluies" selon l'ONF. La canopée (parinari et tachigali) y est très ouverte et elle se repère donc en photo-interprétation. Cependant, on peut se demander quelle est l'influence de la saison dans la signature spectrale de ce type de végétation. On observe que leur signature spectrale dans la bande verte est très élevée en fin d'année (251 en moyenne), pour un minimum de 185 en fin de saison des pluies (juin). Dans le moyen infrarouge le maximum est de 116 pour septembre 2012, et 122 pour juin 2012. Contrairement au palmier bâche et pinot, le parinari a donc un moyen infrarouge en baisse en juin et septembre, ce qui permettrait de mieux le différencier.

3.5. Les différents problèmes rencontrés

3.5.1. Détection des palmiers pinot

Avec l'algorithme GMM, sur certaines images les palmiers pinots ne sont pas détectés, et ce malgré des ROI bien ciblés. Sur la scène 687-342 par exemple, sur 2870 pixels de pinot, aucun n'a été classé comme tel. Le kappa entre deux images d'une même scène augmente sensiblement lorsque cette végétation est détectée, passant ainsi de 0,92 à 0,98 ($\Delta \kappa = 0,06$). Cependant avec l'algorithme Random Forest, les palmiers pinots ont bien été détectés sur la scène 687-342 et ce avec les mêmes parcelles d'entraînement.

3.5.2. Difficulté à seuiller l'intensité des parinaris

La principale difficulté de ce travail a été de bien seuiller l'intensité des parinaris. En effet, le faible nombre de relevés terrain, qui représente la moyenne des parinaris sur une parcelle de 100m², et le fait que les parinaris sont parfois minoritairement présents sur ces zones, ont rendu difficile l'estimation de leur présence par pixel. La création des parcelles d'entraînement, primordial pour ce travail, a été facilitée en forçant les contrastes de l'image depuis Qgis afin de bien différencier les zones potentielles de parinaris avec la forêt. Il a donc fallu plusieurs essais en vérifiant visuellement et statistiquement les résultats pour valider une méthodologie.

3.5.3. Les variations d'éclairement

Au début du stage, n'ayant pas les librairies compatibles permettant de faire tourner les scripts Qgis avec Random Forest, les différents essais ont été effectués avec l'algorithme GMM. Il se trouvait que dans les scènes avec les parinaris, il y a avait une confusion entre les zones d'ombre et de bas-fond avec la forêt de ce type. Cela s'explique par une différence d'éclairement en fonction du relief. Ce phénomène s'appelle le BRDF (bidirectional reflectance distribution function). La seule méthode trouvée pour éviter à l'algorithme de se tromper était d'ajouter comme bande l'ombrage du MNT car ce type de végétation se situe sur les forêts de flat, autrement dit des pénéplaines, zones marécageuses et plates. L'algorithme GMM a donc montré ici sa limite sur une image seulement traitée en Top Of Atmosphere.

Une méthode expliquée notamment par Olivier Hagolle (2013) du CESBIO permet d'enlever l'effet du BRDF, c'est-à-dire d'enlever les différences liées à l'ombre de l'ubac et à la surexposition de l'adret. Avec la réussite de l'algorithme Random Forest il n'y a pas eu besoin d'effectuer la mise en place complexe de ce traitement, d'autant plus que la qualité du modèle numérique de terrain en Guyane est faible (30m).

3.6. Validation et comparaison avec le précédent travail

Les tests statistiques ont été réalisés à partir du logiciel R et de sa librairie pROC, permettant de mesurer l'efficacité de la classification à partir d'une courbe ROC.

La distance entre chaque point de référence et le polygone de végétation basse le plus proche est obtenu à partir d'un outil SIG. Le test ROC calcule ensuite la spécificité (pourcentage des objets détectés correspondant effectivement aux objets observés) et la sensibilité (pourcentage des objets observés effectivement détecté). A partir de ces données, l'AUC (Area Under the Curve, soit l'Aire Sous la Courbe) est calculée et représente la qualité de la prédiction (plus l'AUC se rapproche de 1 plus le modèle est bon). Le résultat d'AUC est caractérisé par une distance optimale (distance entre objets détectés et points GPS) correspondant au point le plus proche de l'idéal (le maximum obtenu en faisant la somme de la sensibilité avec la spécificité).

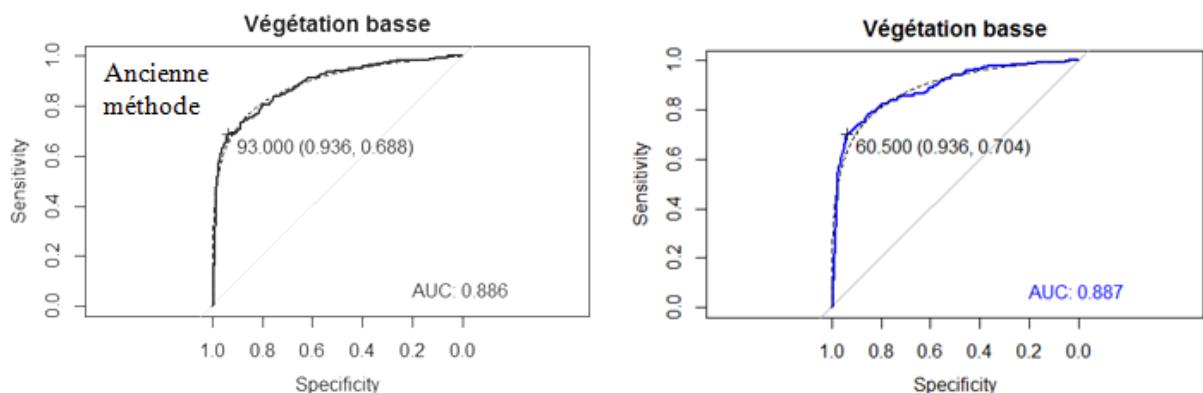


Figure 28 : Courbe ROC comparant les résultats des végétations basses

Les points de référence terrain ont été essentiellement pris à partir de l'hélicoptère, créant des décalages de plusieurs centaines de mètres entre les coordonnées relevées et les points observés du fait de la hauteur de vol, des parallaxes et de la vitesse. La plupart des

points ont été recalés a posteriori, mais ils ne peuvent pas être considérés comme géographiquement précis.

La courbe ROC des végétations basses montre une AUC identique à l'ancienne méthode (0.886 contre 0.887 pour la nouvelle), mais cependant on constate une meilleure précision pour la nouvelle méthode (60,5m contre 93m). Cela signifie qu'en moyenne l'erreur est de 60,5m. On peut expliquer ces résultats par des relevés terrains qui représentent parfois des parcelles de faible importance (100m² par exemple) qui ne sont donc pas représentées par cette cartographie. La végétation basse la plus proche ne sera donc pas celle observée sur le terrain mais une autre. Par exemple on observe que la moins bonne prédiction est à plus de 3km, ce qui explique un écart de 60m en moyenne.

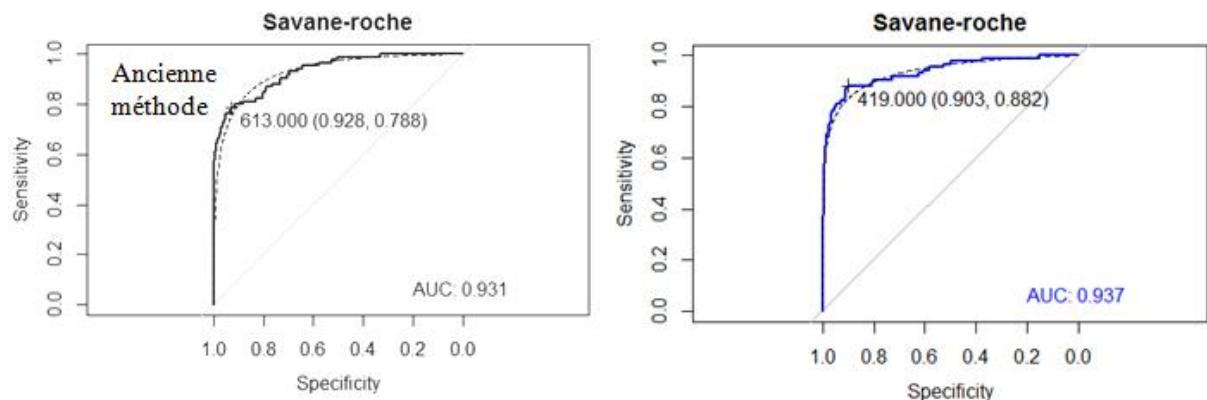


Figure 29 : Courbe ROC comparant les résultats des savanes-roche

Cette conclusion est d'autant plus vraie avec les résultats de la savane-roche. Ici la nouvelle méthode est sans appel de meilleure qualité. L'AUC est très proche (0.931 contre 0.937) mais la précision est améliorée à 419m contre 613m pour les anciens résultats. Le nombre de savanes-roche étant moins important, s'il en manque une dans la cartographie alors l'écart avec la savane-roche la plus proche sera d'autant plus grand.

La nouvelle cartographie qui regroupe les végétations comme les palmiers pinots, bâches, ou les parinaris, montre qu'elle est autant, voire plus précise que l'ancienne.

3.7. Les limites de la cartographie des végétations particulières

Plusieurs limites inhérentes à la méthodologie et au capteur sont à noter sur le résultat final. Elles peuvent être liées au manque de données terrain, au mauvais géoréférencement de l'image, et à la difficulté de bien séparer spectralement les végétations à partir d'une image à une date unique. Un fichier vectoriel regroupant les quelques incohérences de la cartographie finale a été créé dans le but de référencer les erreurs potentielles et de les corriger ou de les invalider à l'avenir.

3.7.1. La juxtaposition des images

L'ensemble du travail se faisant en raster et de manière automatique, les images ont été regéoréférencées quand les différences de géolocalisation étaient trop grandes entre elles. Cependant lors du manque suffisant de points d'amér en forêt, il s'est avéré que le repositionnement des images ne permettait pas d'améliorer la géoréférencement initial. Il en

réulte, lors de la combinaison des images et des scènes entre elles, des chevauchements d'une même végétation. Ces erreurs semblent cependant assez peu fréquentes sur l'ensemble du parc.



Figure 30 : Erreur de juxtaposition des images

Travailler sur des images mieux géoréférencées ou à plus grande emprise spatiale permettrait d'éviter de reproduire ce type d'erreur.

3.7.2. La présence de palmiers bâche

Le flat de la waki est une pénéplaine où se trouve l'intégralité des végétations particulières du parc. C'est donc aussi l'endroit où les confusions sont les plus susceptibles de se produire, notamment entre les végétations très proches comme le palmier pinôt ou le palmier bâche. C'est pour cette dernière essence que le risque de confusion avec les autres essences est le plus élevé. Le nombre très limité de parcelles d'entrainement (15) ne permet pas d'avoir toute l'amplitude spectrale afin de bien la séparer du reste des végétations. Il se trouve donc que sur les scènes de la Waki (690-344 et 689-344 de préférence) il n'a pas été possible de vérifier suffisamment bien la présence de ces palmiers. Quelques photos de survols ont cependant infirmé leur présence, et d'autres l'ont affirmée. Ces deux scènes sont donc certainement celles où il est important lorsqu'il y aura un prochain survol ou une mission terrain de prendre de nouveaux relevés.

3.8. Carte d'identité des végétations et des capteurs

Ce travail se basant sur les images du satellite SPOT-5 désormais inactif, il ne sera plus possible de reproduire à l'identique la même méthodologie. Le parc dispose facilement des images Sentinel-2 et grâce au projet SEAS d'images SPOT-6, Pléiades.

Les conditions particulières de la Guyane (forte nébulosité, grandeur du territoire) rendent le travail compliqué et certains éléments sont donc à privilégier. Par exemple, pour mieux discerner les végétations, il a été remarqué que le nombre de bandes infrarouges était primordial. Compte tenu de la nébulosité, afin d'avoir plusieurs images par année, il faut donc la résolution temporelle la plus forte. Un autre problème vient de la grandeur du territoire. Avec SPOT-5, il y avait besoin de 23 scènes différentes à agglomérer, et de 55 images au total. Le manque de point d'amer montrait des différences de géoréférencement

pouvant aller jusqu'à plus de 100m : plus l'emprise d'une image est importante, moins il y aura des erreurs dues à la juxtaposition des images et des scènes.

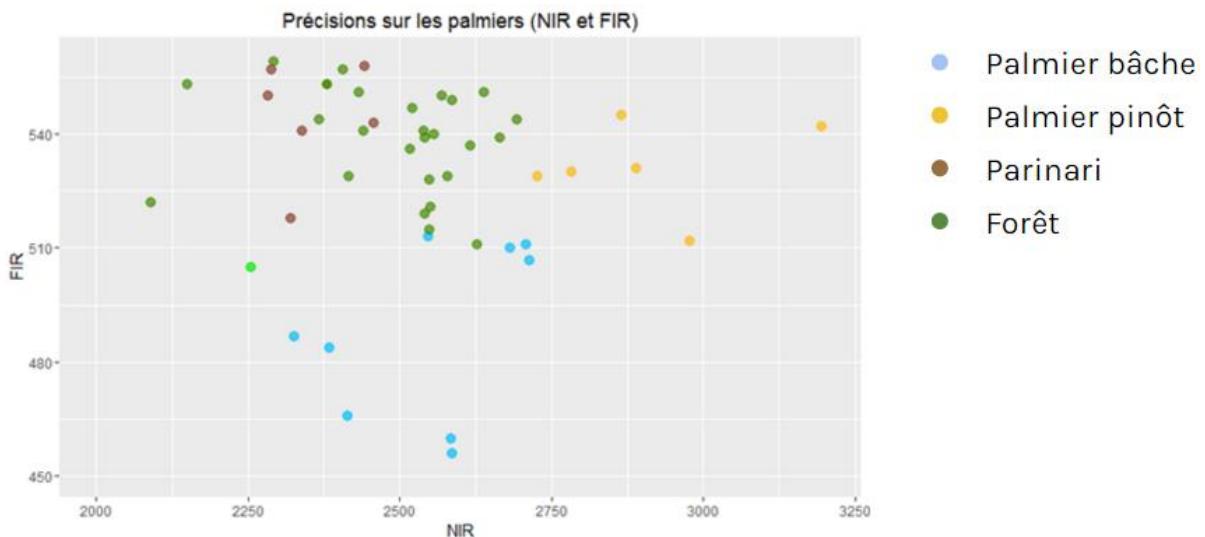


Figure 31 : réflectance des palmiers avec le proche (NIR) et lointain infrarouge (FIR),
Sentinel-2

En analysant la réflectance des palmiers avec les différents capteurs, il s'avère que le proche infrarouge, le moyen et le lointain sont les seuls qui peuvent différencier correctement ces végétations. Sans infrarouge lointain, des confusions existent entre les palmiers bâches, la forêt et les palmiers pinôts. L'alternative est donc d'avoir un nombre important d'images sur une année afin de suivre l'évolution sans moyen ou lointain infrarouge, mais cela nécessite une méthodologie plus complexe qui n'est pas encore mise au point.

Avec 13 bandes spectrales, une résolution temporelle de 5 jours, une emprise de 250x200km, et sa facilité d'acquisition en ligne, il apparaît que Sentinel-2 est la meilleure des solutions dans le futur pour poursuivre ce travail.

4. Bilan et perspectives

4.1. Conclusion

L'objectif de réaliser une cartographie des végétations particulières à l'échelle du Parc amazonien de Guyane est accompli. L'une des principales difficultés a été de compenser le manque de données terrain et la forte nébulosité. Il a été mis en avant la qualité et la robustesse de l'algorithme Random Forest dans la classification des essences sélectionnées. En effet, comparé au Modèle de Mélange Gaussien (GMM) qui a une moyenne kappa de 0,87 sur l'ensemble des images, Random Forest s'illustre par un résultat de grande qualité avec en moyenne 0,97. Cet excellent taux est toutefois à nuancer dans les scènes qui comportent un nombre important de végétations différentes. En effet, les palmiers pinôts et bâches sont les deux végétations qui se confondent le plus avec la forêt.

Au niveau de la méthodologie, le nombre important d'images nécessaires à la réalisation de ce travail a nécessité la réalisation de plusieurs scripts pour traiter en lot ces données. Après avoir généré un masque à partir de l'indice de nébulosité NDCI, un fichier vecteur contenant les parcelles d'entraînement a été créé pour chaque image. Pour compenser le manque de données terrains sur certaines scènes, un travail plus précis de photo-interprétation a été nécessaire soit en cherchant manuellement les végétations, soit en se servant des classifications des scènes voisines pour voir si certaines essences se trouvaient en bordure d'image et pouvaient donc servir comme nouvelle parcelle d'entraînement. Par la suite, pour chaque scène SPOT-5 l'ensemble des images a été aggloméré en utilisant *mapcalc* (outil Grass disponible dans Qgis) afin de pouvoir selon les images choisir quel type de végétation garder en priorité, ou quel type éviter. Enfin une fusion de l'ensemble des 23 scènes s'est faite puis le masque des zones agricoles et orpaillées du Parc amazonien de Guyane a été appliqué. Des écarts dus au mauvais géoréférencement des images ont été constatés lors de la fusion des images, mais il n'a pas été possible de les réaligner car il est très difficile voire impossible de trouver des points d'amér suffisamment fiables en forêt.

Afin de permettre au Parc amazonien de Guyane et à d'autres géomaticiens de reproduire ce type de classification, un plugin Qgis nommé *dzetsaka* a été publié dans les dépôts officiels. Il est le premier à supporter l'algorithme Random Forest sur ce logiciel grâce à l'implémentation réalisée par Mathieu Fauvel de la librairie Scikit-Learn. L'ensemble des scripts Qgis et des scripts R utilisés pour réaliser ce travail sont disponibles sur github.com/lennepkade/ScriptsParcAmazonien.

Ce stage rend compte de la complexité de la caractérisation des essences forestières à petite échelle par télédétection. La proximité spectrale des palmiers bâches, palmiers pinôts ou encore des parinaris montre que la méthodologie pourrait être améliorée, soit avec la création d'images multi-temporelles, mais cela signifie disposer d'images sans nuages, soit avec l'apport de nouveaux capteurs incluant plus de bandes rouges et infrarouges comme Sentinel-2.

Pour poursuivre et améliorer ce travail, il apparaît que le satellite Sentinel-2 tire son épingle du jeu et est donc à privilégier, car il est de résolution spatiale égale à SPOT-5 et est doté d'une plus grande résolution temporelle (ce qui permet de multiplier les passages afin de capturer les zones fortement ennuagées), d'une résolution spectrale incluant plusieurs rouges et le proche, moyen et lointain infrarouge.

4.2. Bilan personnel

Ce stage a été synonyme de nombreuses découvertes et d'approfondissement des connaissances. La télédétection mélange le travail informatique avec le terrain et j'ai pu me familiariser avec les essences sur lesquels je travaillais, chose que je n'avais jamais faite auparavant. J'ai ainsi pu découvrir la richesse et l'immensité de la forêt amazonienne de Guyane et réaliser la difficulté de ce travail tant par l'ampleur du territoire que par la très forte nébulosité. La Guyane est un territoire encore peu connu car très difficile d'accès, c'est donc un très beau challenge que j'ai pu réaliser grâce à Pauline Perbet et Pierre Joubert, que je ne saurais trop remercier.

Compliqué par bien des aspects, cet exercice m'a permis d'expérimenter les problèmes rencontrés sur des champs nouveaux de recherche. Le fait de devoir produire le meilleur des résultats possibles m'a poussé à revoir mes méthodes, à essayer des solutions inspirées par divers travaux (apport des indices de végétation par exemple). C'est aussi la première fois que je réalise un travail en télédétection complet et innovant devant résoudre des problèmes spécifiques. Même si tout n'est pas parfait, ce qui s'explique par les nombreuses limites (grandeur du territoire, nombreuses scènes, nébulosité importante, mauvais géoréférencement), la méthodologie que j'ai développée a non seulement permis de cartographier pour la première fois de manière précise les forêts de palmiers pinots, palmiers bâches ou de parinaris ; elle est aussi facilement reproductible grâce aux outils comme dzetsaka. J'ai également pu découvrir et apprendre le langage R, très utilisé pour les statistiques, qui m'a permis de faire des rapports sur les différents capteurs et d'établir une carte d'identité des végétations par capteur.



Figure 32 : Présentation du 1^{er} septembre 2016 au Parc amazonien de Guyane

C'était avec enthousiasme que j'ai pu, le 1^{er} septembre 2016, présenter le résultat de mes travaux devant une vingtaine de personnes de divers organismes comme l'ONF, la DEAL (Direction de l'Environnement, de l'Aménagement et du Logement de la Guyane), le WWF, la CTG (Collectivité Territoriale de Guyane) ou encore la réserve naturelle nationale des Nouragues. Une belle émotion après six mois de travail et des difficultés qui paraissaient insurmontables (je pense notamment aux parinaris qui ont beaucoup complexifié ma tâche).

Bibliographie

Brondizio, Eduardo S., Emilio F. Moran, Paul Mausel, and You Wu. "Land Use Change in the Amazon Estuary: Patterns of Caboclo Settlement and Landscape Management." *Hum Ecol Human Ecology* 22.3 (1994): 249-78. Print.

Carreiras, João M.b., José M.c. Pereira, and Yosio E. Shimabukuro. "Land-cover Mapping in the Brazilian Amazon Using SPOT-4 Vegetation Data and Machine Learning Classification Methods." *Photogrammetric Engineering & Remote Sensing Photogramm Eng Remote Sensing* 72.8 (2006): 897-910. Print.

Catalogue Des Habitats Forestiers De Guyane. Cayenne (Guyane): Direction Régionale ONF Guyane, 2015. Print.

Cherrington, Emil A., Nicolas Barbier, Pierre Ploton, Gregoire Vincent, Daniel Sabatier, Uta Berger, and Raphael Pelissier. "Equatorial Forests Display Distinct Trends in Phenological Variation: A Time-Series Analysis of Vegetation Index Data from Three Continents." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* (2016): 1-7. Print.

Demaze, Moïse Tsayem. "Le Parc Amazonien De Guyane Française : Un Exemple Du Difficile Compromis Entre Protection De La Nature Et Développement." *Cybergeo Cybergeo* (2008). Print.

Fauvel, Mathieu, Clement Dechesne, Anthony Zullo, and Frederic Ferraty. "Fast Forward Feature Selection of Hyperspectral Images for Classification With Gaussian Mixture Models." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 8.6 (2015): 2824-831. Print.

Foody, G. M., and R. A. Hill. "Classification of Tropical Forest Classes from Landsat TM Data." *International Journal of Remote Sensing* 17.12 (1996): 2353-367. Print.

Fromard, F., C. Vega, and C. Proisy. "Half a Century of Dynamic Coastal Change Affecting Mangrove Shorelines of French Guiana. A Case Study Based on Remote Sensing Data Analyses and Field Surveys." *Marine Geology* 208.2-4 (2004): 265-80. Print.

Ghiyamat, Azadeh. "Influence of Tree Species Complexity on Discrimination Performance of Vegetation Indices." *EuJRS European Journal of Remote Sensing* (2016): 15. Print.

Girard, Michel C., Colette M. Girard, and Dominique Courault. *Traitemet Des Données De Télédétection*. Paris: Dunod, 1999. Print.

Gond, Valéry, Vincent Freycon, Jean-François Molino, Olivier Brunaux, Florent Ingrassia, Pierre Joubert, Jean-François Pekel, Marie-Françoise Prévost, Viviane Thierron, Pierre-Julien Trombe, and Daniel Sabatier. "Broad-scale Spatial Pattern of Forest Landscape Types in the Guiana Shield." *International Journal of Applied Earth Observation and Geoinformation* 13.3 (2011): 357-67. Print.

Granville, Jean-Jacques De, and Marc Gayot. *Guide Des Palmiers De Guyane*. ONF. 2014. Print.

Guitet, Stéphane, Jean-François Cornu, Olivier Brunaux, Julie Betbeder, Jean-Michel Carozza, and Cécile Richard-Hansen. "Landform and Landscape Mapping, French Guiana (South America)." *Journal of Maps* 9.3 (2013): 325-35. Print.

Hagolle, Olivier. "La correction des variations d'éclairement dues au relief". Séries temporelles. Web. 20 septembre 2013.

Herrault, Pierre-Alexis, David Sheeren, Mathieu Fauvel, and Martin Paegelow. "Vectorisation Automatique Des Forêts Dans Les Minutes De La Carte D'état-major Du 19esiècle." *Revue Internationale De Géomatique Rev. Int. Geomat.* 25.1 (2015): 35-51. Print.

Marshak, A., Y. Knyazikhin, A. B. Davis, W. J. Wiscombe, and P. Pilewskie. "Cloud-vegetation Interaction: Use of Normalized Difference Cloud Index for Estimation of Cloud Optical Thickness." *Geophys. Res. Lett. Geophysical Research Letters* 27.12 (2000): 1695-698. Print.

Morton, Douglas C., Jyoteshwar Nagol, Claudia C. Carabajal, Jacqueline Rosette, Michael Palace, Bruce D. Cook, Eric F. Vermote, David J. Harding, and Peter R. J. North. "Amazon Forests Maintain Consistent Canopy Structure and Greenness during the Dry Season." *Nature* 506.7487 (2014): 221-24. Print.

Pavé, Alain. "Evaluating Biodiversity: The Example of French Guiana." *On the Origins and Dynamics of Biodiversity: The Role of Chance* (2010): 139-51. Print.

Pennec, Alexandre, Valéry Gond, and Daniel Sabatier. "Tropical Forest Phenology in French Guiana from MODIS Time Series." *Remote Sensing Letters* 2.4 (2010): 337-45. Print.

Saatchi, Sassan, Wolfgang Buermann, Hans Ter Steege, Scott Mori, and Thomas B. Smith. "Modeling Distribution of Amazonian Tree Species and Diversity Using Remote Sensing Measurements." *Remote Sensing of Environment* 112.5 (2008): 2000-017. Print.

Saatchi, Sassan, Wolfgang Buermann, Hans Ter Steege, Scott Mori, and Thomas B. Smith. "Modeling Distribution of Amazonian Tree Species and Diversity Using Remote Sensing Measurements." *Remote Sensing of Environment* 112.5 (2008): 2000-017. Print.

Samanta, Arindam, Sangram Ganguly, Eric Vermote, Ramakrishna R. Nemani, and Ranga B. Myneni. "Why Is Remote Sensing of Amazon Forest Greenness So Challenging?" *Earth Interactions Earth Interact.* 16.7 (2012): 1-14. Print.

Sheeren, D., M. Fauvel, C. Planque, J. Willm, and J. F. Dejoux. "Tree Species Discrimination in Temperate Woodland Using High Spatial Resolution Formosat-2 Time Series." *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)* (2015). Print.

Somers, B., and G.p. Asner. "Tree Species Mapping in Tropical Forests Using Multi-temporal Imaging Spectroscopy: Wavelength Adaptive Spectral Mixture Analysis." *International Journal of Applied Earth Observation and Geoinformation* 31 (2014): 57-66. Print.

Stach, N., A. Salvado, M. Petit, J. F. Faure, L. Durieux, C. Corbane, P. Joubert, D. Lasselin, and M. Deshayes. "Land Use Monitoring by Remote Sensing in Tropical Forest Areas in Support of the Kyoto Protocol: The Case of French Guiana." *International Journal of Remote Sensing* 30.19 (2009): 5133-149. Print.

Table des illustrations

| | |
|--|----|
| Figure 1 : Les différents parcs nationaux. Source : Parc Nationaux de France | 8 |
| Figure 2 : Les deux grandes zones du Parc amazonien de Guyane | 9 |
| Figure 3 : Organigramme du Parc. Source : PAG | 11 |
| Figure 4 : De gauche à droite : palmiers pinôts (<i>Euterpe oleracea</i>), savane-roche, végétation basse (cambrouse de bambous <i>Lasiacis ligulata</i>). Crédits : Daniel Sabatier (IRD)..... | 13 |
| Figure 5 : Canopée de la forêt amazonienne. Photo : Daniel Sabatier (IRD) | 14 |
| Figure 6 : La nébulosité moyenne au Parc. Donnée source de EarthEnv | 15 |
| Figure 7 : Date d'acquisition des images SPOT-5..... | 15 |
| Figure 8 : Nombre d'images utilisées pour chaque scène SPOT-5 | 16 |
| Figure 9 : Nombre de données terrain (pour 100km ²) | 17 |
| Figure 10 : Utilisation d'une tablette/GPS pour vérifier les végétations détectées..... | 18 |
| Figure 11 : De grandes difficultés d'accès aux végétations | 19 |
| Figure 12 : Schéma d'organisation des images..... | 20 |
| Figure 13 : Exemple de synthèse de 4 images d'une même scène | 23 |
| Figure 14 : Logo de dzetsaka | 25 |
| Figure 15 Fenêtre principal de dzetsaka | 25 |
| Figure 16 : Exemple de résultat de classification au Sud-ouest..... | 28 |
| Figure 17 : Validation par photographie héliportée | 30 |
| Figure 18 : Exemple de résultat de classification dans la pénéplaine de la Waki | 31 |
| Figure 19 : Courbe ROC des Parinaris | 32 |
| Figure 20 : Validation par photographie héliportée | 33 |
| Figure 21 : Les scènes utilisées pour étudier l'effet biogéographique | 35 |
| Figure 22 : Évolution de la bande verte selon la latitude pour la végétation basse (SPOT-5) 35 | 35 |
| Figure 23 : Évolution de la bande verte selon la latitude pour la végétation basse (SPOT-6) 36 | 36 |
| Figure 24 : Phénologie des végétations avec SPOT-5 | 36 |
| Figure 25 : Évolution de la réflectance du Palmier Pinôt | 37 |
| Figure 26 : Évolution de la réflectance du Palmier Bâche..... | 37 |
| Figure 27 : Évolution de la réflectance du Parinari | 38 |
| Figure 28 : Courbe ROC comparant les résultats des végétations basses..... | 39 |
| Figure 29 : Courbe ROC comparant les résultats des savanes-roche | 40 |
| Figure 30 : Erreur de juxtaposition des images | 41 |
| Figure 31 : réflectance des palmiers avec le proche (NIR) et lointain infrarouge (FIR), Sentinel-2 | 42 |
| Figure 32 : Présentation du 1 ^{er} septembre 2016 au Parc amazonien de Guyane | 44 |

Table des tableaux

| | |
|---|----|
| Tableau 1 : Aperçu des végétations particulières sous SPOT-5 | 14 |
| Tableau 2 : Détail des bandes spectrales de SPOT-5..... | 16 |
| Tableau 3 : Nombre de données terrain par type de végétation..... | 18 |
| Tableau 4 : Extrait des valeurs des ROI des 55 images | 22 |
| Tableau 5 : Matrice de confusion de la scène 688-345..... | 29 |
| Tableau 6 : Matrice de confusion de la scène 690-344..... | 31 |
| Tableau 7 : Accord Global (OA) et indice Kappa par algorithme en combinant les bandes spectrales (SB) avec les indices, Image du 13/07/2008 | 33 |
| Tableau 8 : Accord Global (OA) et indice Kappa par algorithme en combinant les bandes spectrales (SB) avec les indices, Image du 21/06/2012..... | 34 |
| Tableau 9 : Accord Global (OA) et indice Kappa par algorithme en combinant les bandes spectrales (SB) avec les indices, Image du 12/09/2012..... | 34 |

Annexes

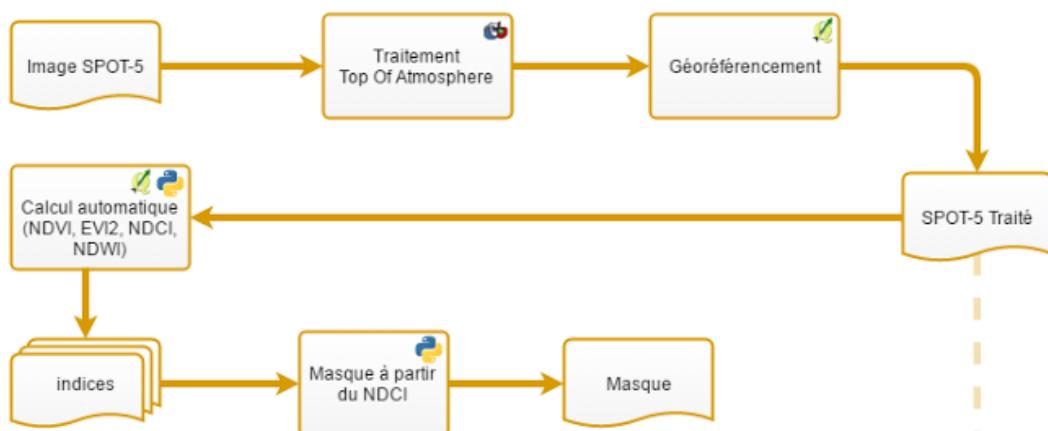
Annexe 1 : Procédure de cartographie des végétations particulières



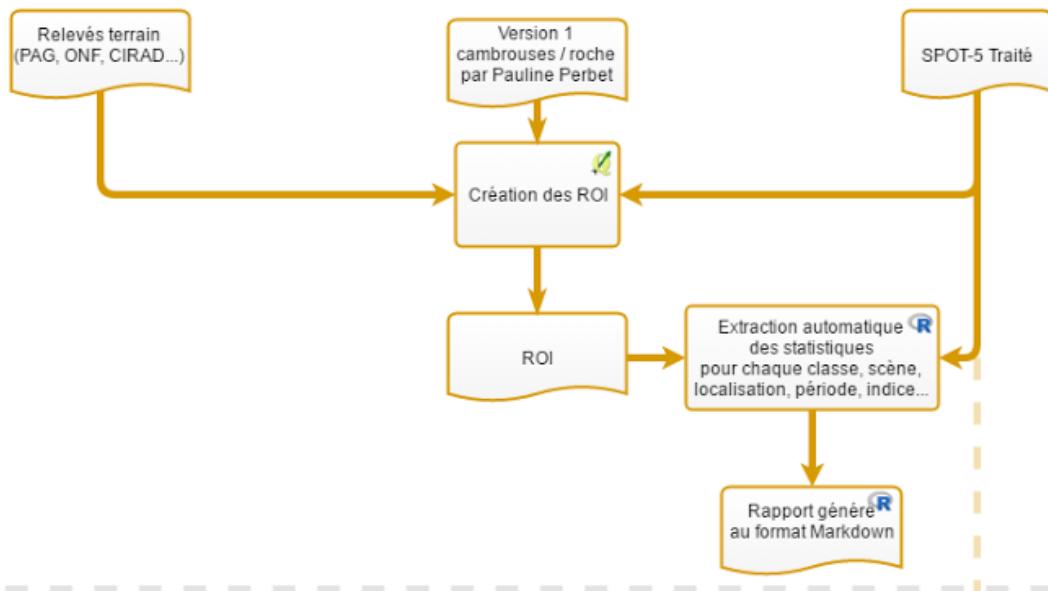
Procédure de cartographie des végétations particulières

Nicolas Karasiak

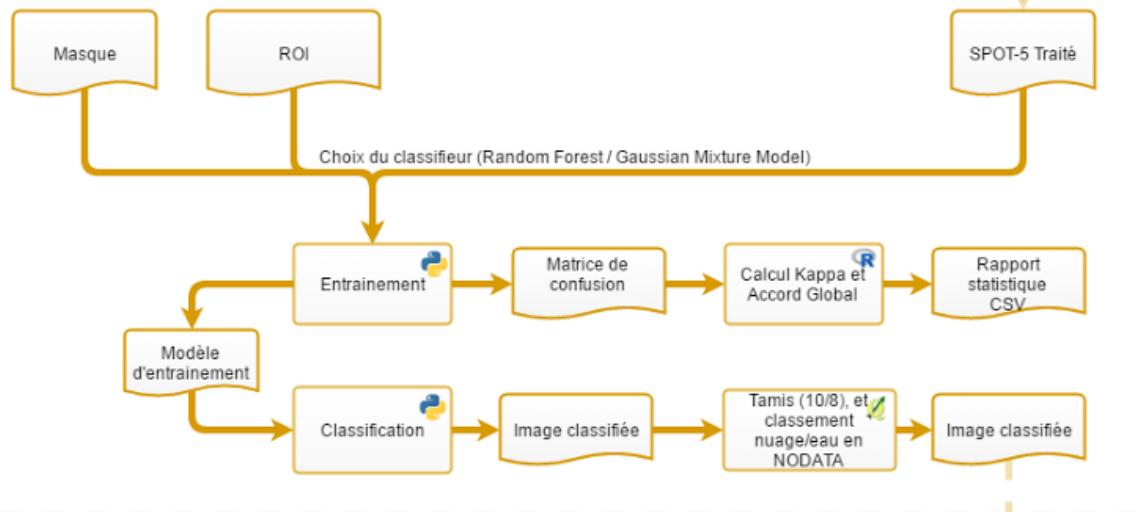
1. Traitement d'images



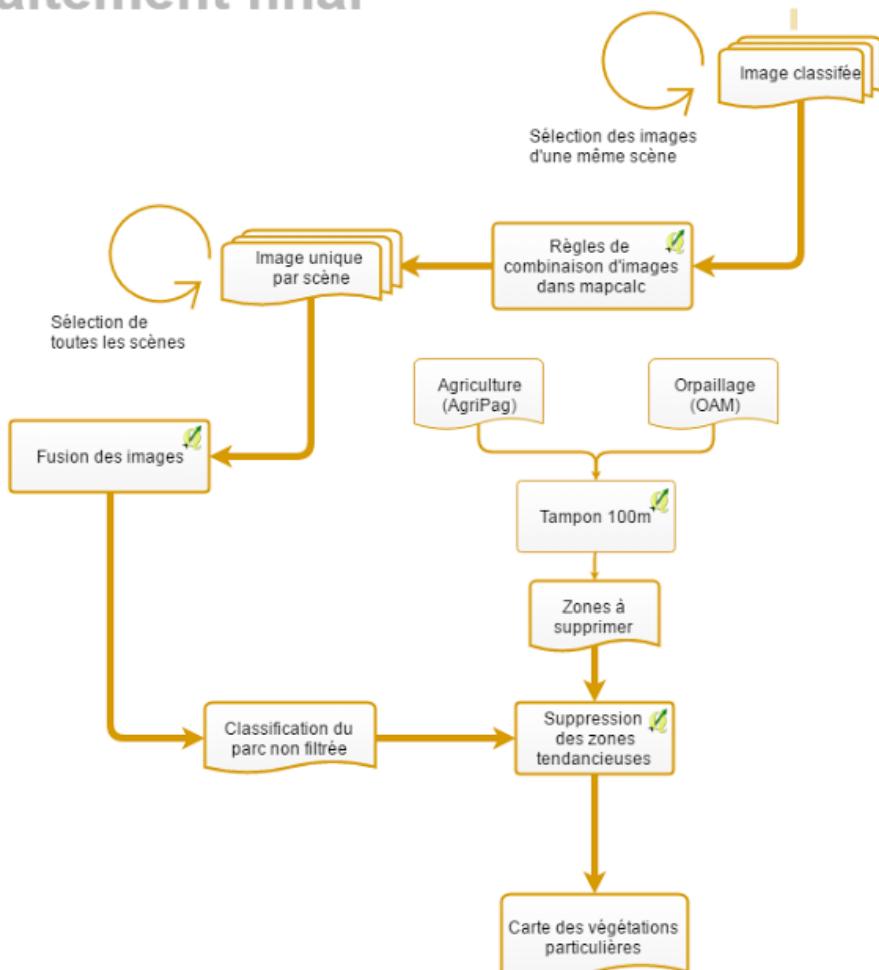
2. Identification des végétations



3. Classification par image



4. Traitement final

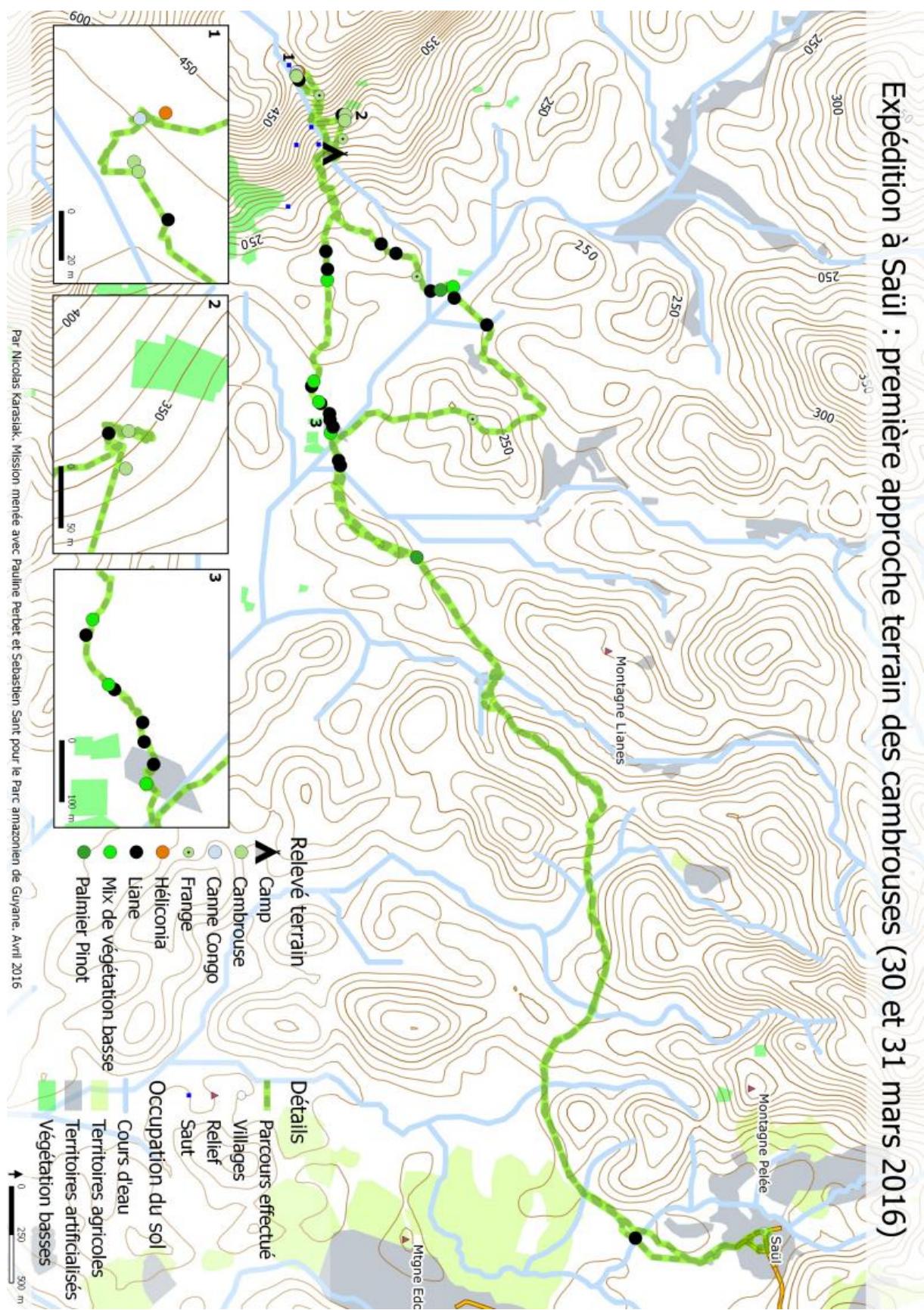


Détails :

Fichier
 Fichiers
 Procédure
 Qgis
 Python
 R (R Studio)
 Monteverdi (Orfeo ToolBox)

Expédition à Saül : première approche terrain des cambrousse (30 et 31 mars 2016)

Annexe 2 : Carte de l'expédition à Saül



Annexe 3 : Script runStatsBatch.R pour calculer les valeurs de pixel de l'ensemble des ROI

```
1  # _____
2  # | _ \                                         ( )
3  # | |_) |-----|-----|-----|-----|-----|
4  # | __/ \ | ' \ / | / \ | / \ | / \ | / \ | / \ |
5  # | | | | | | | | | | | | | | | | | | | | | | |
6  # | | \_ \_ \_ \_ | \_ \_ \_ \_ | \_ \_ \_ \_ | \_ \_ \_ \_
7  # | | | | / \ |
8  # | | | | | | | | | | | | | | | | | | | | | | |
9  # | | | | | | | | | | | | | | | | | | | | | | |
10 # | | | | | | | | | | | | | | | | | | | | | | |
11 # \_ \_ \_ \_ | \_ \_ \_ \_ | \_ \_ \_ \_ | \_ \_ \_ \_ |
12 # | | | | | | | | | | | | | | | | | | | | | |
13 # | | | | | | | | | | | | | | | | | | | | | |
14
15 # R Script to compute SPOT5 stats from ROI
16 # This script parse subfolders, find tif and shp with same name, and extract mean Value on each band
17 # Sample : find file in folder 680-345/yourfile.tif, the script will find yourfile.shp, and if you
18 # had some indices, find yourfile_INDICE.tif and link them to yourfile.shp
19
20 require(sp)
21 require(raster)
22 require(maptools)
23 require(rgeos)
24 require(gtools)
25
26 initwd <- 'Y:/VegetationsParticulieres/SPOT5/SPOT5_TOAgeoref'
27 setwd(initwd)
28
29
30 folders=list.dirs(getwd(),recursive=FALSE)
31
32 for (folder in folders){
33   setwd(folder)
34
35   # Get all tif from folder
36   inRaster <- list.files(folder,pattern='*.tif$')
37
38   # Order them to have main image first
39   inRaster <- inRaster[order(nchar(inRaster))]
40
41   for(rasterLoop in inRaster){
42     print(paste('calculating ',rasterLoop))
43     # get fileName with no extension and/or no indice
44     fileName <- strsplit(strsplit(rasterLoop, "\\."[[1]], "\\_")[[1]])
45     inShape <- paste(fileName[1],'.shp',sep='')
46
47     openRaster <- raster(rasterLoop)
48
49     avoidImage <- FALSE
50
51     # if indice image (type : image_ndvi.tif), take same shp as original image
52     # fileName[2] is the indice name, or it's NA
```

```

53   if (!is.na(fileName[2])){
54
55     # if mask, avoid calc
56     if(fileName[2]=='mask'){
57       avoidImage <- TRUE
58     }
59     else{
60       indiceImage <- TRUE
61     }
62     indiceName <- fileName[2]
63
64   }
65   else{
66     indiceImage <- FALSE
67     inShape <- paste(fileName[1],'.shp',sep='')
68   }
69
70   # If same scene but indices, performs stat with same shp
71   # if shp with samename exist, perform stat
72   # if shp doesn't exists or avoidImage is TRUE, do not perform stat
73
74   if (file.exists(inShape) & avoidImage==FALSE)
75   {
76     inShape <- readShapePoly(inShape)
77
78     csvToWrite <- paste(fileName[1],'.csv',sep='')
79     if(indiceImage){
80
81       mydf=read.csv2(csvToWrite)
82     }
83     else{
84       mydf <- data.frame(inShape)
85       # Save image name in image column
86       mydf['imageName'] <- fileName[1]
87     }
88
89     # For each band, write zonal stat
90     for(i in 1:openRaster@file@nbands){
91       openBand <- raster(rasterLoop,band=i)
92       if(indiceImage){
93         band=indiceName
94       }
95       else{
96         if(i==1){band='V'}
97         else if(i==2){band='R'}
98         else if(i==3){band='NIR'}
99         else if(i==4){band='MIR'}

```

```

100      }
101      print(paste('calculating',band))
102
103      #assign band name with min/max like NIRmin, NIRmax...
104
105      #fill table
106      meanB <- extract(openBand,inShape, fun = mean)
107      mydf[band] <- round(meanB,0)
108  }
109
110  write.csv2(mydf,csvToWrite)
111
112  }
113  else{
114      print('No vector with the same name as your main image or image to avoid')
115  }
116  }
117 }
118
119 ### Merge all CSV
120 # Go back to root
121 setwd(initwd)
122
123 # Get all csv in all subfolders
124 inCsv <- list.files(recursive=TRUE,pattern='*.csv$')
125 outCsv <- 'zonalStat.csv'
126 # save it at root with name zonalStat.csv
127 if(file.exists(outCsv)){
128     file.remove(outCsv)
129 }
130
131 # create empty data frame
132 df<-data.frame()
133
134 # loop throught csv and save it in df
135 for(i in inCsv){
136     print(i)
137     temp <-read.csv2(i)
138
139     #using smartbind from gtools to combine if different column number
140     df<-smartbind(df, temp)
141     rm(temp)
142 }
143
144 write.csv2(df,outCsv)

```

Annexe 4 : Diagramme GANTT du stage

