

## Table des matières

---

Remerciements.....	VII
Table des matières.....	IX
Acronymes .....	XI
Table des annexes .....	XII
Table des figures.....	XIII
Table des tableaux.....	XIII
1. Introduction .....	1
2. Contexte.....	2
2.1. Occupation du sol à la Réunion.....	2
2.2. Télédétection optique .....	2
2.3. Télédétection radar.....	3
2.4. Modélisation.....	4
2.4.1. PLS (Partial least square).....	4
2.4.2. Réseaux de neurones.....	4
2.4.3. Random Forest .....	5
2.4.4. SVM (Séparateur à vastes marges).....	6
3. Matériels et méthodes.....	7
3.1. Zone d'étude.....	7
3.2. Données et logiciels utilisés .....	7
3.2.1. Données radars.....	7
3.2.2. Données vectorielles .....	8
3.2.3. Mesures terrain.....	8
3.2.4. Logiciels utilisés .....	8
3.2.5. Détermination des classes d'occupation de la sole agricole .....	9
3.3. Méthode.....	9
3.3.1. Prétraitements .....	9
3.3.2. Calcul des variables radar .....	10
3.3.3. Modélisation .....	15
4. Résultats.....	18
4.1. Modèle de classification de la sole agricole.....	18
4.1.1. Modèle à 64 variables .....	18
4.1.2. Modèle à 15 variables différentes par date .....	18

4.1.3	Modèle à 15 variables identiques pour toutes les dates.....	19
4.2	Modèle de cartographie des classes générales d'occupation du sol.....	20
4.3.	Modèle de classification en polarisation double .....	21
4.4	Comparaison des différents modèle .....	22
5.	Discussion.....	23
6.	Conclusion.....	24
	Bibliographie.....	25
	Annexes.....	28

## Acronymes

---

<b>ACP :</b>	Analyse en composante principale
<b>ASF :</b>	Alaska Satellite Facility
<b>BOS :</b>	Base d'Occupation des Sols
<b>CCT :</b>	Centre Canadien de Télédétection
<b>CIRAD :</b>	Centre de coopération Internationale en Recherche Agronomique pour le Développement
<b>DAAF :</b>	Direction de l'Alimentation, de l'Agriculture et de la Forêt
<b>ESA :</b>	Agence Spatiale Européenne
<b>INSEE :</b>	Institut National de la Statistique et des Etudes Economiques
<b>MNT:</b>	Modèle Numérique de Terrain
<b>MOS :</b>	Mode d'Occupation des Sols
<b>PAC:</b>	Politique Agricole Commune
<b>PLU:</b>	Plan Local d'Urbanisme
<b>RADAR:</b>	RAdio Detection And Ranging
<b>RMSE:</b>	Root Mean Square Error
<b>RPG :</b>	Registre Parcellaire Graphique
<b>RVI:</b>	Radar Vegetation Index
<b>SAR :</b>	Synthetic Aperture RADAR
<b>SAR :</b>	Schéma d'Aménagement Régional
<b>SAU :</b>	Surface Agricole Utile
<b>SCOT :</b>	Schéma de COhérence Territoriale
<b>SEAS-OI :</b>	Surveillance de l'Environnement Assistée par Satellite dans l'Océan Indien
<b>SMO :</b>	Sequential Minimal Optmization
<b>SVM :</b>	Séparateur à Vaste Marges
<b>UNESCO :</b>	Organisation des Nations Unies pour l'Education, la Science et la Culture
<b>UTM:</b>	Universal Transverse Mercator
<b>WEKA:</b>	Waikato Environment for Knowledge Analysis

## Table des annexes

---

<b>Annexe 1</b> : Répartition de la SAU de La Réunion en 2013.....	P 28
<b>Annexe 2</b> : Résumé des critères d'exclusion appliqués à la BOS afin d'obtenir la base d'apprentissage .....	P 29
<b>Annexe 3</b> : Répartition interclasses pour les 2 premières composantes de l'ACP.....	P 30
<b>Annexe 4</b> : Répartition des individus pour les 2 premières composantes de l'ACP .....	P 31
<b>Annexe 5</b> : Corrélations entre variables radar pour les 2 premières composantes de l'ACP .....	P 32
<b>Annexe 6</b> : Variables sélectionnées à partir du premier Random Forest .....	P 33
<b>Annexe 7</b> : Les 15 variables uniques sélectionnées à partir du second Random Forest .....	P 34

## Table des figures

---

<b>Figure 1</b> : Longueurs d'ondes radar .....	P 3
<b>Figure 2</b> : Organisation d'un réseau de neurones.....	P 5
<b>Figure 3</b> : Arbre de décision élémentaire .....	P 5
<b>Figure 4</b> : Exemple d'un séparateur à vastes marges .....	P 6
<b>Figure 5</b> : Localisation de la zone d'étude .....	P 7
<b>Figure 6</b> : Répartition des images RADARSAT-2 obtenues, sur l'année 2014.....	P 8
<b>Figure 7</b> : Organigramme général de la méthodologie radar. ....	P 14
<b>Figure 8</b> : Comparaison de l'indice de Kappa entre le modèle à 64 et celui à 15 variables .....	P 18
<b>Figure 9</b> : Indice de Kappa pour le modèle à 15 variables selon le nombre de dates. ....	P 19
<b>Figure 10</b> : Indice de Kappa pour le modèle à 15 variables uniques selon le nombre de dates.....	P 19
<b>Figure 11</b> : Matrice de confusion du modèle de 4 classes d'occupation basé sur 4 dates. ....	P 20
<b>Figure 12</b> : Indice de Kappa pour le modèle général selon le nombre de dates.....	P 21
<b>Figure 13</b> : Indice de Kappa pour le modèle en polarisation double selon le nombre de dates.....	P 21
<b>Figure 14</b> : Indice de Kappa pour les différents modèles selon le nombre de dates. ....	P 22

## Table des tableaux

---

<b>Tableau 1</b> : Résumé des différents logiciels utilisés et leurs utilisations.....	P 9
<b>Tableau 2</b> : Présentation des différentes classes d'occupation de la sole agricole. ....	P 9



# 1. Introduction

---

La Réunion, espace insulaire volcanique de 2 512 km<sup>2</sup>, est située dans l'océan Indien à 55°30' de longitude Est et 21°09' de latitude Sud. Son relief affirmé contraint les dynamiques territoriales qui se concentrent principalement sur les 40 % de l'île aménageables. Zones urbaines, agricoles, activités touristiques, portuaires et industrielles se concentrent essentiellement dans les « Bas », une bande littorale qui s'étire jusqu'à 500 m d'altitude, alors que les Hauts regroupent des territoires à dominante rurale tournés vers l'élevage et la conservation d'un patrimoine exceptionnel classé par l'UNESCO en 2010.

Dans un contexte où les enjeux de l'occupation de l'espace et la concurrence entre filières agricoles sont omniprésents, la pression foncière particulièrement forte impacte les espaces apparaissant comme inutilisés. Il devient donc nécessaire, de mettre en place des suivis de la consommation de ces espaces à la Réunion.

Le Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) et la Société d'Aménagement Foncier et d'Etablissement Rural (SAFER) travaillent depuis quelques années à l'élaboration de méthodes de traitement d'images satellitaires afin d'identifier les terres cultivées à la Réunion par télédétection.

Ces terres cultivées, forment un parcellaire agricole très fragmenté, avec des parcelles dont la surface moyenne est inférieure à l'hectare (0.72 hectare, [INSEE, 2007](#)). Ce parcellaire associé à une couverture nuageuse très présente et un développement phénologique souvent différent d'une parcelle à l'autre, fait de la Réunion, un espace dont la cartographie de l'occupation est très difficile. Les méthodes de télédétection optique à moyenne et haute résolution spatiale développées en pays tempérés sont peu adaptées à ce contexte.

Pour pallier ces limites le CIRAD explore depuis peu, deux nouvelles méthodes de caractérisation de l'occupation de la sole agricole. La première méthode s'appuie sur l'analyse orientée objet d'images à très haute résolution spatiale offerte par le capteur Pléiades (50 cm de résolution).

La deuxième méthode s'appuie sur l'utilisation des nouveaux radars à synthèse d'ouverture (SAR). Cette technique consiste à générer une image en haute résolution de la surface terrestre à partir d'un traitement effectué sur la différence de phase d'un ensemble d'ondes émises par l'antenne du satellite. Ils permettent de caractériser les propriétés géométriques (surface, volume, structure) du couvert végétal, très utile pour cartographier les cultures.

Ce stage a pour objectif de tester le potentiel des images radar polarimétriques provenant du satellite canadien à synthèse d'ouverture RADARSAT-2 afin de caractériser l'occupation de la sole agricole d'un milieu tropical très hétérogène, avec un relief contrasté qu'est l'île de la Réunion.

Dans une première partie, nous présenterons l'état des connaissances dans ce domaine. Puis, nous présenterons la méthodologie abordée pour répondre aux objectifs, celle-ci étant divisée en deux étapes : une partie consistant au développement d'un modèle de classification utilisant les variables radar afin de caractériser la sole agricole et une autre partie consistant à la mise en œuvre opérationnelle du modèle. Enfin, dans une dernière partie, nous présenterons les résultats obtenus tout en mettant en avant les limites et les perspectives de cette étude.

## 2. Contexte

---

### 2.1. Occupation du sol à la Réunion

Avec 44 000 hectares de terres agricoles soit environ 18% de la surface de l'île, la Réunion reste un territoire marqué par l'agriculture. Même si ce domaine d'activité génère aujourd'hui moins de volume financier que d'autres secteurs comme le tourisme ou les services, l'agriculture occupe quand même près de 15 000 personnes dont 7 600 chefs d'exploitation ([AGRESTE, 2013](#)) (soit 4,3 % de la population active occupée, taux supérieur à la métropole).

L'agriculture locale, caractérisée par un parcellaire agricole très fragmenté, avec des exploitations dont la taille moyenne avoisine les 6 hectares (Direction de l'Alimentation, de l'Agriculture et de la Forêt (DAAF), 2013), reste dominée par la culture de canne qui s'étend sur plus de 24 499 ha (soit 58 % de la Surface Agricole Utile, SAU). Toutefois, d'autres activités la complètent comme l'élevage (bovin, ovin, porc, avicole, caprin et cervidé) sur une dizaine de milliers d'hectares (soit 26 % de la SAU ([AGRESTE, 2013](#))), ou encore le maraîchage, les vergers et certaines cultures traditionnelles en déclin comme la vanille, le géranium ou le vétiver (utilisés dans les parfums) ([Annexe 1](#)). Ces cultures sont réparties selon des gradients altitudinaux et topographiques : la canne se concentre principalement sur les zones basses (entre 0 et 600 mètres d'altitude à l'Est et 800 mètres à l'Ouest) alors que l'élevage, le maraîchage ou encore les cultures traditionnelles occupent plutôt les parties hautes de l'île, au-delà de 600-800 m (Martignac 2006).

Pour répondre à la demande croissante de production agricole, il est nécessaire de documenter l'occupation du sol. Les nouvelles technologies d'observation de la surface terrestre peuvent apporter une contribution primordiale.

### 2.2. Télédétection optique

La télédétection regroupe l'ensemble des techniques d'acquisition de données sur un territoire, sans qu'aucun contact direct avec celui-ci ne soit nécessaire (Girard and Girard 1999). Il s'agit donc de données, la plupart du temps des images, acquises depuis un appareil aérospatial comme un avion, un drone, un ULM, un satellite, etc. Placés en orbite autour de la terre, les satellites sont capables de fournir des données sur l'ensemble de la surface terrestre de manière régulière, grâce à des revisites fréquentes et à moindre coût (Bappel 2005). Ces avantages ont fait de la télédétection un outil incontournable pour étudier l'occupation du sol à toutes les échelles.

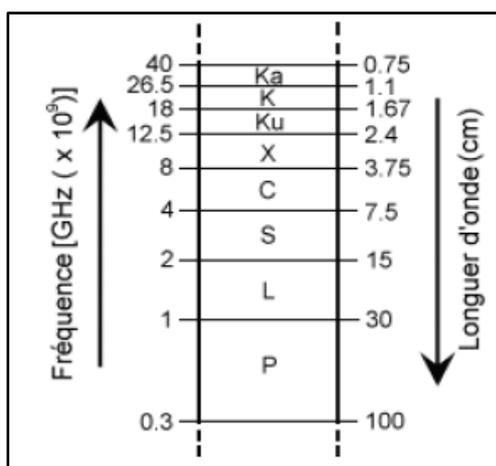
Les capteurs optiques sont capables d'enregistrer la réponse spectrale des surfaces observées dans le domaine du visible mais également dans d'autres parties du spectre électromagnétique riches en information (proche infrarouge, moyen infrarouge, etc). Chaque grand type d'occupation du sol ayant une signature spectrale différente (zone bâtie, végétation, zones humides), l'information radiométrique est donc particulièrement intéressante pour l'étude de l'occupation du sol (Guerschman, Paruelo et al. 2003).

Néanmoins, ces capteurs optiques sont très sensibles aux effets météorologiques. En effet, de par leur longueur d'onde, il est impossible pour un capteur optique d'obtenir une image de la surface terrestre sous couvert nuageux, limitant ainsi les prises de vue dans un milieu tropical tel que la Réunion. De plus, du fait d'une saisonnalité moins marquée (cycles culturels faibles synchronisés), les séries temporelles censées fournir de l'information, comme cela peut être effectué en milieu tempéré, y sont moins efficaces.

Tous ses facteurs, sont des contraintes qui nous ont poussés à privilégier l'utilisation des satellites Radar SAR.

### 2.3. Télédétection radar

Développé à partir des années 1970, la technologie radar (RADio Detection And Ranging) est une technologie permettant d'envoyer et de recevoir des signaux situés dans la longueur d'onde des micro-ondes. Les scientifiques ont divisé cette longueur d'onde en 8 « bandes », correspondant à des fréquences différentes, facilitant ainsi son utilisation (Figure 1).



**Figure 1 : Longueurs d'ondes radar (source : Centre canadien de télédétection, CCT).**

Les micro-ondes, de longueur d'onde plus grande que les ondes visibles et infrarouges, ne sont pas sensibles à la diffusion atmosphérique qui affecte les ondes plus courtes. Elles peuvent ainsi, traverser la couche nuageuse, la brume, la poussière et la pluie fine. Cette propriété permet la détection dans presque toutes les conditions atmosphériques, et donc l'acquisition de donnée sur la surface terrestre en tout temps, de jour comme de nuit.

Cette onde électromagnétique est caractérisée par une amplitude (taille de l'onde), une phase (position de l'onde à l'instant t) et une orientation verticale ou horizontale.

L'orientation des ondes émises et reçues permet de caractériser la polarisation qui est une propriété d'une onde électromagnétique plane (Maghsoudi 2011). On distingue 3 types de polarisation utilisés par les satellites radar SAR :

La polarisation simple, est l'utilisation d'une seule combinaison (émission H/ réception H (HH), HV, VH ou VV). Les travaux de Lee (Lee, Ainsworth et al. 2011) présentent son utilisation afin d'étudier l'occupation du sol. Permettant d'obtenir une zone d'étude plus vaste que les autres polarisations, elle permet de comparer les structures présentes au sein de cet espace, tout en caractérisant l'évolution de la végétation grâce à la grande répétitivité temporelle de ce mode.

La polarisation double, correspond à l'emploi de deux combinaisons (VV/VH, HH/HV, etc.). (Ainsworth, Kelly et al. 2009), présentent les différents avantages (coût moindre, répétitivité et fauchée plus importantes) de ce mode pour l'identification des cultures en milieu tempéré.

La polarisation complète, correspond à l'utilisation des 4 combinaisons provenant des composantes verticale et horizontale (HH/HV/VH/VV). Fournie par quelques satellites (RADARSAT-2, TERRASAR-X et ALOS-2), elle est l'objet d'un grand nombre de travaux récents (Jedlovec 2009), (Engelbrecht, Kemp et al. 2013). Elle fournit une information plus complète que les polarisations simple et double permettant une classification et une caractérisation des cultures plus fines (Lee, Grunes et al. 2001).

A l'heure actuelle, l'ensemble de ces polarisations est utilisé par la communauté scientifique. Néanmoins, très peu de travaux présentent l'utilisation de la polarimétrie, pour la classification et la caractérisation de l'occupation du sol dans un milieu tropical, tel que la Réunion. Les principaux

travaux effectués dans ce domaine datent de 2015. Les travaux réalisés dans la Péninsule de Leizhou en Chine, démontrent le potentiel de la polarisation complète afin de caractériser un espace agricole dominé par la canne à sucre en milieu subtropical (Li, Chen et al. 2015).

En parallèle à la polarimétrie, les scientifiques ont mis en avant l'emploi de la rétrodiffusion correspondant à l'analyse de l'amplitude du signal dans une combinaison radar (HH, HV, VH ou VV). Elle est utilisée à la Réunion dans les travaux de Baghdadi (Baghdadi et al. 2010) qui présentent la mise en œuvre de la rétrodiffusion pour le suivi de la croissance de la canne à sucre par le capteur TERRASAR-X. Il démontre ainsi l'utilité de la rétrodiffusion pour le suivi de l'état d'une culture (haute résolution, forte répétitivité temporelle).

L'information physique obtenue par ces méthodes reste complexe. Afin de rendre cette information facilement interprétable, des outils du domaine de la fouille de données tels que les modèles de classifieurs aident à traduire ces variables polarimétriques en variables « agronomiques » interprétables (classes d'occupation du sol).

## **2.4. Modélisation**

Depuis plusieurs années maintenant, un ensemble très large de classifieurs a imprégné le domaine de l'analyse spatiale et plus précisément de la télédétection (Tuia, Ratle et al. 2009), (Clevers, Van der Heijden et al. 2007).

### **2.4.1. PLS (Partial least square)**

Héritée de l'approche linéaire déterministe, la PLS développée par Wold (Wold 1968), est basée sur la combinaison de la régression linéaire et de l'ACP (Analyse en composante principale). C'est un modèle linéaire qui spécifie la relation entre une ou des variables dépendantes (réponses) et un ensemble de variables prédictives. Généralement utilisée dans la construction des modèles de prédiction, notamment lorsque les variables explicatives sont fortement corrélées entre elles (Barker and Rayens 2003), elle permet de prendre en compte un grand nombre de variables d'entrées.

En revanche, la PLS ne se prête pas à la modélisation statistique déterministe. Dans un cadre similaire, quand il s'agit d'interpréter et d'extraire les variables les plus pertinentes parmi un très grand nombre, elle s'avère peu utile et difficilement exploitable.

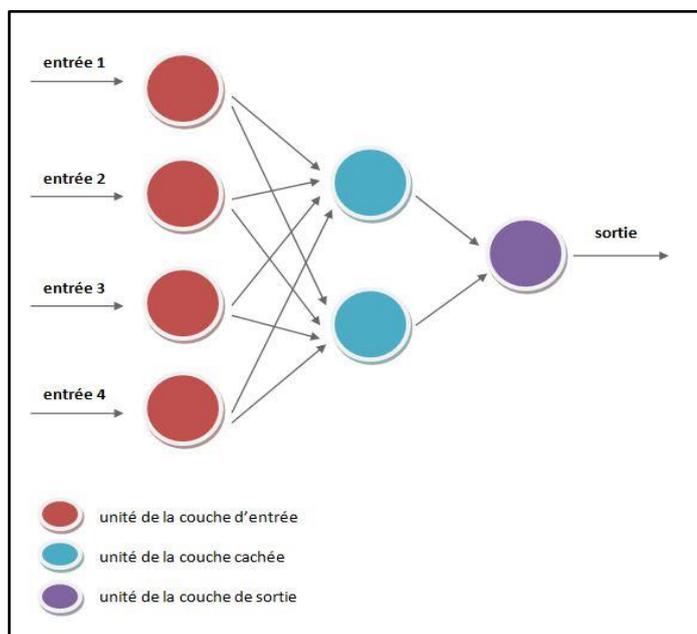
### **2.4.2. Réseaux de neurones**

Introduits pour la première fois par (McCulloch and Pitts 1943), les réseaux de neurones sont une technique de modélisation et de prévision permettant de simuler des fonctions particulièrement complexes (Han, Kamber et al. 2006). Un réseau de neurones est constitué par un assemblage de neurones élémentaires (ou formels) sous la forme de couches successives de plusieurs neurones.

Le neurone est constitué par une fonction élémentaire (seuil, linéaire, sigmoïde,...) qui traduit une valeur d'entrée en variable de sortie normée. Chaque neurone reçoit une variable d'entrée unique constituée d'une combinaison linéaire des variables d'entrée du système (pour la couche de neurones d'entrée) ou des sorties des neurones de la couche amont (Figure 2).

Les entrées arrivent au neurone par l'intermédiaire d'une connexion à laquelle est attribué un poids. Le calcul des coefficients de pondération est réalisé par la phase d'apprentissage. La sortie d'un

réseau de neurones peut ainsi reproduire une fonction complexe par association (combinaisons linéaires et enchainements) de fonctions élémentaires.



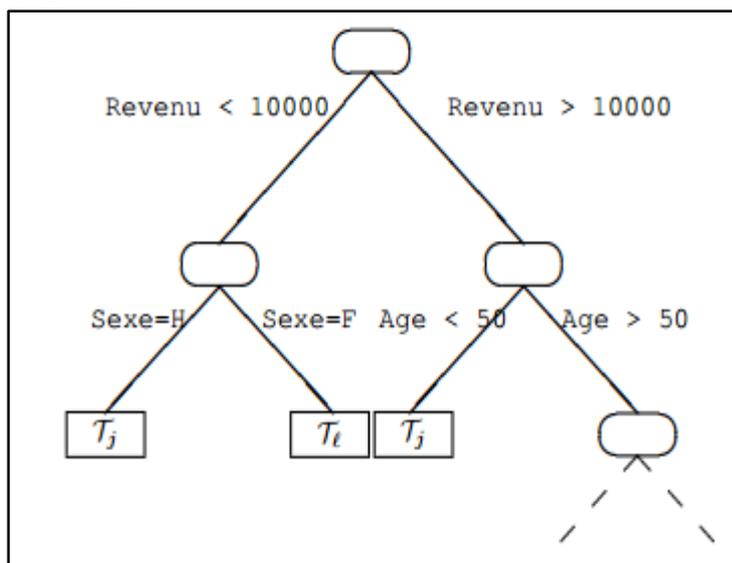
Très utilisés en télédétection (Atkinson and Tatnall 1997), les réseaux de neurones présentent une forte tolérance au bruit provenant des données satellitaire (Han, Kamber et al. 2012).

Toutefois, ils impliquent un temps d'apprentissage pouvant être assez long, et nécessitent également la détermination empirique d'un certain nombre de paramètres. De plus, ils sont incapables de fournir une explication des phénomènes appris (Han, Kamber et al. 2012).

**Figure 2 : Organisation d'un réseau de neurones (Source : Wikiversité).**

### 2.4.3. Random Forest

La technique des arbres de régression, inventée par Leo Breiman (Olshen and Stone 1984) est une méthode statistique basée sur la construction d'un arbre discriminant les échantillons en différents groupes leur attribuant une valeur ou une classe. Ces groupes sont construits à partir de règles binaires en s'appuyant sur les variables prédictives qui peuvent être quantitatives ou qualitatives (Figure 3).



Cette technique a été réadaptée en intégrant le Bagging (Breiman 2001). Cette méthode, ou voting pour la classification, permet d'améliorer les performances prédictives des arbres de décisions en construisant, non plus un seul arbre, mais plusieurs dizaines, voire des centaines (forêts aléatoires) dont les réponses sont combinées (moyennées) pour obtenir une classification plus robuste.

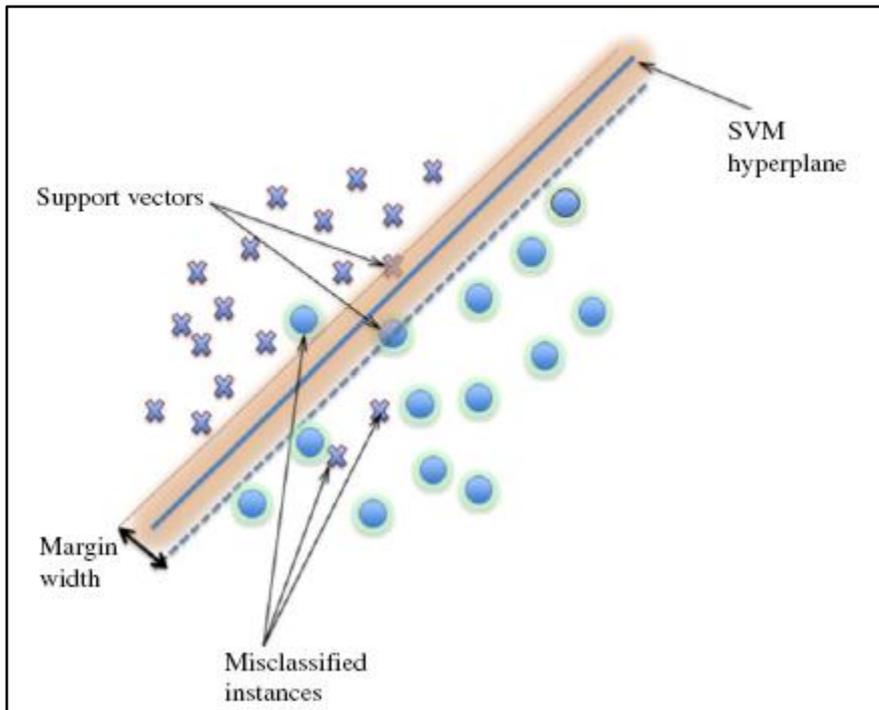
**Figure 3 : Arbre de décision élémentaire (Source : Wikistat).**

De plus en plus utilisé en télédétection (Pal 2005), la technique du Random Forest fonctionne bien dans des situations où le nombre de variables d'entrée explicatives est très important. Il nous renseigne également sur la participation de chacune des variables à la classification, grâce à une

notion d'importance. Toutefois, ce modèle perd en simplicité du fait du nombre de paramètres à régler pour obtenir la meilleure classification possible.

#### 2.4.4. SVM (Séparateur à vastes marges)

Introduits par Vapnik (Vapnik, Boser et al. 1992), les SVM sont un ensemble de techniques d'apprentissage supervisées, destinées à résoudre des problèmes de discrimination et de régression.



Les SVM sont basés sur la recherche d'un hyperplan avec les marges les plus larges possibles (Figure 4) qui, lorsque c'est possible, séparent correctement les données en classes dont les éléments sont le plus éloigné possible des autres les observations (Han, Kamber et al. 2012).

**Figure 4** : Exemple d'un séparateur à vastes marges (Source: (Mountrakis, Im et al. 2011)).

Ils fournissent des modèles très précis et fiables pour la prédiction de variables continues ou discrètes. A l'inverse des autres modèles, les SVM sont aussi moins sensibles au sur-ajustement (Han et al, 2012).

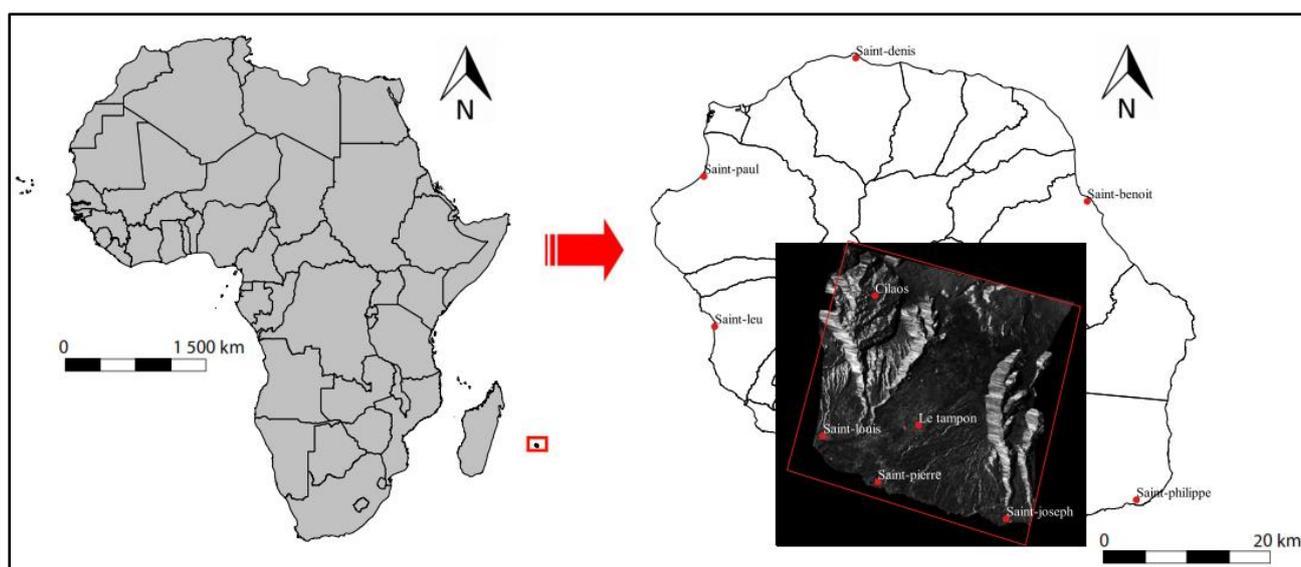
Nous avons décidé d'utiliser le classifieur Random Forest afin de sélectionner les variables explicatives à l'aide de la fonction d'importance. Pour les classifications, nous utiliserons le classifieur SVM, car ce dernier, de par ces caractéristiques reste le plus utilisés dans les études récentes menées en télédétection (Mountrakis, Im et al. 2011).

### 3. Matériels et méthodes

#### 3.1. Zone d'étude

Nous avons choisi une zone d'étude située dans le Sud-Ouest de l'île de la Réunion (Figure 5), contenue dans l'emprise d'images RADARSAT-2 en polarisation complète. Cet espace, représentatif de la diversité des cultures de La Réunion, a été choisie situé à proximité de la station de recherche du CIRAD de Saint-Pierre afin de faciliter les observations terrain. Cette zone, s'étend de l'ouest de Saint-Louis (55°24' de longitude Est), à l'est de Saint-Joseph (55°39' de longitude Est). En latitude, elle couvre du sud de Saint-Pierre (21°22' de latitude Sud) au Nord de la Plaine des Cafres (21°7' de latitude Sud).

La zone s'étend sur environ 700 km<sup>2</sup> (environ 25 km par 25 km) avec une altitude variant entre 0 et 2 500 mètres. Des images du satellite canadien RADARSAT-2 ont été obtenues par l'intermédiaire de la station de réception SEAS-OI (Surveillance de l'Environnement Assistée par Satellite pour l'Océan Indien) de Saint-Pierre de la Réunion, à la fois station de réception (SPOT 4 et 5 et RADARSAT-2) et centre de recherche en télédétection.



**Figure 5 : Localisation de la zone d'étude, Source : Image RADARSAT-2 de SEAS-OI (RADARSAT-2 Données et Produits © MacDONALD, DETTWILER AND ASSOCIATES LTD. (2014) – Tous droits réservés. RADARSAT est une marque officielle de l'Agence spatiale canadienne).**

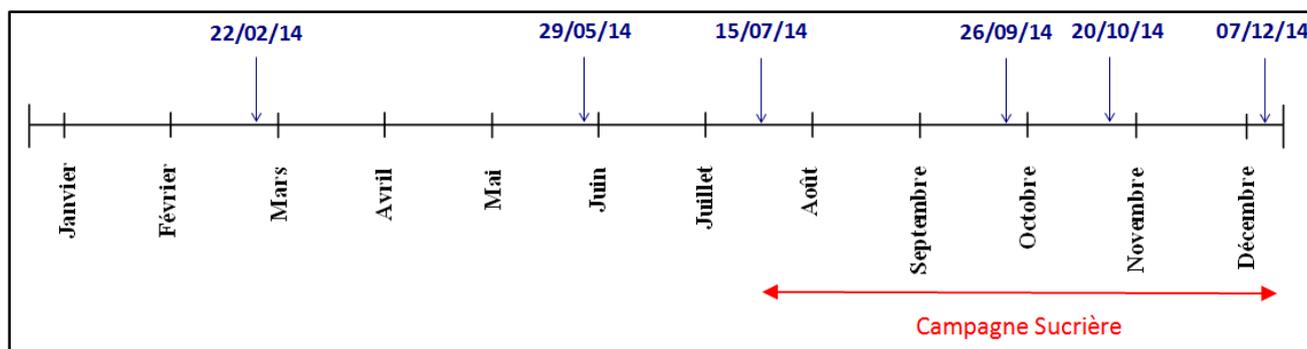
#### 3.2. Données et logiciels utilisés

##### 3.2.1. Données radars

Les données acquises par SEAS-OI sont disponibles gratuitement pour les institutions publiques et les laboratoires de recherche du sud-ouest de l'Océan Indien, dont le CIRAD.

Nous avons obtenu 13 images polarimétriques (en polarisation complète) acquises entre le 22 février 2014 et le 7 décembre 2014, en mode Fine Quad avec un angle d'incidence de 38°. Néanmoins, par

un souci de parcimonie, seulement 6 de ces images acquises à des périodes marquantes de l'année ont été traitées (Figure 6).



**Figure 6 : Répartition des images RADARSAT-2 obtenues, sur l'année 2014.**

### 3.2.2. Données vectorielles

Outre la donnée radar, nous disposons d'une base de données d'occupation des sols (BOS) élaborée à partir du registre parcellaire graphique, RPG. Ce dernier respecte les normes de la directive européenne, imposée par la politique agricole commune, PAC (Leo and Lemoine 2000), où la culture majoritaire de chaque îlot cultural est identifiée. Cette BOS fut réalisée, en 2013 par la DAAF.

### 3.2.3. Mesures terrain

Des observations du type d'occupation du sol ont été effectuées tout au long de cette étude afin de construire une base d'apprentissage et de validation pour développer le modèle de classifieur. Ces relevés de terrain ont été réalisés tous les 24 jours lors du passage satellite, à raison de deux journées par passage. Trois journées d'observations ont permis de valider cette base d'apprentissage. Au total, sur l'année 2014-2015, 29 journées terrain ont été effectuées.

### 3.2.4. Logiciels utilisés

Les logiciels utilisés (tableau 1) appartiennent tous à la catégorie des logiciels libres. Ils ont été choisis afin de faciliter le transfert futur de la méthode vers un opérateur local.

Concernant les traitements radar, nous avons calculé les variables polarimétriques des images en polarisation complète à l'aide du logiciel PolSARPro V4.2 et V5.0. Nous avons calculé les variables polarimétriques des images en polarisation double à l'aide du logiciel Sentinel 1.0.1.

Nous avons utilisé les logiciels Weka et Rstudio, pour réaliser les traitements car ces derniers permettent l'obtention de résultats de manière facile et rapide.

**Tableau 1 : Résumé des différents logiciels utilisés et leurs utilisations.**

<b>Logiciel</b>	<b>Version</b>	<b>Utilisation</b>
<b>ASF MapReady</b>	3.1	Correction et Géo-référencement des images radar
<b>PolsarPro</b>	4.2 et 5.0	Traitements des images en polarisation complète
<b>Sentinel-1 toolbox</b>	1.0.1	Traitements des images en polarisation double
<b>SAGA GIS</b>	2.1	Extraction statistique
<b>Qgis</b>	2.6.1	Visualisation et mise en forme de la base de données
<b>Rstudio</b>	3.0.1	Statistiques (ACP, Random Forest)
<b>Weka</b>	3.6	Classification (SVM (SMO))

### 3.2.5. Détermination des classes d'occupation de la sole agricole

Nous avons dans un premier temps classé la sole agricole afin de réaliser un modèle d'occupation agricole de notre zone d'étude. Dans un second temps, nous avons effectué une classification afin de différencier les zones agricoles, des zones dites « non-agricoles » (bâti, végétation naturelle).

Afin d'établir les grandes classes de notre première base d'apprentissage, nous nous sommes basés sur la BOS. Cette base d'occupation du sol était composée à l'origine de 19 classes d'occupation. Néanmoins, pour un souci de simplification et de réalisme, nous les avons regroupées en 6 grandes classes (Tableau 2).

**Tableau 2 : Présentation des différentes classes d'occupation de la sole agricole (Source : BOS 2013).**

<b>Nouvelles classes d'occupation</b>	<b>Anciennes classes d'occupation</b>	<b>Part de la SAU</b>
<b>Canne à sucre</b>	Canne à sucre	<b>58%</b>
<b>Prairie</b>	Elevage, Ensilage, Fourrage, Prairie	<b>26%</b>
<b>Maraîchage</b>	Aromate, Curcuma, Diversification, Fleur, Géranium, Légume, Mais, Vanille, Vétiver	<b>16%</b>
<b>Verger</b>	Fruit, Letchi, Verger	
<b>Ananas</b>	Ananas	
<b>Banane</b>	Banane	

## 3.3. Méthode

### 3.3.1. Prétraitements

#### *Mise en forme de la base d'apprentissage*

Nous avons vérifié et nettoyé la base d'apprentissage. En effet, la BOS, correspond à une déclaration annuelle volontaire réalisée par les agriculteurs de l'île. Or cette dernière est représentée par îlot cultural et peut parfois être synthétique ou obsolète pour certains ilots.

A l'aide du logiciel Qgis, nous avons affecté à l'ensemble des ilots un certain nombre de critères d'exclusion (Superficie, forme, situation de la parcelle) à partir d'une image très haute résolution Pléiades, permettant de minimiser les erreurs provenant de la BOS ([Annexe 2](#)). Nous avons

également effectué une vérification terrain de certaines parcelles de la sole agricole afin d'obtenir une base de données la plus fiable possible.

### ***Correction et Géo-référencement***

Les images fournies en format SLC (Single Look Complex), ne sont ni calibrées, ni géo-référencées. Nous avons orthorectifié la matrice de cohérence T3 (propre aux images en polarisation complète) en WGS 84, UTM (Universal Transverse Mercator) 40° Sud. Pour ce faire, nous avons utilisé le logiciel américain ASF MapReady et le MNT (modèle numérique de terrain) Litto3d à 5 mètres de résolution.

### ***Filtrage***

Toutes les images radar contiennent une certaine quantité de ce que l'on appelle le chatoisement (ou speckle) radar. Le chatoisement est produit par l'interférence aléatoire, constructive et destructive, provenant de la diffusion multiple qui se produit dans chaque cellule de résolution. Ce dernier provoque une variation de l'intensité pixel par pixel, se manifestant comme une texture poivre et sel sur les images (Lee and Pottier 2009).

Le chatoisement est essentiellement une sorte de bruit qui dégrade la qualité de l'image et peut rendre l'interprétation visuelle ou numérique plus difficile. Ainsi, il est généralement préférable de réduire le chatoisement avant l'interprétation et l'analyse des images. Pour se faire, il existe deux techniques permettent de le réduire :

- le traitement multi-visée, qui consiste en l'acquisition de plusieurs images distinctes d'une même scène.
- le filtrage spatial, qui consiste à extraire une fenêtre de quelques pixels au-dessus de chaque pixel de l'image, pour y réaliser une opération mathématique qui remplacera la valeur existant du pixel.

Nous nous intéresserons ici à la méthode du filtrage spatial. Cette dernière nous a permis de filtrer la matrice de cohérence T3 afin de limiter l'influence du chatoisement sur l'information extraite. Pour ce faire nous avons utilisé un filtre Lee Refined, avec une fenêtre de 3\*3, car ce dernier reste le plus utilisé et le plus performant à l'heure actuelle pour l'imagerie radar.

### **3.3.2 Calcul des variables radar**

Une fois la matrice de cohérence (T3) filtrée, nous avons pu extraire grâce à cette dernière un ensemble d'indicateurs radar correspondant aux variables d'entrée de notre modèle de classification.

### ***Décompositions***

Issues des travaux de Chandrasekhar, puis formalisées pour la première fois par Huynen (Huynen 1970), les décompositions radar ont pour but de fournir une interprétation des contraintes physiques présentées dans les images radar. Elles exploitent l'information additionnelle extraite des matrices de cohérence (T3) et de covariance (C3) pour obtenir le plus d'informations possibles des images polarimétriques (Lee and Pottier 2009).

Depuis leur formalisation, ces décompositions n'ont cessé d'évoluer. A ce jour, il en existe 4 types principaux :

- Les décompositions basées sur la dichotomie de la matrice de Kennaugh K (Zebker and Van Zyl 1991) (Hyunen, Holm & Barnes, Yang).
- Les décompositions basées sur la décomposition de la matrice de cohérence T3 et de covariance C3 (Freeman & Durden, Yamaguchi, Dong).
- Les décompositions utilisant l'analyse des valeurs et vecteurs propres des matrices de covariance et cohérence (Cloude, Holm, Van Zyl, Cloude & Pottier).
- Les décompositions employant la décomposition cohérente de la matrice de diffusion S (Krogager, Cameron, Touzi).

Dans cette étude, nous nous concentrons sur les 3 derniers types, les plus utilisés en imagerie radar soit 5 décompositions au total.

- **Freeman & Durden**

Cette méthode permet de mettre en place un modèle basé sur 3 mécanismes de diffusion sans mesures de terrain: la diffusion de volume (ou de la canopée), la diffusion en double rebonds et la diffusion de surface (représentant la rugosité) (Lee and Pottier 2009). Elle permet de déterminer pour chaque pixel la puissance totale de ces différentes diffusions.

- **Yamaguchi**

La décomposition de Yamaguchi, basée sur les travaux de Freeman & Durden ainsi que sur l'expansion de la matrice de covariance (Yamaguchi, Moriyama et al. 2005), met en évidence un nouveau mécanisme de diffusion : la diffusion « Helix ». Elle représente la puissance du mécanisme principal de diffusion (volume, surface ou double rebonds) dans les 4 composantes de polarisation pour chaque élément de l'image.

- **Cloude & Pottier**

La décomposition de Cloude & Pottier se base sur la décomposition des valeurs et vecteurs propres des matrices de cohérence et de covariance (Cloude and Pottier 1996). Elle permet de calculer trois mécanismes :

- l'entropie, H, représentant le degré d'incohérence de la diffusion dominante de l'image.
- l'angle Alpha, variant entre  $0^\circ$  et  $90^\circ$ , permet d'identifier les différents types de diffusion ( $0^\circ$  pour la diffusion de surface,  $45^\circ$  pour la diffusion de volume et  $90^\circ$  pour la diffusion en double rebonds).
- l'anisotropie, A, complémentaire à l'entropie, représente la quantité de mélange entre le deuxième et troisième mécanisme de diffusion.

- **Van Zyl**

La décomposition de Van Zyl, correspond à la première décomposition de Cloude (Lee and Pottier 2009) adaptée aux images radar polarimétriques. Basée sur la décomposition des valeurs propres de la matrice de covariance, elle permet de mettre en évidence 3 mécanismes de diffusion : la diffusion de volume (canopée), la diffusion de surface (rugosité) et la diffusion en double rebond.

- **Krogager**

La décomposition de Krogager, se base sur la décomposition de la matrice de diffusion S, correspondant à la diffusion d'une cible « simple », afin d'extraire 3 composantes cohérentes. Ces trois composantes sont, la décomposition de type Sphère (représentant une diffusion de volume ou de la canopée), la décomposition Diplane (représentant une diffusion de surface, la rugosité) et la décomposition Hélix (identique à la décomposition de Yamaguchi).

### *Texture*

Elle est décrite comme l'effet visuel créé par la fréquence spatiale et l'amplitude des changements des niveaux de gris (texture fine ou rugueuse, texture spongieuse ou ouateuse, etc.) ou par l'arrangement d'agrégats d'objets (de brillance et de dimension quelconque) formant des motifs structurés ou non (texture géométrique ou aléatoire, etc.).

Simonett et Davis (1983) sont parmi les premiers à définir la notion de la texture sur des images radar. Leur discussion se limite à des images radar des milieux naturels. Ils reconnaissent alors trois composantes de texture. La micro-texture est inhérente à l'imagerie radar et est associée au bruit du chatolement. La méso-texture est produite par des hétérogénéités spatiales s'étendant sur plusieurs pixels (plusieurs unités de micro-texture). Finalement, la macro-texture est reliée surtout à la présence des accidents topographiques (alternances pentes éclairées-ombres) ; elle est clairement observable lorsque, sur la même image, des unités géomorphologiques différentes sont présentes (Codjia 2014).

Ces différentes textures, sont regroupées selon plusieurs paramètres basés sur leur similitude et leur formulation mathématique. En analysant l'espace d'étude, nous avons décidé de nous concentrer sur uniquement trois de ces paramètres :

- Les paramètres moyenniers qui visuellement qualifie une texture brillante ou sombre. Nous retiendrons uniquement la Moyenne.
- Les paramètres de dissemblance et de ressemblance, nous retiendrons dans cette catégorie le Contraste et la Dissimilarité.
- Les paramètres de désordre dans lesquels nous retiendrons uniquement l'Entropie.

Ces textures ont ensuite été appliquées aux différents indicateurs radar.

### *Indices radar*

Nous avons calculé en plus des décompositions à partir des matrices de cohérence filtrées, des indices et des éléments radar nous paraissant intéressants à aborder :

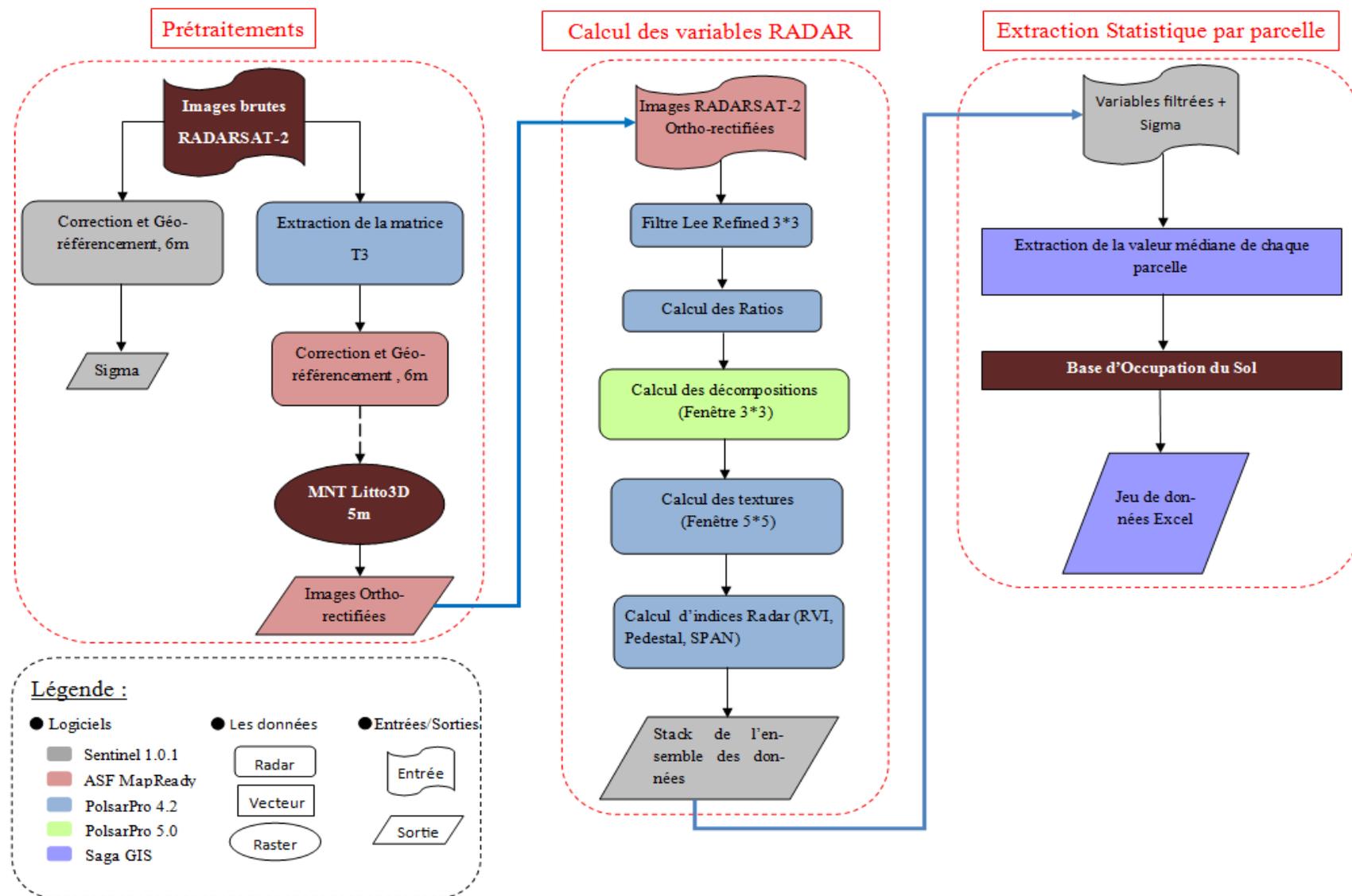
- L'indicateur RVI (Radar Vegetation Index) qui est un indice de végétation basé sur les coefficients de rétrodiffusion des différentes polarisations ( $RVI = \frac{8\sigma_{HV}}{\sigma_{HH} + \sigma_{VV} + 2\sigma_{HV}}$ ), principalement utilisé pour mettre en évidence l'eau dans les plantes (Kim, Jackson et al. 2012).
- le Pedestal Height qui est un indicateur du degré de polarisation d'un élément (McNairn, Duguay et al. 2002).
- le SPAN qui représente la puissance totale des 9 éléments de la matrice pour chaque pixel de l'image.

### *Variables d'intensité*

Nous avons calculé les ratios d'intensité (HH/VV, VH/HH et VH/VV) à partir du logiciel PolSarPro, ainsi que le coefficient de rétrodiffusion (normalisé) d'une cible radar, (Sigma) pour chacune des polarisations (HH, VV, VH, HV) à partir du logiciel Sentinel-1 Toolbox.

### *Extraction de l'information parcellaire*

Nous avons extrait les valeurs médianes, étant plus robustes aux valeurs extrêmes, de chacune de ces variables à l'intérieur de chaque parcelle de la base de données BOS.



**Figure 7 : Organigramme général de la méthodologie radar.**

### 3.3.3. Modélisation

L'information physique extraite des variables radar, est très riche et difficilement interprétable en l'état. Afin de pouvoir interpréter cette information, plusieurs étapes furent nécessaires.

#### *Exploration de la base d'apprentissage*

Nous avons exploré la base afin de construire le classifieur le plus adapté. Nous avons ainsi réalisé, une ACP afin d'observer les corrélations entre les variables et les individus. Cette première analyse nous permet également de déterminer si l'ACP permet de construire un classifieur efficace de notre base d'apprentissage (en attribuant un poids aux différentes composantes de cette dernière).

3 étapes ont été nécessaires :

- Une analyse de la répartition des individus en fonction des différentes classes. La représentation des individus ([Annexe 3](#)) montre une confusion entre certaines classes d'occupation. C'est pourquoi nous décidons de distinguer 4 classes d'occupation (Banane/Canne/Prairie/Autre) représentant l'ensemble de la SAU.
- Une analyse de la répartition des individus en fonction des composantes de l'ACP. Les résultats ([Annexe 4](#)), nous ont permis de déterminer que ces composantes n'étaient pas suffisantes (ne permettaient de séparer les différentes classes d'occupation) afin de réaliser une classification conforme aux exigences de la sole agricole.
- Une analyse des corrélations entre variables. Le résultat ([Annexe 5](#)) nous a permis de mettre en évidence les corrélations existantes entre les variables et ainsi de déterminer le classifieur SVM, comme étant le plus efficace ([voir partie 2.4.4](#)) pour classer la base d'apprentissage.

#### *Elaboration du modèle de classification de la sole agricole*

L'exploration de la base d'apprentissage effectuée, nous avons déterminé que l'ACP ne suffisait pas afin d'établir une classification de la sole agricole conforme aux attentes. La base d'apprentissage dont nous disposons, est composée d'un grand nombre de variables explicatives fortement corrélées entre elles. Il a donc été nécessaire d'utiliser une méthode s'articulant autour de 2 classifieurs répondant aux critères de cette base d'apprentissage (Random Forest et SVM) afin d'obtenir un modèle le plus opérationnel possible.

- **Classification avec l'ensemble des variables**

Nous avons effectué une première classification de la base d'apprentissage, regroupant l'ensemble des 64 variables explicatives, grâce à l'algorithme SMO (Sequential minimal optimization) qui reprends le principe des SVM ([voir partie 2.4.4](#)) implémenté par John Platt (Platt 1999) avec le logiciel Weka (Waikato Environment for Knowledge Analysis).

Nous avons estimé la précision de cette classification par une validation croisée. Elle consiste à diviser aléatoirement notre base d'apprentissage en 10 sous-jeux de données d'apprentissage et de validation. Chaque sous-jeu est divisé en 2 parties, 90% pour l'apprentissage et 10% pour la validation.

Les objectifs de cette validation croisée consistent à optimiser les paramètres du modèle, et à calculer l'erreur de prédiction, afin de la minimiser. Cette erreur de prédiction est estimée par la moyenne des RMSE (Root Mean Square Error) obtenue sur les différents sous-jeux de données de validation (Han, Kamber et al. 2012).

Parallèlement, nous avons calculé la matrice de confusion de notre modèle. Elle fournit le taux de réussite de la classification. Elle permet d'obtenir l'indice de Kappa, qui est une synthèse du taux de réussite obtenue pour notre modèle.

- **Classification avec 15 variables différentes par date**

Nous avons cherché ensuite à réduire le nombre de variables explicatives afin de diminuer le temps de calcul. Pour cela, nous avons réalisé un Random Forest ([voir partie 2.4.3](#)) afin d'estimer l'importance de chacune des variables dans la classification. Nous avons identifié les 15 variables les plus explicatives pour chaque date (15 variables différentes selon chaque date) que nous avons retenues dans une deuxième phase.

Nous avons ainsi pu effectuer une nouvelle classification à partir des jeux de données composés de ces 15 variables différentes pour chacune des dates (soit un total de 33 variables différentes ([Annexe 6](#))).

- **Classification avec 15 variables identiques pour l'ensemble des dates**

Le modèle de classification établi, nous devons le rendre opérationnel, en réduisant notamment le temps de calcul nécessaire afin de le mettre en place. Pour se faire, nous avons cherché à réduire le nombre de variables explicatives afin de diminuer le temps de calcul.

A l'image de la précédente classification, nous avons réalisé un Random Forest sur chacune des 6 dates. Nous avons pu ainsi, identifier les 15 variables les plus explicatives (coefficient de rétrodiffusion, élément de la matrice de cohérence, décomposition et textures) ([Annexe 7](#)) avec un temps calcul limité.

Une fois les 15 variables identiques pour l'ensemble des date obtenues, nous avons réalisé une nouvelle classification pour observer les différences de précision pouvant exister entre les 2 modèles (celui avec 15 variables uniques et celui avec 15 variables différentes par dates) en comparant le pourcentage de classification, ainsi que l'indice de Kappa de ces derniers.

### ***Elaboration d'un modèle de cartographie des classes générales d'occupation du sol***

Nous avons aussi réalisé un modèle de cartographie des classes générales d'occupation du sol permettant de discriminer les zones agricoles du reste de l'occupation du sol (le bâti et la végétation naturelle). Dans les zones où l'information parcellaire reste limitée, en se basant sur une segmentation, cette méthode permet d'identifier et d'extraire la sole agricole des autres occupations générales du sol. Une fois la classe sole agricole extraite, il est possible de réaliser un modèle de classification de cette dernière.

Afin de mettre en place ce modèle pouvant intéresser certains décideurs, nous avons réalisé une classification sur une nouvelle base d'apprentissage, avec l'algorithme SMO, représentant les 3 classes d'occupation du sol (bâti, végétation naturelle, surface agricole).

### *Extension à la polarisation double*

Pour réduire les coûts de production, nous avons cherché à adapter notre méthode de classification à l'aide d'images dont l'accès est libre et gratuit. Nous avons donc opté pour des images provenant du satellite radar européen SENTINEL-1 qui est également un satellite en Bande C (comme RADARSAT-2). Les images obtenues grâce à ce satellite sont des images en polarisation double. Il a donc été nécessaire d'adapter le modèle avec des indicateurs calculés à partir de polarisation double et non complète.

En nous basant sur des travaux effectués dans ce domaine (Ainsworth, Kelly et al. 2009), et afin d'obtenir un modèle le plus opérationnel possible nous avons opté pour l'extraction des indicateurs simple en double polarisation. Nous avons donc extrait à l'aide du logiciel SENTINEL-1 Toolbox les coefficients de rétrodiffusion (Sigmas), l'angle d'indidence projeté ainsi qu'un ensemble de textures regroupé en 4 classes :

- Les paramètres moyenniers qui visuellement qualifient une texture brillante ou sombre, nous retiendrons classerons la Moyenne.
- Les paramètres de dissemblance et de ressemblance, nous retiendrons dans cette catégorie le Contraste, la Dissimilarité et l'Homogénéité.
- Les paramètres de désordre dans lesquels nous retiendrons l'Entropie, l'Energie et le deuxième moment angulaire (ASM) souvent comparée à l'énergie.
- Les paramètres d'écart à la centralité dans lesquels nous retiendrons la Variance, la Corrélacion et le Maximum

Une fois ces indicateurs calculés, nous avons appliqué les mêmes méthodes utilisées pour l'élaboration des précédents modèles de classification.

## 4. Résultats

Nous présentons dans cette partie uniquement les résultats essentiels et les plus marquants. Les résultats secondaires, sont présentés en annexes.

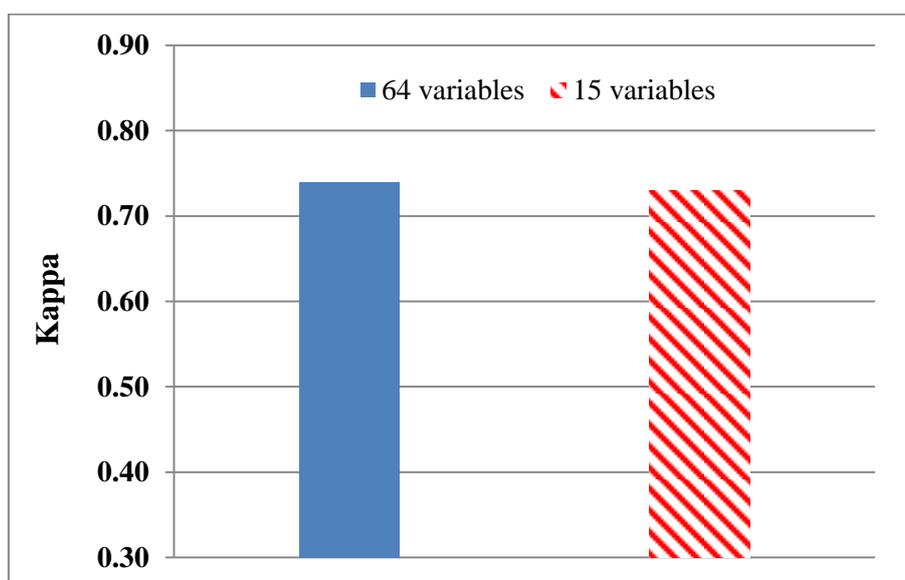
### 4.1. Modèle de classification de la sole agricole

#### 4.1.1 Modèle à 64 variables

La figure 8, compare la précision de classification des cultures par 2 modèles, l'un utilisant une base d'apprentissage complète (64 variables) et l'autre, une base d'apprentissage réduite (15 variables), à la date du 20 octobre 2014.

Ce résultat montre une précision légèrement meilleure pour le modèle à 64 variables avec un Kappa de 0.74 par rapport au modèle à 15 variables possédant un Kappa de 0.73.

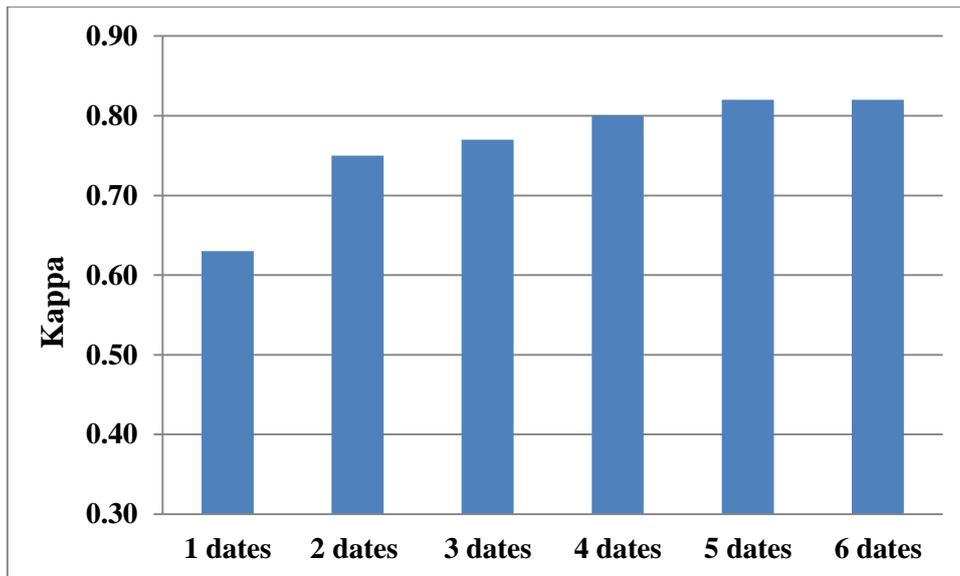
Toutefois, l'écart entre ces 2 modèles reste très faible. C'est pourquoi, nous avons décidé de conserver le modèle à 15 variables différentes pour chacune des dates car ce dernier demande un temps de calcul (calcul des variables et du modèle) moins important.



**Figure 8** : Comparaison de l'indice de Kappa entre le modèle à 64 et celui à 15 variables

#### 4.1.2 Modèle à 15 variables différentes par date

Une fois le choix du modèle établi, une classification fut réalisée à partir de ce modèle. La figure 9 présente les résultats du premier modèle à 15 variables en fonction du nombre de dates. Nous observons une augmentation de la précision avec le nombre de dates. La précision maximale est atteinte pour un modèle à 5 dates avec un Kappa atteignant 0.82. Au-delà, nous observons une stagnation de l'indice de Kappa. Nous retenons donc le modèle à 5 dates pour discriminer les classes Banane/Canne/Prairie/Autre.

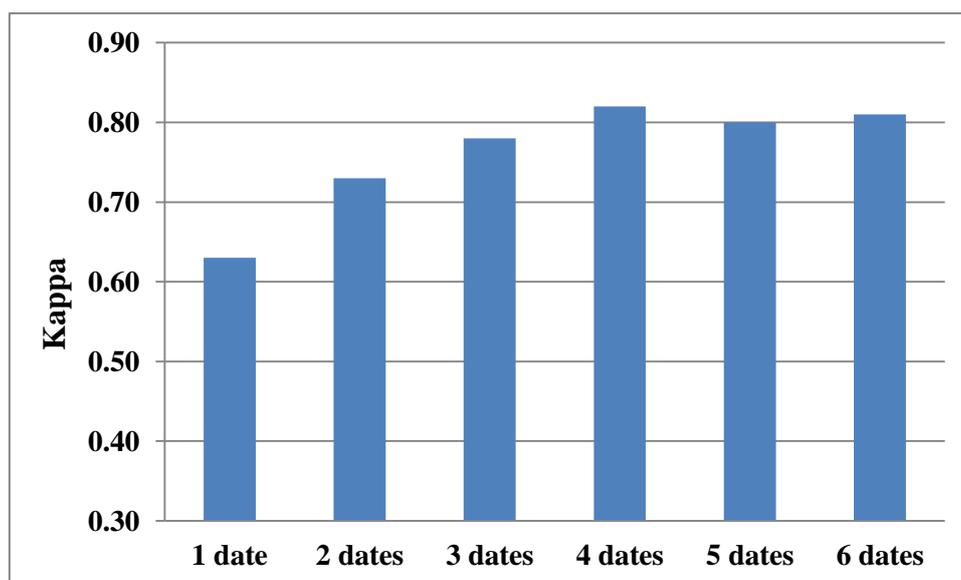


**Figure 9** : Indice de Kappa pour le modèle à 15 variables selon le nombre de dates.

#### 4.1.3 Modèle à 15 variables identiques pour toutes les dates

Il a été nécessaire par la suite de rendre opérationnel notre modèle. Ce dernier, à l'inverse du premier modèle à 15 variables, utilise des variables identiques d'une date à l'autre. La figure 11, présente donc les résultats de ce second modèle à 15 variables, en fonction du nombre de dates. Nous observons, à l'image du premier modèle, une augmentation de sa précision en avec le nombre de dates. La précision maximale est atteinte pour un modèle à 4 dates avec un Kappa atteignant 0.82. Nous observons également, une stagnation de cette évolution à partir de la quatrième date. Nous décidons de retenir le modèle à 4 dates dont la précision reste meilleure.

En comparant les 3 premières classifications, nous pouvons observer des précisions assez semblables pour ces 3 modèles. Toutefois, pour un souci d'opérationnalité, nous préférons nous orienter vers le modèle à 15 variables uniques dont le temps de calcul est moins important.



**Figure 10** : Indice de Kappa pour le modèle à 15 variables uniques selon le nombre de dates.

La matrice de confusion (figure 11), présente les résultats de ce modèle à 15 variables identiques pour 4 classes d'occupation du sol et 4 dates. Nous observons des différences de précision entre les classes. La précision des classes Canne et Prairie avoisine les 90% avec une confusion s'effectuant principalement avec la classe d'occupation « Autre » dont la précision avoisine les 83%. La plus faible précision, est atteinte pour la classe Banane dont la précision globale avoisine les 60%, cela étant due à une forte confusion avec la classe d'occupation « Autre ».

Notre modèle final, atteint donc une précision globale de 88% et un Kappa de 0.82, présentant ainsi son efficacité pour la classification des cultures tropicales.

		Objets de référence				Total	Précision producteur (%)	Précision utilisateur (%)
		autre	banane	canne	prairie			
Objets classés	autre	513	10	67	34	624	83.824	82.212
	banane	17	21	0	0	38	65.625	55.263
	canne	49	1	782	13	845	90.719	92.544
	prairie	33	0	13	426	472	90.063	90.254
	Total	612	32	862	473	1979	Précision globale = 88.02% Kappa = 0.82	

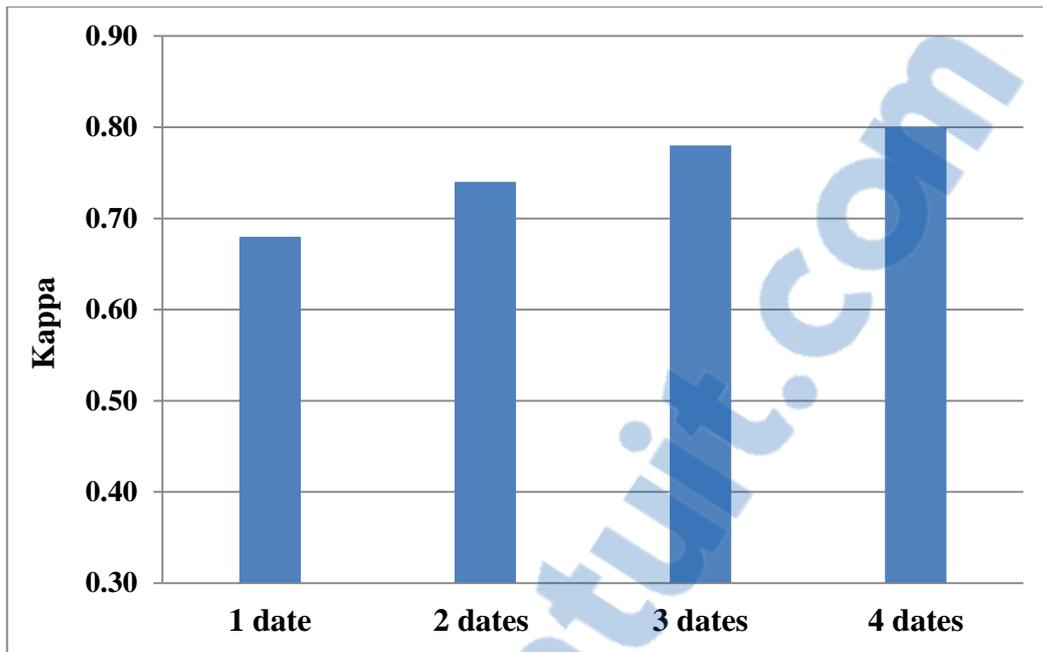
**Figure 11** : Matrice de confusion du modèle de 4 classes d'occupation basé sur 4 dates.

## 4.2 Modèle de cartographie des classes générales d'occupation du sol

La figure 12, présente les résultats du modèle de cartographie des classes générales d'occupation du sol en fonction du nombre de dates pour 3 classes d'occupation (Bâti, végétation naturelle, cultures). Nous observons une augmentation de sa précision en fonction de l'ajout de dates. La précision maximale est atteinte pour un modèle à 4 dates avec un Kappa atteignant 0.80.

Cette précision, pourra permettre à l'aide d'une segmentation d'extraire la classe cultures des 2 autres classes et ainsi d'appliquer le modèle de classification de la sole agricole sur cette dernière.

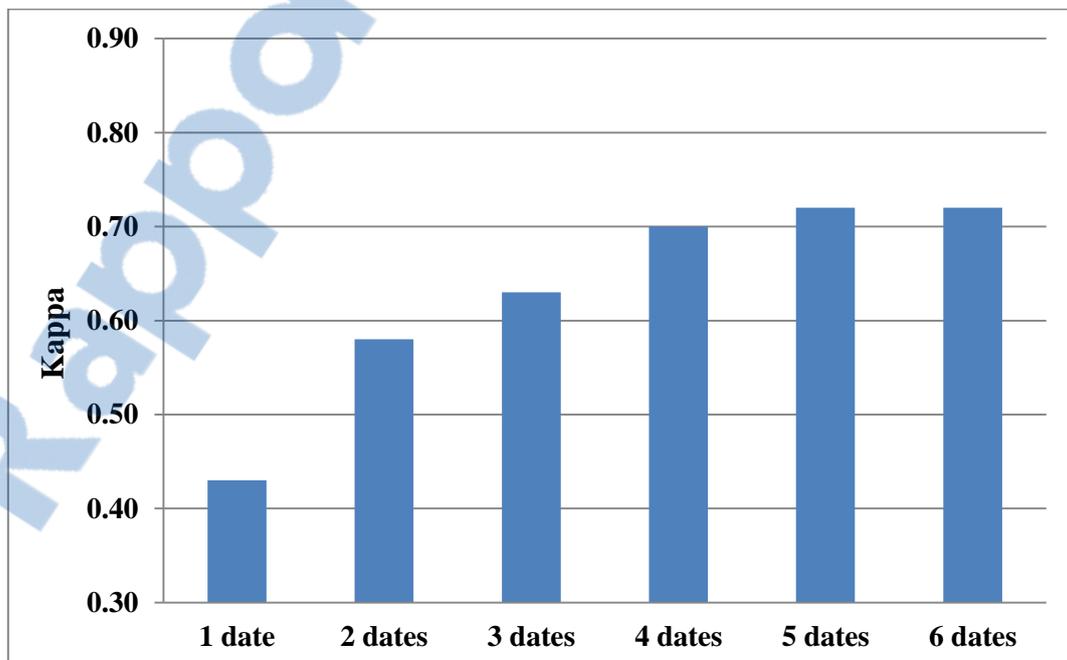
Ce modèle permet d'envisager un possible transfert de compétences de notre méthode dans des zones géographiques où l'information est limitée.



**Figure 12 :** Indice de Kappa pour le modèle général selon le nombre de dates.

### 4.3. Modèle de classification en polarisation double

La figure 13 présente les résultats du modèle basé sur des indicateurs mesurés en polarisation double, en fonction du nombre de dates. Nous observons comme précédemment, une augmentation de sa précision en fonction du nombre de dates. La précision maximale est atteinte pour un modèle à 5 dates avec un Kappa atteignant 0.72. Nous retenons donc le modèle à 5 dates, distinguant les classes Banane/Canne/Prairie/Autre.



**Figure 13 :** Indice de Kappa pour le modèle en polarisation double selon le nombre de dates.

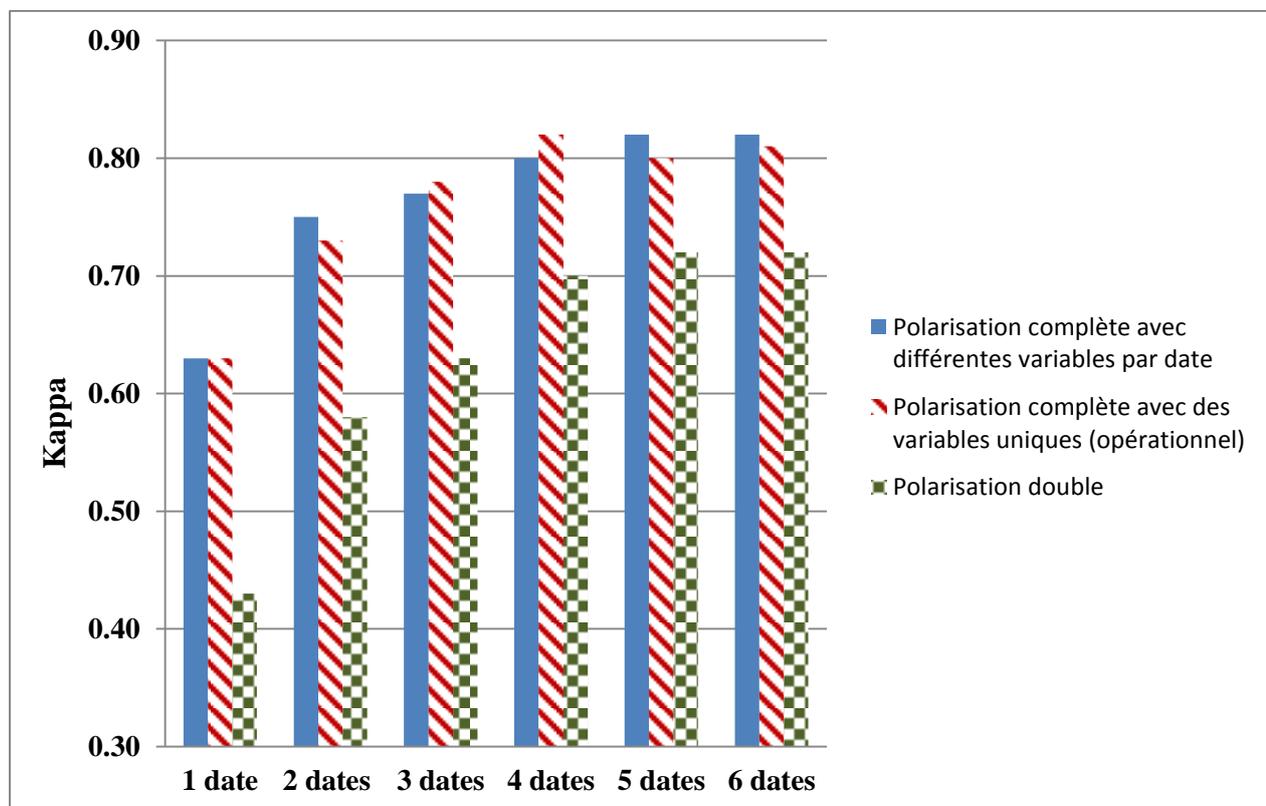
## 4.4 Comparaison des différents modèle

La figure 14, présente une comparaison des résultats des 3 modèles de classification des cultures : le modèle en polarisation complète à 15 variables différentes, le modèle en polarisation complète à 15 variables identiques et le dernier en polarisation double, en fonction du nombre de dates. Nous observons, comme attendu, que la précision augmente avec le nombre de dates.

Nous constatons que les 2 modèles basés sur la polarisation complète obtiennent une précision supérieure avec un Kappa maximal de 0.82. La précision moins importante du modèle en polarisation double peut s'expliquer par le fait que cette dernière fournit moins d'informations que la polarisation complète. Toutefois, au vu du temps de calcul, il est préférable d'utiliser le deuxième modèle (le modèle opérationnel) qui demande moins d'indicateurs radar (moins de calcul d'indicateurs) et est donc moins long à mettre en œuvre.

De plus, à la différence des images en polarisation complète, les images en polarisation double sont disponibles librement. Nous pouvons donc nous poser la question de l'opérationnalité de ce modèle par rapport à un modèle en polarisation complète qui nécessite l'achat d'images.

En effet, malgré des résultats moins importants, les caractéristiques de la polarisation double (grande surface d'étude : Toute l'île, coût moins élevé) font de cette dernière un outil intéressant pouvant permettre de développer une donnée préliminaire pour la mise en place du mode d'occupation des sols



**Figure 14** : Indice de Kappa pour les différents modèles selon le nombre de dates.

## 5. Discussion

---

### *Approche nouvelle*

L'objectif de cette étude, est d'élaborer une méthode d'identification et de cartographie du Mode d'Occupation des Sols (MOS) par imagerie radar grâce à des images SAR en polarisation complète.

A ce jour, la cartographie du MOS à la Réunion, est effectuée par photo-interprétation. Cette cartographie nécessite énormément de temps de travail (environ 800 heures pour la Réunion). L'approche présentée, permet aux décideurs d'identifier les grands types d'occupation agricole (Canne à Sucre, Prairie, Maraichage, Banane, Ananas, Verger) avec une précision intéressante de 88%, tout en réduisant le temps nécessaire à sa réalisation.

Même si pour l'heure, la finesse des classes ne respecte pas les prérequis pour l'élaboration des documents d'occupation du sol tel que le Schéma d'aménagement régional (SAR), les résultats obtenus restent satisfaisants et innovants dans un espace d'étude où le relief, le parcellaire très hétérogène rendent la cartographie par satellites radar très difficile (notamment pour la haute résolution spatiale).

En effet, cette approche préliminaire démontre le potentiel de ces images pour la classification des cultures tropicales. Ce sont les premiers travaux par imagerie SAR recensés à la Réunion sur cette thématique. En effet, jusqu'à présent les principaux travaux à La Réunion sur le MOS ont été réalisés à partir de satellites optiques à haute résolution spatiale (El Hajj, Bégué et al. 2009) voire très haute résolution spatiale (Bellon de la Cruz 2014). Malgré la bonne performance de ces travaux, ils restent toutefois confrontés aux limites des capteurs optiques ([voir Partie 2.2](#)).

Aucun travail à ce jour ne présente la cartographie du MOS par polarimétrie radar. Néanmoins, des travaux proches de ceux entrepris durant ce stage ont été réalisés cette année en Chine. Les travaux de (Li, Chen et al. 2015), présentent l'utilisation de la polarisation complète de TERRASAR-X pour la caractérisation de l'état de la canne à sucre dans un milieu subtropical (péninsule de Leizhou dans le sud de la Chine) à partir du capteur radar TERRASAR-X (bande X). Toutefois, ces derniers restent focalisés sur la caractérisation de la culture de la canne à sucre (se concentrant uniquement sur la période de croissance de la canne à sucre, le tallage) en délaissant la caractérisation des autres cultures tel l'eucalyptus, la banane...

### *Limites et contraintes*

Au cours de cette étude, nous avons utilisé un certain nombre de données produites par différents partenaires. Si l'utilisation de ces données est essentielle, elle peut aussi entraîner des biais (erreurs de digitalisation, problème de projection, erreur sur les traitements réalisés). C'est pourquoi, nous avons dû faire attention quant à l'utilisation et l'analyse de ses données.

Nous avons été également confrontés à un problème au niveau de l'opérationnalité du modèle. En effet, afin de rendre notre modèle le plus opérationnel possible, il est nécessaire de réduire le temps et les coûts nécessaires pour sa mise en place.

C'est pourquoi, nous avons développé un modèle basée sur des images Sentinel-1 disponibles gratuitement à l'inverse des images RADARSAT-2 dont la disponibilité est gérée par l'agence canadienne MacDonald, Dettwiller and Associate LTD. Cependant, la précision des résultats obtenue reste moins importante que celle obtenue grâce aux images RADARSAT-2. Toutefois, cela reste

encourageant pour des perspectives futures d'utilisation de la polarisation double. En effet comme nous l'évoquions précédemment, certes la précision de cette polarisation reste plus faible mais le coût et la fauchée de l'image est intéressante pour la mise en place d'une approche préliminaire pour aider la photo-interprétation mis en place actuellement.

### *Perspectives*

Les résultats que nous avons obtenus jusqu'à présent restent très satisfaisants. Néanmoins, il reste nécessaire pour la suite de ce projet de vérifier si les modèles mis en place que ce soit en polarisation double ou complète permettent d'aider les décideurs dans l'identification du mode d'occupation des sols.

Il sera nécessaire également de réaliser un modèle de cartographie des classes générales d'occupation du sol basé sur la polarisation double afin d'observer si cette dernière permet également à l'image de la polarisation complète de séparer ces grandes classes d'occupation.

Enfin, il nécessitera l'export des différents modèles effectuer jusqu'alors dans des espaces différents que celui abordé à travers notre zone d'étude afin de tester la fiabilité du modèle.

## **6. Conclusion**

---

Cette étude avait pour but d'évaluer le potentiel des images satellitaires SAR pour la cartographie de l'occupation de la sole agricole à la Réunion. Le choix a été fait d'utiliser des produits RADARSAT-2 en polarisation complète couvrant une zone dans le sud-ouest de la Réunion, représentative des enjeux fonciers de l'île.

La méthodologie développée a permis de discriminer efficacement les 2 cultures majoritaires de la sole agricole (canne à sucre, prairie) par une modélisation statistique, utilisant des classifieurs (SVM et Random Forest). Ces 2 cultures, représentant environ 90% de la SAU réunionnaise, ont été bien discriminées les unes des autres avec un kappa atteignant 0.82 et une précision globale de 88%.

Les résultats démontrent ainsi la performance des images SAR (RADARSAT-2) en polarisation complète pour discriminer les différentes cultures des milieux agricoles tropicaux fragmentés, caractérisés par des systèmes cultureux très hétérogènes et des parcelles de petite taille, tels que celles du territoire réunionnais.

Afin d'étendre l'opérationnalité de l'approche étudiée, une méthode basée sur des images disponibles librement (Sentinel-1) en polarisation double a également été testée. Les résultats, certes de précision moindre restent encourageants pour la suite du projet.

Le MOS actualisé est une donnée fortement attendue par les collectivités locales qui doivent élaborer les documents réglementaires tels que le SAR, SCOT (Schéma de Cohérence Territoriale) et PLU (Plan Local d'Urbanisme). Il est donc nécessaire d'affiner et de fiabiliser la méthodologie présentée tout en continuant à diriger les efforts de recherche vers de nouvelles méthodologies qui profitent des technologies de pointe, contribuant ainsi à une meilleure gestion du développement territorial.

## Bibliographie

---

- Ainsworth, T. L., J. P. Kelly, et al.** (2009). "Classification comparisons between dual-pol, compact polarimetric and quad-pol SAR imagery." Isprs Journal of Photogrammetry and Remote Sensing 64(5): 464-471.
- AGRESTE** (2013). Recensement agricole 2013. <http://www.agreste.agriculture.gouv.fr/en-region/reunion>
- Atkinson, P. M. and A. Tatnall** (1997). "Introduction neural networks in remote sensing." International Journal of remote sensing 18(4): 699-709.
- Baghdadi, N., R. Cresson, et al.** (2010). "Multitemporal Observations of Sugarcane by TerraSAR-X Images." Sensors 10(10): 8899-8919.
- Bappel, E. A.** (2005). Apport de la télédétection aérospatiale pour l'aide à la gestion de la sole cannière réunionnaise Doctorat, Université de la Réunion.
- Barker, M. and W. Rayens** (2003). "Partial least squares for discrimination." Journal of chemometrics 17(3): 166-173.
- Bellon de la Cruz, B.** (2014). Evaluation du potentiel des images satellitaires à très hautes résolution spatiale et leur analyse orientée objet pour la cartographie du mode d'occupation du sol à la Réunion.
- Breiman, L.** (2001). "Random forests." Machine learning 45(1): 5-32.
- Clevers, J., G. Van der Heijden, et al.** (2007). "Estimating grassland biomass using SVM band shaving of hyperspectral data." Photogrammetric Engineering & Remote Sensing 73(10): 1141-1148.
- Cloude, S. R. and E. Pottier** (1996). "A review of target decomposition theorems in radar polarimetry." Geoscience and Remote Sensing, IEEE Transactions on 34(2): 498-518.
- Codjia, C.** (2014). "Caractérisation des occupations du sol en milieu urbain par imagerie radar."
- El Hajj, M., A. Bégué, et al.** (2009). "Integrating SPOT-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices—The case of sugarcane harvest on Reunion Island." Remote sensing of environment 113(10): 2052-2061.
- Engelbrecht, J., J. Kemp, et al.** (2013). The phenology of an agricultural region as expressed by polarimetric decomposition and vegetation indices. International geoscience and remote sensing symposium. Melbourne.
- Girard, M.-C. and C. M. Girard** (1999). "Traitement des données de télédétection."
- Guerschman, J., J. Paruelo, et al.** (2003). "Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data." International Journal of remote sensing 24(17): 3381-3402.

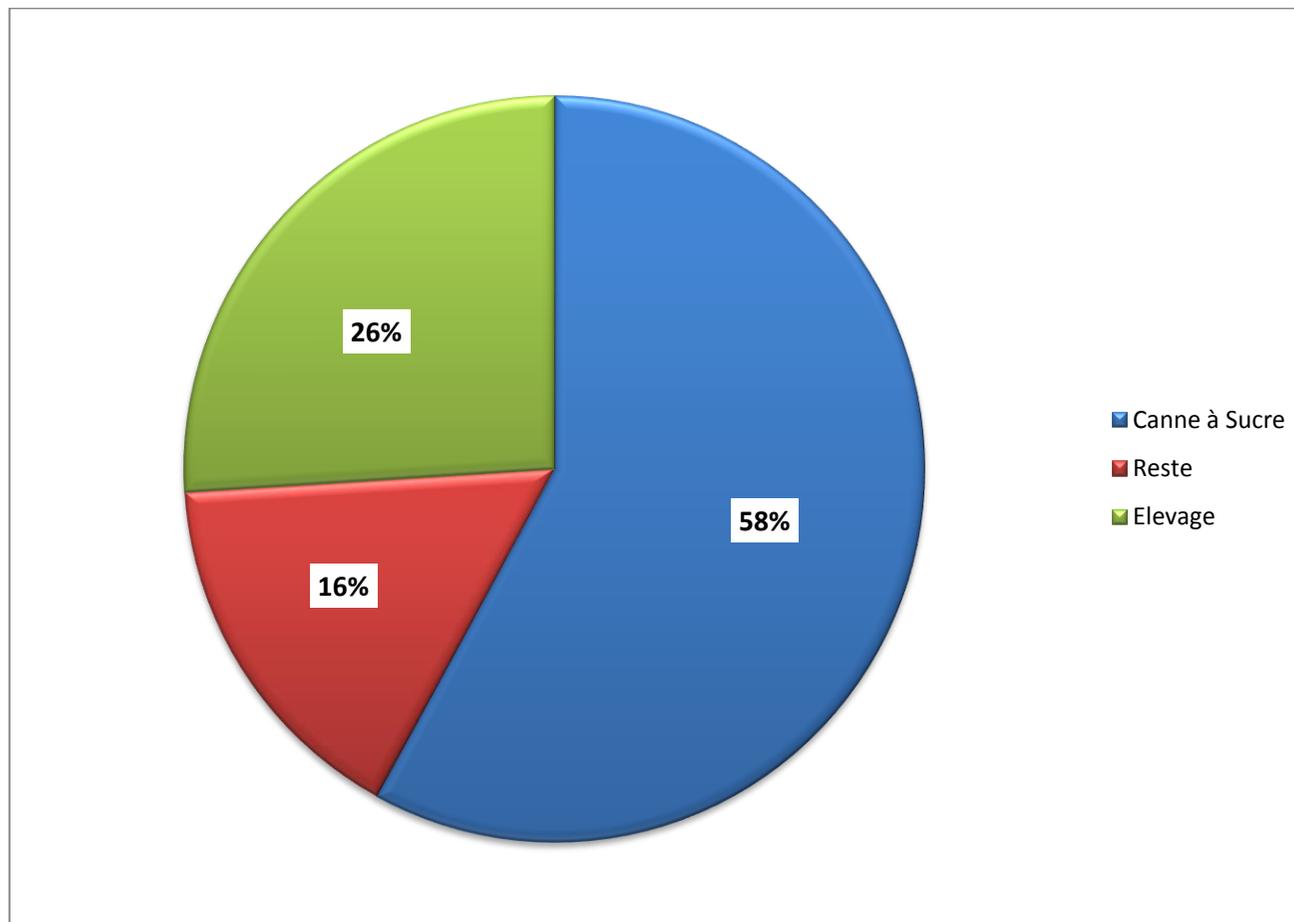
- Han, J., M. Kamber, et al.** (2006). Data mining, southeast asia edition: Concepts and techniques, Morgan kaufmann.
- Han, J., M. Kamber, et al.** (2012). Data Mining : Concepts and techniques, Morgan Kaufmann.
- Huynen, J. R.** (1970). Phenomenological theory of radar targets, TU Delft, Delft University of Technology.
- INSEE** (2007). Revue économique de la réunion N°4 Hors série- Juillet 2008. [http://www.insee.fr/fr/themes/document.asp?ref\\_id=13287&reg\\_id=24](http://www.insee.fr/fr/themes/document.asp?ref_id=13287&reg_id=24)
- Jedlovec, G.** (2009). Advances in Geoscience and Remote Sensing, InTech.
- Kim, Y., T. Jackson, et al.** (2012). "Radar vegetation index for estimating the vegetation water content of rice and soybean." Geoscience and Remote Sensing Letters, IEEE 9(4): 564-568.
- Lee, J.-S. and E. Pottier** (2009). Polarimetric radar imaging: from basics to applications, CRC press.
- Lee, J. S., T. L. Ainsworth, et al.** (2011). "Terrain Categorization Based on Scattering Mechanisms for Single-Pol High-Resolution Terrasar-X Images." 2011 Ieee International Geoscience and Remote Sensing Symposium (Igarss): 3606-3609.
- Lee, J. S., M. R. Grunes, et al.** (2001). "Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR." Ieee Transactions on Geoscience and Remote Sensing 39(11): 2343-2351.
- Leo, O. and G. Lemoine** (2000). "Land Parcel Identification Systems in the Frame of Regulation (EC) 1593."
- Li, H., J. Chen, et al.** (2015). "Sugarcane Mapping in Tillering Period by Quad-Polarization TerraSAR-X Data."
- Maghsoudi, Y.** (2011). Analysis of Radarsat-2 Full Polarimetric Data for Forest Mapping, Ph. D. dissertation, Calgary University, Calgary, The Canada.
- Martignac, C.** (2006). Maintien d'une filière dominante ou projet de territoire: le cas du sucre à La Réunion, Université Paul Valéry-Montpellier III.
- McCulloch, W. S. and W. Pitts** (1943). "A logical calculus of the ideas immanent in nervous activity." The bulletin of mathematical biophysics 5(4): 115-133.
- McNairn, H., C. Duguay, et al.** (2002). "The effect of soil and crop residue characteristics on polarimetric radar response." Remote sensing of environment 80(2): 308-320.
- Mountrakis, G., J. Im, et al.** (2011). "Support vector machines in remote sensing: A review." ISPRS Journal of Photogrammetry and Remote Sensing 66(3): 247-259.
- Olshen, L. and C. J. Stone** (1984). "Classification and regression trees." Wadsworth International Group 93(99): 101.

- Pal, M.** (2005). "Random forest classifier for remote sensing classification." International Journal of remote sensing 26(1): 217-222.
- Platt, J.** (1999). "Fast training of support vector machines using sequential minimal optimization." Advances in kernel methods—support vector learning 3.
- Simonett, D. S. and R. E. Davis** (1983). "Image analysis-active microwave." Manual of Remote Sensing: theory: theory, instruments and techniques: 1125-1179.
- Tuia, D., F. Ratle, et al.** (2009). "Active learning methods for remote sensing image classification." Geoscience and Remote Sensing, IEEE Transactions on 47(7): 2218-2232.
- Vapnik, V. N., B. E. Boser, et al.** (1992). A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, ACM.
- Wikistat.** Arbres binaires de décision. <http://wikistat.fr/pdf/st-m-app-cart.pdf>
- Wikiversité.** Réseaux de neurones (Graphique). [http://fr.wikiversity.org/wiki/Réseaux\\_de\\_neurones/Les\\_neurones\\_en\\_réseaux](http://fr.wikiversity.org/wiki/Réseaux_de_neurones/Les_neurones_en_réseaux)
- Wold, H. O. A.** (1968). Nonlinear estimation by iterative least square procedures.
- Yamaguchi, Y., T. Moriyama, et al.** (2005). "Four-component scattering model for polarimetric SAR image decomposition." Geoscience and Remote Sensing, IEEE Transactions on 43(8): 1699-1706.
- Zebker, H. A. and J. J. Van Zyl** (1991). "Imaging radar polarimetry: A review." Proceedings of the IEEE 79(11): 1583-1606.

## Annexes

---

**Annexe 1 : Répartition de la SAU de La Réunion en 2013, Source : Agreste 2013**



## Annexe 2 : Résumé des critères d'exclusion appliqués à la BOS afin d'obtenir la base d'apprentissage

Classes	Critères d'exclusion					
	Bâti	Superficie	Forme de la parcelle	Situation de la parcelle	Culture	Présence de bâtiments pour l'agriculture
Ananas						
Banane						
Canne						
Maraichage						
Prairie						
Verger						

Critère appliqué pour cette classe
  Critère non appliqué pour cette classe

**Bâti** : Suppression ou modification des parcelles comprenant entièrement des zones de bâti. Pour la classe maraichage suppression supplémentaire pour les parcelles intersectées avec ce bâti.

**Superficie** : Suppression des parcelles dont la taille est inférieure à 1 000 m<sup>2</sup>. Division des parcelles de canne à sucre dont la taille est supérieure à 10 hectares à l'aide de la couche shapefile comprenant l'ensemble des parcelles de canne à sucre dans le Sud de l'île.

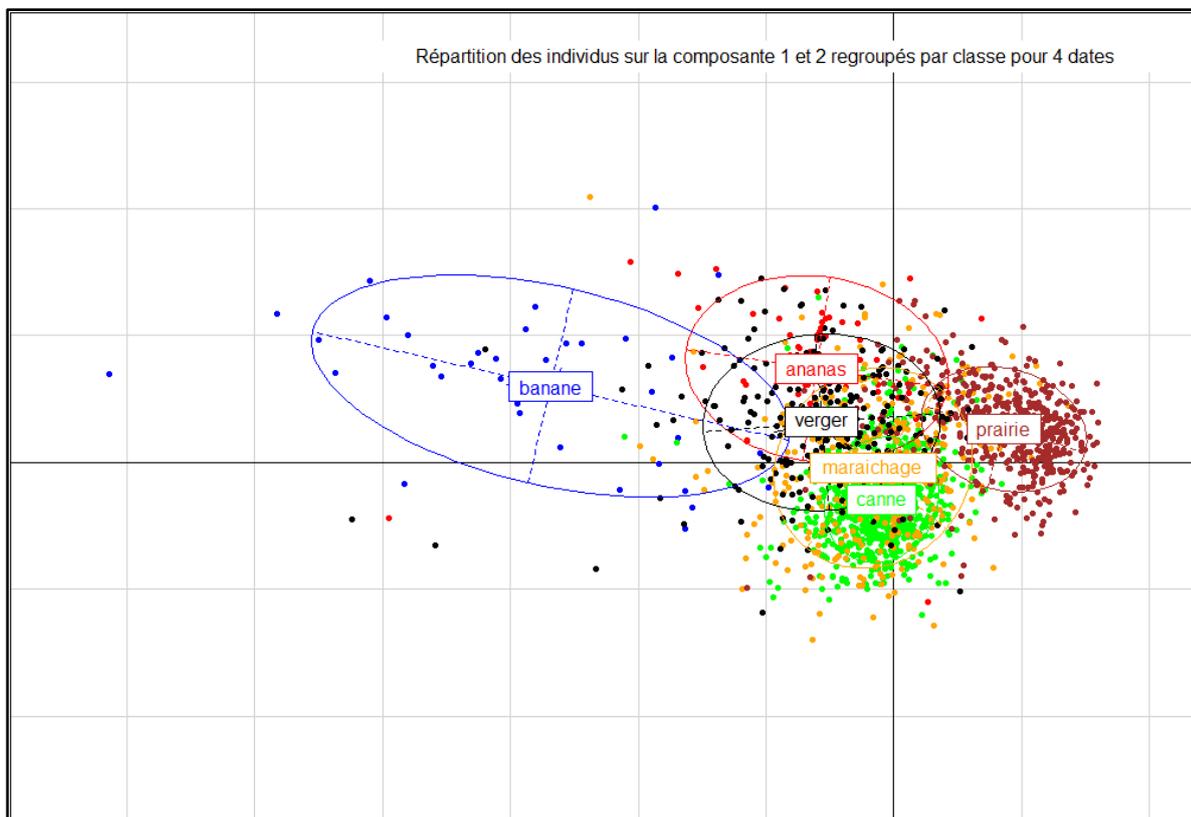
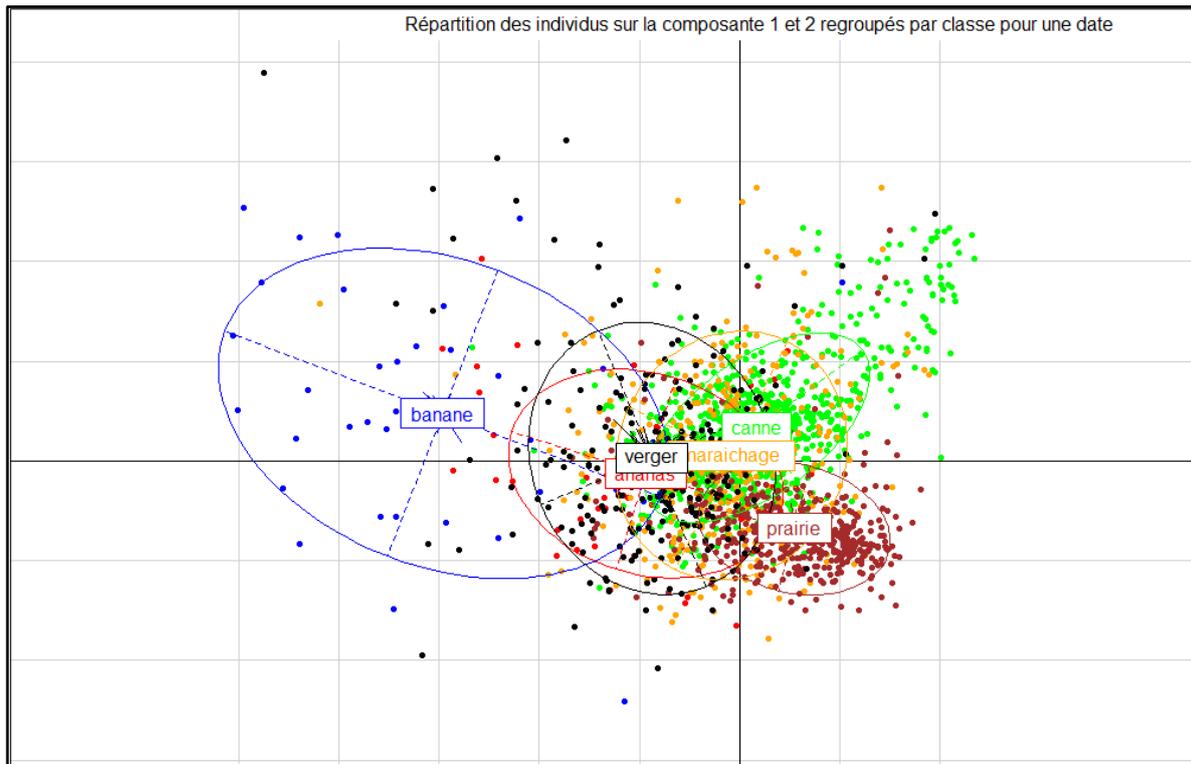
**Forme de la parcelle** : Suppression des parcelles ayant une forme pouvant poser problème pour l'analyse radar en fonction de la pente et de l'orientation de la parcelle par rapport au signal incident (position des parcelles les unes par rapport aux autres).

**Situation de la parcelle** : Suppression des parcelles localisées à proximité des ravines, des zones de bâti afin d'éviter les effets de bords ainsi que les effets liés aux bruits.

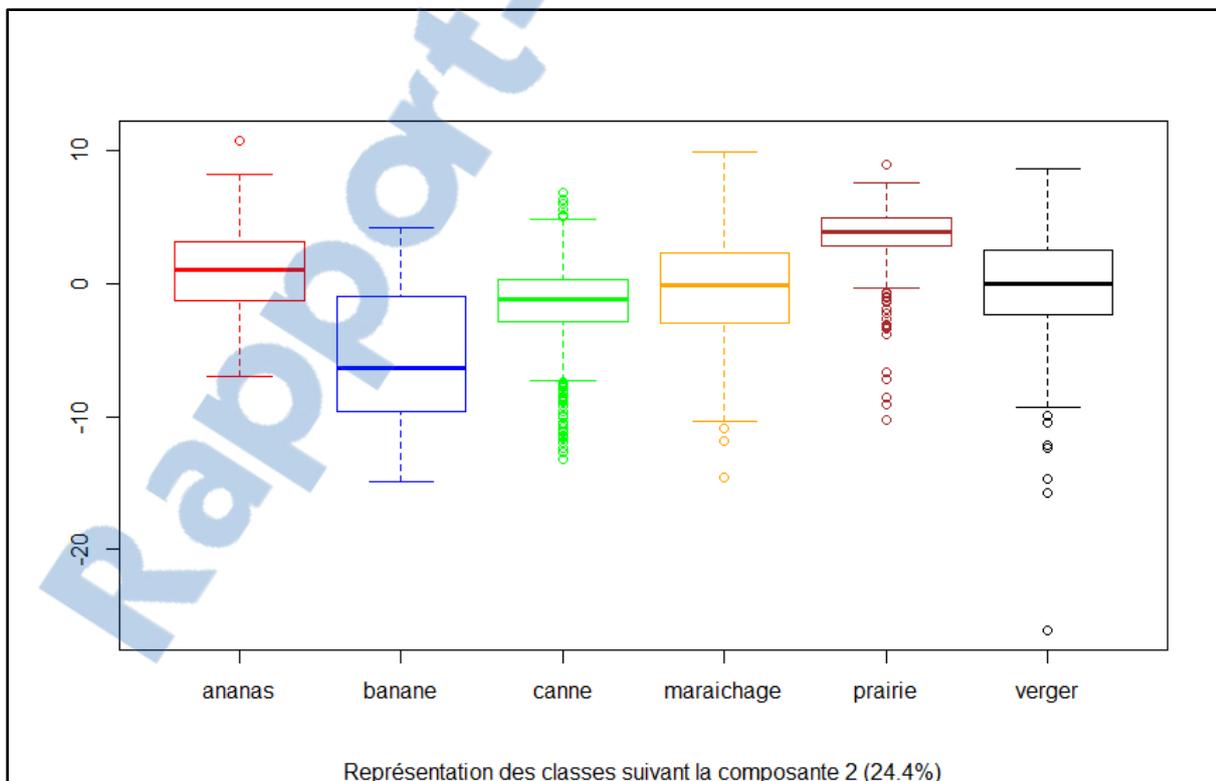
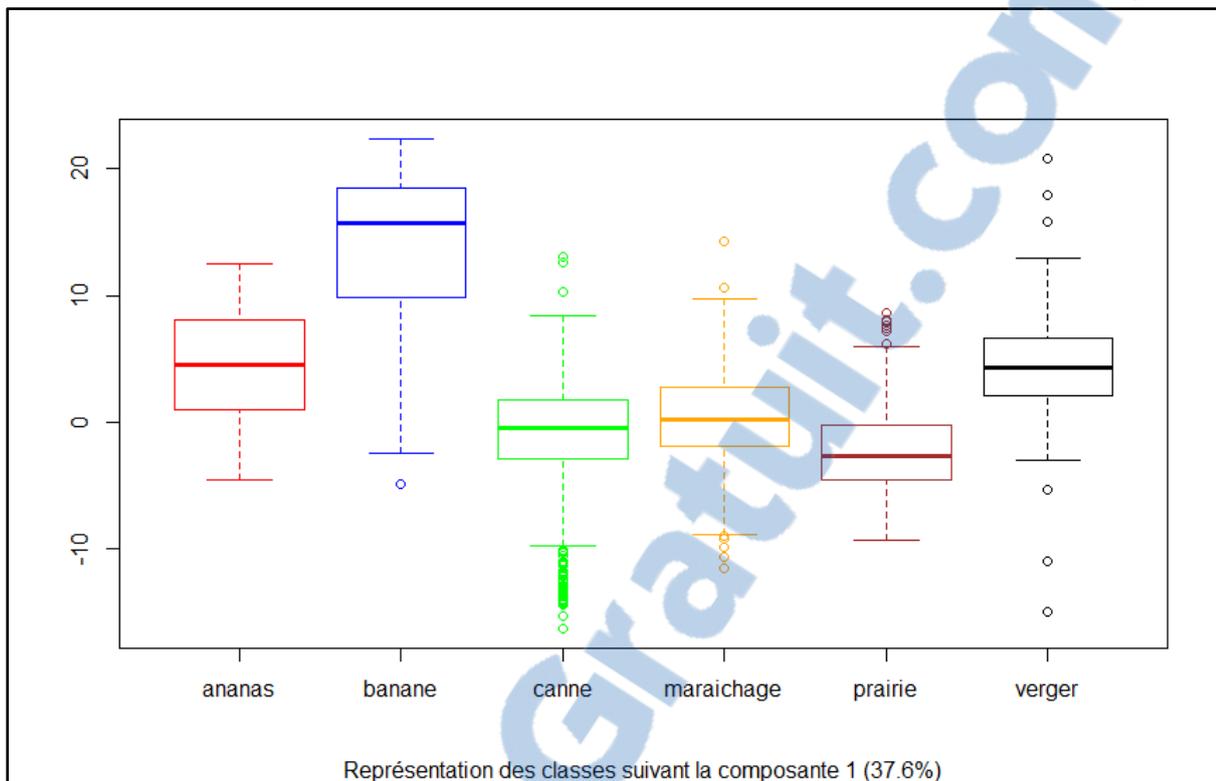
**Culture** : Suppression ou modification (changement de contours) des parcelles dont l'activité agricole recensée par la DAAF est différente de celle observée sur le terrain (ex : présence de forêt dans les parcelles de prairie ou de maraichage).

**Présence de bâtiments pour l'agriculture** : Uniquement pour la classe maraichage. Suppression des parcelles possédant des serres sur leur superficie.

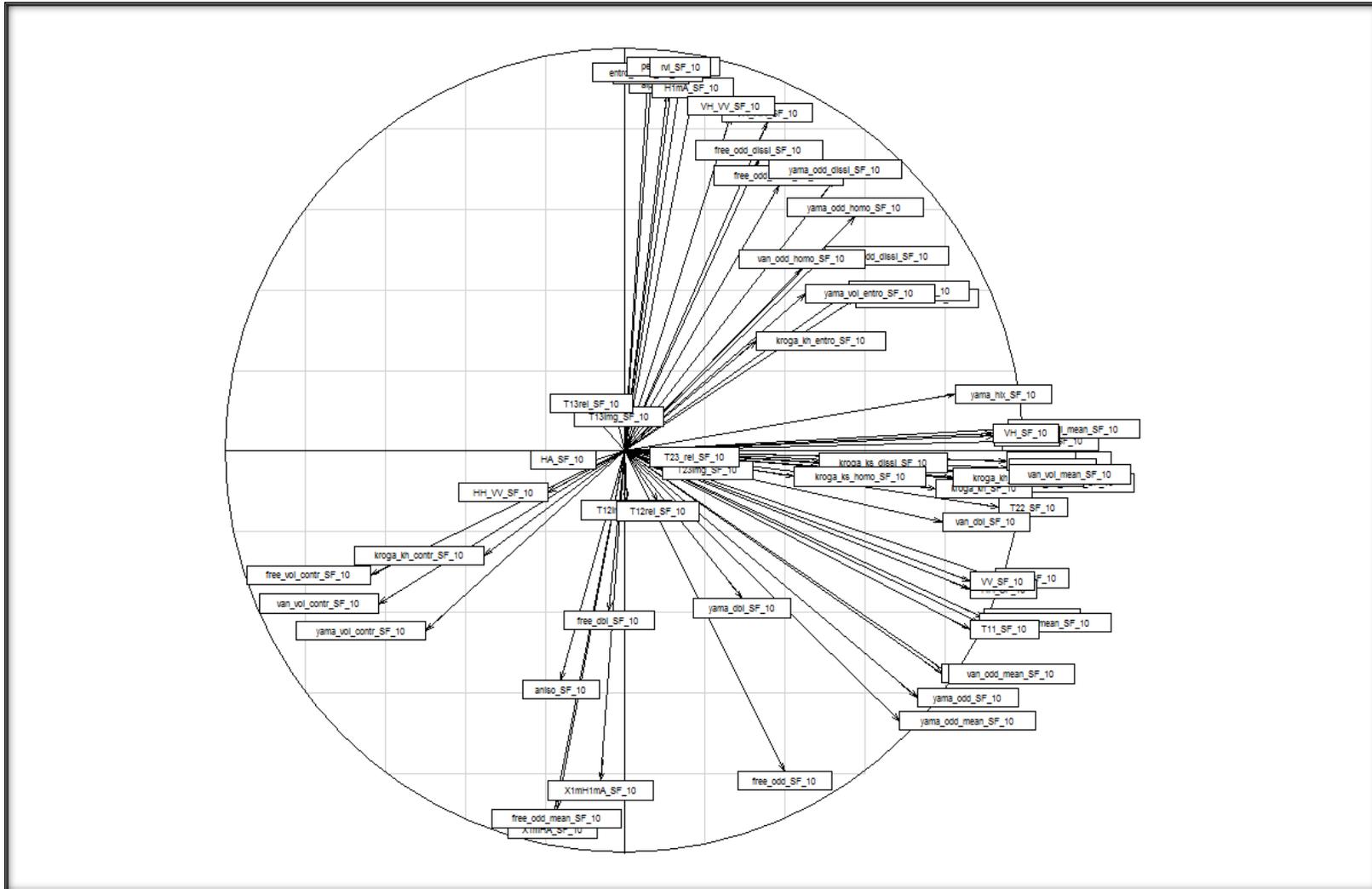
### Annexe 3 : Répartition interclasses pour les 2 premières composantes de l'ACP



**Annexe 4 : Répartition des individus pour les 2 premières composantes de l'ACP**



**Annexe 5 : Corrélations entre variables radar pour les 2 premières composantes de l'ACP**



## **Annexe 6 : Variables sélectionnées à partir du premier Random Forest**

<b>Les variables les plus significatives</b>	<b>Nombre d'apparition (en date)</b>
Sigma HH	6
T12 imaginaire	6
T12 réelle	5
T13 réelle	5
T23 réelle	5
Sigma VV	5
Alpha (Cloude et Pottier)	4
Krogager Surface, Texture: Homogénéité	4
T13 imaginaire	4
Freeman_Durden Surface	3
Freeman_Durden odd, Texture: Homogénéité	3
Freeman_Durden volume, Texture: Contraste	3
HH/VV (Ratio)	3
Sigma HV	3
Sigma VH	3
Yamaguchi Surface, Texture: Moyenne	3
Entropie, Texture: Moyenne	2
Freeman_Durden Surface, Texture: dissimilarité	2
Freeman_Durden Surface, Texture: Moyenne	2
Krogager Hélix, Texture: Contraste	2
Krogager Surface, Texture: Dissimilarité	2
T23 imaginaire	2
VH/VV (Ratio)	2
Yamaguchi Surface, Texture: Homogénéité	2
Entropie (Cloude et Pottier)	1
H*(1-A) (Cloude et Pottier)	1
Krogager Hélix	1
Van Zyl Surface, Texture: Moyenne	1
Van Zyl Volume, Texture: Contraste	1
VH/HH (Ratio)	1
Yamaguchi Doubles rebonds	1
Yamaguchi Surface	1
Yamaguchi Surface, Texture: Dissimilarité	1

**Annexe 7 : Les 15 variables uniques sélectionnées à partir du second Random Forest**

<b>Variables choisies</b>
<b>Entropie (Cloude &amp; Pottier), Texture : mean</b>
<b>Freeman-Durden surface</b>
<b>Freeman-Durden surface, Texture : homogénéité</b>
<b>Sigma HH</b>
<b>Ratio HH/VV</b>
<b>Sigma HV</b>
<b>T12 imaginaire</b>
<b>T12 réel</b>
<b>T13 imaginaire</b>
<b>T13 réel</b>
<b>T23 réel</b>
<b>Sigma VH</b>
<b>Ration VH/VV</b>
<b>Sigma VV</b>
<b>Yamaguchi surface, Texture: mean</b>



**Diplôme** : Master Géographie-aménagement  
**Spécialité** : Géomatique  
**Spécialisation / option** : Image  
**Enseignant référent** : Corgne Samuel

**Auteur(s)** : Denize Julien

**Organisme d'accueil** : Cirad

**Date de naissance\*** : 21/07/1992

**Adresse** : 7 chemin de L'Irat, Ligne Paradis,  
97410 Saint Pierre, Réunion

**Nb pages** : 36 pages      **Annexes** : 7 pages

**Année de soutenance** : 2015

**Maître de stage** : Todoroff Pierre

**Titre** : Développement d'un modèle de classification par imagerie SAR pour la cartographie des cultures tropicales : Exemple de l'île de la Réunion

**Title** : Development of an SAR imaging classification model for mapping of tropical crops: Example of Reunion Island

**Résumé** :

A La Réunion, île tropicale de 2 512 km<sup>2</sup> contraint par un relief accidenté, les filières agricoles se font concurrence sur un parcellaire très morcelé. Les enjeux de l'occupation du territoire se concentrent dans les rares espaces aménageables des « bas » de l'île. Cette étude, s'attache à fournir des références fiables et actualisées aux décideurs, aménageurs du territoire et institutions, par une méthode de télédétection permettant l'identification et la cartographie des cultures tropicales. Dans cet objectif, nous évaluons le potentiel de l'imagerie SAR par polarimétrie associée à une modélisation statistique basée sur des classifieurs. Une méthode a été élaborée et testée sur une zone d'étude de 25\*25 km grâce à une série de 6 images RADARSAT-2 de 2014 en polarisation complète. Un ensemble d'indicateurs radar (indices, décomposition, texture, etc.) ont été calculés à partir de la matrice de cohérence. Des séparateurs à vastes marges (SVM), ont permis la classification de ces indicateurs à partir des îlots cultureux recensés en 2013. La méthode obtient une précision globale de 88% pour l'identification des quatre cultures majoritaires. Toutefois, certaines confusions persistent, notamment entre les cultures maraîchères et la banane. La prochaine étape consistera à optimiser la procédure en réduisant le nombre de variables explicatives pour délivrer un produit opérationnel sur la Réunion. La méthode sera reproductible afin de minimiser les coûts de production.

**Abstract** :

In Reunion, a tropical island of 2 512 km<sup>2</sup> constrained by a rugged relief, agricultural sectors are competing in highly fragmented agricultural land. The issues of land use are concentrated in scarce convertible areas along the coast. This study aims to provide reliable and updated references to policymakers, planners and institutions, by a remote sensing method for the identification and mapping of tropical crops. For this purpose, we assess the potential of polarimetric SAR imagery associated with statistical modeling based on classifiers. A method has been developed and tested on a study area of 25\*25 km thanks to 6 RADARSAT-2 images in 2014 in full-polarization. A set of radar indicators (indices, decomposition, texture, etc.) was calculated from the coherency matrix. Support vector machines (SVM), allowed the classification of these indicators from fields recorded in 2013. The method shows an overall accuracy of 88% for the identification of the four major crops. However, some confusion persists, especially between vegetable crops and bananas. Further investigation will aim to optimize the processing by reducing the number of explanatory variables with the objective of an operational product in Reunion. The method will be reproducible in order to minimize production costs.

**Mots-clés** : RADARSAT-2, SENTINEL-1, polarisation complète, polarisation double, SVM, forêts aléatoires, mode d'occupation des sols

**Key Words**: RADARSAT-2, SENTINEL-1, full-polarization, Dual-polarization SVM, Random Forest, Land Use