THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT - AGROPARISTECH

N° : 2021AGPT0005

En Géomatique

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau – n°584
Portée par l'Université de Montpellier

Unité de recherche TETIS - Territoires, Environnement, Télédétection et Information Spatiale

Télédétection et intégration de connaissances via la modélisation spatiale pour une cartographie plus cohérente des systèmes agricoles complexes :

Application aux Hautes Terres, à Madagascar

Présentée par Arthur CRESPIN--BOUCAUD

Le 23 mars 2021

Sous la direction d'Agnès BÉGUÉ
et Valentine LEBOURGEOIS, Danny LO SEEN et Mathieu CASTETS

Devant le jury composé de

Agnès BÉGUÉ, Chercheuse (HDR), CIRAD	Directrice de thèse
Delphine BURGER-LEENHARDT, Directrice de Recherche, INRAE	Rapporteure
Samuel CORGNE, Professeur des Universités, Université Rennes 2	Rapporteur
Carmen GERVET, Professeur des Universités, Université de Montpellier	Présidente du jury
Valentine LEBOURGEOIS, Chercheuse, CIRAD	Membre du jury
David SHEEREN, Maître de conférences, INP-ENSAT	Membre du jury
Danny LO SEEN, Chercheur, CIRAD	Invité

AgroParisTech

UNIVERSITÉ DE MONTPELLIER



Ce doctorat a été réalisé au sein de l'UMR TETIS et du DP SPAD. Il a été financé par le CNES et le CIRAD.

UMR TETIS (Territoires, Environnement, Télédétection et Information Spatiale)
AgroParisTech/CIRAD/INRAE
CIRAD - Département Environnements et Sociétés (ES)
Maison de la Télédétection
500 Rue Jean-François Breton
34093 Montpellier Cedex 5
France

REMERCIEMENTS

La couverture de cette thèse le montre bien : ce travail n'a pu être effectué seul. Ce fut, certes, un travail solitaire, mais bien accompagné par de nombreuses personnes qui ont permis l'écriture de ce manuscrit, mais surtout qui ont réfléchi avec moi, qui ont pris le temps de discuter, de m'expliquer des concepts, des méthodes, ou encore de m'aider à prendre du recul, à relativiser et à me détendre.

Tout d'abord, je tiens à remercier mes encadrants pour tout ce que vous avez fait. Une thèse est à la fois une aventure scientifique et humaine et vous avez largement participé sur ces deux points. Dans les moments difficiles de la thèse vous avez su être présents avec vos attentions, vos remarques sympathiques et surtout je pense que vous avez réussi à me faire comprendre les apports de cette thèse. Je pense que sans vous je ne serai pas allé bien loin...

Merci Valentine de m'avoir accompagné durant cette thèse. Notre chemin a été mouvementé, mais tu as toujours su être là lorsque j'en avais besoin. Tu m'as « obligé » à poser mes idées, à expliquer ce que je pensais de manière claire et surtout à structurer ma pensée pour qu'on produise ensemble ce travail. Je te dis simplement merci pour tout car c'est parfois plus clair que de longues phrases. J'ai beaucoup appris à tes côtés et je pense que cette expérience a été complètement partagée dans les bons moments et ceux plus difficiles. Merci de m'avoir poussé à fond, cette thèse te doit énormément ! Et à bientôt pour des suites, on a plein de choses à faire encore ! :)

Merci Danny pour les discussions très riches que nous avons pu avoir. Je vais être honnête, au début de mon doctorat j'étais parfois perdu dans ce que tu me disais, tu étais à un niveau de réflexion tellement supérieur à mes préoccupations de « mais le code ne fonctionne pas ». Tes remarques m'ont permis de me dire que la modélisation c'est avant tout une réflexion et surtout de percevoir à la fin de ma thèse l'ensemble des possibilités qu'offre la modélisation. La richesse de nos discussions va me manquer, et ce sur tous les sujets. Ce fut un grand plaisir de partager ce travail avec toi.

Agnès, tu as suivi cette thèse d'un oeil bienveillant avec toute ton expérience qui a permis de poser une structure cohérente à ce manuscrit et surtout à me motiver dans les moments difficiles de la rédaction où ton aide et tes corrections m'ont reboosté alors que je repartais de chez toi avec un chapitre complet à modifier. Merci pour ta bienveillance sans limites.

Pour finir, merci Mathieu pour tout. Tu as réussi à devenir un de mes encadrants officiels alors que l'on partageait le même bureau et les mêmes pauses. Tu sais que ta contribution est majeure et à quel point je t'en suis reconnaissant.

Je tiens à remercier mes rapporteurs, Samuel Corgne et Delphine Leenhardt, pour avoir accepté d'évaluer ce travail, pour votre lecture approfondie de mon manuscrit et vos commentaires qui m'ont permis de mieux identifier les lacunes et les points forts de ce travail. Merci aussi à David

Sheeren et Carmen Gervet d'avoir accepté d'être membres de mon jury. Merci à vous quatre pour les discussions que l'on a pu avoir durant la soutenance, ce fut un moment très intéressant et agréable.

Ma reconnaissance va également à Xavier Augusseau, Damien Arvor et Jean-Stéphane Bailly pour l'intérêt que vous avez porté à ce travail, pour vos conseils et pour vos remarques lors de mes comités de suivi individuel. Merci Xavier d'avoir être attentif à mon moral lors des confinements, ça m'a beaucoup touché.

Merci aussi à Nathalie et Sylvain qui, il y a un peu moins de 5 ans ont cru en moins et ont commencé à me former à la recherche. Je vous dois beaucoup pour votre confiance.

Je tiens aussi à remercier toutes les personnes qui ont pris le temps de discuter de mon sujet de thèse au cours de réunions, de présentations, de conférences, toujours avec des propositions intéressantes que j'ai essayé d'intégrer dans cette thèse. Merci à l'équipe STRUDEL de l'IGN est surtout à Paul, merci à Elisabetta Carfagna pour les conseils et merci aux chercheurs qui ont croisé mon chemin.

Comme l'indique le titre de ma thèse, j'ai passé un moment à Madagascar et c'est pourquoi je souhaite remercier les personnes qui ont m'ont aidé à préparer ma mission, à comprendre les enjeux socio-économico-géographico-agronomiques du Vakinankaratra et découvrir Antsirabe et ses alentours.

Merci Éloise et Valérie pour le travail que l'on a effectué ensemble sur le terrain, pour votre connaissance des plantes, les nombreuses conversations que l'on a pu avoir lors des trajets en voiture, des repas et pour m'avoir permis de découvrir la culture malgache. Ce fut de super moments.
Merci Stéphane de m'avoir aidé à préparer ma première mission et de m'avoir accompagné pour la découverte du travail de terrain ainsi que tous les bons restos d'Antsirabe. Je pense que j'ai hérité de ta rigueur au travail... Merci pour les discussions sur la télédétection et les randos et pour t'avoir recroisé à la Réunion, puis à la MTD.
Merci aux membres du DP SPAD, Bertrand pour l'accueil, pour avoir accepté de transporter les tablettes et t'être fait arrêté à la douane, Alain pour ta connaissance très poussée du terrain et des enjeux, pour avoir pris le temps de me recevoir plusieurs fois et de discuter science, Paolo pour les cafés agrémentés, Sarah et Yann pour les randos et les moments chez Sam, Patrice pour les discussions sur les cultures associées et les bons moments passés ensemble, ce fut agréable de travailler avec toi ! Merci Jean-François, pour tes remarques, pour ton regard finalement bienveillant sur mes travaux et pour avoir partagé des discussions variées. J'espère pouvoir retravailler avec toi sur des aspects d'analyse spatiale et socio-économique.

Merci à tous les collègues et personnes qui sont devenus des amis lors de mon passage à la MTD. Grâce à vous l'ambiance était toujours au top ! Merci Annie pour tes petites attentions qui font que la vie logistique d'un petit thésard est facile et agréable, Isa pour ta bonne humeur au quotidien et nos parties de coinche. Merci Christian, il y en a peu des comme des toi. Merci Raffael pour tous les coups de main pour Moringa et m'aider pour le développement en Python. Maguelonne,

merci de m'avoir faire découvrir la fouille de données et pour ton efficacité. Merci Pascal pour les très nombreux coup de main sur Ocelet, tu m'as appris avec Mathieu et Danny à comprendre sa philosophie et son code. Audrey, Bastien, Clovis, Vincent, Benjamin, Alexia, Gab pour les pauses (en dehors et à la MTD) et les discussions.
Pour les thésards, post-doc passés et suivants, Béa, Marc, Jacques, Guilhem, Flo, Fred, Sarah, merci pour votre accueil, les bières, la découverte de Montpellier et tous les bons moments passés ensemble et à venir. Guilhem et Marc désolé de vous avoir empêché de bosser lors de mes nombreuses pauses dans votre bureau, je pense que sans ça vos thèses auraient été encore meilleures. Merci Seb pour tous tes coups de main le soir pour déboguer du code et pour tous les moments hors du boulot, tu as été présent en permanence et pour tout. Courage à vous Dav, Rodrique, Martin, Marie and co, avec vous la relève est assurée !

Et plus éloigné de ce travail, mais toujours là pour me permettre de prendre de la distance, de penser à autre chose, de prendre le temps de m'écouter parler de ma thèse, d'aller faire des randos, des jeux de société et boire des coups, merci à tous mes amis, vous avez largement contribué à cette thèse (au moins du point de vue humain). Maxime, ces 2 ans de colocs, nos discussions sur tous les sujets, les jeux de société, l'aide que tu as fournie sur ma thèse ont été précieux. Thibaud, Aurélie, Axel, Tiphaine, Robin, Geoffroy, Aliénor, Clément, Maxime et Maxime, Joana, Damien, Mathilde, Florent, Léa, Delphine, Enrico, Charlotte, Camille, Eliott, Lucas, Pauline et tous ceux qui savent que je les ai oubliés, car je suis un boulet merci !
Merci aussi à ma famille, je pense que vous n'en avez pas vraiment conscience, mais vous m'avez permis d'aller jusqu'ici et vous avez toujours été là pour m'encourager et vous assurer que tout allait bien. Vous êtes super et vous comptez beaucoup pour moi.

Pour terminer merci Sarah pour tes relectures, nos discussions sur le machine learning et surtout pour tout le reste à côté et à tes côtés.

TABLE DES MATIÈRES

Remerciements i

Table des figures xii

Table des tableaux xvii

Glossaire xxi

Acronymes xxii

Introduction générale 1

1 Contexte et problématique de la thèse **5**

1 L'agriculture familiale 6

1.1 Un système à multiples facettes 6

1.1.1 Définitions 6

1.1.2 La diversité de l'agriculture familiale 9

1.2 Agriculture familiale et sécurité alimentaire 10

1.2.1 La sécurité alimentaire 10

1.2.2 Un système vulnérable... 11

1.2.3 ...mais avec des forces et une certaine résilience 12

1.3 Deux modes de représentation 13

1.3.1 La cartographie 13

1.3.1.1 Occupation des sols et usage des sols 13

1.3.1.2 L'échelle d'analyse spatio-temporelle 14

1.3.1.3 L'état actuel de la cartographie de l'agriculture familiale 16

1.3.2 Typologie des exploitations 17

2 Les outils d'analyse spatio-temporelle 18

2.1 Télédétection et agriculture familiale 18

2.1.1 Une offre variée d'images 19

2.1.2 Des méthodes de classification 20
2.1.2.1 Classification non supervisée et supervisée 21
2.1.2.2 Classification par pixel et approche objet 22
2.1.3 La télédétection et l'agriculture familiale 23
2.2 Modélisation spatio-temporelle de l'agriculture familiale 25
2.2.1 La modélisation spatio-temporelle des systèmes 25
2.2.2 La modélisation de la dynamique de l'occupation et de l'usage des sols 26
2.2.2.1 La dynamique des systèmes 26
2.2.2.2 Les approches individu-centrées 27
2.2.2.3 Les autres approches en modélisation spatio-temporelle 28
2.2.3 La modélisation spatiale de l'agriculture familiale 29
2.2.4 Des approches encore peu appliquées aux espaces tropicaux . . 29
2.3 Combinaison télédétection / modélisation spatio-temporelle 30
3 Objectifs de la thèse et démarche adoptée 32

2 Modèle conceptuel **37**

1 Pourquoi un modèle conceptuel 38
2 Les niveaux d'organisation des systèmes complexes agricoles 39
2.1 L'objet 40
2.2 Le groupement d'objets de niveau 1 41
2.3 Le groupement d'objets de niveau 2 43
3 Description du modèle conceptuel 44
3.1 Principes de base du modèle 44
3.1.1 Les Vecteurs de Probabilité d'Appartenance aux Classes (VPAC) 45
3.1.2 Connaissances et règles 46
3.1.3 Observation et possibilité 47
3.2 Principaux modules du modèle conceptuel 47
3.2.1 Rôle et complémentarité des modules 48
3.2.2 Module 0 : Télédétection, classification supervisée 51
3.2.3 Module 1 : modélisation du fonctionnement temporel (histoire des objets) 52

3.2.4 Module 2 : Règles spatiales (contexte spatial des objets) 53
3.2.5 Module 3 : Règles temporelles 53
3.2.6 Module 4 : Modélisation de la répartition spatiale des classes . 54
3.2.7 Attribution d'une classe finale en fonction des VPAC : probabilité et méthodes de fusion 55

3 Zone d'étude, matériel et méthodes 59

1 Madagascar et les Hautes Terres 60
1.1 Les enjeux de l'agriculture à Madagascar 60
1.2 Le Vakinankaratra 62
1.3 Zone d'étude 65
2 Données 67
2.1 Données satellitaires et prétraitements 68
2.1.1 Image mono-date annuelle à très haute résolution spatiale : Pléiades 68
2.1.2 Les séries temporelles à haute résolution spatiale 69
2.1.2.1 Sentinel-2 69
2.1.2.2 Venμs 69
2.2 Les données de terrain 71
2.2.1 Relevés *in-situ* d'occupation et d'usage des sols 71
2.2.2 Méthode d'échantillonnage 72
2.2.3 Entretiens et enquêtes de terrain 76
2.2.3.1 Entretiens avec des scientifiques 76
2.2.3.2 Enquêtes avec des agriculteurs 76
2.2.4 Entretiens avec des responsables locaux 77
2.3 Données support pour l'application des connaissances 77
2.3.1 Topographie et relief 78
2.3.2 Carte pédologique 78
2.3.3 Réseau routier 79
2.3.4 Données administratives 80
2.3.5 Bâtiments 80
3 Outils 81

3.1 Moringa . . . 81
3.2 Ocelet . . . 81
4 Méthodes d'évaluation . . . 82

4 Méthode d'application des modules à la région des Hautes Terres 87

1 Télédétection . . . 88
1.1 Traitement des images satellite . . . 89
1.1.1 Image à très haute résolution spatiale . . . 89
1.1.1.1 Pré-traitement . . . 89
1.1.1.2 Segmentation . . . 89
1.1.1.3 Textures . . . 92
1.1.2 Séries temporelles d'images satellitaires . . . 93
1.1.2.1 Pré-traitements . . . 93
1.1.2.2 Extraction des variables . . . 94
1.1.3 Statistiques zonales . . . 95
1.2 Classification des données de télédétection . . . 95
1.2.1 Partitionnement de la base de données de vérité terrain . . . 95
1.2.2 Modèle de forêts aléatoires (*random forest*) . . . 96
2 De la connaissance à la probabilité d'appartenance : les règles expertes . . . 97
2.1 Sélection des connaissances à formaliser . . . 98
2.2 Formalisation des connaissances en règles . . . 100
2.3 Attribution de probabilités aux règles (issues de la connaissance) . . . 101
3 Application des modules . . . 103
3.1 Préparation de données support pour l'application des connaissances . . . 104
3.2 Sélection des segments . . . 107
3.2.1 Sélection selon la fiabilité des $VPAC_{RF}$ . . . 107
3.2.2 Sélection des segments de cultures associées à partir des $VPAC_{RF}$ 108
3.2.3 Sélection des segments non-agricoles . . . 110
3.3 Calcul des probabilités de présence de culture par segment . . . 110
3.3.1 Standardisation des résultats et création des $VPAC_{RS}$ . . . 114

3.3.2 Fusion des probabilités de la télédétection du module 2 : règles spatiales . . . 115
3.4 Application du module 4 : modélisation de la répartition spatiale des classes 116
4 Synthèse des expérimentations menées . . . 119
4.1 Analyse de la « cohérence agricole » des cartes . . . 119
4.2 Analyse de la classification des données de télédétection . . . 120
4.3 Résultats de la combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ . . . 121

5 Résultats d'applications 123

1 Résultats de la classification des données de télédétection . . . 124
1.1 Impact des données satellitaires utilisées sur la précision globale . . . 124
1.2 Analyse des résultats de classification par classe . . . 125
1.3 Segmentation et classification . . . 128
2 Application du module 2 : règles spatiales . . . 130
2.1 Probabilités issues du module des règles spatiales . . . 130
2.2 Choix des paramètres du module . . . 132
2.3 Vers une améliorations des résultats ? . . . 133
2.3.1 Analyse quantitative des résultats . . . 133
2.3.2 Indépendance des méthodes . . . 137
2.4 Incidence de l'application du modèle sur la répartition spatiale des classes 138
2.4.1 Incidence sur les probabilités des segments . . . 142
2.4.1.1 Distribution des probabilités par classe . . . 142
2.4.1.2 Évolutions des probabilités des segments de la BD de vérité terrain . . . 143
2.5 Synthèse et résultats spatiaux qualitatifs . . . 146
2.5.1 Carte d'occupation et d'usage des sols produite avec les $VPAC_{Fusion}$ 146
2.5.2 Savane et cultures pluviales . . . 147
2.5.3 Vergers et légumes . . . 150
2.5.4 Cuvette d'Ambohibary . . . 151
2.5.5 Autres classes et localisations . . . 153
2.6 Synthèse des résultats qualitatifs et quantitatifs . . . 153
3 Résultats du module modélisation de la répartition spatiale des classes . . . 154

3.1 Changements généraux . . . 156
3.2 Plateaux et pentes, une place pour les cultures pluviales ? . . . 157
3.3 Évolution du Sud : les communes d'Alakamisy et d'Ambano . . . 158

6 Discussion **161**

1 Bilan général de l'application des modules . . . 163
1.1 Rappel des limites existantes de la caractérisation des espaces agricoles complexes et des concepts principaux de l'approche développée . . . 163
1.2 Une application du modèle conceptuel sans temporalité . . . 165
1.3 Résultats globaux de l'application des modules spatiaux . . . 166
2 Application : pistes et limites . . . 170
2.1 Une meilleure exploitation des images de télédétection . . . 170
2.2 L'aspect multiscalaire du modèle . . . 171
2.3 L'impact de la qualité des données sur les règles et les résultats . . . 172
2.4 Impact des choix pour la formalisation des connaissances et pour l'utilisation des VPAC . . . 175
2.5 Validation du modèle . . . 178
3 Quels changements pour le modèle conceptuel ? . . . 179
3.1 Un modèle trop centré sur Madagascar, mais déjà pluridisciplinaire . . . 179
3.2 Penser une nouvelle approche thématique . . . 180
4 Perspectives . . . 182

Conclusion générale . . . 185

Bibliographie . . . 187

Annexe A Bandes spectrales **221**

Annexe B Indices spectraux **223**

Annexe C Sélection d'ouvrages techniques et scientifiques utilisés pour l'identification et la définition des connaissances **225**

Annexe D Questionnaire socio-économique utilisé lors de la saison culturale 2017-2018 auprès des agriculteurs **227**

Annexe E Résumé des connaissances issues des enquêtes auprès des agriculteurs en 2017-2018 pour les « principales » cultures 233

Annexe F Résumé des connaissances identifiées par communes à partir des questionnaires réalisés en 2019 239

Annexe G Position topographique ou relief 269

Annexe H Production de la carte du temps de trajet à pied entre un segment et un bâtiment 271

Annexe I Matrices de confusion 275

Annexe J Résumé des règles utilisées dans le modèle par type de connaissance 279

Annexe K Implémentation avec Ocelet du module 4 : modélisation de la répartition spatiale des classes 285

Annexe L Article : Agriculturally consistent mapping of smallholder farming systems using remote sensing and spatial modelling 287

TABLE DES FIGURES

1 Carte de la taille des champs dans le monde, réalisée à dire d'expert par photo-interprétation d'environ 14 000 extraits d'images à très haute résolution spatiale mis à disposition sur la plateforme de *crowd sourcing* GEO-Wiki (Fritz et al., 2015). 7

2 Carte de la sous-nutrition en 2020 selon le Programme Alimentaire Mondial (https://www.wfp.org/publications/hunger-map-2020.) 11

3 Typologie et définition des systèmes de culture selon Bégué et al. (2018) 15

4 Carte des moyens de subsistance à Madagascar selon l'USAID en 2017, consolidation du travail effectué en 2013 (USAID and NET, 2013). 17

5 Relations selon les satellites optiques (symbole orange) ou radar (symbole bleu), entre la résolution spatiale et la fauchée (a) et entre la fréquence de revisite et la résolution spatiale (b), figure de Tonneau et al. (2020) 19

6 Organisation de la thèse 34

7 Schéma des différents niveaux d'organisation spatiale et leurs imbrications dans l'objectif de caractériser l'occupation ou l'usage des sols d'un objet. Trois niveaux sont représentés pour caractériser l'agriculture familiale, et trois principaux contextes ont été retenus comme ayant une influence majeure sur les stratégies et choix des agriculteurs dans leur choix de la mise en culture d'un champ. 40

8 Schéma résumant les principaux concepts mobilisés dans le modèle conceptuel. . 45

9 Schéma général du modèle conceptuel. 50

10 Schéma représentant l'enchaînement des modules produisant des VPAC. 51

11 Schéma des étapes de la modélisation de la répartition spatiale des cultures. . . . 56

12 Carte de la région du Vakinakaratra, et localisation du site JECAM ainsi que de la zone d'étude. 63

13 Photographies de la zone d'étude, prises dans les communes d'Alakamisy, de Soanindrariny, d'Ambano et de Tritriva. 66

14 Fauchée des différentes images satellitaires utilisées. 68

15 Extraits d'images satellitaires Pléiades (acquise le 01/02/2019), Venμs (acquise le 02/02/2019) et Sentinel-2 (acquise le 03/02/2019). En jaune sont représentés les contours des relevés d'occupation et d'usage des sols effectués sur le terrain. . . . 70

16 Cartographie des points GPS relevés entre le 31/01/2018 et le 10/04/2019. Sources : limites administratives et routes OpenStreetMap 2020. 75

17 Schéma des étapes du traitement des données de télédétection. 88

18 Schéma des étapes de l'application du modèle conceptuel. 106

19 Sélection des segments traités par les modules de modélisation, basée sur le $VPAC_{RF}$. 108

20 Sélection des segments pouvant être affectés aux classes d'associations de cultures. pCulture signifie la probabilité de la culture. Riz p. est l'acronyme de riz pluvial. Dans notre cas seuil1 = 0,4, seuil2 = 0,6 et seuil3 = 0,8. 109

21 Forêt et savane. Polygones mal classés (en rouge) et correctement classés (en vert) à partir du traitement des données de télédétection. Les bordures délimitent la segmentation de ces polygones. 128

22 Classification de l'occupation des sols à partir des données de télédétection. . . . 129

23 Probabilités d'appartenance des segments à différentes cultures issues du module règles spatiales pour les classes : (**a**) *maïs*, (**b**) *verger*, (**c**) *pomme de terre*, et probabilités issues de la classification des données de télédétection pour la classe (**d**) *verger*. L'impact de certaines règles d'importance est visible sur ces cartes (par exemple, sur la carte (a), l'importance de la route nationale 7 qui crée un corridor nord-sud ayant une probabilité d'appartenance plus élevée). 131

24 Cartes des segments ayant changé de classe après l'application du module des règles spatiales. À gauche, selon la méthode de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la probabilité maximale et à droite, par la moyenne pondérée. . . . 139

25 Distribution des probabilités par classe et pour l'ensemble des segments. **RF** correspond à classification des données de télédétection. **Max P.** et **Moy P.** correspondent respectivement aux classifications issues de l'application du module des règles spatiales et de la combinaison des des $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la probabilité maximale et de la moyenne pondérée. 143

26 Distributions des probabilités de chaque classe pour les segments bien classés et mal classés. **RF** correspond à classification des données de télédétection et **Moy P.** à la classification issue de l'application du module des règles spatiales et de la méthode de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la moyenne pondérée (paramètres $\Delta RF = 0,4$ et $coef\ wRS = 1$). La classe *sol nu agricole* n'est pas représentée car aucun segment n'a été attribué à cette classe. 145

27 (**a**) Carte de l'occupation des sols obtenue après application du module des règles spatiales avec une combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la moyenne pondérée paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$) et (**b**) carte des probabilités les plus élevées pour la méthode de la moyenne pondérée (paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$). 147

28 Cartes de l'occupation des sols selon (a) la classification des données de télédétection, l'application du module des règles spatiales avec combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par (b) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0.2$ et $coef\ wRS = 1$, (c) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0.4$ et $coef\ wRS = 1$ et (d) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0.5$ et $coef\ wRS = 5$. 149

29 Exemples d'erreurs de classification des données de télédétection dans la commune d'Alakamisy où les parcelles de *légumes* (orange) sont classées comme *verger* (couleur vert), alors que seuls quelques pieds de pommiers sont présents. Ce cas est typique et peut être corrigé par le passage de la classe *verger* à la classe *légumes* ou *pomme de terre* grâce au module 2 : règles spatiales. Plus occasionnellement, ce changement peut être en faveur des cultures potagères. Les autres classes présentes sur la carte (b) sont : *pomme de terre* (marron) et *cultures potagères* (mauve) et *bâti* (noir). 151

30 Cartes de l'occupation et de l'usage des sols de la cuvette d'Ambohibary, selon (a) la classification des données de télédétection, (b) l'application du module des règles spatiales en combinant les $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0,2$ et $coef\ wRS = 1$, (c) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0,4$ et $coef\ wRS = 1$ et (d) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0,5$ et $coef\, wRS = 5$ 152

31 Comparaison entre une image Pléiades, la classification des données de télédétection et la fusion entre les classification des données de télédétection et du module des règles spatiales. 155

32 (a) Carte de l'occupation et de l'usage des sols après l'application du module de modélisation de la répartition spatiale des classes avec un seuil de fiabilité de 0,8 et un seuil de variabilité de 0,01. (b) Carte des segments ayant changé de classe entre l'application du module des règles spatiales (paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$) et l'application du module de modélisation de la répartition spatiale des classes avec un seuil de fiabilité de 0,8 et un seuil de variabilité de 0,01. . 157

33 Carte des types de reliefs de notre zone d'étude selon la méthode (Weiss, 2001). Les classes sont décrivent dans le tableau 23 270

34 Schéma du calcul du temps de trajet selon l'équation de Tobler (20) appliqué aux pixels d'un MNT selon un voisinage de Moore. Pour une facilité de lecture, les flèches d'application de la fonction de coût du voisinage de Moore d'ordre 2 ne sont représentées que un point de départ. 273

35 Carte du temps de trajet entre un segment et le bâtiment le plus proche. 274

TABLE DES TABLEAUX

1 Critères pour différencier la forme d'agriculture d'entreprise de la forme d'agriculture familiale, tableau issu de Bélières et al. (2013) 8

2 Synthèse des données utilisées. 67

3 Répartition de l'ensemble des relevés de terrain par classe pour la saison culturale 2018-2019 (polygones > 200m²). 73

4 Nombre d'échantillons relevés par campagne de mesure, et taille des polygones. Au total, 2617 relevés ont été effectués sur le terrain, 126 ont été rajoutés par photo-interprétation et 201 polygones ont été supprimés, car jugés trop petits. 74

5 Évaluation de la segmentation avec comme paramètre d'échelle **250**, comme paramètre spectral **0,3** et comme paramètre spatial **0,7**. La précision est calculée pour l'ensemble des segments qui intersectent un polygone de vérité terrain (rappel = 1). 91

6 Évaluation de la segmentation en terme de précision et de F-mesure, en fonction de la valeur du rappel choisi. 91

7 Tableau récapitulatif du paramétrage de l'algorithme de forêts aléatoires et des paramètres sélectionnés. 97

8 Données utilisées dans le modèle. 104

9 Variables et seuils utilisés pour sélectionner les segments non-agricoles. 110

10 Correspondance entre la nomenclature et les connaissances utilisées pour la création des onze scénarios. La catégorie « Autre » correspond aux classes *culture boisée de rente*, *savane*, *sols rocheux*, *bâti*, *eau*, *herbes et autres cultures fourragères* et *sol nu agricole*. Le nombre maximum de connaissances mobilisées est de 8 (classe *légumes*). 113

11 Répartition des cultures dans les Fokontany Tsaratanana, Alakamisyet et Andohafarihy du Sud de la commune d'Alakamisy selon le responsable d'une association d'agriculteurs et selon le responsable du développement rural de la commune. Il est important de préciser que selon les personnes enquêtées, le maïs pur est très rare, il est majoritairement associé au riz pluvial ou aux pommes de terre (principalement en bordure de champ). Sources enquêtes auprès des responsables locaux, Annexe F. 117

12 Extrait de la monographie de la commune rurale Alakamisy - Ambano Antsirabe II (Alakamisy, 2017). Dans le texte, la superficie de la commune est estimée à 42 km^2 alors qu'en la mesurant à partir de la couche vectorielle de la commune, son aire est de 68,43 km^2. Les calculs de pourcentage de surface ne sont pas compris dans la monographie. Les arbres sont recensés par pieds malgré le fait que la monographie indique leur superficie en hectares. Même si cela n'est pas précisé dans la monographie, les surfaces cultivées sont estimées à l'année, la somme peut donc être supérieure à 100 %, ce qui signifie que des parcelles sont en culture au moins pour la saison culturale principale et pour une autre saison culturale. 117

13 Nombre de Fokontany en fonction de leur aire dans la zone d'étude. 118

14 Comparaison des résultats pour chaque type de combinaison de données satellitaires et auxiliaires. **BS** signifie bandes spectrales, **I** signifie indices spectraux, **VA** signifie variables auxiliaires (textures, pente, altitude, exposition ; voir chapitre 3, section 2.3 Données support pour l'application des connaissances) 125

15 Métriques d'évaluation par classe et globales pour la classification des données de télédétection, q correspond au désaccord de quantité, a au désaccord de répartition. Les résultats ont été calculés par aire et sont exprimés en pourcentage. La couleur verte correspond à la catégorie 1, la couleur jaune à la catégorie 2 et la couleur rouge à la catégorie 3. Pour une meilleure lisibilité, les résultats sont exprimés en pourcentage et q et a ont été multipliés par 100. 127

16 Résultats de la précision globale (%) à l'issue de l'application du module des règles spatiales (combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ en un $VPAC_{Fusion}$ incluse) en fonction de la variation des paramètres ΔRF et de la pondération (équation 14 pour la méthode de la moyenne pondérée) avec un seuil de fiabilité de 1 et un $coef\ classe$ de 0,5. La précision globale de la classification des données de télédétection est de 81,93 %. 133

17 Métriques d'évaluation par classe et globales pour la classification des données de télédétection combinées avec le module des règles spatiales à l'aide de la méthode moyenne pondérée (Moy.P.) (paramètres $\Delta RF = 0,5$ et $coef\ wRS = 1$) et de la méthode de la sélection de la probabilité maximale (P. Max.) (paramètre $\Delta RF = 0,2$). q correspond au désaccord de quantité et a au désaccord de répartition. Les trois catégories définies dans la classification des données de télédétection sont représentées par des couleurs : ma couleur verte correspond à la catégorie 1 (F-mesure 1 - 0,8), la couleur jaune à la catégorie 2 F-mesure 0,8 - 0,5) et la couleur rouge à la catégorie 3 (F-mesure 0,5 - 0). Pour une meilleure lisibilité, les résultats sont exprimés en pourcentage et q et a ont été multipliés par 100. 136

18 Métriques d'évaluation globales pour la classification des données de télédétection combinées avec le module des règles spatiales à l'aide de la méthode moyenne pondérée (Moy.P.) (paramètres $\Delta RF = 0,5$ et $coef\ wRS = 1$) et de la méthode de la sélection de la probabilité maximale (P. Max.) (paramètre $\Delta RF = 0,2$). Les résultats sont exprimés en pourcentage. 137

19 Résultats du test du $\chi 2$ de McNemar avec un degré de liberté de 1. 137

20 Répartition des aires de chaque classe en pourcentage, selon la classification des données de télédétection et l'application du module des règles spatiales en combinant les $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la probabilité maximale ($\Delta RF = 0,2$) et par la méthode de la moyenne pondérée ($\Delta RF = 0,2$ et $coef\ wRS = 1$). 140

21 Synthèses des bandes spectrales des satellites utilisés. En gras les bandes spectrales qui n'ont pas été utilisées. 221

22 Sélection d'ouvrages techniques et scientifiques utilisés pour l'identification et la définition des règles 225

23 Classes de la position topographique de de chaque pixel selon la méthode Weiss (2001) 269

24 Correspondance entre les classes et les codes pour les matrices de confusion. . . . 275

25 Matrice de confusion de la classification des données de télédétection. Les résultats sont exprimé en pourcentage d'aire. La correspondance des classes est présentée dans le tableau 24. 276

26 Matrice de confusion de la fusion des $VPAC_{RF}$ (classification des données de télédétection) et $VPAC_{RS}$ (module 2 : règles spatiales en un $VPAC_{Fusion}$ selon la méthode de la moyenne pondérée avec les paramètres $\Delta RF = 0,5$ et $coefwRS = 1$. La correspondance des classes est présentée dans le tableau 24. 277

27 Règles basées sur la pédo-géomorphologie 279

28 Règles basées sur la distance à un cours d'eau 279

29 Règles basées sur l'altitude 280

30 Règles basées sur la pente 280

31 Règles basées sur la géomorphologie 281

32 Règles basées sur l'exposition 282

33 Règles basées sur le temps de déplacement depuis un bâtiment 282

34 Règles basées sur la distance à une route 282

35 Règles basées sur la distance à la route nationale 7 283

36 Règles basées sur la distance visuelle 283

Glossaire

cohérence : L'analyse des résultats a pour principal objectif d'évaluer et d'appréhender si à l'issue de l'utilisation de la modélisation spatiale, la caractérisation de la zone d'étude est « **agriculturellement cohérente** ». Cette cohérence est principalement thématique et peut être discutée à la fois par des résultats qualitatifs et par l'analyse qualitative des cartes produites. L'évaluation de la cohérence est subjective car c'est le regard du modélisateur-thématicien qui va estimer, selon des indicateurs définis, si une carte est cohérente par rapport aux stratégies des agriculteurs et par rapport à une autre carte (ici celle issue du traitement des données de télédétection).. xxi, 119, 120, 167, 168, 185

cohérent *voir* cohérence. 120, 142, 156, 167–169, 185

contexte : Pour un objet, l'ensemble des propriétés qui le caractérisent peuvent être résumées par le terme **contexte**. Ce contexte peut être biophysique, temporel, socio-économique ou climatique.. 41, 44, 45

fiabilité : Ces probabilités peuvent être assimilées à un indicateur de la **fiabilité** d'appartenance ou non d'un objet à une classe donnée. Si pour un objet, la probabilité d'une classe donnée est proche de 0, alors il est peu probable que l'objet présente cette classe d'occupation ou d'usage des sols. Si la probabilité est proche de 1, alors cette classe pourra être affectée à l'objet avec une grande fiabilité. Dans ces deux cas on peut considérer que la probabilité de cette classe est fiable dans l'inclusion ou l'exclusion. Plus la probabilité d'une classe de est proche de 1 ou de 0, plus la probabilité de voir l'objet appartenir ou ne pas appartenir à la dite classe sera fiable.

Pour une plus grande clarté de ce manuscrit, la définition utilisée de la **fiabilité** correspond à : lorsque la probabilité du VPAC est élevée, alors le segment peut être considéré comme fiable.. xxi, 45, 46, 49, 55, 58, 118, 121, 142, 144, 145, 153, 154, 159, 166–168, 172, 177, 285

fiable *voir* fiabilité. 46, 107, 118, 122

Fokontany : Selon l'article 2 du décret N° 2004-299 du 3 mars 2004, structurant administrativement les Fokontany , "Le Fokontany est une subdivision administrative de base au niveau de la Commune". Le Fokontany, selon l'importance des agglomérations, comprend des hameaux, villages, secteurs ou quartiers. xviii, 65, 67, 77, 80, 98, 99, 101–103, 105, 116–119, 122, 150, 154, 156–159, 168, 170, 171, 174, 178, 182, 239, 240, 242, 285

VPAC : Vecteur de Probabilité d'Appartenance aux Classes. Chaque segment est caractérisé par une propriété *d'occupation du sol possible*. Cette possibilité est représentée par un vecteur de probabilité d'appartenance à une classe **VPAC** qui contient des éléments ordonnés et indexés selon les types d'occupation du sol définis. C'est une représentation simplifiée de la possibilité d'attribution d'un segment à une classe définitive. A un instant t, la classe définitive correspond à l'élément disposant de la probabilité maximum dans le VPAC.

Ces VPAC peuvent être déclinés selon leur utilisation ou les processus affectant les éléments le composant.

$VPAC_{RF}$ correspond aux VPAC résultant des processus de traitement des données de télédétection.

$VPAC_{RS}$ correspond aux VPACs résultant des processus de modélisation de règles avec Ocelet.

$VPAC_{final}$ correspond aux VPACs définitifs, dont la classe majoritaire est désignée comme la classe d'occupation ou d'usage du sol du segment. xxi, xxiii, 45–48, 51–58, 115, 119, 130, 137, 138, 142, 146, 162, 164, 165, 167, 168, 170–172, 175–179, 183

Acronymes

ASP Anomaly hot Spots of Agricultural Production. 62

B bleu. 221

BRI Brightness Index. 94, 223

CIRAD Centre de coopération Internationale en Recherche Agronomique pour le Développement. 76

CNES Centre National d'Études Spatiales. 20, 69

DP SPAD Dispositif de Recherche et d'Enseignement en Partenariat sur les Systèmes de Production d'Altitude et Durabilité. 60, 76

ESA European Space Agency. 19, 69, 163

EVI Enhanced Vegetation Index. 25

FAO Food and Agricultural Organization. 6, 7, 9, 16, 103

FEWS NET Famine Early Warning Systems Network. 16, 62

FIFAMANOR Fiompiana Fambolena Malagasy Norvéziana. 243

FOFIFA Centre National de Recherche Appliquée au Développement Rural. 242

GEOGLAM Group on Earth Observations Global Agricultural Monitoring. 23

GLCM matrice de co-occurences des niveaux de gris. 92, 93

GPS système mondial de positionnement. 72

HLPE High Level Panel of Experts (Groupe d'experts de haut niveau). 7

ISA Isreal Space Agency. 69

JECAM Joint Experiment for Crop Assessment and Monitoring. xiii, 63, 65, 71, 73, 81

JICA Agence japonaise de coopération internationale. 243

MAJA MACCS-ATCOR Joint Algorithm. 69

MDA Minimum Decrease In Accuracy. 121, 124

MIR moyen infrarouge. 69, 93–95, 221

mNDVI modified Normalized Difference Vegetation Index. 94, 223

MNDWI modified Normalized Difference Water Index. 95, 223

MNT modèle numérique de terrain. xvi, 51, 77, 78, 86, 89, 94, 95, 104, 269, 272, 273, 279, 280, 282, 283

NASA National Aeronautics and Space Administration. 20

NDRE Normalized Difference Red-Edge. 94, 223

NDVI Normalized Difference Vegetation Index. 25, 53, 94, 171, 223

NDWI Normalized Difference Water Index. 25, 94, 223

OBIA object-based image analysis. 20, 22, 24, 32, 37, 44, 45, 47, 51, 58, 164, 167, 171, 185

OMP Ocelet Modelling Platform. 28, 81

OTB Orfeo ToolBox. 81, 89

PAM Programme Alimentaire Mondial. 62

PIR proche infrarouge. 69, 93–95, 221

R rouge. 94, 95, 221

SMA système multi-agents. 27, 28, 31

SRTM Shuttle Radar Topography Mission. 78, 89

TF-IDF Term Frequency-Inverse Document Frequency. 177

Theia Pôle de données et de services surfaces continentales. 69

THRS très haute résolution spatiale. 20, 22, 24, 25, 68, 74, 80, 81, 86, 88, 89, 92, 94, 95, 97, 122, 128, 146, 148, 150, 151, 159, 168, 170, 171, 173, 178, 183

TOA top of atmosphere. 89

USAID United States Agency for International Development. 16

USDA United States Department of Agriculture. 8

USGS United States Geological Survey. 19, 89, 94, 279, 280, 282, 283

V vert. 94, 221

$VPAC_{Fusion}$ vecteur de probabilité d'appartenance aux classes correspondant au VPAC définitif après l'exécution du module 2 : règles spatiales, dont la classe majoritaire est désignée comme la classe d'occupation ou d'usage du sol du segment. Ce VPAC peut correspondre au $VPAC_{RF}$ si les segments n'ont pas été sélectionnés pour être modifiés, ou à la fusion du $VPAC_{RF}$ et du $VPAC_{RS}$ voir VPAC.. xviii, xx, 105, 114–116, 118–123, 132, 133, 141–144, 148, 154, 158, 167, 277, 285

$VPAC_{RF}$ vecteur de probabilité d'appartenance aux classes issu de l'algorithme de forêt aléatoire, voir VPAC. xiv–xvi, xviii–xx, xxiii, 88, 96, 97, 104, 105, 107–110, 115, 119, 121–123, 128, 130, 132, 133, 137–140, 142, 143, 145–149, 151–153, 164, 167, 168, 170, 177, 178, 181, 271, 277

$VPAC_{RS}$ vecteur de probabilité d'appartenance aux classes issu du module des règles spatiales, voir VPAC. xiv–xvi, xviii–xx, xxiii, 102, 103, 105, 114, 115, 118, 119, 121, 122, 128, 130, 132, 133, 135, 137–140, 142, 143, 145–147, 149, 152, 153, 164, 165, 167, 168, 174, 177, 178, 277

WGS 84 World Geodetic System 1984. 67

INTRODUCTION GÉNÉRALE

Le plan d'action de la Décennie des Nations Unies pour l'agriculture familiale (2019-2028) s'inscrit dans les Objectifs du développement durable. Il a pour but de mettre l'agriculture familiale au centre de la lutte contre la pauvreté et la faim, tout en préservant la biodiversité et les ressources naturelles (FAO and FIDA, 2019). Depuis 2014, année internationale de l'agriculture familiale, ce mode de production est au centre des politiques nationales et internationales. En effet, l'agriculture familiale est la principale forme de production alimentaire mondiale. Elle est également l'une des principales composantes économiques et agricoles des espaces tropicaux et subtropicaux, approvisionnant à la fois leurs marchés domestiques et les marchés internationaux, créant des emplois et réduisant la pauvreté et les inégalités géographiques et socio-économiques (HLPE, 2013).

Bien que sensible aux risques climatiques et environnementaux, l'agriculture familiale est mieux intégrée dans les territoires, est plus soucieuse de l'environnement, et potentiellement plus résiliente que d'autres modèles agricoles. C'est un élément majeur de la sécurité alimentaire à l'échelle mondiale, et en particulier dans les pays des Suds où elle est majoritaire. Pour soutenir l'agriculture familiale, l'élaboration de politiques publiques agricoles, mais aussi environnementales et territoriales nécessite une bonne connaissance de ces systèmes agricoles à l'échelle nationale (Sourisseau et al., 2014). Malheureusement, les données sur les systèmes de culture et leur production ne sont pas toujours disponibles (Bélières et al., 2013), en particulier dans les pays des Suds où agriculture familiale est souvent synonymes de « petite agriculture », et donc synonyme d'une agriculture diversifiée, dynamique, difficile à qualifier et quantifier.

Depuis une quarantaine d'années, la télédétection a montré ses capacités uniques pour cartographier, caractériser et suivre les systèmes agricoles, à la fois sur de grandes étendues et dans le temps. La multiplication des capteurs d'observation de la Terre et l'augmentation du volume et de la diversité des données disponibles (séries temporelles d'images satellites multispectrales à résolution décamétrique, images métriques en tout point du globe, données optiques et radar ; Tonneau et al. (2020)) ont permis le développement de nouvelles méthodes de traitement de données qui permettent de détecter et caractériser des éléments du paysage de plus en plus petits, et à une fréquence plus élevée.

Pour passer de l'image satellite à une carte d'occupation ou d'utilisation des sols, des classifications supervisées sont généralement utilisées. L'algorithme de classification utilise des données d'apprentissage, étiquetées, et attribue aux pixels, ou groupes de pixels (objet), de l'image, la classe qui leur correspond en fonction de leurs attributs spectraux et temporels (et spatiaux pour les objets). L'augmentation de la fréquence de revisite des satellites, de leur résolution spatiale et de leur richesse spectrale permet d'accroître le nombre d'attributs utilisés et conduit à une meilleure classification des pixels ou objets. Avec un temps de revisite de cinq jours et treize bandes spectrales, la mission Sentinel-2 de l'Agence Spatiale Européenne représente un immense

progrès pour la caractérisation des surfaces agricoles. La qualité des cartes d'occupation et de l'usage des sols produites avec ces nouvelles données est bonne pour les espaces d'agriculture intensive, mais reste insatisfaisante pour les systèmes agricoles complexes tels que l'agriculture familiale tropicale. L'utilisation des données de télédétection pour caractériser ces systèmes agricoles se heurte à une contrainte double : les contraintes inhérentes à la télédétection et celles liées aux caractéristiques de l'agriculture familiale tropicale (Begue et al., 2016). Pour la télédétection, la nébulosité limite les acquisitions d'images optiques ; la taille des champs peut être du même ordre de grandeur que celle d'un pixel, ce qui conduit à des pixels mixtes, difficiles à classer ; la variabilité spatiale intra-parcellaire et inter-parcellaire des cultures liées à la l'hétérogénéité des milieux et à la diversité des pratiques culturales vont également avoir un impact sur la distribution des propriétés spectro-temporelles des pixels et donc rendre difficile leur attribution à une classe unique (Lebourgeois et al., 2017) ; enfin la synchronisation des phénologies des agrosystèmes et des écosystèmes peut également induire des confusions dans la classification (Xiong et al., 2017).

De nouvelles approches ont été conçues pour prendre en compte ces contraintes. Elles sont basées par exemple sur le développement de méthodes de classification plus efficientes, comme l'apprentissage profond (ex. (Benedetti et al., 2018)), ou sur l'utilisation de sources de données plus diversifiées, telles que des données multi-capteurs (multi-bandes, multi-résolution ; ex. Forkuor et al. (2014); Ienco et al. (2019)) ou des données multi-sources (utilisation de données environnementales par exemple ; e.g. Lebourgeois et al. (2017); Dupuy et al. (2020a)). Malgré une amélioration des résultats de classification de l'occupation et de l'usage des sols, des progrès restent possibles. En effet, ces méthodes, bien que s'appuyant sur une expertise pour le choix des données, laissent peu de place à l'utilisation de connaissances thématiques pour améliorer la cartographie.

L'utilisation des connaissances agricoles est une alternative nouvelle et peu explorée pour améliorer la cartographie des systèmes d'agriculture familiale. Ces connaissances peuvent être d'ordre biophysique (par exemple les limites d'altitude pour un type de culture), ou d'ordre agronomique (par exemple un choix de pratique culturale selon une tradition). Les méthodes de modélisation des dynamiques spatiales semblent être un bon moyen d'intégrer des connaissances qui représentent des processus dynamiques et multi-scalaires, et une alternative à des méthodes mathématiques (chaînes de Markov ; ex. Bailly et al. (2018)). La construction de règles permet de formaliser et de retranscrire des connaissances issues d'autres sciences (ex. El Hajj et al. (2009)). L'agriculture familiale est un cas d'étude intéressant pour tester l'utilisation des connaissances dans la classification de l'utilisation agricole des sols, car les agriculteurs gèrent les ressources naturelles et les écosystèmes, partagent des savoirs traditionnels et aménagent leur environnement. L'objectif général de ce travail de thèse est d'explorer de nouvelles pistes méthodologiques pour combiner données de télédétection et modélisation des dynamiques spatiales pour produire des cartes thématiquement et spatialement cohérentes sur l'occupation et l'usage agricole des sols en milieux tropicaux complexes. Pour cela, cette thèse vise à répondre à deux sous-objectifs :

- produire un modèle conceptuel alliant à la fois des méthodes de traitement de données de télédétection et des méthodes de modélisation spatiale pour cartographier l'usage des sols ;

- appliquer le modèle conceptuel à un cas d'agriculture familiale tropicale.

Le cas d'étude est les Hautes Terres du Vakinankaratra, à Madagascar. Cette région est représentative de l'agriculture familiale tropicale, avec des paysages fortement hétérogènes. Le Vakinankaratra dispose de la seconde plus haute densité de population de l'île et est composé de zones de montagnes et de plaines, alternant entre cultures pluviales et cultures irriguées. Les cultures sont variées : si les bas-fonds sont dominés par le riz irrigué, les pentes et plateaux sont occupés par une importante variété de cultures pluviales (maïs, riz pluvial, soja, taro) qui peuvent être associées entre elles. Les exploitations sont de petite taille et le travail de la terre est principalement manuel. Les données utilisées pour ce travail sont une série temporelle d'images Sentinel-2 et Venμs acquise entre septembre 2018 et juin 2019 sur une zone de 620 km², et des enquêtes de terrain. Les outils mis en œuvre sont la chaîne Moringa (Gaetano et al., 2019) pour la partie traitement des images et la plateforme Ocelet (Degenne and Lo Seen, 2016) pour la modélisation spatio-temporelle des territoires.

La thèse est organisée en six chapitres. Le premier chapitre « Contexte et problématique de la thèse » présente les enjeux de l'agriculture familiale dans le monde, et l'état de l'art des méthodes existantes en télédétection et modélisation spatio-temporelle pour la cartographie et le suivi de ces systèmes complexes. Le chapitre 2 « Modèle conceptuel » présente la démarche adoptée et décrit les principaux modules spatiaux et temporels qui le composent. Les chapitres 3, 4 et 5 répondent au deuxième sous-objectif de la thèse et présentent le cas d'application du modèle conceptuel à une zone d'étude des Hautes Terres du Vakinankaratra. Le chapitre trois « Zone d'étude, matériel et méthodes » décrit le contexte agricole de la région du Vakinankaratra et la zone d'étude, les données (données satellites, données de terrain et données de support à l'application des règles) et les outils Moringa et Ocelet utilisés respectivement pour la télédétection et la modélisation. Le chapitre 4 « Méthode d'application des modules à la région des Hautes Terres » est focalisé sur la description de l'adaptation du modèle conceptuel au cas d'étude, sur la production des données nécessaires à l'application des règles et sur l'application des modules spatiaux à la zone d'étude. Le chapitre 5 « Résultats d'application » présente l'analyse quantitative et qualitative des résultats cartographiques obtenus. Sont présentés les résultats de la classification des données de télédétection et ceux de l'approche combinée télédétection-modélisation avec application du module des règles spatiales. Une analyse qualitative de l'application du module de modélisation de la répartition spatiale des classes est également fournie. Enfin, le chapitre 6 résume les résultats obtenus et fait le bilan du modèle conceptuel développé et de son application à un système d'agriculture familiale tropicale. Les limites de l'approche les voies d'amélioration sont discutées. Pour conclure, deux perspectives sont proposées à ce travail, l'une de développement méthodologique et l'autre pour une mise en application à plus grande échelle des résultats obtenus.

Chapitre 1

Contexte et problématique de la thèse

Table des matières

1. L'agriculture familiale
   1.1. Un système à multiples facettes
   1.2. Agriculture familiale et sécurité alimentaire
   1.3. Deux modes de représentation
2. Les outils d'analyse spatio-temporelle
   2.1. Télédétection et agriculture familiale
   2.2. Modélisation spatio-temporelle de l'agriculture familiale
   2.3. Combinaison télédétection / modélisation spatio-temporelle
3. Objectifs de la thèse et démarche adoptée

Le premier chapitre de ce manuscrit présente le contexte général dans lequel s'inscrit notre travail, à la fois de manière thématique et de manière méthodologique. Dans la première section, nous aborderons les questions de l'agriculture familiale dans le monde et dans les espaces tropicaux au travers de la définition de ce mode de production agricole. À partir de la définition retenue, nous présentons un état des lieux de l'agriculture familiale et son lien avec la sécurité alimentaire. Nous terminerons cette section par les modes de représentation de cette agriculture. Dans la seconde section, nous aborderons les outils pour caractériser et représenter l'agriculture familiale en abordant à la fois leurs points forts et leurs faiblesses. Ensuite, nous exposerons les deux principaux outils pour représenter l'agriculture familiale : la télédétection et la modélisation des dynamiques spatiales. L'utilisation conjointe de méthodes de traitement de données de télédétection et de modélisation spatiales sera ensuite abordée. La section trois présentera la problématique et les objectifs de ce travail de recherche ainsi que les moyens mis en place pour y répondre.

1 L'agriculture familiale

La première sous-section présente les différentes définitions de l'agriculture familiale et la diversité des systèmes agricoles et des systèmes de culture qui les compose. La deuxième sous-section met en perspective ce mode de production agricole et la sécurité alimentaire, montrant la vulnérabilité de ce système, mais aussi sa résilience pour la lutte contre l'insécurité alimentaire. La troisième sous-section aborde les outils et définitions mobilisés pour caractériser l'agriculture familiale. Cette caractérisation passe sa cartographie, la définition de classes d'occupation et d'usage des sols et un choix d'échelle d'analyse spatio-temporelle. L'état de la cartographie de l'agriculture familiale est ensuite présenté, ainsi que les typologies d'exploitations agricoles pour représenter autrement sa diversité.

1.1 Un système à multiples facettes

1.1.1 Définitions

L'agriculture familiale (*Family Farming* ou *Family Agriculture* en anglais, parfois confondue avec l'appellation *smallholder agriculture*) n'a pas de définition claire et acceptée par la communauté scientifique.

Pour certains auteurs, l'agriculture familiale peut être présentée comme l'opposée du mode de production agricole extensif et intensif (Foley et al., 2011; Tilman et al., 2011), mais cette définition est simpliste et ne représente pas sa diversité et sa complexité. D'autres définitions se rapportent à des critères de taille, proches du concept de *smallholder agriculture*, comme celle de Wiggins (2009) qui définit l'agriculture familiale comme l'inverse des grandes exploitations en matière de taille et de rendement à l'hectare. Lorsque des auteurs utilisent le critère de taille pour définir une petite exploitation familiale, il est mis en perspective avec des critères régionaux ou nationaux[1] pour conserver sa cohérence (Jayne et al., 2016). Ainsi, en Asie, une exploitation d'une taille de 1 ha ou 2 ha sera considérée comme de la petite agriculture (et majoritairement familiale), mais pour le Brésil ce seuil passe à 50 ha. La figure 1 présente une cartographie de la taille des champs dans le monde, les plus petits étant principalement situés en zones subtropicales et en Asie.

La FAO and FIDA (2019) définissent l'agriculture familiale (y compris toutes les activités agricoles pratiquées dans un cadre familial) comme « un moyen d'organiser la production agricole, forestière, halieutique, pastorale et aquacole qui est gérée et conduite par une famille et repose principalement sur le travail familial, tant celui des femmes que celui des hommes. Exploitations agricoles et familles évoluent l'une avec l'autre et l'une par l'autre. Elles combinent des fonctions économiques, environnementales, sociales et culturelles ». Cette définition englobe une très grande variété de types d'exploitation avec comme critère principal la famille, donc une limitation de la main d'œuvre sur l'exploitation. Pour Delfosse (2015) la définition de la FAO est acceptable, car elle englobe la complexité de l'agriculture familiale et s'adapte aux différentes problématiques

[1] http://www.fao.org/family-farming/data-sources/dataportrait/farm-size/en/

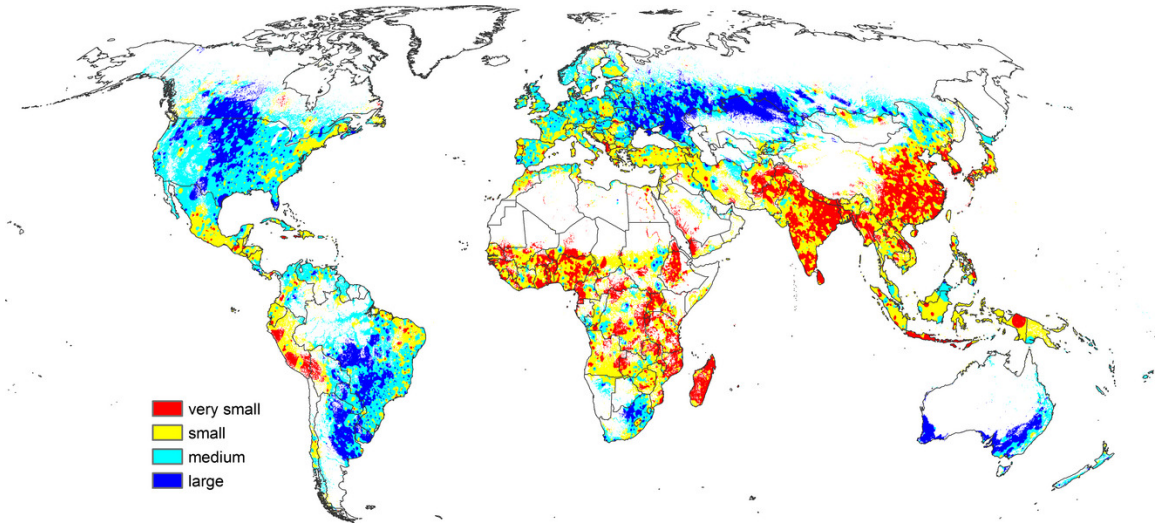


Figure 1 Carte de la taille des champs dans le monde, réalisée à dire d'expert par photo-interprétation d'environ 14 000 extraits d'images à très haute résolution spatiale mis à disposition sur la plateforme de *crowd sourcing* GEO-Wiki (Fritz et al., 2015).

selon l'échelle spatiale des objets d'étude. L'une des limites de cette définition est son critère qualitatif qui ne permet pas de distinguer de manière homogène l'agriculture familiale d'autres systèmes de production agricole entre des pays (tels que ceux présentés tableau 1). L'HLPE (2013) vise à être plus précis en la définissant comme une exploitation agricole gérée par une famille et employant sa propre force de travail ; la part d'autoconsommation est importante et l'exploitation est de petite taille, c'est-à-dire qu'elle est à la limite de sa capacité de production pour subvenir aux besoins du ménage. Les activités de la petite exploitation agricole regroupent les productions végétales et animales, la pêche et/ou la chasse et la transformation ou vente des produits. La part agricole est souvent complétée par des activités de services ou de travaux manuels (construction par exemple), que ce soit lors des périodes de soudure ou tout au long de l'année.

En opposition aux définitions uniquement basées sur un critère de taille de l'exploitation et en complément des définitions de la FAO et de l'HLPE, des auteurs ont un autre positionnement. Cousins (2013) propose de s'intéresser aux causes sociales et politiques pour définir l'agriculture familiale, mettant en valeur qu'elle n'est pas homogène et uniforme, mais qu'elle comprend un large panel de pratiques culturales au sein de mêmes systèmes agricoles. Bélières et al. (2013) caractérisent les formes de productions agricoles selon leur type d'organisation. Ils distinguent les exploitations de type « entrepreneurial » des exploitations de type « familial ». L'agriculture « patronale » est à mi-chemin entre ces deux formes d'organisation. Les critères de différenciation des trois principaux types d'agriculture (tableau 1), ne prennent pas en compte la taille de l'exploitation qui, prise comme seul critère, n'est pas discriminante et est porteuse de préjugés. Ces auteurs proposent cette définition de l'agriculture familiale : « L'agriculture familiale (*family farming*) désigne une des formes d'organisation de la production agricole regroupant des exploitations caractérisées par des liens organiques entre la famille et l'unité de production et par la mobilisation du travail familial excluant le salariat permanent. Ces liens se matérialisent par l'inclusion du capital productif dans le patrimoine familial et par la combinaison de logiques do-

mestiques et d'exploitation, marchandes et non marchandes, dans les processus d'allocation du travail familial et de sa rémunération, ainsi que dans les choix de répartition des produits entre consommations finales, consommations intermédiaires, investissements et accumulation ». Selon cette définition, l'agriculture patronale et l'agriculture familiale sont les formes les plus répandues dans le monde, l'agriculture d'entreprise représentant seulement moins 10 % des exploitations des pays industrialisés[2].

Table 1 Critères pour différencier la forme d'agriculture d'entreprise de la forme d'agriculture familiale, tableau issu de Bélières et al. (2013)

	Agricultures d'entreprises	⟷	Agricultures familiales
Critères / Idéaux-types	**Formes entrepreneuriales**	**Formes patronales**	**Formes familiales**
Main-d'œuvre	Exclusivement salariés	Mixte, présence de salariés permanents	Dominance familiale, pas de salariés permanents
Capital	Actionnaires	Familial ou association familiale	Familial
Management	Technique	Familial / technique	Familial
Statut juridique	Société anonyme ou autres formes sociétaires	Statut d'exploitant, formes associatives	Informel ou statut d'exploitant
Statut foncier	Propriété ou faire-valoir indirect formel	Propriété ou faire-valoir indirect, formel ou informel	
Consommation	Sans objet	Résiduelle	Autoconsommation partielle à dominante

Smallholder agriculture

Le terme *smallholder agriculture* pouvant être traduit par « petite exploitation agricole », ne définit une exploitation que par sa taille. Ce critère de taille peut être indexé sur la taille du domaine cultivé, du cheptel ou bien sur des critères économiques comme la valeur des ventes. Ce terme est très employé dans la littérature anglo-saxonne pour des raisons historiques : il permettait de différencier les exploitations industrielles coloniales des exploitations indigènes (Wickizer, 1960). Il a ensuite été maintenu pour désigner les exploitations non industrielles. Cette définition historique n'est plus d'actualité et recouvre les définitions d'agriculture familiale, patronale et une minorité d'agriculture d'entreprise (Bélières et al., 2013). Ce terme est largement utilisé dans certains domaines scientifiques, tels que la télédétection ou modélisation spatiale, pour parler à tort de l'agriculture familiale.

[2]En 2018, selon l'USDA, 98 % des fermes sont considérées comme des exploitations familiales aux États-Unis d'Amérique. Les exploitations qui ont des revenus inférieurs à 350 000 USD représentent 90 % des fermes américaines.

1.1.2 La diversité de l'agriculture familiale

L'agriculture familiale est un mode de production agricole qui correspond à plusieurs types de **systèmes agricoles**. La FAO définit un système agricole comme « une population de systèmes d'exploitations individuels comparables en termes de bases de ressources, de structure d'entreprise, de moyens d'existence et de contraintes des ménages, et pour lesquels il serait approprié d'appliquer des stratégies et des interventions de développement analogues »[3]. Ainsi, il existe différents systèmes agricoles dans le monde selon les ressources naturelles disponibles (climat, pédologie, situation géographique) et les types d'activités agricoles pratiquées dans la région concernée. Ces systèmes peuvent être imbriqués au sein d'un même territoire, les exploitations pouvant correspondre à différents types de ménages et différents systèmes de culture (Kuivanen et al., 2016; Alvarez et al., 2018; Schoneveld et al., 2019).

Au sein d'un même système agricole, des **systèmes de culture** peuvent être différenciés. Un système de culture est défini par Sebillotte (1990) comme « l'ensemble des modalités techniques mises en œuvre sur des parcelles cultivées de manière identique. Chaque système étant défini par : la nature des cultures et leur ordre de succession, les itinéraires techniques appliqués à ces différentes cultures, ce qui inclut le choix des variétés ». Par conséquent, ces systèmes de culture connaissent une forte hétérogénéité pour l'agriculture familiale, en raison de la diversité des ressources mobilisables par les exploitations, et de l'environnement mis en culture.

Bien que répondant à plusieurs caractéristiques communes, l'agriculture familiale est majoritairement hétérogène vis-à-vis de nombreux critères tels que l'accès à la terre, la fertilité du sol ou le groupe socioculturel des agriculteurs (Tittonell, 2014). Cette hétérogénéité, ainsi que l'importante diversité des espèces cultivées au sein d'une même exploitation, contribue à la complexité et à la richesse spécifique de l'agriculture familiale (Ricciardi et al., 2018).

Chaque système agricole peut être mis en relation avec la prévalence de la pauvreté dans les zones concernées. En Afrique sub-saharienne, 15 principaux systèmes agricoles sont identifiés, dont 6 en lien avec une pauvreté importante, représentant 58 % des terres. Le système agricole basé sur l'irrigation, représentant 1 % des terres, est le seul à être considéré comme peu corrélé avec la pauvreté. En Asie du Sud, 11 principaux systèmes agricoles sont identifiés. Leur lien avec la pauvreté est variable au sein d'un même système agricole, mais existe dans la quasi-totalité de ces systèmes, dont ceux basés sur l'irrigation (Dixon et al., 2001).

[3] http://www.fao.org/farmingsystems/description_fr.htm

L'agriculture familiale dans le monde

Au Brésil, 85 % des exploitations agricoles sont de forme familiale et possèdent 25 % de la superficie agricole. Paradoxalement, leurs capacités de production sont supérieures à celles des exploitations agro-industrielles vis-à-vis de certains types de production tels que le maïs, le manioc, le haricot ou le lait (Bélières et al., 2013).
Au Sénégal, 95 % des exploitations sont considérées comme familiales et produisent 60 % de la nourriture consommée dans le pays. Certaines cultures sont privilégiées par l'agriculture familiale, car adaptées à ce mode de production qui est principalement manuel. Le maraîchage par exemple, difficilement mécanisable, est produit à 80 % par des exploitations inférieures ou égales à 20 ha dont 91 % sont familiales (Sall et al., 2010).

1.2 Agriculture familiale et sécurité alimentaire

1.2.1 La sécurité alimentaire

Agriculture familiale et insécurité alimentaire (majoritairement la sous-nutrition) sont observées dans les mêmes zones de la planète[4] situées dans les zones tropicales et équatoriales (figure 2). Dans ces espaces géographiques, l'agriculture est l'une des principales composantes économiques, ainsi que la principale productrice de nourriture. Dans les pays en voie de développement, 80 % des denrées alimentaires consommées sont issues d'environ 500 millions d'exploitations possédant moins de 2 ha[5] (Lowder et al., 2016; Dieppe and World Bank, 2020).

La sous-alimentation n'est pas répartie socialement de manière égale dans les pays en développement. Elle touche environ 50 % des agriculteurs pauvres, 22 % des paysans ne possédant pas de terres, 20 % des urbains pauvres et 8 % des personnes vivant de manière traditionnelle en zone rurale (Collin, 2012). Cette sous-alimentation est liée aux conditions d'accès à la nourriture, que ce soit l'accès à de bonnes terres agricoles ou bien le revenu nécessaire à l'achat d'aliments (FAO et al., 2020).

La croissance démographique entraîne le besoin d'augmenter la production par personne. Mais, la possibilité de cultiver de nouvelles terres pour une personne ou pour un groupe est parfois compliquée, car régie par l'héritage, et les terres à mettre en valeur deviennent de plus en plus rare (Burnod et al., 2016). Les personnes ne travaillant pas dans le secteur primaire ou n'étant pas, ou plus, en âge de travailler, augmentent plus rapidement que la croissance de la productivité de la terre et la productivité du travail (Bélières et al., 2013). Par exemple, en Afrique Sub-Saharienne, les terres agricoles sont saturées et la superficie disponible par actif diminue. Dans certains pays tels que le Sénégal, la croissance de l'agriculture familiale est moins rapide que la croissance de la population, entraînant un déséquilibre entre la consommation et la production (Sall et al., 2010).

[4] Ici ne sera pas abordée la question de sécurité alimentaire pour l'agriculture familiale dans les pays industrialisés, car les problématiques sont différentes. À la sous-alimentation, s'ajoutent les questions de sur-alimentation, de qualité et de sûreté de l'alimentation dans ces pays.

[5] https://www.un.org/sustainabledevelopment/fr/hunger/

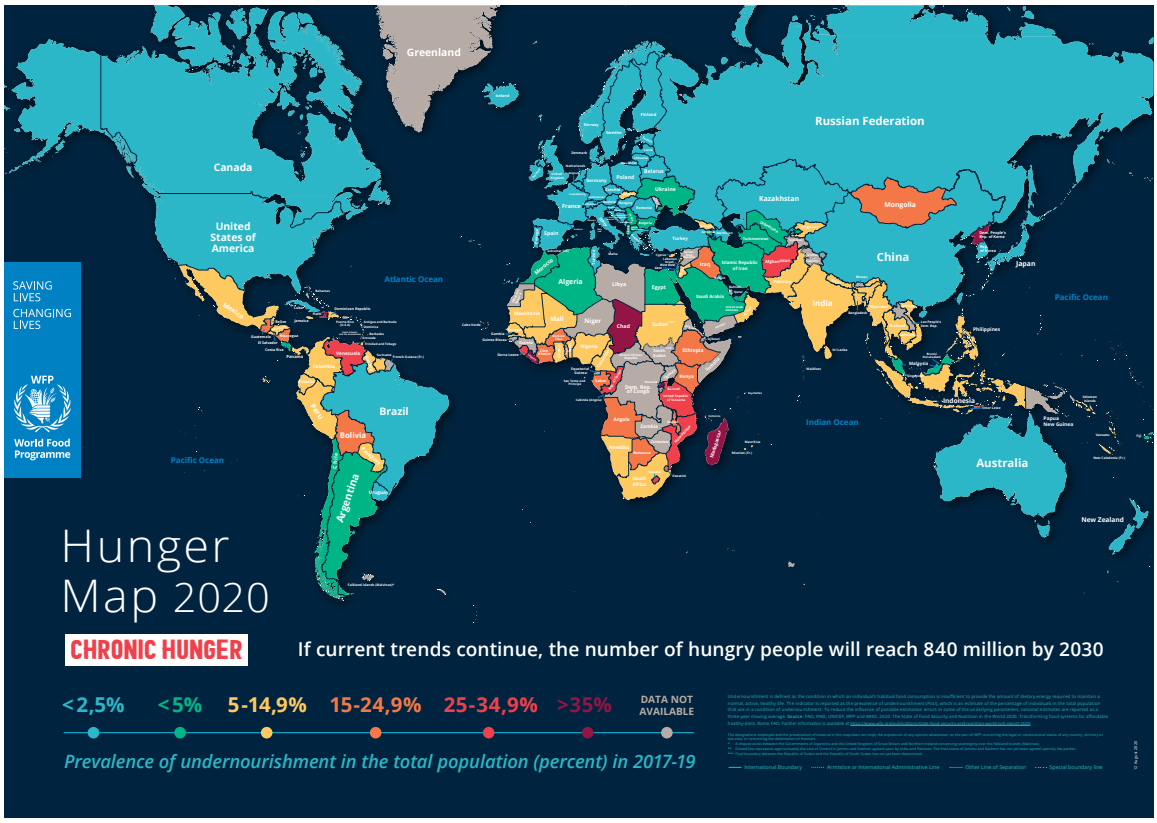


Figure 2 Carte de la sous-nutrition en 2020 selon le Programme Alimentaire Mondial (https://www.wfp.org/publications/hunger-map-2020.)

Actuellement, malgré les problèmes d'insécurité alimentaire dans les zones sub-tropicales, l'agri-culture familiale est l'un des principaux moyens de lutte pour préserver ou améliorer la sécurité alimentaire. Ce type d'agriculture est le plus présent, produisant la majorité de la production agricole, et les possibilités d'amélioration de ces systèmes agroalimentaires sont encore possibles (Bélières et al., 2013).

1.2.2 Un système vulnérable...

L'agriculture familiale est sensible à différents types de risques en raison de la faiblesse de ses capitaux, qu'ils soient financiers ou fonciers. Le risque domestique, tels que de mauvaises relations intrafamiliales ou la maladie ou un décès dans la famille, peut par exemple fortement diminuer la force de travail disponible et donc limiter les surfaces mises en culture. Le risque politique avec la favorisation d'autres modèles agricoles peut limiter l'accès à certaines ressources aux petites exploitations (irrigation, infrastructures, débouchés commerciaux). La faible rémunération du travail familial a eu comme avantage de réduire les coûts de production, mais devient une vulnérabilité face à la hausse du coût de la vie (Bélières et al., 2013). L'agriculture familiale est également vulnérable face aux risques économiques nationaux (surestimation de la production) (David-Benz et al., 2014), climatiques (inondations, sécheresses, retard de la saison de pluies), et aux mauvais choix stratégiques qui peuvent créer une spirale d'appauvrissement (Holzmann and Jørgensen, 2001; Morton, 2007; HLPE, 2013).

Le changement climatique a et va avoir un effet sur l'agriculture au niveau mondial (Lobell and Field, 2007; IPCC, 2019), mais celui-ci est particulièrement important pour l'agriculture familiale tropicale (Morton, 2007). En effet, ce type d'agriculture est majoritairement pluvial (Wani et al., 2009; Sathyan et al., 2018) donc fortement impacté par les variations intra-saisonnières des régimes de précipitation, et ce sur tout type de topographie (collines, plaines inondables) avec leurs propres aléas tels que les inondations ou les mouvements de terrain (Morton, 2007). La faible capacité d'adaptation aux changements climatiques de ce système (Adger et al., 2003; Harvey et al., 2014; Holland et al., 2017), et sa dépendance à des biens et services écosystémiques qui se raréfient (déforestation, perte de biodiversité) (Shiferaw et al., 2019) mettent en péril l'agriculture familiale. Par exemple, l'actuelle épidémie de SARS-CoV-2 a un effet négatif sur la sécurité alimentaire, principalement pour l'agriculture familiale où les systèmes agricoles peuvent être soumis à certains risques (diminution d'acheteurs sur les marchés, diminution des prix de vente) (WFP, 2020).

1.2.3 ...mais avec des forces et une certaine résilience

L'agriculture familiale est un moyen de lutte contre l'insécurité alimentaire. La productivité de l'agriculture familiale est généralement plus importante que celle des exploitations industrielles et mieux adaptée à certaines cultures telles que les fruits et légumes (Dan, 2006; Larson et al., 2012; HLPE, 2013; WB, 2016). L'agriculture familiale est résiliente aux chocs, car elle dispose de plusieurs sources d'approvisionnement financier. Les principales sources financières sont la vente de produits agricoles et les activités non agricoles qui permettent, lorsque cela est possible, d'avoir deux revenus indépendants et donc une sûreté économique face aux diverses perturbations. Les agricultures familiales permettent d'alimenter les espaces ruraux et urbains avec une alimentation diversifiée et de qualité. Étant à petite échelle et sans hiérarchies, ces systèmes peuvent s'adapter à de nouvelles demandes de consommation (Bélières et al., 2013). L'autoconsommation des produits et les sources d'approvisionnement proches permettent une relative stabilité face à l'instabilité des marchés internationaux (Altieri and Miguel A, 2008; HLPE, 2013).

C'est principalement l'augmentation de la productivité de l'agriculture familiale qui permet une réduction de la pauvreté (Lipton, 2005). Cette augmentation de la productivité nécessite un investissement dans l'agriculture et les premiers investisseurs sont les exploitations familiales (FAO, 2015). Le changement de système de production vers des systèmes plus intensifs et à valeur ajouté (irrigation) ou de production animale est moins coûteux en investissement pour les petites exploitations que pour de grandes. L'investissement peut prendre diverses formes, que ce soit de manière financière (achat de terres), de main d'œuvre (terrassement) ou d'amélioration de la fertilité (réutilisation de nutriments) (HLPE, 2013).

La structure sociale de l'agriculture familiale permet une diversité d'investissements et de soutiens en période difficile (Reardon and Vosti, 1995). La famille peut intervenir de manière humaine avec le prêt de capital humain, d'un soutien financier ou pour permettre l'accès à certaines ressources naturelles. Les liens de réciprocité qui s'appuient sur la parenté et la proximité sociale permettent un partage de connaissance et d'expérience accumulées, des actions collectives (déci-

sions et règles, main d'œuvre) et une diversité de variétés culturales et de races animales (HLPE, 2013).

Les exploitations sont relativement indépendantes entre elles, en dehors de la proximité sociale, ce qui limite l'effet d'une mauvaise gestion de ressources sur de larges zones. Les chefs d'exploitation font leurs choix agronomiques et techniques en lien avec les stratégies de la sécurité alimentaire et disposent de ressources sociales, éco-géographiques ou économiques (Gafsi et al., 2007; Janicot et al., 2011).

1.3 Deux modes de représentation

En raison des risques de plus en plus importants qui pèsent sur l'agriculture familiale tropicale, et des enjeux sociaux qui y sont associés, il est nécessaire de pouvoir caractériser cette agriculture dans toute sa complexité (Alvarez et al., 2018), par la cartographie (approche « territoriale ») ou des typologies d'exploitation (approche « structurale »).

1.3.1 La cartographie

La cartographie des systèmes de culture permet de suivre les cultures dans le temps, c'est-à-dire la production potentielle et les pratiques culturales. La caractérisation spatiale de l'agriculture est utilisée, par exemple, pour estimer l'utilisation des ressources en eau, la présence de pollution, ou la déforestation au profit des terres agricoles (Grau et al., 2005; Leenhardt et al., 2010; Temme and Verburg, 2011). Des indicateurs peuvent être dérivés des cartes, tels que l'intensification agricole ou la diversification des cultures. Ces données sont essentielles pour l'évaluation et le renforcement de la sécurité alimentaire (HLPE, 2017). Les cartes produites sont aussi utilisées pour la prospective territoriale, l'aménagement du territoire, les évaluations d'impacts ou la prévention des risques, particulièrement bénéfique pour l'agriculture familiale qui souffre d'importants manques de données (HLPE, 2013). Pour comprendre les enjeux de la cartographie de l'agriculture familiale, il est important de définir certaines notions : la différence entre l'occupation et l'usage des sols, ainsi que l'échelle spatio-temporelle.

1.3.1.1 Occupation des sols et usage des sols

La caractérisation de la surface se fait à plusieurs échelles, principalement à des échelles administratives telles que l'échelle nationale ou internationale, avec des résolutions spatiales différentes (ex. U.S. National Land Cover Dataset, Countryside Survey au Royaume-Uni, Carte d'occupation des sols en France, CORINE Land Cover pour l'Europe). Ces cartographies confondent souvent les concepts d'**occupation des sols** et d'**usage des sols**. Ces deux définitions sont distinctes, mais leur différenciation peut parfois être difficile. Fisher and Comber (2005) les définissent ainsi :

- l'occupation des sols est ce qui est observé directement par imagerie aérienne ou satellite (herbe, béton, eau, arbre),

- l'usage des sols représente la manière dont sont utilisées les surfaces ; sa définition nécessite donc une analyse socio-économique de l'activité exercée (culture irriguée, prairie, bâtiments industriels, etc.).

À travers ces définitions, on peut souligner que la majorité des cartes produites sont à la fois des cartes d'occupation des sols et d'usage des sols. L'usage des sols demande un vocabulaire spécifique qui peut être connoté politiquement, culturellement ou écologiquement, là où l'occupation des sols tend vers une objectivation de l'information. Une surface en herbe (concept d'occupation des sols) peut ainsi être affectée à différentes classes d'usage des sols telles qu'une prairie pâturée, un terrain de football ou une zone résidentielle. La majorité des classes d'occupation des sols regroupent donc plusieurs éléments d'usage des sols (Comber et al., 2005). Le mélange des concepts d'occupation et d'usage des sols peut créer des problèmes de cartographie. Par exemple pour l'agriculture familiale la forêt peut être utilisée pour le pâturage, alors que pour les exploitations industrielles ces forêts sont exploitées pour le bois (Hoeschele, 2000).

Les concepts d'occupation des sols et d'usage des sols sont donc profondément liés, mais dépendent de l'objectif de la cartographie. La cartographie de l'occupation des sols est souvent utilisée dans les modèles biophysiques, alors que celle de l'usage des sols soutient par exemple l'application de politiques publiques ou d'aménagement du territoire. La définition de la nomenclature est donc primordiale pour caractériser spatialement l'agriculture familiale.

1.3.1.2 L'échelle d'analyse spatio-temporelle

Gibson et al. (2000) définissent l'échelle comme « les dimensions spatiales, temporelles, quantitatives ou analytiques utilisées pour mesurer et étudier des phénomènes »[6]. Lévy and Lussault (2013) précisent que le concept d'**échelle géographique** est important lorsqu'il est mis en relation avec d'autres espaces : l'analyse à une échelle régionale n'a pas le même objectif qu'à une échelle nationale et ne donnera pas les mêmes résultats. L'échelle est donc un concept multidimensionnel qui permet différents niveaux d'analyse selon l'échelle choisie. Ce choix d'échelle dépend de la **résolution spatiale** qui détermine la finesse de l'observation d'un processus ou d'un individu, soit « la précision utilisée pour une mesure »[7] (Gibson et al., 2000). Pour étudier spatialement l'agriculture familiale, il faut travailler à grande échelle pour comprendre l'organisation de ces petits exploitants dans des systèmes complexes et avec une résolution fine (les composantes de ce système, notamment les champs, étant souvent de petite taille). Alors que l'analyse d'un bassin de production dans les grandes plaines d'Amérique du Nord pourra être faite à petite échelle, avec une résolution plus grossière (grandes parcelles).

L'**échelle temporelle** détermine la période étudiée. Une carte est une représentation d'un système ou d'un ensemble d'éléments à un instant donné (ex. carte des limites administratives, des infrastructures), ou une représentation synthétique d'une période définie (ex. carte annuelle d'occupation des sols, carte synthétique de l'évolution d'une inondation).

[6] « The spatial, temporal, quantitative, or analytical dimensions used to measure and study any phenomenon »
[7] « The precision used in measurement »

Selon le choix de la période étudiée, la caractérisation d'un système agricole peut se faire de plusieurs manières. Bégué et al. (2018) proposent trois principales temporalités pour la caractérisation des pratiques culturales : 1) la saison culturale principale, 2) l'année, et 3) l'échelle pluriannuelle (figure 3).

1. La saison culturale principale permet de distinguer les cultures uniques (« One crop, one piece of land ») et les associations de culture (« More than one crop, one piece of land ») telle que l'association riz pluvial - maïs par exemple.
2. Au sein d'une année, plusieurs cultures peuvent se succéder sur une même parcelle, on parle alors de culture séquentielle. Cette succession peut être une alternance de culture d'été et de culture d'hiver ou saison des pluies / saison sèche.
3. L'analyse pluriannuelle met en valeur les successions de cultures. Les schémas de rotations culturales peuvent être identifiés (par exemple, alternance entre maïs une année et soja l'année suivante), tout comme l'assolement (rotation spatiale et temporelle entre des cultures) ou encore la monoculture (arboriculture ou succession dans le temps d'une même culture). Cette démarche permet de comprendre les tendances et les schémas d'organisation spatiale et temporelle d'un système.

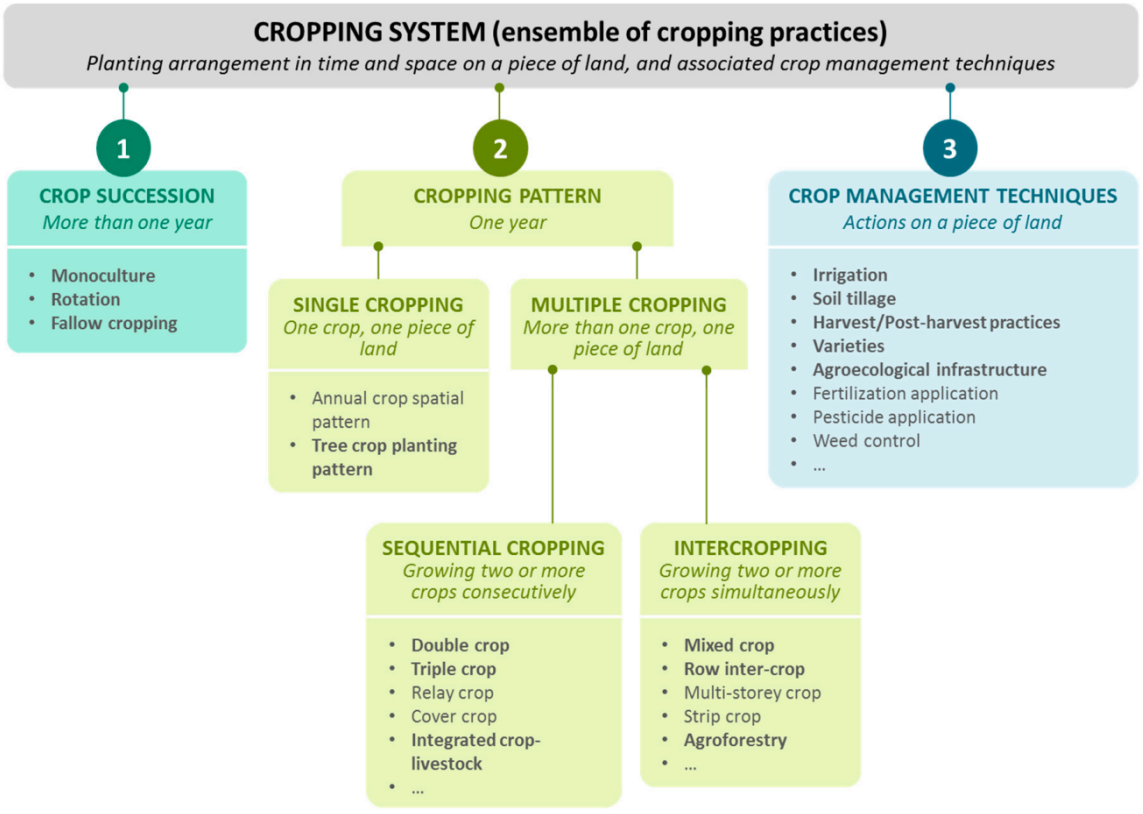


Figure 3 Typologie et définition des systèmes de culture selon Bégué et al. (2018)

1.3.1.3 L'état actuel de la cartographie de l'agriculture familiale

La cartographie des systèmes agricoles à une échelle continentale ou semi-continentale permet d'identifier les systèmes utilisés par l'agriculture familiale, et donc les zones dans lesquelles celle-ci est majoritaire. La FAO est à l'origine de plusieurs cartographies de ces systèmes. Par exemple, Dixon et al. (2001) ont cartographié les grands systèmes agricoles pour six régions du monde où l'agriculture familiale est la principale utilisatrice de surface agricole, totalisant 72 systèmes agricoles différents adaptés à chaque région à l'aide de données biophysiques et socio-économiques. Dans l'ensemble, la résolution spatiale de ces cartes est assez large afin de généraliser les systèmes agricoles à des régions.

À l'échelle nationale, l'USAID et le FEWS NET produisent des zonages des moyens de subsistance de 35 pays souffrant de l'insécurité alimentaire. Ces cartes sont produites pour identifier les zones qui ont des accès aux marchés et des moyens de production de nourriture similaires. L'élaboration des zonages et des typologies a été effectuée selon « l'approche de l'économie des ménages » afin de comprendre le fonctionnement de ceux-ci. Cette approche se base sur une compréhension et une classification des ménages selon des critères définis, sur des expertises locales pour la compréhension et la spatialisation de l'information (telles que des cartographies participatives basées sur l'avis des experts locaux), ainsi que la possibilité d'intégration des données agroclimatiques, d'occupation des sols, biophysiques et socio-économiques selon leur disponibilité (Holzmann et al., 2008; Grillo and Holt, 2009). Ces cartes sont accompagnées d'une explication complète du fonctionnement de chaque système cartographié. Par exemple pour Madagascar (Figure 4), 27 moyens de subsistance sont cartographiés et analysés (USAID and NET, 2013)

Ces cartes à petite échelle permettent d'avoir une idée générale des moyens de production d'une zone géographique, mais ne reflètent pas la réalité locale (Nachtergaele and Biancalani, 2013). Les cartes produites à l'échelle nationale ne permettent pas l'identification précise et le suivi des systèmes de culture pour une utilisation opérationnelle tels que l'aménagement du territoire ou l'identification de zones d'intervention prioritaire par exemple pour la sécurité alimentaire (Fisher and Comber, 2005; Leenhardt et al., 2010).

Pour des besoins spécifiques, des cartes de système de culture peuvent être produites à l'échelle locale. Cette opération nécessite des données adaptées dont la plus utilisée est l'imagerie satellitaire, qui permet de couvrir de larges zones dans une temporalité définie (voir partie 2.1). La cartographie permet une représentation fiable d'un territoire, de ses systèmes de cultures ou de ses systèmes agricoles, mais ne permet pas l'identification des exploitations (appartenance d'un champ à un propriétaire défini) ou de leur typologie.

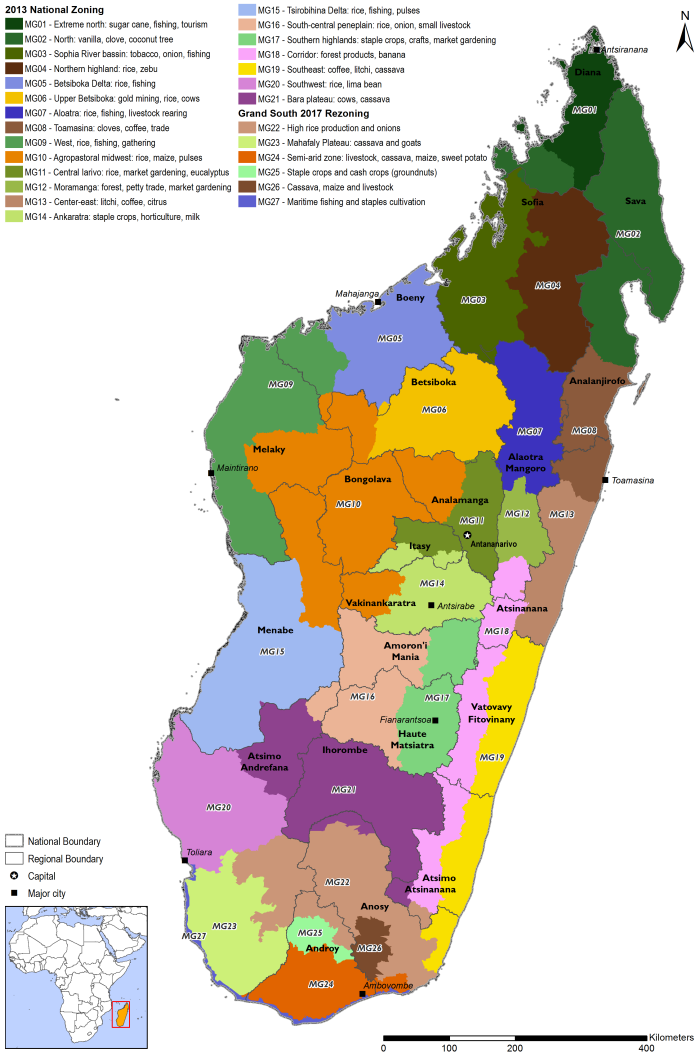


Figure 4 Carte des moyens de subsistance à Madagascar selon l'USAID en 2017, consolidation du travail effectué en 2013 (USAID and NET, 2013).

1.3.2 Typologie des exploitations

La diversité de l'agriculture familiale a poussé de nombreux auteurs à créer des typologies afin de mieux les analyser et comprendre leur complexité. La compréhension de ces systèmes agricoles permet de mieux évaluer les moyens de subsistance et de production mis en place (Tittonell, 2014) et les risques associés. Cette évaluation permet de cibler des interventions sur des exploitations spécifiques (Kuivanen et al., 2016; Hauswirth et al., 2015), d'identifier des politiques publiques adaptées, d'évaluer leurs impacts ou encore de modéliser ces systèmes (Roy Chowdhury, 2010; Jahel et al., 2018).

Deux principales méthodes de création de typologies existent (Matus et al., 2013). La méthode déductive ou qualitative se base sur les connaissances expertes qui permettent de définir une typologie a priori. Cette typologie peut être validée a posteriori sur le terrain, à l'aide d'enquêtes ou d'entretiens. La seconde méthode est basée sur des données statistiques (enquêtes, données administratives) et des groupements selon certains critères afin de créer des groupes homogènes

ayant une faible intra-variabilité (par exemple une analyse multivariée suivie d'une classification).

À partir d'une typologie basée sur des enquêtes et sur la taille des exploitations, donc suffisamment ouverte pour intégrer différents systèmes agricoles dans cette région du monde, Jayne et al. (2016) montrent l'évolution temporelle de la taille des fermes en Afrique sub-saharienne tend vers une majorité d'exploitations entre 5 et 100 ha. Alvarez et al. (2018) présentent une typologie de fermes en Zambie à partir d'hypothèses basées sur le profil de la ferme (taille de l'exploitation, du troupeau, type de culture), les ressources (espèces cultivées, travail de la terre) et sur des caractéristiques biophysiques, associées à une analyse statistique. Cette approche permet de lier les aspects socioculturels, économiques et biophysiques à une approche statistique et reproductible. Ainsi, des changements d'échelle sont possibles, tout comme des changements de contexte. Ces typologies sont souvent à l'échelle de l'exploitation et sont peu spatialisées.

La construction d'une typologie permet de définir une base pour analyser et modéliser les stratégies et pratiques à la fois spatiales et temporelles d'un type d'exploitation.

Dans cette section nous avons présenté la diversité de l'agriculture familiale dont nous retenons la définition de Bélières et al. (2013). À travers le prisme de l'agriculture familiale, nous avons abordé l'insécurité alimentaire et les forces et faiblesses de cette forme de production ainsi que les deux principaux modes de représentation : la cartographie et la typologie d'exploitation.

2 Les outils d'analyse spatio-temporelle

Comme nous l'avons vu précédemment, la caractérisation des systèmes d'agriculture familiale passe par la cartographie (compréhension du territoire) et par la création de typologie d'exploitation (compréhension de la structure des exploitations). Pour caractériser spatialement l'agriculture familiale, la télédétection et la modélisation des dynamiques spatiales sont deux outils majeurs. La télédétection permet de produire des cartes d'occupation et d'usage des sols, tandis que la typologie des exploitations permet de nourrir la modélisation des dynamiques spatiales afin de comprendre le fonctionnement et la répartition des systèmes agricoles. Ces deux outils sont utilisés pour répondre à des objectifs différents sans être associés, mais sont complémentaires dans leur vocation de caractériser spatialement et temporellement l'agriculture familiale.

2.1 Télédétection et agriculture familiale

Les données de télédétection sont étudiées depuis une quarantaine d'années pour produire de l'information géospatiale sur les territoires. Aujourd'hui, une centaine de satellites d'observation de la Terre sont en orbite et fournissent des données à différentes résolutions spatiales, de 30 centimètres à plusieurs kilomètres, à différentes fréquences de revisite (horaire, journalière, bimensuelle ...) et dans des bandes spectrales variées (optiques, micro-ondes) et complémentaires. L'observation de la Terre reste un outil privilégié pour caractériser les surfaces agricoles aux échelles locale et globale. Le traitement des données de télédétection peut être décliné en méthodes

centrées sur les pixels ou sur des groupements de pixels (objets) de manière supervisée ou non. Cet outil et les méthodes qui lui sont associées font toujours face à difficultés pour caractériser l'agriculture familiale.

2.1.1 Une offre variée d'images

Pour caractériser l'occupation et l'usage des sols agricoles, la télédétection optique (mesure passive de rayonnements visible et infrarouge) est principalement utilisée, mais elle est de plus en plus complétée par de l'imagerie radar (mesure active de rayonnement hyperfréquence) (El Hajj et al., 2018; Sivasankar et al., 2018; Sharma et al., 2019). Progressivement, ces deux familles de satellites ont évolué vers une résolution spatiale de plus en plus fine, une meilleure répétitivité temporelle et une fauchée permettant l'étude de grandes surfaces (Panda et al., 2015). Les principales caractéristiques des satellites d'observation de la terre sont : 1) le nombre de bandes spectrales ainsi que leur résolution (largeur du spectre couvert par une bande spectrale), 2) la résolution spatiale (taille des pixels), 3) la répétitivité temporelle (fréquence d'acquisition des images) et 4) la fauchée (taille de l'emprise au sol de l'image acquise). Ces caractéristiques sont liées entre elles : une résolution spatiale très élevée est associée à une fauchée réduite ; si la répétitivité est importante avec une large fauchée, alors la résolution spatiale sera grossière (figure 5). La diversité actuelle des satellites permet de travailler sur des thématiques variées et à différentes échelles (Oppelt et al., 2015).

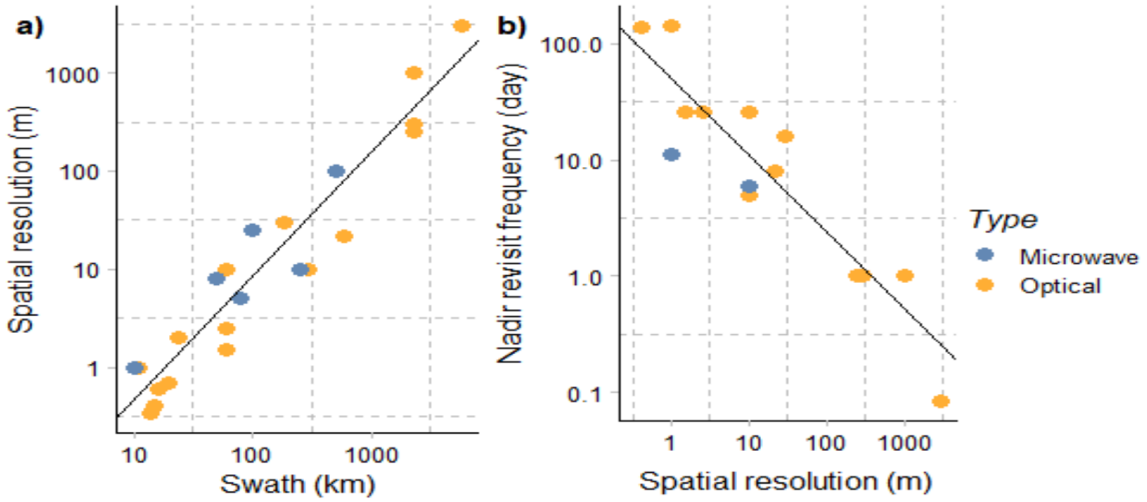


Figure 5 Relations selon les satellites optiques (symbole orange) ou radar (symbole bleu), entre la résolution spatiale et la fauchée (a) et entre la fréquence de revisite et la résolution spatiale (b), figure de Tonneau et al. (2020)

En 2008, l'ouverture des archives Landsat par l'United States Geological Survey (USGS) et la mise à disposition gratuitement des images a changé le paradigme de la distribution des images satellitaires. Les agences spatiales telles que l'European Space Agency (ESA) ont fait le choix de suivre cette dynamique et de rendre disponibles gratuitement les acquisitions des missions « Sentinel » via le programme Copernicus. La disponibilité de séries temporelles sur le temps long et à haute (10 m) et moyenne (100 m) résolution spatiale a des impacts bénéfiques sur la so-

ciété et pour la science (Zhu et al., 2019). Elle permet aussi la production plus rapide de cartes et d'indicateurs spatiaux, particulièrement importants dans le contexte de la lutte contre l'insécurité alimentaire. Les images à très haute résolution spatiale (THRS) (métrique à infra-métrique), suivent malheureusement une politique commerciale, limitant leur utilisation.

Gratuité des données données satellitaires

L'accès aux données satellitaires à haute résolution spatiale (jusqu'à 10 mètres) est aujourd'hui facilité par des initiatives nationales ou internationales, qui proposent des données gratuites (free access) ou en libre accès (open access), par exemple : le Pôle français de données et services pour les surfaces continentales (Theia), le Dispositif Institutionnel National d'Approvisionnement Mutualisé en Imagerie Satellitaire (Dinamis), la Commission européenne (Copernicus, le programme d'observation de la Terre de l'Union européenne), le CNES (Centre national d'études spatiales, France, via sa Plateforme d'exploitation des données Sentinel - PEPS), la National Aeronautics and Space Administration (NASA) aux États-Unis (Tonneau et al., 2019).

La possibilité d'utiliser des séries temporelles et des images à différentes résolutions spatiales et temporelles permet de caractériser des systèmes de la manière la plus adaptée. Pour caractériser un système agricole, Duveiller and Defourny (2010) proposent un cadre conceptuel afin de définir quelle est la meilleure résolution spatiale à choisir selon le système agricole. Le choix des données d'observation de la Terre peut aussi être guidé selon plusieurs critères dépendant des données. Ceux-ci sont 1) la zone à caractériser (Fritz et al., 2015), 2) la période étudiée (Whitcraft et al., 2015a), 3) de la fréquence des observations (Whitcraft et al., 2015c) et 4) de la résolution spatiale et spectrale des données (Whitcraft et al., 2015b). Par exemple Bellón et al. (2017) cherchent à caractériser les systèmes agricoles sur l'ensemble d'un état brésilien. Ils utilisent pour cela des séries temporelles d'images à basse et moyenne résolution MODIS (250 m) sur plus d'une année avec de nombreuses bandes spectrales. À l'opposé, Lelong et al. (2020) caractérisent les systèmes agro-forestiers à une échelle locale à l'aide d'une image monodate THRS Worldview3 ayant 8 bandes spectrales.

2.1.2 Des méthodes de classification

La caractérisation de systèmes de cultures s'appuie principalement sur des techniques de classification qui ont progressivement évolué de façon à prendre en compte la disponibilité croissante des données satellitaires et en s'appuyant sur une plus grande puissance de calcul des ordinateurs. Les classifications sont passées de méthodes statistiques à des méthodes moins dépendantes de la distribution des données, telle que les arbres de décision (Friedl and Brodley, 1997). On distingue deux grands types de classifications, non supervisées et supervisées, et deux familles d'approches, l'approche pixel et l'approche objet (plus connue sous le nom d'object-based image analysis (OBIA)).

2.1.2.1 Classification non supervisée et supervisée

Entre les méthodes de classification supervisées et non supervisées, l'objectif est le même : regrouper de manière cohérente des éléments dans des classes. Cependant, la manière d'y parvenir diffère par un point important : le fait de disposer ou non d'une base de données d'apprentissage, tel que des relevés d'occupation des sols issus d'une campagne de terrain. Les forces et faiblesses des algorithmes de classifications supervisées pour la caractérisation de l'occupation des sols sont décrites par Gómez et al. (2016).

La base de données utilisée pour une classification supervisée est généralement partitionnée en deux parties : 1) un ensemble destiné à l'apprentissage de l'algorithme, et 2) un second destiné à l'évaluation de ce même algorithme. Le but est d'évaluer automatiquement si l'algorithme a bien su apprendre à différencier les classes selon leurs caractéristiques (spectrales par exemple) et à classer correctement de nouveaux éléments. Les principaux indicateurs permettant d'évaluer la précision d'une classification sont le **rappel** soit la part d'éléments pertinents retrouvés par le classifieur pour une classe, celui-ci sera élevé si tous les éléments concernés sont bien classés ; la **précision**, soit la part d'éléments classés sélectionnés qui correspondent bien à cette classe et non à une autre (pureté de la classe) ; et la **F-mesure** qui est la moyenne harmonique de la précision et du rappel.
Parmi les méthodes supervisées utilisées en télédétection pour cartographier les systèmes d'agriculture familiale, nous pouvons citer les **machines à vecteurs de support** (suport vector machine (SVM) en anglais) (Boser et al., 1992) utilisés par Löw et al. (2012) ou **Random Forest** (Breiman, 2001), par exemple utilisé dans l'étude de Lebourgeois et al. (2017). Pour les classifications complexes, il est possible d'assembler plusieurs classifieurs puis d'agréger les résultats pour obtenir un score de classification plus fiable (Kozak et al., 2017; Aguilar et al., 2018). L'apprentissage profond ou *deep learning* est aussi une méthode de classification supervisée utilisée en télédétection, que ce soit pour la fusion de données (Benedetti et al., 2018; Ienco et al., 2019) ou pour la cartographie de l'occupation ou de l'usage des sols (Kussul et al., 2017; Bailly et al., 2018; Interdonato et al., 2019). Du et al. (2012) présentent une revue de l'ensemble des méthodes possibles.

Les méthodes de classification non supervisées ne demandent aucune base de données d'apprentissage (vérité terrain) pour être utilisées. Leur fonctionnement vise à construire des classes cohérentes à partir d'éléments non labellisés, et donc à trouver la meilleure méthode pour partitionner ces données Bellón et al. (2018). Il y a deux principales méthodes pour évaluer cette classification : 1) si une « vérité terrain » est connue, on peut calculer les différences entre ces classes et la partition proposée par le classifieur de la même manière que pour une classification supervisée ; et 2) vérifier et calculer les distances entre les partitions et si celle-ci sont cohérentes et distinctes les unes des autres, la labellisation se faisant à partir de l'interprétation des partitions.

Les méthodes supervisées ont pour avantage de définir exactement les classes que l'on souhaite obtenir, mais demandent des campagnes de terrain coûteuses en temps et en argent, particulièrement pour la caractérisation des espaces agricoles ou les cultures changent à chaque saison culturale Atzberger (2013). Les méthodes non supervisées s'affranchissent des données pour l'apprentis-

sage, mais des données (ex. vérité terrain ou photo-interprétation) sont souvent nécessaires pour la labellisation des classes et une validation complète d'un produit opérationnel.

2.1.2.2 Classification par pixel et approche objet

L'approche basée sur la classification des pixels est la méthode la plus conventionnelle, mais présente des limites pour la cartographie des systèmes complexes. Lorsque différents systèmes de cultures sont présents sur de petites parcelles et que les mesures (satellites, terrain) ne sont pas réparties de manière homogène ou de qualité suffisante, ces contraintes sont importantes (Vieira et al., 2012). Les espaces complexes agricoles ne sont donc pas correctement caractérisés par ce type d'approche (Inglada et al., 2015). Pour pallier cette limite, l'approche object-based image analysis (OBIA) a été développée. Cette approche est basée sur des techniques de segmentation d'une image (Pal and Pal, 1993) qui permettent de regrouper les pixels spatialement et radiométriquement proches en **objets** cohérents (Haralick and Shapiro, 1985; Blaschke et al., 2014). Afin de délimiter des objets complexes, la méthode OBIA se base généralement sur une image ayant une résolution adaptée au système étudié, permettant ainsi de définir le contour des objets de manière fine (Whiteside and Ahmad, 2005). Par exemple, dans un système agricole familial, les pixels des parcelles (souvent sinueuses et de petite taille) peuvent être regroupés sur des images dont la résolution spatiale est assez élevée (THRS) par rapport à la taille de la parcelle (voir encadré La télédétection et l'agriculture familiale).
De plus, avec cette approche, l'ensemble des variables contextuelles, texturales, spectrales et morphologiques peuvent être attribuées à l'objet pour le caractériser (Teluguntla et al., 2015). L'utilisation de la méthode OBIA peut s'effectuer avec une seule résolution spatiale (segmentation et caractéristiques radiométriques), ou bien avec en combinant une image à haute ou très haute résolution spatiale (afin d'effectuer la segmentation) avec des séries temporelles d'images à résolution plus grossière (haute résolution spatiale, moyenne résolution spatiale). Les séries temporelles permettent d'observer le fonctionnement et la dynamique des objets à travers l'évolution de leur radiométrie. Cette méthode a été appliquée avec succès pour différents types d'agricultures dans les études de Castillejo-González et al. (2009), Peña-Barragán et al. (2011), Vieira et al. (2012), Matton et al. (2015), Novelli et al. (2016), et Lebourgeois et al. (2017).

2.1.3 La télédétection et l'agriculture familiale

Malgré un accès accru à l'imagerie satellitaire et aux outils de traitement, l'utilisation des images de télédétection pour créer des produits et des services est encore limitée dans la plupart des pays du continent africain (Tonneau et al., 2019).

GEOGLAM et Sen2Agri

Le projet international Group on Earth Observations Global Agricultural Monitoring (GEOGLAM)[a] lancé en 2011 (Parihar et al., 2012) a pour objectif d'améliorer la sécurité alimentaire des pays à risques via le suivi de l'agriculture à l'échelle nationale ou régionale, le renfoncement des capacités nationales et le suivi des dyamiques agricoles à une résolution spatiale de 10 m à 100 m (Whitcraft et al., 2015a).
En appui à GEOGLAM, le projet Sen2Agri a pour objectif d'utiliser des données Sentinel-2 (avec une résolution spatiale de 10 x 10 m) afin de pouvoir caractériser les cultures pour une grande diversité de systèmes agricoles dont les systèmes agricoles disposant de champs de petite taille.

[a] https://www.earthobservations.org/geoglam.php

Comme nous l'avons vu précédemment, le choix des données d'observation de la Terre doit être guidé selon le type de système agricole à caractériser (Duveiller and Defourny, 2010). En zone intertropicale, où l'agriculture familiale est majoritairement composée de petits champs, la forte nébulosité, la grande diversité des pratiques agricoles et le petit parcellaire sont autant de contraintes pour l'observation de la Terre.

Dans le cas de l'agriculture familiale, système situé majoritairement en zone sub-tropicale (section 1.2), et où la saison culturale principale se déroule durant la **saison des pluies**, donc avec une forte nébulosité et une saturation de l'humidité du sol, les deux techniques de télédétection (optique et radar) présentent des désavantages. Pour les images optiques, la solution est de privilégier des séries temporelles denses afin de limiter l'impact de la nébulosité (Whitcraft et al., 2015c), alors que les images radars la résolution spatiale et le relief restent des obstacles. Une possibilité pour contrer ces difficultés est la combinaison des images optiques et radars (Skakun et al., 2016; Gaetano et al., 2017).

Malgré l'amélioration de l'offre globale en termes d'observation de la Terre et de techniques et méthodes de classification, l'utilisation de la télédétection se heurte toujours à des difficultés pour caractériser les systèmes complexes. Compte tenu de la petite taille des champs en zone inter-tropicale (Fritz et al., 2015), la caractérisation de l'agriculture familiale nécessite l'utilisation d'images acquises à une résolution spatiale fine afin de limiter les pixels dits « mixtes » qui contiennent différents types d'occupation ou d'usage des sols. D'après Duveiller and Defourny (2010), le choix de la résolution spatiale est un compromis entre la taille du pixel et sa pureté selon le type de culture à caractériser.
Si on laisse de côté les capteurs à résolution hectométrique dont la résolution est inadaptée pour ce

type de système, les capteurs à haute et très haute résolution spatiale apportent des informations complémentaires sur ces systèmes :

- les champs individuels de grande taille sont caractérisés à l'échelle régionale à l'aide de capteurs à moyenne (hectométriques) et haute résolution (décamétrique) (Novelli et al., 2016; Griffiths et al., 2019), mais jusqu'à récemment, la fréquence de revisite des satellites à haute résolution spatiale était faible et donc pénalisante pour obtenir des images sans nuages, surtout dans les régions avec une forte nébulosité (Inglada et al., 2015). Depuis 2015, avec une fréquence de 10 jours, puis 5 jours (2017), la mission Sentinel-2 combinée à d'autres missions (Landsat, Venμs) d'augmenter la probabilité d'obtenir des images sans nuages (Gómez et al., 2016).
- à l'échelle locale, la cartographie des surfaces agricoles met en œuvre majoritairement des images à haute résolution spatiale, parfois couplées avec des images THRS (métrique). Même si ce type d'approche est adapté à l'agriculture familiale, il reste tout de même limité, car coûteux, nécessitant de traiter un important volume de données (Lebourgeois et al., 2017; Dupuy et al., 2020a), et non opérationnel sur de grandes surfaces à cause du peu d'acquisitions.

Bontemps et al. (2015) ont montré que la qualité des acquisitions était très variable selon les espaces géographiques à caractériser. En croisant différents systèmes agricoles à travers le monde allant des systèmes agricoles intensifs et extensifs aux systèmes de petites agricultures tropicales, Inglada et al. (2015) montrent la difficulté à correctement caractériser l'agriculture composée de champs de petite taille par rapport aux autres systèmes.

Résolution spatiale et agriculture familiale

D'après Woodcock and Strahler (1987), pour discriminer un objet, un pixel doit avoir une taille au minimum entre 1/2 à 3/4 de la taille de l'objet à caractériser. Ce ratio limite les pixels mixtes et ceux-ci seront situés en bordure des objets.
Exemple théorique pour la caractérisation de l'agriculture familiale :
Le nombre de champs d'une exploitation familiale est variable entre les pays, tout comme l'organisation spatiale d'un champ (carré ou sinueux), mais les données pour prendre en compte ces critères sont rarement disponibles. Pour simplifier, nous postulons que les champs sont carrés et qu'il y en a 7 par exploitations. Au Brésil, pour une exploitation de 50 ha (champ d'environ 7 ha), la résolution spatiale maximale pour éviter les pixels mixtes serait de 229 m. Une exploitation en Asie du Sud-Est avec une surface de 1 ha (champ d'environ 14 ares) la résolution maximale serait de 32,4 m. Pour une exploitation dans les hautes terres de Madagascar où la taille moyenne d'une exploitation est de 0,5 ha (champ d'environ 7 ares), la résolution maximale serait de 22,9 m.

L'utilisation de la méthode OBIA a aussi d'importantes limites. Lorsqu'elle nécessite des images à deux résolutions différentes, son application à de larges zones peut être restreinte (du fait du coût généralement élevé, et de la petite emprise des images THRS), et cette méthode ne résout pas les problèmes d'acquisition dans les zones tropicales. Elle reste tout de même performante pour la caractérisation de l'agriculture familiale (Valero et al., 2016; Inglada et al., 2016).

La caractérisation de l'agriculture familiale se base sur le traitement de bandes spectrales et d'indices. Les principaux indices utilisés sont les indices de végétation tel que le Normalized Difference Vegetation Index (NDVI), le NDWI (Valero et al., 2016; Lebourgeois et al., 2017), ou l'Enhanced Vegetation Index (EVI) (Bellón de la Cruz, 2018). Les indices de texture d'Haralick sont aussi mobilisés lorsque qu'une image THRS est utilisée (Lebourgeois et al., 2017; Dupuy et al., 2020a).[8].

L'agriculture familiale reste donc difficile à caractériser par télédétection. Les principales raisons sont : 1) la petite taille des champs (Fritz et al., 2015), 2) la forte hétérogénéité de l'environnement qui rend difficile l'obtention d'une segmentation avec des objets homogènes (Li et al., 2016; Persello et al., 2019), 3) la forte nébulosité des zones tropicales (Whitcraft et al., 2015c), 4) la variabilité des dynamiques spatio-temporelles pour une même culture, 5) la synchronisation des phénologies des cultures et de la végétation naturelle alentour sur le régime des pluies (Xiong et al., 2017) et 6) les pratiques culturales qui peuvent être diversifiées pour une même culture (Siebert et al., 2010).

2.2 Modélisation spatio-temporelle de l'agriculture familiale

La modélisation spatio-temporelle permet de simuler un environnement dans le temps selon des contraintes, ou règles, utilisées pour recréer les dynamiques, synergies et liens du système analysé (*i.e.* soit le fonctionnement du système). Les principales contraintes de la modélisation sont la compréhension du système étudié, sa retranscription dans un **modèle conceptuel** et enfin son implémentation informatique. À travers cette sous-section, la modélisation spatio-temporelle des systèmes sera abordée à travers les différents paradigmes de la modélisation de l'occupation et de l'usage des sols de manière générale, puis pour l'agriculture familiale et enfin spécifiquement aux espaces sub-tropicaux.

2.2.1 La modélisation spatio-temporelle des systèmes

C'est avec l'avènement de l'informatique dans les années 1960 que la modélisation se formalise et peut être effective. von Bertalanffy (1968) pose des bases de l'**approche systémique** et de la modélisation des systèmes. Il propose de modéliser un **système** en le décrivant **dans son intégralité**, avec ses composants, leurs organisations, et leurs interactions mutuelles. Le besoin de description et de compréhension d'un système dans son ensemble fait de la modélisation une science interdisciplinaire, alliant souvent informaticiens et thématiciens (Shubik, 1960; Guetzkow, 1962; Beshers, 1965; Dutton and Starbuck, 1971). Ce paradigme demande donc d'analyser un système de manière interdisciplinaire, puis de concevoir un modèle conceptuel, et finalement de l'implémenter de manière informatique (Chery and Smektala, 2004).

Les systèmes ne sont pas des structures isolées, mais intègrent souvent des sous-systèmes et peuvent être compris dans des systèmes plus généraux. On peut hiérarchiser les systèmes et structurer leurs interactions, qu'elles soient hiérarchiques et/ou d'adjacence (von Bertalanffy, 1968). Un

[8]Les indices de végétation fréquemment mobilisés sont détaillés dans le Chapitre 4, section 1

système n'est donc pas un ensemble, mais un système en soi, avec sa propre dynamique et identité qui font qu'un système doit être vu dans sa globalité et non uniquement comme la somme des éléments qui le composent. Souvent les systèmes modélisés sont complexes, car composés de processus, pas nécessairement nombreux, mais qui sont connectés entre eux, sans être forcément redondants Le Moigne (1977).

Les modèles spatiaux sont construits pour l'analyse de l'organisation spatiale d'un territoire. Brunet (1980) définit le concept d'**organisation spatiale** comme « un ensemble plus ou moins cohérent de lieux mis en relation. Cet ensemble est à la fois différencié et intégré en aires et réseaux (lignes et points) dans lesquels circulent des flux ». Cette définition comprend les trois éléments principaux d'un modèle : **l'espace, l'interaction spatiale qui définit des états, et la temporalité par les interactions** (Ruas and Sanders, 2015). Ils sont utilisés pour étudier un système, que ce soit les dynamiques qui ont permis l'émergence de ce système, ou bien ses évolutions possibles.

2.2.2 La modélisation de la dynamique de l'occupation et de l'usage des sols

La modélisation spatiale a évolué au cours des changements de paradigmes, des objets d'étude et des disciplines, changeant la façon d'appréhender un système.

De nombreuses études ont entrepris de classer les méthodes de modélisation d'un espace (Ford and Ford, 1999; Agarwal et al., 2002; Degenne, 2012; Kelly , Letcher; Laniak et al., 2013; Elsawah et al., 2017; Müller-Hansen et al., 2017). De ces études on peut différencier 1) les approches descendantes, visant à étudier la « dynamique des systèmes », 2) les approches ascendantes qui modélisent le comportement d'individus dans un système, et 3) les approches qui diffèrent de ces courants principaux.

2.2.2.1 La dynamique des systèmes

La « dynamique des systèmes » (Forrester, 1969) se base sur des relations entre stocks, reliés par des réservoirs. Les stocks vont varier et l'étude de boucles de rétroactions est possible. La dynamique des systèmes permet de correctement modéliser des évolutions globales et les dynamiques de changements, mais la modélisation de ces éléments dans un espace géographique est difficile. L'espace géographique rajoute une autre dimension qui peut fragmenter les stocks, et donc augmenter fortement la complexité du modèle. Par exemple, l'une des limites de la spatialisation de ces modèles est la limitation du nombre de zones à modéliser (Voiron and Chéry, 2005). Cette approche est toujours utilisée aujourd'hui avec des modèles tels que Stella (Costanza, 1987; Costanza et al., 1997) pour la modélisation des dynamiques de changement d'usage des sols par les exploitations familiales (Garedew et al., 2012), ou Vensim (Eberlein and Peterson, 1992) pour l'étude de boucles de rétroaction de changements d'occupation et d'usage des sols (Verburg, 2006), ou pour la compréhension des facteurs de changements de l'usage des sols dans systèmes agro-pastoraux au Sahel (Rasmussen et al., 2012).

La modélisation d'un espace à l'aide de la dynamique des systèmes peut se faire de plusieurs manières. Muetzelfeldt and Massheder (2003) décomposent le système en plusieurs éléments disposant de modèles stock-flux et de fonctions de désagrégation. L'approche développée par Maxwell and Costanza (1997) et Costanza and Voinov (2004) consiste à décomposer l'espace en cellules et à inclure dans chacune d'elle un modèle stock-flux dont la dynamique peut varier dans l'espace. Les cellules voisines sont connectées entre elles permettant l'échange de flux (Sanders and Sanders, 2004). Elsawah et al. (2017) présentent les avantages et inconvénients d'un ensemble de modèles selon plusieurs cas d'études. Ces approches sont utilisées pour modéliser les changements d'occupation ou d'usage des sols (Sandker et al., 2010; Neuwirth et al., 2015, 2016).

2.2.2.2 Les approches individu-centrées

La modélisation « individu-centrée » est une approche explicitement spatiale qui s'intéresse à la modélisation de chaque individu dans un système, c'est-à-dire aux comportements des éléments d'un système (Degenne, 2012). Cette approche se décompose en deux principales formes de modélisation : 1) les automates cellulaires et 2) les système multi-agents (SMA).
En 1966, Von Neumann (1966) a développé un système auto-réplicatif (les automates cellulaires) qui est basé sur : 1) une grille, 2) un ensemble d'états possibles pour une cellule et 3) une ou plusieurs règles pour définir le changement d'état d'une cellule[9]. Ce type de modélisation est appliqué à différents domaines comme l'épidémiologie (White et al., 2007), la croissance des villes (Langlois and Phipps, 1997; Colonna et al., 1998), à celui des espaces protégés (Islam et al., 2018) ou explicitement aux changements d'usage ou d'occupation des sols (Houet et al., 2014). Ces modèles sont limités par la taille unique de leurs cellules, ce qui ne permet pas l'étude de phénomènes spatiaux de différentes échelles sur une même grille, ainsi que par l'uniformité des règles pour chaque cellule.

Les SMA sont basés sur le concept d'acteurs développé par Hewitt (1977). Selon Hewitt (1977) un acteur est une entité qui peut communiquer avec d'autres entités. Gasser et al. (1987) transforment les acteurs en agents et développent le concept de système multi-agents. À la différence des acteurs, les agents ne font pas que communiquer, mais ont des objectifs propres et donc une autonomie relative, leur permettant de se déplacer, d'interagir avec d'autres agents ou avec l'environnement par exemple. Les acteurs peuvent donc être des individus, mais aussi des entités d'un autre niveau (de groupe) (Bousquet and Le Page, 2004).

Le terme modélisation « ascendante » est utilisé, car tous les agents ou cellules emmagasinent des états, variables et autres données qui peuvent être agrégés pour créer des indicateurs spatiaux, temporels ou de comportement.

[9] Le *jeu de la vie* est l'un des exemples les plus connus d'automate cellulaire (Gardner, 1970) et de nombreuses implémentations sont disponibles (par exemple : https://www.dcode.fr/jeu-de-la-vie)

2.2.2.3 Les autres approches en modélisation spatio-temporelle

D'autres méthodes ont été développées afin de modéliser des systèmes d'une manière différente des approches descendantes et ascendantes. Ici sont présentés uniquement les concepts développés pour la modélisation de l'occupation des sols.

Les approches « descendantes » et « ascendantes » divergent par leur façon et de représenter un système, mais ne sont pas nécessairement opposées. La combinaison de ces approches est utilisée dans certaines études comme la prédiction de changement d'usage des sols via la combinaison de la dynamique des systèmes et d'automates cellulaires par Xu et al. (2016). Il existe un logiciel commercial, Anylogic (Borshchev and al, 2004) qui combine trois approches de modélisation : la dynamique des systèmes, la modélisation en SMA, et la modélisation d'événements discrets. Malheureusement ce programme est propriétaire et ne permet pas une visibilité complète sur les processus de modélisation.

Les outils CLUE (Veldkamp and Fresco, 1996), CLUE-S (Verburg et al., 2002), SELES (Fall and Fall, 2001) et LANDIS (Mladenoff, 2004) ont été conçus afin d'utiliser plusieurs grilles, ou vecteurs, représentant un espace géographique où chaque cellule représente une portion de surface caractérisée par une occupation du sol. Ces modèles combinent des approches ascendantes et descendantes, avec l'utilisation de processus au niveau de l'usage des sols, mais aussi la désagrégation de processus globaux à l'échelle des individus (Verburg and Overmars, 2009). L'objectif de ces méthodes est d'associer changement d'occupation des sols et facteurs de changements.

Le langage L1 et la plateforme DYPAL (Gaucherel et al., 2012), servent à modéliser la dynamique des paysages à l'aide de règles génériques qui prennent compte de l'organisation spatiale des éléments et qui peuvent modifier leur état géométrique. Le langage permet de décrire l'évolution du paysage de la manière souhaitée, par exemple via des automates cellulaires (Houet et al., 2014).

Ocelet est un langage métier dédié à la modélisation de dynamiques spatiales qui comprend une plateforme de modélisation appelée Ocelet Modelling Platform (OMP) (Degenne, 2012; Castets, 2015; Degenne and Lo Seen, 2016). Cette approche a été conçue pour modéliser un espace géographique à l'échelle souhaitée et sans contraintes spatiales, cela par son intégration directe des formats vectoriels ou de pavages réguliers (ex. raster). Les entités sont distribuées dans l'espace, ou non, et leurs interactions sont modélisées à l'aide de graphes.

Ces outils sont utilisés pour modéliser des dynamiques de l'occupation ou de l'usage des sols (Verburg and Veldkamp, 2005; Manson, 2005; Mertens and Lambin, 2000; Muller and Zeller, 2005; Schneider and Gil Pontius, 2001; Meiyappan et al., 2014; Houet et al., 2014), mais il existe peu de modélisation spatio-temporelle explicitement centrée sur l'agriculture familiale (Jahel, 2016).

2.2.3 La modélisation spatiale de l'agriculture familiale

Parmi les approches spatio-temporelles centrées sur l'agriculture familiale, deux types de modélisations peuvent être différenciés : 1) la modélisation de l'exploitation agricole, et 2) la **modélisation de la dynamique d'un système agricole**.

La modélisation de l'exploitation agricole (ex. organisation et production modélisées à l'aide de systèmes de culture) est généralement pensée uniquement à l'échelle de l'exploitation et est peu spatialisée (Belhouchette et al., 2011; van Ittersum et al., 2013; Leite et al., 2014).

La modélisation des systèmes agricoles modélise principalement les changements et facteurs de changements. Ces modèles visent surtout l'étude des dynamiques des systèmes agricoles. Dans cette partie, nous nous intéressons uniquement aux modèles spatiaux traitant de l'agriculture familiale. Contrairement à l'échelle de l'exploitation, la modélisation d'un système agricole permet d'analyser la répartition spatiale des cultures. Par exemple, elle permet de mettre en relation un système agricole avec des pratiques culturales ou avec la gestion des ressources naturelles (Jouve, 2006). La répartition des types de cultures est difficile à modéliser sur de grandes étendues, car l'échelle optimale est celle de la parcelle ou du groupe de parcelles. Étendre cette modélisation à un territoire complet est délicat, car cela demande un nombre important de données (Leenhardt et al., 2010). La caractérisation des systèmes de cultures dans un contexte d'agriculture familiale est donc soumise à la finesse de l'information disponible, pouvant permettre de modéliser certains types de cultures. Les autres pratiques culturales sont quant à elles plus difficilement mobilisables, principalement par manque de données spatialisées nécessaires à cette modélisation ou par manque de connaissances sur ces pratiques. Des typologies d'exploitations sont souvent utilisées pour faire des études rétrospectives ou d'évolution possible de ces exploitations ou de l'espace agraire (Jahel, 2016; Holland et al., 2017; van der Laan et al., 2018). Ces modèles ont donc une représentation spatiale théorique des exploitations et ne représentent pas la réalité du terrain, souvent par manque de données spatialisées.

La majorité des modèles spatiaux traitent de l'agriculture familiale sous l'aspect du changement d'occupation et d'usage des sols. Les principales applications sont par exemple les facteurs de déforestation liés à l'agriculture familiale (Roy Chowdhury, 2006; Wyman and Stein, 2010), ou pour la prospective territoriale (Evans et al., 2011; Xie et al., 2014; Chávez et al., 2014).

2.2.4 Des approches encore peu appliquées aux espaces tropicaux

Parmi les quelques modèles spatialisés des exploitations à une échelle fines, peu sont appliqués aux espaces agricoles sub-tropicaux. Bien que les modèles d'exploitation soient rarement spatialisés, ce champ de recherche commence à être exploré avec des méthodes variées. Par exemple, (Eigner and Nuppenau, 2019) cherchent à modéliser les changements de cultures, de taille de champs et d'exploitation à l'aide d'un modèle spatio-temporel appliqué en Allemagne. Des modèles de bio-économie spatialisée comme MOSAICA (Chopin et al., 2015b) peuvent être utilisés pour modéliser spatialement l'impact des politiques publiques sur l'agriculture l'échelle ré-

gionale, ici la Guadeloupe[10]. D'autres modèles sont basés sur des connaissances expertes pour analyser l'influence des choix des agriculteurs sur la répartition des cultures (Jahel et al., 2018).

Au sein des modèles de changements d'occupation et d'usage des sols, les données satellitaires sont utilisées pour cartographier l'occupation des sols d'une saison culturale ou d'une année. Lorsque ces cartes sont utilisées pour des projections à moyen ou long terme, la validité et l'évaluation du modèle peuvent être limitées (Kozak et al., 2017).

Les zones tropicales sont particulièrement difficiles à cartographier par télédétection (voir section 2.1.3), ce qui rend difficile l'application des modèles dans ces espaces. Aux difficultés de la télédétection, s'ajoute aussi une faible disponibilité de données thématiques fiables sur certains pays, empêchant l'usage de certains processus de modélisation dépendants de ces données[11]. En faisant abstraction de la dépendance aux données des modèles en général, les modèles de dynamiques spatiales peuvent être fiables dans la représentation de processus, mais ont comme point négatif une forte inexactitude spatiale.

2.3 Combinaison télédétection / modélisation spatio-temporelle

Afin d'améliorer la caractérisation de l'occupation des sols ou la connaissance de l'usage des sols, on retrouve dans la littérature quelques travaux combinant la télédétection avec des modèles spatiaux, bien que ces méthodes sont encore expérimentales. En effet, la caractérisation de l'occupation ou de l'usage des sols des systèmes complexes tels que l'agriculture familiale par télédétection ne prend pas en compte les dynamiques et les stratégies des acteurs de ces espaces. La modélisation spatio-temporelle quant à elle reste peu appliquée à la thématique de l'agriculture familiale et, lorsque c'est le cas, elle souffre d'un manque de précision spatiale. Parmi les méthodes actuellement développées, la piste privilégiée des chercheurs est de s'appuyer sur des outils techniques, basés principalement sur des méthodes statistiques d'intégration de données de télédétection dans des modèles, ou d'utilisation de modèles d'apprentissage profond sur des données temporelles.

En dehors de la thématique de l'agriculture familiale et des problématiques d'occupation ou d'usage des sols qui lui sont liées, le lien entre modélisation spatio-temporelle et télédétection existe depuis plusieurs années en se basant sur deux principales méthodes: le **forçage** et l'**assimilation**.

Le **forçage** consiste à utiliser des données de télédétection comme variables d'entrée dans un modèle et est souvent utilisé pour l'étude des changements d'occupation et d'usage des sols. Pour les modèles dynamiques, cette méthode permet de réactualiser une variable du modèle à un instant donné. Par exemple, en mettant à jour des données de précipitation afin de modéliser si des

[10] Ce système n'est pas nécessairement familial, mais est composé d'une majorité de petites exploitations d'une taille moyenne de 4 ha

[11] La critique de l'article et de la production cartographique de Hansen et al. (2013) par Tropek et al. (2014) met l'accent sur plusieurs réflexions à avoir par rapport au processus de modélisation. Que ce soit la définition de l'objet étudié, ici de forêt, de toutes les analyses et productions qui en découlent ou bien de la fiabilité des données support. Il faut rester particulièrement vigilant à l'application d'un modèle à grande échelle et à l'analyse des résultats.

plantes peuvent pousser, ou si cette zone subit une sécheresse. L'utilisation de données satellitaires ou issues d'instruments de mesure ponctuels n'étant pas continue, les modèles interpolent les données entre deux acquisitions.

Des recherches sont toujours en cours pour améliorer la caractérisation spatiale des systèmes complexes, principalement à l'aide de méthodes de forçage. Barcza et al. (2009); Górecki and Łuczak (2013) et Górecki and Łuczak (2015) développent dans leurs approches l'intégration de séries temporelles ou des structures de classifications temporelles. Bailly et al. (2018) proposent l'intégration de structures temporelles dans des classifieurs, à l'aide de chaînes de Markov pour les rotations de cultures sur une nomenclature fine, dans une zone complexe, mais avec des données abondantes (images satellitaires optique et radar, données terrain, délimitation des champs). D'autres utilisent des processus stochastiques (Janssen and Middelkoop, 1992; Lu and Weng, 2007), ou encore des modèles chaotiques (Mangiarotti et al., 2018) pour cartographier l'occupation des sols. D'autres méthodes sont aussi présentes dans la littérature telles que la reclassification d'images à l'aide de modèles probabilistes (Sweeney et al., 2015) ou l'utilisation de données statistiques pour cartographier les rotations culturales au cours d'une année (Qiu et al., 2003).

Vannier (2011) compare ainsi télédétection et modélisation spatiale sous l'angle de l'expertise plutôt que d'un aspect uniquement technique. L'une des perspectives de ses travaux est la combinaison des méthodes de traitement de données de télédétection et des méthodes utilisées en modélisation spatio-temporelle experte (type SMA) afin d'obtenir une représentation plus réaliste du territoire, pouvant résulter sur un modèle plus fiable.

L'assimilation vise à estimer des variables grâce à la combinaison optimale de valeurs a priori et d'observations. Cette méthode est principalement utilisée dans les domaines de la climatologie et de l'océanologie (Suzuki and Zupanski, 2018). Elle est aussi utilisée depuis quelques années dans les modèles de cultures, mais peu pour les modèles des dynamiques paysagères (Jahel, 2016).

Parmi les quelques études sur la combinaison télédétection - modélisation spatio-temporelle, Jahel (2016) présente une méthode d'assimilation des données de télédétection pour la caractérisation de l'occupation des sols basée sur le satellite MODIS pendant 15 ans. La méthode est appliquée aux tendances d'indice de végétation puis, l'évolution de l'espace est analysée avec une typologie espace cultivé / espace non cultivé et les facteurs de choix des agriculteurs burkinabè. L'analyse de l'évolution des types de culture n'est pas effectuée en raison des résultats non significatifs. Un autre exemple, moins lié aux dynamiques spatio-temporelles, mais appliquées aux aires cultivées et celui de Lambert et al. (2018) qui cartographient des cultures de la la ceinture du coton au Mali (coton, maïs, millet, sorgho, arachides et autres) et réaffectent la classe des objets en fonction d'un filtre basé sur les aires cultivées par culture.

Il y a encore peu de recherches sur des méthodes combinant la télédétection et la modélisation spatiale basées sur des connaissances expertes. Ces approches peuvent être une alternative aux approches du type « boite noire » que peuvent être l'apprentissage profond ou des modèles chaotiques.

Nous avons présenté dans cette section deux outils pour analyser de façon spatio-temporelle l'agriculture familiale : la télédétection et la modélisation spatio-temporelle. Les données d'observation de la Terre sont diverses et différentes méthodes pour traiter ces données sont toujours développées, mais l'agriculture familiale reste difficile à caractériser par télédétection. La modélisation spatio-temporelle est utilisée pour simuler un environnement ou des dynamiques à partir d'un système étudié. Cet outil est beaucoup utilisé pour la compréhension de processus ou pour la prospective territoriale. Malgré la complémentarité de ces outils, ils sont encore peu associés dans la littérature.

3 Objectifs de la thèse et démarche adoptée

L'agriculture familiale est un système complexe qui est largement étudié d'un point de vue économique, sociologique, géographique et agronomique. Ce système, et ses nombreuses variations, contribue majoritairement à la sécurité alimentaire mondiale, particulièrement dans les pays en voie de développement, où il est particulièrement vulnérable. La compréhension de la structure et des dynamiques de ces systèmes est importante pour caractériser et améliorer leur résilience et pour promouvoir de nouvelles stratégies d'adaptation aux changements. Cette compréhension passe par une représentation spatiale des systèmes telle que délivrée par l'observation de la Terre, ainsi que par la connaissance de leur fonctionnement et de leur état.

Le progrès des missions satellitaires a permis une offre accrue en matière d'images satellitaires, avec une grande variété de résolution spatiale et de résolution temporelle. L'avènement des séries temporelles ayant des résolutions spatiales de plus en plus élevées et la conception de nouvelles méthodes de classification (tel que la méthode OBIA) permettent de mieux caractériser les systèmes complexes agricoles, mais des limites demeurent. L'agriculture familiale tropicale reste difficile à cartographier et à caractériser en partie à cause de la nébulosité pour les images optiques, de la saturation du sol en eau pour les images radar, de la taille des champs (fréquemment inférieures à la taille d'un pixel), et des difficultés de discrimination entre les cultures et la végétation naturelle et des pratiques culturales variées.

La caractérisation de l'agriculture familiale par le biais de la modélisation spatio-temporelle a comme avantage la compréhension de l'organisation de ces systèmes, que ce soit à l'aide de typologies ou de simulations à l'échelle d'une exploitation, mais ces structures sont rarement spatialisées. La modélisation spatio-temporelle de l'agriculture familiale a comme objectif principal la prospective territoriale et souffre donc de fortes inexactitudes spatiales. Dans les systèmes tropicaux, très peu d'études s'intéressent aux dynamiques spatiales de l'agriculture familiale, en partie par manque de données spatialisées.

Face aux limites des méthodes traditionnellement utilisées pour comprendre et caractériser les systèmes complexes, certains chercheurs développent de nouvelles méthodes pour mieux caractériser ces systèmes. Actuellement, les méthodes développées en télédétection utilisent principalement des modèles qui apprennent directement des données (apprentissage automatique avec de l'intelligence artificielle), mais qui laissent peu de place à l'expertise thématique. L'expertise thé-

matique, qui peut être représentée à l'aide de la modélisation spatio-temporelle, reste cependant une piste à poursuivre et à développer pour une meilleure compréhension des systèmes complexes (Vannier, 2011).

Ainsi, la problématique principale de cette thèse est :

Comment améliorer la caractérisation des espaces façonnés par l'agriculture familiale en zone tropicale ?

Pour répondre à cette problématique générale, l'objectif de ce doctorat est le **développement** et l'**évaluation** d'une nouvelle démarche de caractérisation des systèmes et types de culture, basée sur des images de télédétection et la modélisation d'informations environnementales et socio-économiques. L'utilisation conjointe de méthodes de télédétection et de modélisation des dynamiques spatiales nécessite l'élaboration d'un **modèle conceptuel** qui permet l'utilisation de méthodes variées ainsi que de l'intégration de données hétérogènes.

Pour traiter cette problématique, ce travail de thèse s'attache à répondre aux questions de recherche suivantes :

- Comment organiser un modèle conceptuel, alliant à la fois des méthodes de traitement de données de télédétection et des méthodes de modélisation spatiale ?
- Comment appliquer le modèle conceptuel développé à un cas d'agriculture familiale ?

Le développement de cette nouvelle démarche passe par la construction d'un **modèle conceptuel** qui doit répondre à certaines contraintes :

- le modèle conceptuel doit permettre la **combinaison** de données et de méthodes de télédétection et de modélisation spatio-temporelle ;
- le modèle conceptuel et son application doivent s'appuyer sur des connaissances thématiques/expertes, qui peuvent être hétérogènes, multi-échelles et multi-sources. Ces connaissances doivent être **formalisées** pour être intégrées dans le modèle ;
- le modèle conceptuel doit être **générique** et **reproductible** afin de pouvoir être appliqué à des systèmes complexes dans différents espaces géographiques et pour différentes thématiques ;
- le modèle conceptuel doit permettre d'**améliorer** la caractérisation de l'occupation et l'usage des sols, ainsi que la **fiabilité** des produits thématiques.

La création d'un modèle conceptuel vise à lever les limites des méthodes existantes de caractérisation spatiale et thématique des systèmes complexes tropicaux. Pour éprouver ce modèle, il sera appliqué à l'agriculture familiale tropicale des Hautes Terres de Madagascar. Cette région est particulièrement intéressante pour développer une approche combinant télédétection et modélisation spatio-temporelle à partir de règles expertes, car c'est un système fortement anthropisé

qui dépend de facteurs environnementaux, socio-économiques et géographiques. Pour cela, deux sources de données sont combinées :

1. des données de télédétections à deux résolutions temporelles et spatiales. Les images à haute résolution spatiale, mais peu fréquentes, sont utilisées pour définir la structure spatiale de l'espace agricole, alors que des images à basse résolution spatiale, mais plus fréquentes, apportent l'information temporelle liée au fonctionnement des objets étudiés ;
2. les données servant à la modélisation de dynamiques spatiales, intégrant des connaissances a priori, de différentes disciplines (agronomiques, géographiques, socio-économiques) pour mieux interpréter les structures et les signatures temporelles des objets spatiaux en prenant en compte les stratégies et pratiques des acteurs du système étudié.

La figure 6 présente l'organisation et la démarche générale de ce doctorat. Cette thèse vise à construire un modèle conceptuel, afin d'améliorer la compréhension et la caractérisation des systèmes complexes (**chapitre 2**), et à l'appliquer aux Hautes Terres de Madagascar. Le **chapitre 3** présente la zone d'étude ainsi que les données et outils utilisés, du **chapitre 4** qui présente l'application du modèle sur la zone d'étude et les résultats sont donnés en **chapitre 5**. La démarche, le modèle conceptuel et l'application de celui-ci à la région des hautes terres sont discutés dans le **chapitre 6** afin de mener une analyse sur le travail de ce doctorat et avancer un ensemble de propositions visant à améliorer le travail théorique et appliqué.

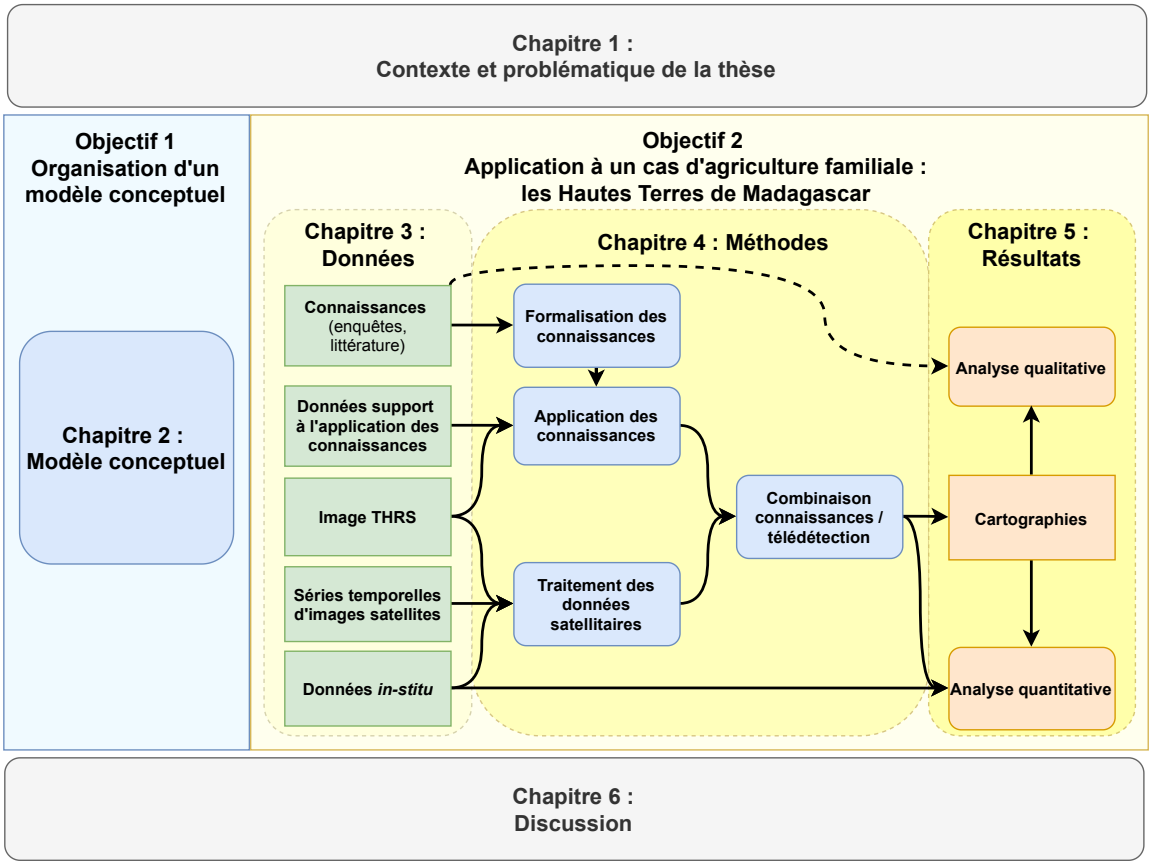


Figure 6 Organisation de la thèse

Les travaux de cette thèse ont été valorisés dans un article de conférence à comité de lecture :

- Crespin-Boucaud Arthur, Lebourgeois Valentine, Lo Seen Danny, Castets Mathieu, Bégué Agnès. Agriculturally consistent mapping of smallholder farming systems using remote sensing and spatial modelling. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2020, XLII–3/W11, 35–42.

Les travaux de collecte de données d'occupation et d'usage des sols sont valorisés dans un article soumis à *Scientific Data* :

- Jolivot Audrey, Lebourgeois Valentine, Ameline Mael, Andriamanga Valérie, Bellón Beatriz, Castets Mathieu, Crespin-Boucaud Arthur, Defourny Pierre, Diaz Santiana, Dieye Mohamadou, Dupuy Stéphane, Ferraz Rodrigo, Gaetano Raffaele, Gely Marie, Jahel Camille, Lelong Camille, le Maire Guerric, Leroux Louise, Lo Seen Danny, Muthoni Martha, Ndao Babacar, Newby Terry, Lira Melo de Oliveira Santos Cecília, Rasoamalala Eloise, Simoes Margareth, Tiaw Ibrahima, Timmermans Alice, Tran Annelise, Bégué Agnès, Harmonized in-situ JECAM datasets for agricultural land use mapping and monitoring in tropical countries. *Scientific Data* (Soumis)

Deux articles sont en cours de préparation.
Un premier article "Analyse des règles expertes par fouille de données et analyse de sensibilité du modèle de caractérisation de l'occupation et de l'usage des sols à partir d'une base de données de vérité terrain" :

- Crespin-Boucaud Arthur, Lebourgeois Valentine, Tesseire Maguelonne, Bégué Agnès, Lo Seen Danny, Evaluationg and ranking of expert rules for cropping pratices. *Intelligent Data Analysis* (en préparation)

Un second article sur l'approche et les résultats de l'application des modules spatiaux tel que présenté dans cette thèse :

- Crespin-Boucaud Arthur, Lo Seen Danny, Lebourgeois Valentine, Castets Mathieu, Bégué Agnès, Rule-based improvement of satellite image classification for agriculturally consistent mapping of smallholder farming systems: case study of Madagascar highlands. *International Journal of Applied Earth Observation and Geoinformation* (en préparation)

Chapitre 2

Modèle conceptuel

Table des matières

1. Pourquoi un modèle conceptuel
2. Les niveaux d'organisation des systèmes complexes agricoles
   2.1. L'objet
   2.2. Le groupement d'objets de niveau 1
   2.3. Le groupement d'objets de niveau 2
3. Description du modèle conceptuel
   3.1. Principes de base du modèle
   3.2. Principaux modules du modèle conceptuel

« Comment combiner télédétection et modélisation des dynamiques spatiales ? »
Cette question est encore peu abordée dans la littérature, les réponses de la recherche sont variées et aucune n'est encore définitive. Dans ce deuxième chapitre, nous proposons un modèle conceptuel pour intégrer des méthodes de traitement de données de télédétection et de modélisation spatio-temporelle afin d'améliorer la caractérisation de l'agriculture familiale en zone tropicale. La première section de ce chapitre présente le choix de construction du modèle autour de la thématique de l'agriculture familiale tropicale. La deuxième section fait l'inventaire des niveaux d'organisation spatiale utilisés pour définir l'occupation et l'usage des sols : le champ, l'agriculteur et le village. Ces niveaux ont une dimension à la fois spatiale et temporelle. La troisième section décrit la démarche et le fonctionnement du modèle conceptuel. Le modèle présenté dans ce chapitre est séparé en cinq modules. Un module d'initialisation du modèle, correspondant au traitement des données de télédétection à partir d'une approche OBIA (module 0) et quatre modules mobilisant à la fois des règles spatio-temporelles et des données satellitaires et thématiques : 1) la modélisation du fonctionnement temporel du système, 2) l'application des règles spatiales, 3) l'application des règles temporelles et, 4) la modélisation de la répartition spatiale des classes.

1 Pourquoi un modèle conceptuel

L'idée d'un modèle conceptuel a été abordée dans le chapitre 1. Nous commençons ici par expliquer le choix de consacrer tout un chapitre de cette thèse à la présentation et à l'élaboration d'un modèle conceptuel. Les systèmes complexes sont très diversifiés, il est donc impossible d'élaborer un modèle conceptuel applicable à toutes les situations. Chaque système doit être analysé individuellement et un modèle doit être conçu pour la compréhension de chaque système. Pour éviter de concevoir un modèle pour chaque cas d'étude, l'utilisation de généralisations pour modéliser les processus est néanmoins possible (Le Moigne, 1977). Le modèle conceptuel développé dans cette thèse est présenté en lien avec l'agriculture familiale, principalement tropicale, car il vise à répondre aux problématiques qui découlent de ce type d'agriculture. La démarche d'explication l'accompagnant a pour but d'amener une réflexion de la part du lecteur afin que l'idée développée puisse être ré-utilisée, modifiée, adaptée et puisse faire l'objet d'une démarche critique visant à l'améliorer.

Ce modèle conceptuel a été développé en s'appuyant sur un système complexe adaptatif : l'agriculture familiale. Comme nous l'avons introduit dans le chapitre 1 ce système a de fortes capacités à évoluer par mutation ou par apprentissage. Ces évolutions incluent à la fois les conséquences des choix des agriculteurs face à des événements exogènes (retard de la saison des pluies, variation des prix), mais aussi les effets du changement climatique qui modifient l'environnement, et donc les processus en cours. Le système de l'agriculture familiale est ainsi le résultat de modifications, d'adaptations, de réflexions à la fois spatiales et temporelles. Ces changements progressifs, liés au contexte environnemental, économique et social, ont peu à peu fait évoluer ce système jusqu'à celui que l'on peut actuellement observer et continuent à le transformer.

Les choix spatiaux des agriculteurs entérinent les principaux facteurs de l'emplacement d'une culture. Tout d'abord, les facteurs biophysiques permettent de déterminer les sites qui peuvent être cultivés et ceux qui sont moins propices à l'agriculture ou à certains types de cultures (ex. type de sol, pente). Le choix du type de culture se fait en fonction d'un contexte culturel (ex. tradition familiale, habitude agricole) qui dépend aussi d'éléments spatiaux (ex. possibilité d'irrigation, route passante, proximité de marchés, point de collecte). En plus de ces facteurs, les antécédents culturaux permettent de déterminer quel type de système de culture est pratiqué sur une parcelle.

Ce système, comme la plupart des systèmes, est en évolution et en changement permanent. Cela nécessite d'adopter une approche spatio-temporelle ainsi que des processus pouvant être amplifiés, diminués, voire supprimés selon le système et la période étudiée.

Le modèle conceptuel développé dans ce chapitre est conçu de manière à :

- produire des cartes d'occupation et d'usage des sols fiables et facilitant la compréhension et la caractérisation du système analysé ;
- ne pas être aussi complexe à comprendre que le système étudié. Cet objectif vise à éviter que le modèle ne soit trop spécifique à la problématique étudiée, ainsi qu'à l'objet d'étude et aux caractéristiques spatio-temporelles qui lui sont liées ;

- avoir une approche multi-scalaire, intégrant des processus locaux et généraux, que ce soit à l'échelle temporelle ou à l'échelle spatiale ;
- développer des processus qui mobilisent à la fois des concepts physiques tels que des indices issus d'une image satellitaire, et des connaissances sur les pratiques culturales qui participent à déterminer l'usage ou l'occupation des sols ;
- évaluer les résultats du modèle de façon quantitative, mais également qualitative (thématique) afin d'éviter les biais des méthodes d'évaluation de la télédétection ou de la modélisation.

Ce modèle conceptuel a pour objectif de proposer une nouvelle approche afin de combiner à la fois 1) des informations issues du domaine de la télédétection, qui sont des données d'observation de la Terre principalement analysées par des méthodes de classification automatiques qui laissent l'intelligence artificielle sélectionner les variables les plus pertinentes pour l'attribution de classes d'occupation ou d'usage des sols ; et 2) des connaissances expertes, émanant d'observations terrain et de raisonnements contextualisés, et nécessitant relativement peu de données. Pour être appliquées, ces connaissances, formalisées en règles expertes ou processus spatio-temporels, peuvent parfois nécessiter des données (ex. pour appliquer une règle liée à une distance aux cours d'eau, il faut la carte des cours d'eau), mais de manière explicite et avec des règles ou processus transparents pour le chercheur.

2 Les niveaux d'organisation des systèmes complexes agricoles à prendre en compte dans l'élaboration du modèle conceptuel

Comme introduit dans le chapitre 1, l'agriculture familiale est variée et complexe à analyser. Elle présente des caractéristiques que l'on ne peut ignorer lors de la mise en oeuvre du modèle conceptuel. Notamment, plusieurs niveaux d'organisation peuvent être utilisées pour définir un type d'occupation ou d'usage des sols dans un contexte agricole (figure 7). Ces niveaux sont imbriqués entre eux selon une hiérarchie qui peut changer selon les contextes. Les principaux niveaux d'organisation spatiale sont : 1) l'objet (le champ), 2) le groupement d'objets de niveau 1 (l'agriculteur et son exploitation), 3) le groupement d'objets de niveau 2 (groupe social ou village). Ces niveaux structurent les trois sous-sections suivantes.
D'autres niveaux de groupement peuvent être ajoutées selon les besoins et le contexte comme une échelle régionale ou nationale. Les principales temporalités d'analyse pour caractériser un champ sont celles décrites par (Bégué et al., 2018) (chapitre 1, section 1.3.1.2 L'échelle d'analyse spatio-temporelle). Les niveaux d'organisation spatiale et les temporalités d'analyse sont intimement liées et il est donc difficile de les dissocier. L'analyse et la caractérisation d'un espace agricole complexe doivent être multiscalaires pour prendre en compte les différents contextes des objets.

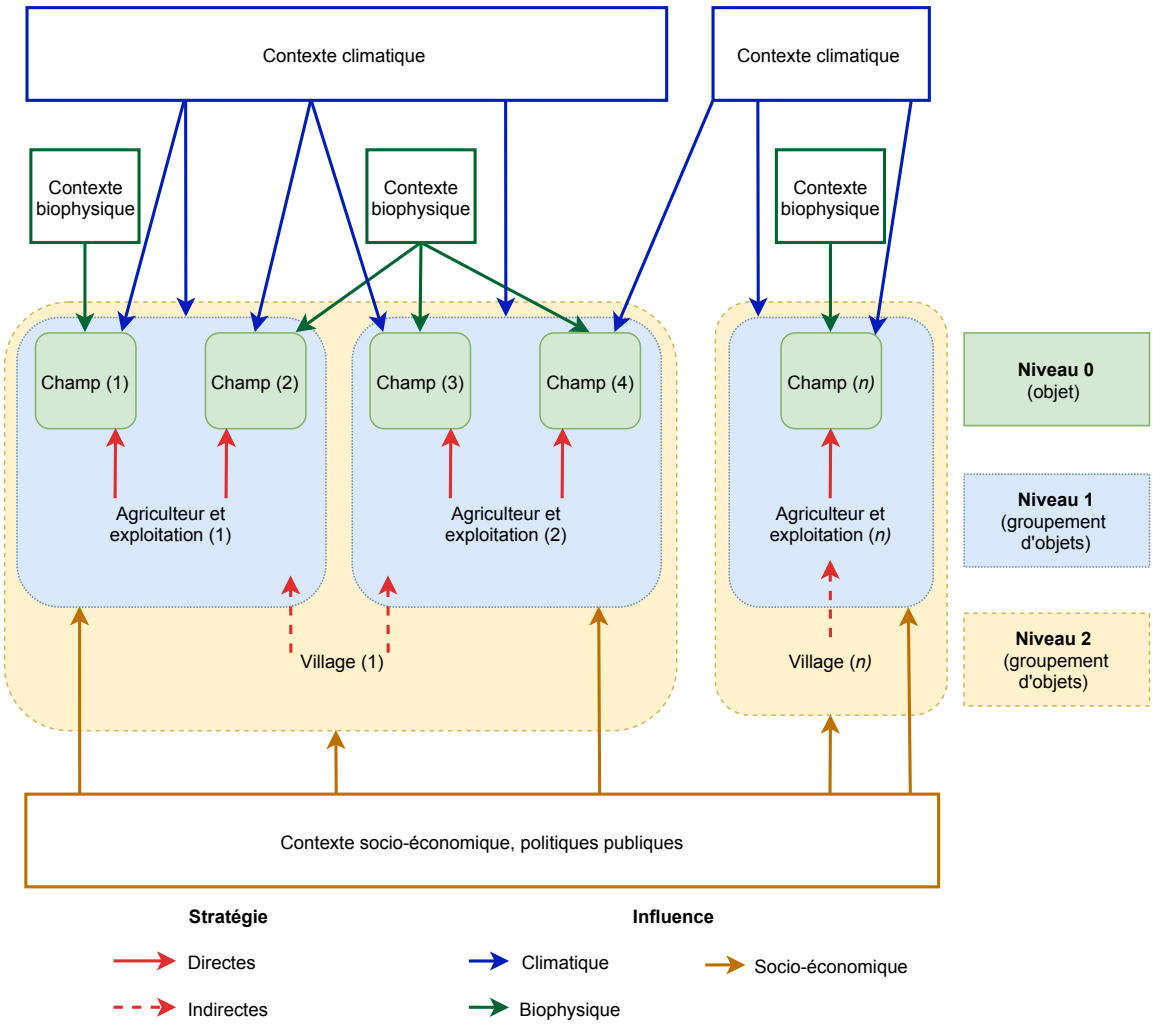


Figure 7 Schéma des différents niveaux d'organisation spatiale et leurs imbrications dans l'objectif de caractériser l'occupation ou l'usage des sols d'un objet. Trois niveaux sont représentés pour caractériser l'agriculture familiale, et trois principaux contextes ont été retenus comme ayant une influence majeure sur les stratégies et choix des agriculteurs dans leur choix de la mise en culture d'un champ.

2.1 L'objet

Pour le niveau d'organisation spatial de l'objet (ex. champ), le milieu biophysique est primordial pour le choix du type de système de culture pratiqué (Chopin and Blazy, 2013). Ainsi, l'altitude, la pédologie, la présence de cours d'eau ou de sources, la pente ou encore la position topographique sont des facteurs déterminant un ensemble d'occupations et d'usages du sol possibles. Par exemple, en zone tropicale ou sub-tropicale, une culture de riz irrigué se trouvera presque exclusivement située dans un bas-fond, alors qu'un verger de pommiers nécessitera une certaine altitude et une exposition minimale pour être productif. Certains facteurs, tels que l'altitude ou l'exposition, sont relativement immuables. D'autres, tels que la fertilité du sol, peuvent au contraire évoluer dans le temps, que ce soit pour des raisons climatiques, environnementales ou liées aux choix et pratiques de l'agriculteur. Pour caractériser un objet (*i.e.* lui attribuer une classe d'occupation ou

d'usage des sols) en utilisant à la fois des données biophysiques (qui peuvent correspondre à des données satellitaires) et des données socio-économiques il faut opérer en plusieurs phases : 1) la **délimitation du contour de l'objet** qui peut être effectuée à l'aide de données existantes comme un parcellaire, par photo-interprétation ou par une segmentation automatique d'une image satellitaire à une résolution spatiale adaptée (où les objets sont directement visibles), l'ensemble de ces délimitations correspondant à la structure de la zone d'étude ; 2) l'analyse de **la situation biophysique de chaque objet** par le biais de données thématiques adaptées (ex. altitude, relief, données climatiques, phénologie) ; et 3) l'observation du **fonctionnement temporel de l'objet**, c'est-à-dire l'évolution de l'état de l'objet dans le temps, que ce soit par des effets climatiques, météorologiques ou par les décisions des agriculteurs sur la gestion de leur champ sur plusieurs années, qui peuvent être approchés avec des données temporelles telles que des séries temporelles d'images satellite (données optiques, de précipitation ou autre).

Pour un objet, l'ensemble des propriétés qui le caractérisent peuvent être résumées par le terme « contexte ». Ce contexte peut être biophysique, temporel, socio-économique ou climatique. Dans la suite du modèle, nous chercherons à caractériser chaque objet en fonction de son contexte par l'application de règles et de modélisations spatio-temporelles (traduisant les stratégies et pratiques des agriculteurs) qui auront pour objectif d'identifier les types d'occupation/usage des sols possibles. Ces stratégies dépendent de fonctionnements opérant à différents niveaux d'organisation spatiale : celui du champ (objet - niveau 0), mais également celui de l'exploitation (groupement d'objets de niveau 1) ou d'un niveau encore supérieur qui peut être administratif ou non (ex. village, groupement d'exploitations, groupe social).

L'objet est le niveau central de ce modèle conceptuel en étant caractérisé tout au long des étapes de modélisation. L'ensemble des objets représente la structure spatiale de la zone étudiée. Chaque objet peut être caractérisé par des propriétés servant à définir l'occupation et l'usage des sols.

2.2 Le groupement d'objets de niveau 1

Un groupement d'objets est composé d'un ensemble d'objets regroupés selon une ou plusieurs caractéristiques thématiquement cohérentes. Pour le niveau d'organisation du groupement d'objets (ex. exploitation), les choix des acteurs déterminent l'occupation ou l'utilisation du sol possible pour chaque objet. À partir du contexte des objets, une réflexion est menée par l'acteur au niveau du groupement d'objets, afin d'adapter les stratégies mises en place de manière cohérente aux objets dans leur ensemble. Les stratégies sont déterminées à partir de l'analyse du contexte des objets, qui peut être à la fois climatique et socio-économique (l'analyse du contexte biophysique est majoritairement effectuée au niveau de l'objet) (Figure 7). Dans un système agricole, les stratégies visent aussi à répartir correctement les systèmes de cultures dans les champs disponibles pour optimiser la diversité et la productivité des cultures, ce niveau d'organisation représente cette réflexion. Dans le cas de l'agriculture familiale, ces groupements peuvent correspondre à l'ensemble des champs d'un même exploitant, ou les champs appartenant à une famille. **Les choix**

agronomiques stratégiques sont effectués selon plusieurs facteurs (socio-économiques, biophysiques, climatiques), qu'ils soient internes à l'exploitation (pédologie, accès, parasites, capital de travail, nombre de champs) ou externes (prix, climat, choix des agriculteurs voisins). Une fois les stratégies définies, elles sont appliquées aux objets (champs). Les champs font donc partie intégrante de l'exploitation qui applique sur chacun d'entre eux une stratégie agronomique (Bureau et al., 2001). Dans le cas de l'agriculture familiale tropicale, les champs d'une exploitation sont généralement diversifiés pour répondre aux besoins de la famille (Bélières et al., 2013). Ce niveau permet d'analyser les systèmes de culture qui sont le produit entre les facteurs biophysiques des champs et le type de mise en valeur agricole défini par les stratégies des agriculteurs.

Par exemple, un bas-fond pourra être valorisé lors de la saison culturale principale par des cultures devant être irriguées ou bien ayant besoin de se situer sur sol hydromorphe (ex. riz irrigué, cultures maraîchères). Pour la saison culturale secondaire (contre-saison), le bas-fond pourra être mis en culture avec des pommes de terre ou des tomates ou bien servir de pépinière pour le riz irrigué. Si un agriculteur possède 4 champs en bas-fond, alors l'un d'entre eux servira probablement de pépinière en contre-saison. Pour l'application de stratégies, une typologie de groupements peut être créée afin de déterminer les grands types de stratégies pouvant être mis en place au niveau d'organisation d'un groupement. En agriculture, le type d'exploitation a un impact important sur les systèmes de culture mis en place, selon son capital, le bétail et les terres possédées (Jouve, 2006; Jahel et al., 2018). **La modélisation des choix d'une exploitation est particulièrement difficile. En effet, à moins d'avoir le cadastre de la zone, il est impossible de connaître les champs qui lui sont liés et, dans ce cas, le niveau d'organisation de l'exploitation ne pourra être utilisé dans la modélisation. Il en est de même si l'agriculteur ne possède qu'un champ**.

De manière spatio-temporelle, l'évolution d'un champ est déterminée par les choix de l'agriculteur en fonction de l'itinéraire technique du champ (combinaison logique et ordonnée des techniques culturales) au sein de l'exploitation. L'histoire du champ est importante (précédentes cultures, anciens itinéraires techniques), mais peut être difficilement compréhensible pour la caractérisation de grands territoires à la cause la démultiplication d'histoires différentes. Une histoire simplifiée, ou probable, peut être identifiée ou modélisée afin de créer cette information spatio-temporelle (Sorel et al., 2010; Dury et al., 2012).

L'intégration de l'échelle temporelle est importante, que ce soit pour caractériser une constante temporelle ou bien pour déterminer des changements. La connaissance de l'historique du contexte de l'objet et de son occupation ou de son usage passé est primordiale. En agriculture, l'itinéraire technique détermine d'évolution de la parcelle et l'ensemble des choix possibles de l'exploitant, toujours subordonnés aux contraintes des champs (Jouve, 2006). Les successions culturales et les choix de répartitions spatiales sont donc le résultat de processus de décisions stratégiques, logiques et non aléatoires (Leenhardt et al., 2010). Par exemple, la succession d'une même culture sur un même champ aura pour conséquence la diminution de la productivité du champ, l'appauvrissement du sol et l'augmentation du risque de parasites. Les cultures associées et les successions culturales sont donc choisies et placées d'après cette logique spatiale et tem-

porelle (Autfray et al., 2014). Généralement cette séquence de succession est fixe, prise en compte à l'échelle de l'exploitation et de son supra-groupement selon un rapport de coûts et d'externalités.

Le niveau de groupement de niveau 1 (ex. exploitations) permet de mieux identifier les principaux systèmes de culture et de les répartir conformément à des stratégies plus générales ou selon des contraintes biophysiques qui peuvent limiter la présence de certains systèmes de culture. Par exemple, si le sol n'est pas suffisamment fertile pour cultiver des légumes et que l'agriculteur ne souhaite pas ou ne peut pas investir dans des fertilisants, certains systèmes de cultures ne peuvent être présents.

Le groupement d'objets de niveau 1 permet de modéliser les stratégies des agriculteurs à la fois temporelles et spatiales selon leur comportement. Ce niveau peut être celui de l'exploitation et des typologies peuvent être nécessaires pour mieux comprendre et modéliser les décisions des agriculteurs.

2.3 Le groupement d'objets de niveau 2

Le groupement d'objets de niveau 2 définit un niveau d'organisation englobant le premier niveau de groupement (niveau 1) et celui de l'objet (niveau 0) (Figure 7). Ce niveau peut être administratif (ex. village) ou sociologique (ex. groupement d'exploitations lié à une association ou une famille dont plusieurs membres possèdent leurs champs). Les groupements de niveau 2 subissent les contraintes de contextes extérieurs qui peuvent être socio-économiques, climatiques, ou encore de groupements plus important tels que les politiques publiques d'un État ou bien de zones protégées (Flichman, 2011). L'objectif de ce groupement d'objets est similaire à celui de niveau 1. Dans cette section, nous décrivons son application à l'agriculture familiale en considérant le village comme un groupement d'objet de niveau 2.

À l'échelle d'un village ou d'une structure en groupe d'agriculteurs, les stratégies sont souvent proches et les principaux systèmes de culture peuvent être identifiés. **Ce regroupement permet d'identifier des terroirs agronomiques**[1] **qui peuvent être diversifiés au sein d'un village ou d'un groupement de villages**. Dans les systèmes d'agriculture familiale, la conduite des systèmes de culture se fait à la fois de manière collective (ex. date collective pour débuter les semis, etc.) et individuelle (ex. priorisation de champs, densité de semis, etc.) (Jouve, 2006). Les exploitations ont une certaine diversité de systèmes de cultures sur leurs champs et cette diversité se retrouve aussi à l'échelle d'un village ou d'un groupe familial. Des villages peuvent adopter des systèmes agricoles différents, ou présenter des spécialisations agricoles en fonction de facteurs biophysiques (pédologie, altitudes) ou économiques (marchés, collecteurs). Il est possible de résumer statistiquement ces groupements. Par exemple, un pourcentage théorique de répartitions entre types de cultures peut être estimé selon des données socio-économiques, des enquêtes ou des statistiques administratives par village ou autre unité administrative/sociologique (Vintrou

[1] « unité de milieu présentant des caractéristiques écologiques homogènes et faisant l'objet du même type d'exploitation agricole, le terroir viticole en étant l'illustration la plus manifeste » Jouve (2006).

et al., 2012b). La modélisation des systèmes de culture et de leurs répartitions suivant une unité administrative ou sociologique (villages, familles, groupements) est possible en prenant en compte le contexte au niveau des objets et des groupements d'objets de niveau inférieur. Par exemple il est possible de répartir les systèmes de culture en fonction de l'altitude pour un village en plus du pourcentage de chaque type de système de culture présent sur le territoire.

Le groupement d'objets de niveau 2 correspond à un niveau d'organisation administratif ou thématique dans le but de modéliser des stratégies ou des pratiques communes. Avec ce niveau, il est possible répartir spatialement les cultures pour obtenir une meilleure représentation du système agricole dominant.

Dans cette section nous avons abordé trois niveaux d'organisation spatiale qui semblent importants à modéliser pour obtenir une caractérisation de l'occupation et de l'usage des sols plus cohérente pour l'agriculture familiale. Ces trois niveaux sont l'objet (champ), le groupement de niveau 1 (agriculteur et son exploitation) et le groupement de niveau 2 (village ou groupe social).

3 Description du modèle conceptuel

Les travaux en télédétection montrent l'intérêt d'utiliser des approches OBIA pour caractériser l'occupation et l'usage des sols dans les systèmes agricoles complexes (Lebourgeois et al., 2017; Dupuy et al., 2020b). C'est donc sur cette approche que se base le modèle conceptuel, les objets représentant la structure spatiale du système étudié. Le modèle conceptuel s'appuie sur plusieurs concepts qui sont développés dans les paragraphes et sous-sections suivantes. La première sous-section présente les concepts développés et utilisés dans cette approche. La deuxième sous-section présente brièvement le module d'initialisation (module 0, télédétection). Les quatre autre modules du modèle conceptuel sont ensuite développés : 1) la modélisation du fonctionnement temporel du système, 2) l'application des règles spatiales, 3) l'application des règles temporelles et, 4) la modélisation de la répartition spatiale des classes. À l'issue de l'un des modules, ou de plusieurs modules, une classe peut être attribuée à l'objet, comme c'est le cas à l'issue du traitement et de la classification des données de télédétection.

3.1 Principes de base du modèle

La section précédente décrit trois niveaux d'organisation spatiale différents : l'objet (champ), le groupement de niveau 1 (agriculteur et exploitation) et le groupement de niveau 3 (village ou groupe social). Cette sous-section présente la démarche du modèle à travers différents concepts à la fois issus de la télédétection et de la modélisation spatio-temporelle. L'approche se base sur une structure spatiale de la zone d'étude, dont les objets vont être caractérisés selon leur probabilité d'appartenance à une classe. **Les probabilités de toutes les classes sont regroupées dans un**

vecteur par la suite nommé VPAC. Cette démarche s'appuie sur la méthode OBIA (classification d'objets cohérents qui sont caractérisés par des propriétés spectrales ou thématiques, ou par la réponse des objets à des règles et modélisations spatio-temporelles) pour obtenir des VPAC qui représentent l'ensemble des probabilités d'appartenance aux classes pour chaque objet (Figure 8). Dans notre démarche, différents VPAC sont créés en sortie des 5 modules que contient le modèle, et sont ensuite fusionnés afin d'obtenir un VPAC final, traduisant la combinaison des différents modules de télédétection (module 0) et de modélisation (modules 1 à 4). Il est important de noter qu'une carte d'occupation ou d'usage des sols peut être produite à chaque étape, à partir des différents VPAC obtenus, ce qui permet de visualiser les effets de chaque module sur la cartographie.

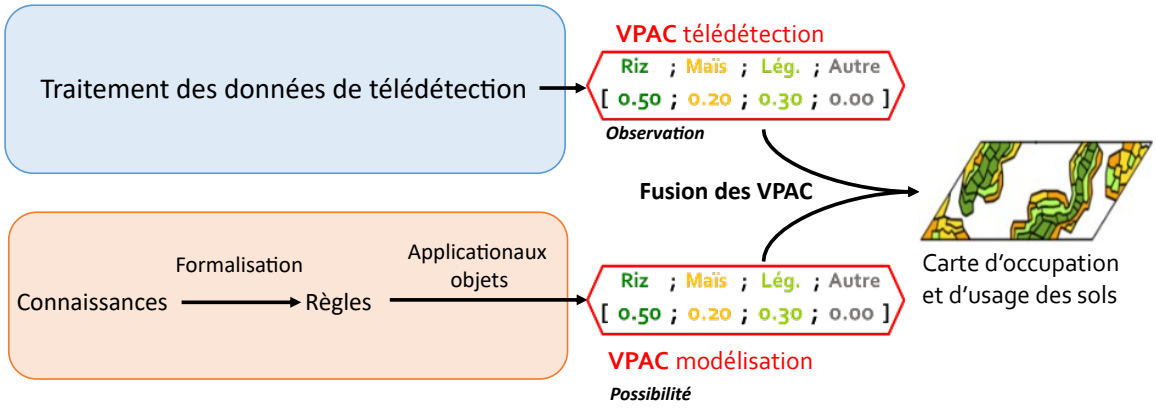


Figure 8 Schéma résumant les principaux concepts mobilisés dans le modèle conceptuel. En bleu, le traitement des données de télédétection (module 0) qui permet d'obtenir un vecteur de probabilité d'appartenance aux classes (VPAC) représentant l'observation de l'occupation et de l'usage des sols. La partie en orange représente la modélisation (modules 1 à 4), dans laquelle la formalisation de connaissances en règles, puis l'application des règles aux objets de la zone d'étude afin d'obtenir des VPAC représentant la possibilité de l'occupation et de l'usage des sols. En encadré rouge, les VPAC sont représentés (ici avec quatre classes dont la somme des probabilités est égale à 1), puis sont fusionnés pour produire une carte de l'occupation et de l'usage des sols.

3.1.1 Les Vecteurs de Probabilité d'Appartenance aux Classes (VPAC)

Dans le modèle, l'objet sert de base à la caractérisation de l'occupation et de l'usage sols. Ce dernier peut être délimité à l'aide de données spatialisées (cadastre, segmentation d'image de télédétection) et dispose de propriétés (biophysiques, climatiques, temporelles, socio-économiques, signature spectrale issue de la télédétection) qui vont déterminer son contexte. Pour affecter l'objet à une classe d'occupation ou d'usage du sol, le modèle se base sur les probabilités d'appartenance de l'objet à chacune des classes. Une probabilité d'appartenance est comprise entre 0 et 1, 0 signifiant que la probabilité que la classe donnée corresponde à l'objet est très faible, alors que 1 signifie que la probabilité que la classe donnée corresponde à l'objet est importante. Toutes les probabilités d'appartenance aux classes pour un même objet sont regroupées au sein d'un **VPAC** (vecteur de probabilités d'appartenance aux classes) dont la somme est égale à 1 (Figure 8).

Ces probabilités peuvent être assimilées à un indicateur de la **fiabilité** d'appartenance ou non

d'un objet à une classe donnée. Si pour un objet, la probabilité d'une classe donnée est proche de 0, alors il est peu probable que l'objet présente cette classe d'occupation ou d'usage des sols, qui pourra donc être écartée de façon certaine. De même, si la probabilité est proche de 1, alors cette classe pourra être affectée à l'objet avec une grande fiabilité. Dans ces deux cas, on peut considérer que la probabilité de cette classe est fiable dans l'inclusion ou l'exclusion. Plus la probabilité d'une classe est proche de 1 ou de 0, plus la probabilité de voir l'objet appartenir ou ne pas appartenir à ladite classe sera fiable. Pour une plus grande clarté de ce manuscrit, la définition utilisée de la **fiabilité** correspond à : **lorsque la probabilité la plus élevée du VPAC est proche de 1, alors l'objet peut être considéré comme fiable**.
Le modèle présenté ci-après est composé de plusieurs étapes, combinant des méthodes d'apprentissage automatique utilisées en télédétection et des méthodes de modélisation spatio-temporelle basées sur des connaissances expertes.

L'issue de la modélisation est d'obtenir une classe finale pour chaque objet ainsi que la probabilité d'appartenance à cette classe, permettant d'estimer sa fiabilité. Pour une plus grande clarté, ces étapes peuvent être regroupés dans des modules thématiques qui peuvent être appliqués de manière itérative, simultanée ou ne pas être appliqués. Chacune des caractérisations effectuées à l'issue des modules peut être évaluée de manière individuelle, ou de manière évolutive tout au long des étapes de caractérisation. La caractérisation finale est atteinte lorsque la fiabilité des objets est suffisante pour ne plus modifier leur classe, ou lorsque l'ensemble des étapes mises en oeuvre pour déterminer la classe d'un objet ne permet plus d'améliorer leur fiabilité.

3.1.2 Connaissances et règles

L'analyse du système agricole permet de connaître son fonctionnement et d'analyser les processus existants à différents niveaux spatiaux et temporels. Ce fonctionnement est représenté par un ensemble de **connaissances** (date de semis, principaux systèmes de cultures, contraintes biophysiques, stratégies villageoises) qui permettent de comprendre et identifier les stratégies culturales des agriculteurs ou d'autres groupes et la situation des objets. Ces **connaissances doivent être formalisées** pour pouvoir être intégrées dans le modèle. Cette formalisation consiste en la création de **règles** qui, lorsqu'elles sont appliquées, permettent de changer les propriétés et les probabilités d'appartenance des objets aux différentes classes des objets. Ces règles sont construites selon la forme **SI condition ALORS action**. La formalisation de connaissances en règles d'allocation spatiale demande de faire face à plusieurs verrous tel que comment traduire des informations qualitatives et peu précises en variables quantitatives ? (ex. issues de discours, d'entretiens) ou comment répartir des systèmes de cultures à partir de connaissances approximatives tout en respectant des assolements statistiquement connu ? (Clavel et al., 2011).

Les connaissances formalisées (*i.e.* les règles) nécessitent aussi des données support qui peuvent être de tout type : spatialisées ou non, temporelles ou statiques, issues ou dérivées des données de télédétection. La possibilité d'appliquer ces règles ainsi que leur importance est dépendante de ces données et de leur qualité. L'application d'une règle ou d'un ensemble de règles permet de

modifier le VPAC d'un objet.

3.1.3 Observation et possibilité

L'utilisation de méthodes de traitement de données de télédétection et de méthodes de modélisation des dynamiques spatiales met en valeur deux concepts clés : l'**observation** et la **possibilité**. La classification des données de télédétection (module 0, d'initialisation) permet d'obtenir un VPAC qui synthétise l'information des images satellitaires pour chaque classe. Ce VPAC représente la probabilité d'**observer** chaque classe pour un objet. L'application des modules peut permettre de produire des VPAC qui correspondent alors à la **possibilité** que l'objet soit caractérisé par cette classe.

Pour initialiser le modèle, la classification des données de télédétection selon une approche OBIA permet de créer, de regrouper les propriétés spectrales des objets et de créer un VPAC qui synthétise le résultat du classifieur. Une fois évalué, le résultat de la classification constitue une entrée dans le modèle. Les objets possèdent alors l'ensemble des propriétés issues des données de télédétection (réflectances, indices spectraux, textures) et données auxiliaires utilisées, ainsi que les probabilités d'appartenance aux classes d'occupation ou d'usage des sols issues de la classification obtenue par télédétection. Suite à ce module d'initialisation, l'intégration de méthodes combinant télédétection et modélisation spatialisée peut être appliquée aux objets par l'application de tout ou partie des modules 1 à 4, et la fusion des VPAC produits avec le VPAC issu du module 0.

3.2 Principaux modules du modèle conceptuel

Le méta modèle de modélisation des dynamiques spatiales et des stratégies culturales peut être découpé en cinq modules que sont : le module **0** : **classification des données de télédétection**, le module **1** : **modélisation du fonctionnement temporel** du système, le module **2** : **règles spatiales**, le module **3** : **règles temporelles** et, le module **4** : **modélisation de la répartition spatiale des classes**. Ces modules peuvent être appliqués à différents niveaux d'organisation spatiale ou temporelles et n'ont pas nécessairement un ordre défini. Dans notre présentation du modèle, nous considérons que le module 0 : télédétection sert à initialiser le modèle et à obtenir un premier VPAC représentant l'observation de l'occupation et de l'usage des sols. Les modules peuvent être appliqués simultanément ou de manière itérative afin que les sorties de chaque module puissent être utilisées comme entrée ou paramètre des autres modules (Figure 9). Ainsi, un VPAC issu d'un module peut être modifié par un autre module (Figure 10). Selon la temporalité d'application du modèle ou la résolution spatiale des objets à caractériser, certains modules peuvent être mis en avant, d'autres être supprimés ou leurs impacts diminués. La dernière sous-section de cette partie présente des méthodes qui peuvent être utilisées pour fusionner les VPAC créés au cours des modules et de la classification des données de télédétection pour déterminer une classe finale.

3.2.1 Rôle et complémentarité des modules

Afin de déterminer la classe occupation ou son usage du sol, il faut prendre en compte les résultats de tous les modules utilisés. Chaque module permet d'établir une caractérisation de la classe de l'objet et d'obtenir un VPAC, chacune ayant ses avantages et ses inconvénients.

Le traitement des données de télédétection (module 0) permet de définir la classe d'un objet à partir de données d'observation de la Terre, donc à un instant, ou un intervalle temporel d'application du modèle défini. L'utilisation de classifieurs spectraux tels que les forêts aléatoires ou les SVM (machine à vecteur de support) ne permet pas de prendre en compte la temporalité des données satellitaires de manière explicite. Ces approches intègrent indirectement la temporalité de l'objet à travers la fréquence des images. C'est-à-dire que chaque bande spectrale ou indice dérivé correspond à une variable, mais ces variables ne sont pas utilisées de manière chronologique. L'objectif de ces classifieurs est de trouver les meilleures variables pour différencier les classes entre elles. Actuellement, les algorithmes de classification traitent les données soit de manière spatiale (classification des valeurs spectrales) soit temporelle (classification des signaux temporels). Aussi, l'expertise (stratégies ou connaissances biophysiques ou thématiques) n'est pas prise en compte, ou faiblement (choix des variables spectels, indices ou textures à intégrer), pour ce module.

La modélisation du fonctionnement temporel d'un objet (module 1) a pour but de représenter correctement les possibilités d'occupation ou d'usage du sol en fonction de connaissances sur les occupations/usages des sols passés de l'objet et de successions culturales théoriques, mais sans prendre en compte sa classe actuelle. Une faible profondeur temporelle peut limiter l'utilisation de ce module qui demande de nombreuses données spatialisées fiables (ex. cartes d'usage des sols des années ou saisons culturales précédentes) pour être appliqué.

L'application des règles spatiales (module 2) permet de caractériser un objet en fonction des stratégies des agriculteurs et des propriétés biophysiques pérennes, parfois immuables (position topographique, pente) de l'objet. Dans ce module, l'aspect temporel n'est pas utilisé.

Les règles temporelles (module 3) permettent d'utiliser des données de télédétection pour caractériser un objet en comparant son signal temporel aux signaux temporels types (*i.e.* de référence) de chaque classe. L'identification d'un signal temporel et la correspondance de ce signal à une classe dans un contexte agricole complexe doit être utilisé avec précaution. Compte tenu de la variabilité de l'agriculture familiale, de la diversité des cultures présentes, des associations de cultures et de la faible taille des champs, ce module peut avoir des résultats limités pour l'identification de cultures saisonnières (variabilité spatiale et temporelle importante à grande échelle). Les séries temporelles utilisées peuvent avoir des lacunes (nuages causant des ruptures temporelles), limitant la continuité du signal temporel. Pour faire face à ces limites, des méthodes ont été développées pour identifier les signaux temporels des cycles de culture (Sakoe and Chiba, 1978; Petitjean et al., 2012; Belgiu and Csillik, 2018).

La modélisation de la répartition spatiale des classes (module 4) permet de modéliser la réparti-

tion de l'occupation et de l'usage des sols d'un groupe social (ex. famille, village) en désagrégeant des données statistiques au niveau d'organisation d'un groupement d'objets vers chaque objet. Cette désagrégation se fait selon une répartition théorique des classes par groupement d'objets. Ce module se base sur la fiabilité de chaque objet et des statistiques pour être utilisé de manière correcte et ne pas perdre la précision spatiale obtenue grâce à la télédétection. Les objets fiables ne seront pas reclassés, alors que ceux avec les probabilités les plus basses pourront être réaffectés à une autre classe. Pour appliquer ce modèle il est nécessaire de connaître les répartitions théoriques à l'échelle d'un groupe.

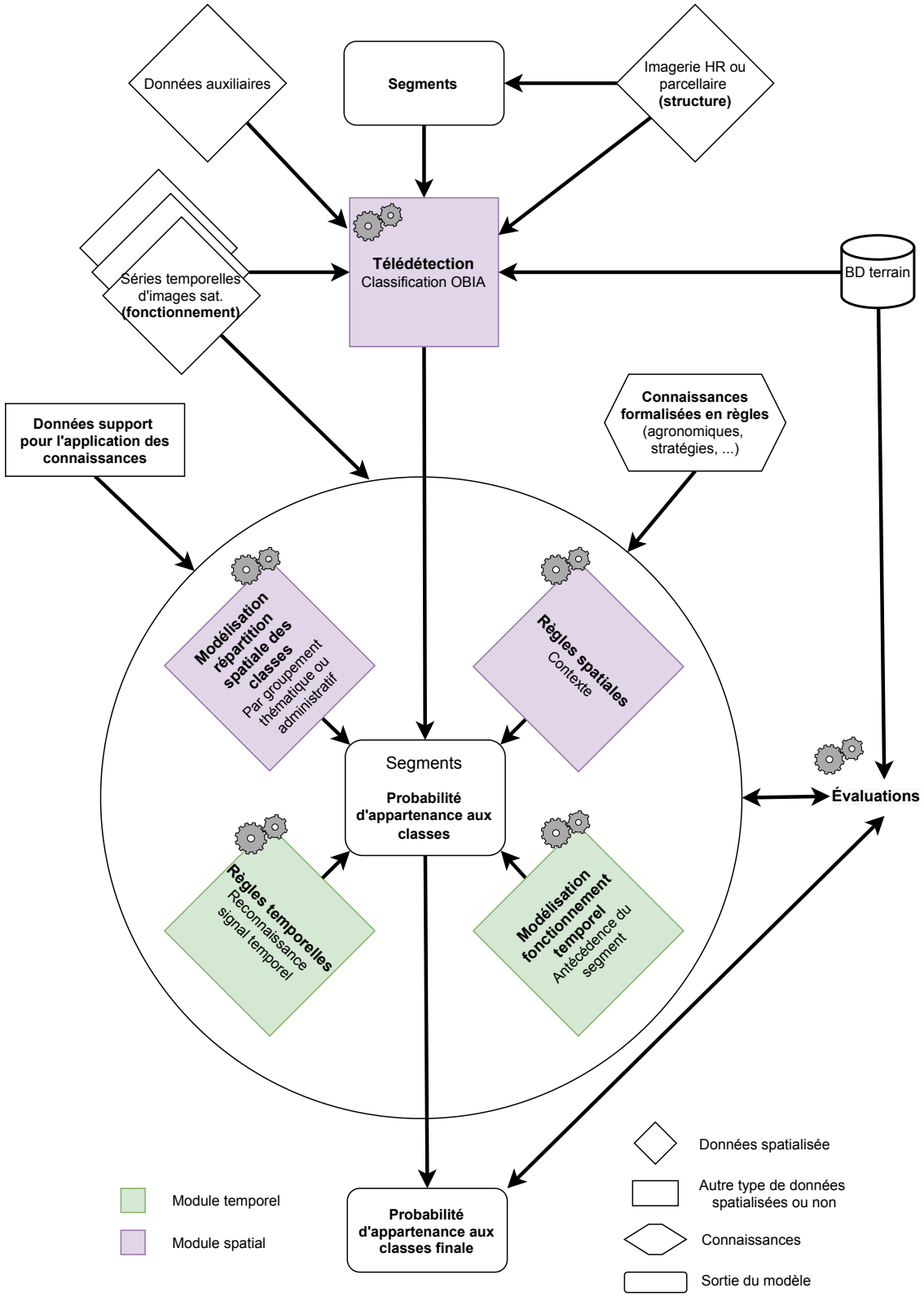


Figure 9 Schéma général du modèle conceptuel. Les différents types de données sont représentés, ainsi que les connaissances formalisées en règles et les modules. Le cercle central représente le coeur du modèle, avec les quatre modules qui le compose. Au centre du cercle, les segments représentent les objets étudiées, selon les modules ils peuvent être analysés seuls ou regroupés selon les différents niveaux d'organisation spatiale (Figure 7)

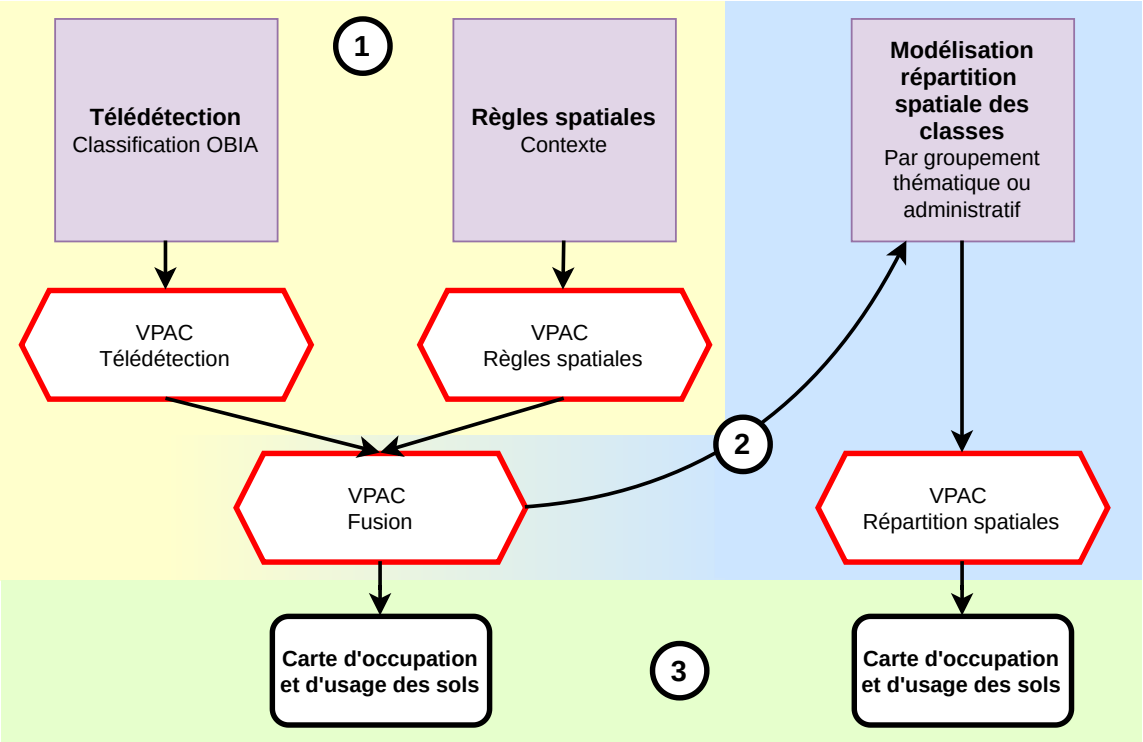


Figure 10 Ce schéma présente l'enchaînement des modules 0, 2 et 4 : **(1)** les module 0 : Télédétection et module 2: Règles spatiale produisent chacun un VPAC (Vecteur de Probabilité d'appartenance aux Classes) pour chaque objet et les VPAC télédétection et VPAC règles spatiales sont ensuite fusionnés pour produire un nouveau VPAC, **(2)**, ce VPAC fusion devient l'entrée du module 4: Modélisation de la répartition spatiale des classes qui permet de produire un nouveau VPAC représentant la répartition spatiale, **(3)** ce VPAC permet d'obtenir une carte d'occupation et d'usage des sols, tout comme le VPAC fusion.

3.2.2 Module 0 : Télédétection, classification supervisée

Pour obtenir une classification spatialement fiable, la classification de données de télédétection est actuellement la méthode automatique qui permet d'obtenir les meilleurs résultats de caractérisation spatiale d'un espace. L'approche objet (object-based image analysis (OBIA)) permet d'intégrer dans un objet l'intégralité des informations spatiales et temporelles nécessaires à la cartographie d'une zone d'étude. Ces informations peuvent être issues de séries temporelles d'images satellitaires, d'images mono-date, de données dérivées de ces images satellitaires (ex. textures, indices) ou encore des données auxiliaires (ex. MNT, pente). Les méthodes de classification à privilégier sont celles permettant de calculer la probabilité d'appartenance à chaque classe et d'obtenir une information sur la fiabilité de classification des objets. Le choix d'une approche supervisée permet d'évaluer les caractérisations tout au long des étapes, ainsi que de conserver un nombre fixe de classes. Ce module permet : 1) d'initialiser le modèle et de synthétiser les informations des données de télédétection dans un VPAC qui correspond aux résultats d'une première classification, et 2) attribuer à chaque objet les variables spectrales (réflectances, indices, textures) issues des images satellitaires utilisées pour cette classification. Cette classification des données de télédétection et le VPAC qui en résulte représentent l'**observation** de chaque classe dans un

objet.

3.2.3 Module 1 : modélisation du fonctionnement temporel (histoire des objets)

La modélisation temporelle permet de prendre en compte l'histoire de l'objet et l'évolution de son occupation ou usage des sols dans le temps. **L'objectif de ce module est de représenter les rotations les plus probables pour un objet, et de modéliser la dynamique de l'occupation et de l'usage des sols pluriannuelle**. L'analyse de son occupation ou usage passé permet de déterminer quelles peuvent être ses classes d'occupation ou d'usage des sols possibles pour la période étudiée. Cette approche a pour but d'identifier les objets pour lesquels la classe d'occupation ou d'usage des sols n'évolue pas dans le temps, et ceux pour lesquels il peut exister une logique de succession des classes (ex. dans le cas de cultures séquentielles ou de rotations culturales). Une fois la logique de succession identifiée pour un objet, des probabilités rassemblées dans VPAC peuvent être modélisées, représentant la possibilité qu'une classe soit présente ou non pour chaque objet. Ces probabilités peuvent signifier que la classe présente de l'objet soit similaire à ses occupations du sol passées, ou que cette classe soit dans une logique de succession (ex. rotations culturales types). Pour appliquer ce module, il est nécessaire de connaître l'antécédence des occupations et usages des sols mais aussi les rotations théoriques pour la zone d'étude afin de modéliser les probabilités. Les règles sont construites à partir de ces rotations théoriques. Pour l'agriculture familiale, ce module peut être appliqué à la caractérisation des espaces agricoles lorsque la temporalité d'application du modèle est annuelle (cultures séquentielles entre saison principale et contre-saison) ou pluriannuelle (rotations culturales).

En effet, dans un cas agricole, la connaissance des systèmes de culture passés permet de déterminer quelles sont les successions culturales possibles. Ce module intègre à la fois des connaissances des stratégies agronomiques pouvant être mises en place par les agriculteurs (itinéraire technique complet, successions culturales théoriques pour un espace défini, favorisation de champs au sein de l'exploitation), et des données permettant de traduire le fonctionnement temporel de chaque objet (telles que des cartes d'occupation et d'usage des sols sur plusieurs saisons ou années) afin d'analyser la variation phénologique de l'objet durant toute la période et de déterminer ses occupations du sol précédentes. La profondeur temporelle d'analyse peut varier, selon les connaissances disponibles et les besoins de la modélisation. Ce module permet d'ajouter des propriétés temporelles passées à chaque objet.

Durant ce module, l'historique des classes d'occupation du sol permet de contraindre à un nombre limité les occupations des sols possibles. Par exemple, en croisant des rotations culturales théoriques et les occupations des sols passées observées par l'imagerie satellitaire, il est possible de modéliser pour chaque type d'occupation des sols leur probabilité d'occurrence. L'ajout des contraintes biophysiques des objets permet de restreindre encore les possibilités de type de culture.

Ce module permet d'obtenir un VPAC synthétisant les possibilités d'occupation et d'usage des sols pour chaque objet en prenant en compte les rotations théorique et l'antécédence de l'occupation et de l'usage des sols.

3.2.4 Module 2 : Règles spatiales (contexte spatial des objets)

L'objectif de l'application de règles spatiales est d'identifier des contextes spatiaux favorables ou défavorables à chacun des systèmes de culture étudiés. L'idée est de modéliser la probabilité d'identifier un type de système de culture donné, puis une probabilité d'appartenance pour chaque classe synthétisée dans un VPAC. Les règles spatiales sont basées sur des connaissances formalisées qui permettent de mettre en lien les caractéristiques biophysiques des objets, les stratégies agronomiques des agriculteurs et les contextes nécessaires à l'implantation de à chaque type de culture. Les objets ne pouvant pas être agricoles sont déterminés de la même manière. Certaines règles appliquées aux propriétés peuvent être déterminantes pour rendre impossible la présence de certaines classes, par exemple un altitude trop importante pour la culture du riz pluvial. Ce sont les règles, appliquées aux propriétés de chaque objet, qui vont déterminer si le contexte de l'objet est favorable ou non à la classe. Plusieurs règles peuvent être mobilisées pour déterminer la probabilité d'appartenance d'un objet à une classe. Par exemple, dans le cas de la caractérisation d'un espace agricole, l'identification des objets non agricoles à l'aide des règles peut être relativement facile à mettre en place (contrainte de pente maximale cultivable, altitude). D'autres objets peuvent quant à eux être considérés comme pouvant convenir à plusieurs types d'occupation ou d'usage du sol, qu'ils soient agricoles ou non lorsque les règles mobilisées ne sont pas exclusives. L'ensemble des règles combinées permettent d'attribuer des probabilités de présence pour chaque classe existante regroupées dans un VPAC. L'application de certaines règles reste dépendante de données thématiques spatialisées, et la limite principale de l'application de ce module est la disponibilité et la qualité des ces données, ainsi que leur fraîcheur. Les propriétés et probabilités extraites de l'application de ces règles spatiales peuvent être utilisées dans les autres modules permettant de les contraindre, de confirmer ou infirmer certaines possibilités de classe d'occupation des sols.

3.2.5 Module 3 : Règles temporelles

Le module des règles temporelles est similaire dans son fonctionnement à celui des règles spatiales. Il s'agit d'analyser les propriétés temporelles (contexte temporel) des objets afin de pouvoir émettre des probabilités d'appartenance à certaines classes. L'objectif de ce module est d'analyser la temporalité des objets au sein de la saison culturale étudiée. Des calendriers culturaux permettent de déterminer des contraintes (ex. plages temporelles) pour le démarrage de certaines cultures ou pour leur période de récolte. Pour caractériser les objets avec ce module, il est nécessaire de mettre en lien les calendriers culturaux issus du terrain avec la croissance des plantes sur les objets étudiés (phénologie). L'analyse de l'évolution de la phénologie d'une plante peut être observée à partir de séries temporelles d'images de télédétection pour une saison culturale. De ces séries temporelle, des indices spectraux peuvent être calculés tels que le NDVI ou d'autres indices de végétation. Les données temporelles de précipitation (provenant d'images satellite comme TRMM - *Tropical Rainfall Measuring Mission* ou TAMSAT - *Tropical Applications of Meteorology using SATellite data and ground-based observations* par exemple, ou bien d'un réseau de pluviomètres) peuvent également aider à adapter les dates des calendriers culturaux pour la période étudiée et

pour un espace défini, tout en prenant en compte ses spécificités qui peuvent être des propriétés de l'objet ou qui ont pu être modélisées dans d'autres modules (orographie, systèmes de cultures de cet espace, stratégies via-à-vis du déroulement de la saison des pluies). La sortie de ce module est un VPAC synthétisant la discrimination temporelle des classes.

Ce module mobilise principalement des données de télédétection (séries temporelles d'images satellite) et des méthodes de classification permettant de différencier les signaux spectraux temporels. La majorité des classifieurs utilisés en télédétection n'utilisent pas la chronologie des données (*ie.* dans le module de classification, les réflectances ou indices spectraux utilisés dans la classification ne sont pas ordonnés temporellement en fonction des dates d'acquisition des images satellitaires dont ils sont issus). Ce module permet d'utiliser certaines méthodes de classification se basant uniquement sur la temporalité des propriétés des objets. Des méthodes de classification des signaux temporels peuvent être utilisées afin de déterminer des signaux temporels proches, identiques ou probables. Par exemple, des signaux temporels types peuvent être connus pour toutes les classes d'occupation des sols à partir des séries temporelles satellitaires. Le signal temporel de chaque objet peut être comparé aux signaux types pour attribuer la classe qui correspond le mieux à une signature temporelle type (en télédétection les travaux de Bellón et al. (2017) ou de Belgiu and Csillik (2018) utilisent des procédés similaires). De manière plus précise, des probabilités peuvent être attribuées selon la ressemblance du signal de l'objet au signal type de chaque classe. La limite principale de ce module est le degré de discrimination des signaux de référence entre les classes. Cela signifie que si deux signaux de référence (soit deux classes) sont temporellement proches, il sera difficile de les différencier, une probabilité similaire sera donc attribuée aux deux classes.

Ce module peut être couplé avec le module de modélisation du fonctionnement temporel afin d'utiliser à la fois les successions culturales théoriques et la classification d'un signal temporel de la phénologie sur plusieurs saisons culturales.

3.2.6 Module 4 : Modélisation de la répartition spatiale des classes

Le module de modélisation de la répartition spatiale des classes permet d'attribuer aux objets au sein d'un espace défini, qui correspond à un groupement d'objets, une classe qui est cohérente avec une répartition théorique des classes pour cet espace. Ce groupement peut être une unité ayant des stratégies communes (niveau 1 : exploitation, famille), ou bien être uniquement une unité qui permet de spatialiser des statistiques (niveau 2 : unité administrative par exemple). Ce module permet de modéliser les stratégies ou d'adapter les statistiques de répartition d'occupation ou d'usage du sol au niveau du groupement. Des données thématiques peuvent permettre de faire varier ces statistiques. Par exemple, si la saison culturale est peu pluvieuse, la part de cultures irriguées sera donc moins importante qu'une année normale pour un système agricole identifié. Ce module vise donc à redistribuer les classes selon les pratiques d'un groupement de niveau 1 ou 2 selon une distribution spatiale théorique des classes. Pour appliquer ce module, il est nécessaire de connaître la distribution théorique des classes, par exemple à partir de bases de données

statistiques ou d'enquêtes. La limite principale pour l'application de ce module est la qualité et la finesse des distributions (ex. représentation des classes de la nomenclature dans les bases de données).

La figure 11 présente le fonctionnement du module de modélisation de la répartition spatiale des objets dans une unité administrative. Les objets pouvant être réaffectés à une classe différente de celle qui leur a été affectée initialement (suite au module de télédétection ou aux autres modules) sont identifiés en fonction d'un seuil de fiabilité (probabilité élevée) qui est défini par analyse du VPAC. Sont considérés comme sûrs, les objets pour lesquelles la probabilité d'appartenance à l'une des classes est supérieure à un seuil. À l'issue de ce module, les objets étant considérés comme fiables auront donc peu de chance de voir leur classe changer. Les objets identifiés comme moins fiables auront une plus grande chance de voir leur classe être modifiée pour correspondre à une répartition spatiale cohérente. Si le niveau de fiabilité d'un objet est inférieur au seuil, alors l'objet sera étiqueté de manière temporaire comme « non fiable ». La répartition des cultures au niveau d'organisation villageoise est comparée avec les statistiques de répartitions théoriques pour déterminer quelles sont les classes déficitaires et les classes excédentaires. Au sein d'une classe excédentaire, les objets les moins fiables sont retirés et étiquetés « non fiables » jusqu'à ce que l'effectif théorique soit atteint. Les objets avec l'étiquette « non fiable » sont ensuite triés afin d'être ré-attribués à la classe dont ils sont les plus proches en termes de probabilité. Lorsqu'une classe atteint son quota, aucun autre objet ne peut être ajouté.

3.2.7 Attribution d'une classe finale en fonction des VPAC : probabilité et méthodes de fusion

Pour attribuer une classe d'occupation ou d'usage des sols à un objet, le modèle s'appuie sur les VPAC. Certains modules peuvent résulter en l'identification de contraintes immuables qui empêchent l'attribution de certaines classes (par exemple un contexte biophysique de sommet de haute montagne sur pente raide empêchera l'attribution de cultures).

Les cinq modules présentés précédemment peuvent produire des résultats différents qui doivent être combinés afin de prendre en compte l'intégralité des propriétés et classes possibles de chaque objet. Durant les modules ou à la fin de l'exécution du modèle, la classe finale de chaque objet peut être déterminée. Pour attribuer une classes parmi celles qui sont possible, le VPAC produit pour chaque module peut être conservé tel quel ou pondéré selon son importance, donnant un poids plus ou moins important à son résultat lors de son agrégation avec un VPAC issu d'un autre module.

Pour agréger les résultats, plusieurs méthodes existent. La fusion coopérative consiste à utiliser des données ou résultats comme paramètres de réglage d'une autre donnée ou d'un autre résultat afin d'obtenir un résultat final plus précis. Dans la littérature, cette méthode peut être utilisée à plusieurs niveaux, que ce soit pour la fusion de données optiques et radar (Giordano, 2015) ou pour segmenter une image (El Merabet, 2013). Dans notre cas, la classification des données de télédétection permet d'obtenir une exactitude spatiale et des résultats souvent acceptables. Les

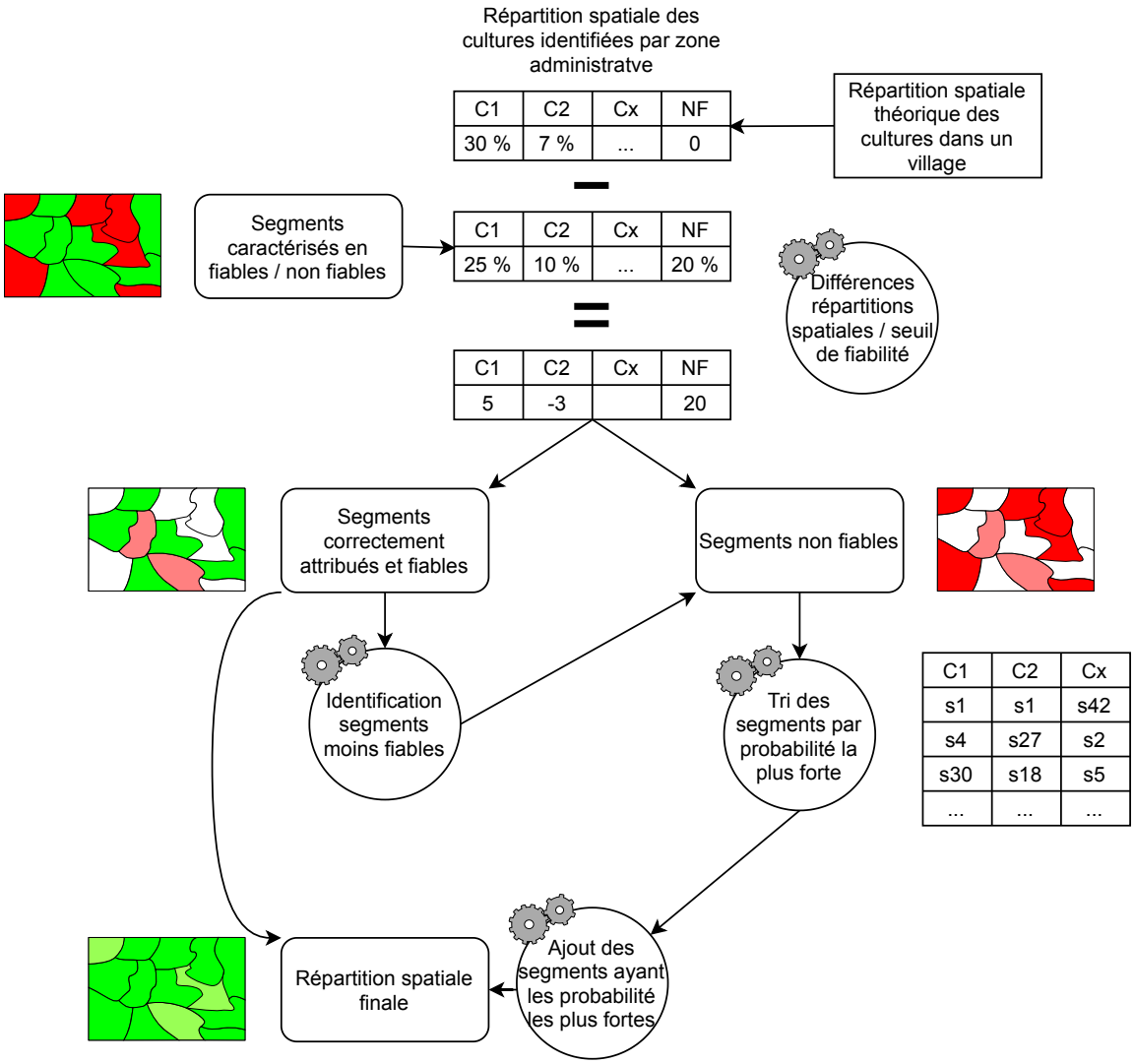


Figure 11 Schéma des étapes de la modélisation de la répartition spatiale des cultures.

approches de modélisation peuvent être utilisées comme paramètres de réglages du résultat de cette classification. **L'objectif de cette approche est d'obtenir une meilleure compréhension des objets que l'on cherche à caractériser et donc obtenir une information plus fiable**, afin de mieux classifier l'occupation ou l'usage des sols.

L'assemblage de ces résultats correspond à l'étape de **haut niveau**, c'est-à-dire l'étape qui vise à réunir les résultats de classifications pour décider de la classe à attribuer (Gamba and Chanussot, 2008). Dans le modèle, ces résultats de classification sont les décisions ou ensembles de probabilités (soit les VPAC) issus des différents modules de modélisation (modules 1 à 4) et de la classification des séries temporelles d'images satellite (module 0). Parmi les méthodes de fusion d'informations, on peut retenir la fusion par vote et la théorie des possibilités.

La fusion par vote permet de déterminer quelle est la classe à attribuer en effectuant un vote entre les résultats des classifieurs. Cette méthode est implicitement utilisée par les modèles de forêt aléatoire, avec un vote à la majorité parmi les classes choisies par chaque arbre pour un objet. C'est dans cette logique que les **méthodes ensemblistes** sont utilisées, en combinant plusieurs

classifieurs, puis en comparant l'ensemble des résultats afin de déterminer quelle classe est la plus probable (Kozak et al., 2017; Aguilar et al., 2018). Dans le cas où les votes entre deux classes sont égaux, c'est le vote majoritaire qui peut être choisi, ou la somme pondérée peut être calculée pour choisir la classe finale. Cette méthode est considérée comme présentant souvent de meilleurs résultats qu'en utilisant uniquement des classifieurs de manière indépendantes (Lam and Suen, 1997). Dans ce travail, nous avons choisi de nous appuyer sur cette méthode. Pour une fusion plus adaptative des VPAC, d'autres méthodes peuvent être utilisées telles que la théorie des possibilités (Zadeh, 1978), ou logique floue.

Une fois la classe finale déterminée, la combinaison des modules peut être évaluée et comparée aux données ou résultats de référence.

Les trois prochains chapitres présentent l'application du modèle conceptuel présenté dans ce chapitre aux Hautes Terres de la région du Vakinankaratra à Madagascar. Seuls les modèles spatiaux (le module 1 : modélisation du fonctionnement temporel et le module 4 : modélisation de la répartition spatiale des classes) sont implémentés. Le chapitre 3 Zone d'étude, matériel et méthodes présente le contexte agricole de la région du Vakinankaratra et la zone d'étude, les données mobilisées (données satellites, données de vérité terrain et données support pour l'application des connaissances), les outils utilisés pour cette thèse (Moringa et Ocelet) et les méthodes d'évaluation des résultats. Le chapitre 4 Méthode d'application des modules à la région des Hautes Terres comprend l'intégralité du déroulement de la classification des données de télédétection et l'application des deux modules spatiaux. Le chapitre 5 Résultats d'applications regroupe l'analyse quantitative et qualitative des résultats de la classification des données de télédétection mises en lien avec l'application du module 1 : règles spatiales, et une analyse succincte et qualitative de l'application du module de modélisation de la répartition spatiale des classes.

POINTS CLÉS DU CHAPITRE 2

- **OBJECTIF** :
  - proposer un **modèle conceptuel** qui combine à la fois **télédétection et modélisation spatio-temporelle** ;
  - proposer une nouvelle approche pour caractériser les systèmes complexes d'**agriculture familiale tropicale** de manière plus fiable et plus cohérente.
- **ÉCHELLES ET CONCEPTS** :
  - 3 principaux niveaux d'organisation spatiale de modélisation ont été définis : l'**objet** (champ), le **groupement de niveau 1** (agriculteur et exploitation) et le **groupement de niveau 2** (village ou groupe social) ;
  - l'**objet** (champ) est central et c'est à ce niveau que la classe d'occupation ou d'usage des sols est attribuée ;
  - 3 concepts principaux : l'approche **OBIA**, le **VPAC** et la **fiabilité** ;
  - différenciation entre l'**observation** d'une classe dans un objet, définie par l'utilisation de données d'observation de la Terre, et la **possibilité** de trouver une classe dans un objet, représenté par la modélisation spatio-temporelle des stratégies des agriculteurs.
- **ORGANISATION DU MODÈLE** :
  - la **classification des données de télédétection** permet d'obtenir un VPAC qui synthétise l'information des données de télédétection ;
  - 4 modules sont proposés pour modéliser et représenter les stratégies et pratiques des agriculteurs ;
  - le module 1 : **modélisation du fonctionnement temporel** simule la probabilité d'observer une classe, selon l'antécédence de l'objet (rotations culturales) ;
  - le module 2 : **règles spatiales** représente la possibilité de trouver une classe en fonction de l'emplacement de l'objet ;
  - le module 3 : **règles temporelles** mobilise à la fois des données de phénologie, de précipitation et les calendriers culturaux pour différencier les différents types de culture et favoriser leur discrimination avec la végétation naturelle ;
  - le module 4 : **modélisation de la répartition spatiale des classes** vise à modéliser les pratiques de groupe (village, groupe social) et à distribuer les cultures selon une répartition théorique dans un espace pour produire une cartographie plus cohérente de l'occupation et de l'usage des sols.

Chapitre 3

Application du modèle conceptuel : Zone d'étude, matériel et méthodes

Table des matières

1. Madagascar et les Hautes Terres
   1.1. Les enjeux de l'agriculture à Madagascar
   1.2. Le Vakinankaratra
   1.3. Zone d'étude
2. Données
   2.1. Données satellitaires et prétraitements
   2.2. Les données de terrain
   2.3. Données support pour l'application des connaissances
3. Outils
   3.1. Moringa
   3.2. Ocelet
4. Méthodes d'évaluation

Après avoir proposé un modèle conceptuel dans le deuxième chapitre, nous présentons dans ce chapitre les moyens mis en œuvre pour appliquer ce modèle aux Hautes Terres de la région du Vakinankaratra, à Madagascar. La première section décrit la situation agricole de Madagascar et ses enjeux, la région du Vakinankaratra qui est une zone de production agricole intensive, et les Hautes Terres où est appliqué le modèle. La deuxième section est consacrée aux données : images de télédétection (Pléiades, Sentinel-2 et Venμs), données collectées sur le terrain (occupation et usage des sols, entretiens et enquêtes), ainsi que données thématiques qui servent à supporter l'application des connaissances. La troisième section présente les deux outils utilisés pour l'ensemble des traitements et modélisations effectués dans cette thèse : la chaîne de traitement de données de télédétection Moringa et la plateforme de modélisation des structures et dynamiques spatiales Ocelet.

1 Madagascar et les Hautes Terres

Madagascar est un pays où l'agriculture familiale a une place majeure, aussi bien dans les zones rurales que dans les villes (Defrise et al., 2019) et a fait l'objet d'études centrées sur le fonctionnement de l'agriculture familiale (Bélières et al., 2013; Andriamanohy et al., 2016; Sourisseau et al., 2016; Raharison et al., 2018). Dans le Vakinakaratra est une région au centre de l'île qui est suivi depuis plusieurs années par le Dispositif de Recherche et d'Enseignement en Partenariat sur les Systèmes de Production d'Altitude et Durabilité (DP SPAD). Le DP SPAD conduit des études à la fois sur les Hautes Terres, espace montagneux, et sur le Moyen Ouest, plaine agropastorale. Dans les Hautes Terres, une zone de cultures intensives et variées (cultures pluviales et irriguées, arboriculture) est suivie depuis 2014 pour développer des méthodes de cartographie de l'usage des sols par télédétection.

Cette section est consacrée à un état des lieux de Madagascar et de ses enjeux à fois économiques, sociaux et environnementaux, puis à la région du Vakinankaratra et enfin à la description de la zone où le modèle conceptuel est appliqué.

1.1 Les enjeux de l'agriculture à Madagascar

Madagascar est la cinquième plus grande île du monde avec une superficie d'environ 587 000 km^2. Située à l'est du continent africain, l'île en est séparée par le canal du Mozambique et est bordée par l'océan Indien à l'Ouest. Le climat de Madagascar est caractérisé par deux saisons, une saison humide de novembre à avril, et une saison sèche de mai à octobre (Nassor and Jury, 1998). Cette situation crée une grande variété de climats, allant d'un climat chaud et humide à l'Est à un climat chaud et sec à l'Ouest, en passant par un climat tropical d'altitude au centre de l'île (Nassor and Jury, 1997).

Selon la FAO and PAM (2013), Madagascar figure parmi les 5 pays les plus vulnérables au changement climatique. Trois à quatre cyclones traversent tous les ans le pays avec comme porte d'entrée la côte est. Ces catastrophes naturelles entraînant destructions et inondations augmentent la vulnérabilité des populations, principalement rurales, et des infrastructures.

Madagascar est un pays en développement avec un Indice de Développement Humain de 0,52, ce qui place le pays au 162^{éme} rang parmi 189 pays (PNUD, 2019). Le pays souffre de défaillances structurelles notables qui ont pour conséquence une pauvreté importante de la population, un fort taux de malnutrition, des infrastructures qui se dégradent et peu de services de collecte de données nationales pour appuyer les politiques publiques.

Selon le rapport préliminaire du recensement de 2019, la population malgache est de 25 680 342 personnes avec 80,5 % de la population vivant en milieu rural (INSTAT, 2019) et environ 40,4 % de la population de moins de 15 ans (Nations, 2020). Madagascar fait partie des 8 pays au monde qui a vu une diminution du revenu par habitant entre 1960 et 2010, alors que le pays n'a pas connu de guerre ou conflit depuis la décolonisation (WB, 2015). Selon le seuil de pauvreté utilisé pour

les comparaisons internationales avec un calcul de parité de pouvoir d'achat de 1,9 $ par jour par personne, le taux de pauvreté s'élève à 77,6 % en 2012, et 97,3 % de la population vit avec moins de 5,5 $ par jour selon la Banque Mondiale[1], plaçant Madagascar comme le pays ayant l'un des plus forts taux de pauvreté mesurés. La pauvreté touche surtout les zones rurales où près de 80 % de la population vit sous le seuil de pauvreté (GBAD and INSTAT, 2013). De plus, l'actuelle épidémie de SARS-CoV-2 a un effet important sur l'économie malgache et augmente le niveau de pauvreté extrême qui était en baisse depuis 3 ans[2].

En 2010, 80,6 % des ménages ont une pratique agricole et 67,8 % sont considérés comme des ménages agricoles. Cette agriculture est principalement familiale, les petites exploitations de moins de 1,5 ha représentent 71,9 % des ménages agricoles, les moyennes exploitations (1,5 ha à 4 ha) constituent 23,3 % de ces ménages et les grandes exploitations 4,8 %. La superficie agricole moyenne par ménage est de 1,4 ha et la superficie médiane est de 1 ha sur l'ensemble de l'île. L'autoconsommation est le principal débouché de l'agriculture avec environ 75 % de la production vivrière des ménages ruraux auto-consommés (INSTAT, 2011).

Selon Burnod et al. (2016), la pression foncière est importante sur les terres arables et de bonnes qualités. La raison principale est la petite taille des exploitations qui diminue de génération en génération avec un morcellement et une redistribution des exploitations agricoles. Cette fragmentation par l'héritage limite l'optimisation par le travail familial ou les économies d'échelle. Cet héritage est de plus en plus limité, car les familles sont nombreuses et, dans certaines régions, aucune nouvelle terre agricole n'est disponible. Les marchés fonciers (location ou achat de terres) ou la migration vers d'autres régions sont donc une solution pour les personnes sans terres ou avec trop peu de terres pour vivre. Ces questions d'accès à la terre sont au cœur des problématiques actuelles, tout comme l'est l'enjeu de la sécurisation de l'exploitation face aux vols (Sourisseau et al., 2016).

Cette pauvreté structurelle entraîne une diminution des infrastructures. En 2014, le réseau routier malgache était estimé à hauteur de 32 000 km dont 13 % bitumés et 8 % considérés comme en bon état. La dégradation du réseau routier est estimée à environ 1 000 km/an pour la période 1970-2000[3]. Cette dégradation est liée au passage des véhicules, aux cyclones et surtout au manque d'entretien des routes et pistes qui doit être effectué par la population locale. Cette faible infrastructure enclave de plus en plus les zones rurales qui ne peuvent plus écouler leur production et bénéficier d'un approvisionnement correct.

[1] https://databank.worldbank.org/data/download/poverty/33EF03BB-9722-4AE2-ABC7-AA2972D68AFE/Global_POVEQ_MDG.pdf

[2] « Les perturbations dans les échanges et les voyages internationaux ainsi que les mesures de confinement décrétées dans le pays devraient provoquer un tassement très net de l'activité, avec une chute attendue du PIB à 1,2 %, très en deçà des prévisions d'avant la crise, qui tablaient sur un rythme de 5,2 %. Dans ce contexte, les populations vulnérables dans les zones urbaines seront particulièrement exposées aux difficultés économiques et aux pièges de la pauvreté. L'effondrement des recettes fiscales et les dépenses liées à la crise sanitaire vont peser sur le déficit budgétaire, provoquant une hausse subite des besoins de financement. » https://www.banquemondiale.org/fr/country/madagascar/overview, consulté le 10/12/2020. « Dans ce contexte, l'extrême pauvreté devrait augmenter en 2020, annulant trois années de baisses consécutives » (WB, 2020)

[3] http://documents1.worldbank.org/curated/en/130911468086644653/pdf/883230REPLACEM00Box385221B00PUBLIC0.pdf

La pauvreté, les problèmes fonciers ou la diminution progressive d'infrastructures fiables sont des déterminants structurels de l'insécurité alimentaire à Madagascar qui ne fait qu'augmenter. À ceux-ci s'ajoutent des facteurs conjoncturels tels que le changement climatique ou la hausse constante du prix du riz qui sont plus difficiles à analyser par manque de données. La climatologie de Madagascar est assez peu étudiée en raison d'un réseau de stations météorologiques au sol peu dense (Macron et al., 2016). Durant la saison des pluies, la forte variabilité des précipitations entraîne régulièrement des inondations ou des retards de la saison des pluies et des sécheresses (Macron et al., 2016). Les températures de l'île risquent d'augmenter de 1,1°C à 2,6°C avec une augmentation des précipitations de janvier à avril et une raréfaction de celles-ci durant l'été à l'horizon 2050 (Tadross et al., 2008), ne faisant qu'augmenter les incertitudes vis-à-vis de l'amélioration de sécurité alimentaire. Les hausses périodiques du prix du riz, principalement durant les périodes de soudure, sont particulièrement imprévisibles, car elles dépendent de plusieurs facteurs ; à titre d'exemple, en 2004 le prix du riz a augmenté de 92 % sur l'année en raison de deux cyclones, de mauvaises estimations de rendements, de la dépréciation de la monnaie nationale et de la hausse des cours mondiaux du riz (Dabat et al., 2006). La hausse des prix en 2010-2011 était aussi liée à des facteurs structurels tels que des surestimations de rendements et des stocks de riz mal estimés (David-Benz et al., 2014). Une stabilisation des prix pour le marché intérieur n'est pas à prévoir à court et moyen termes, car ni les estimations de production agricole ni celles des stocks ne sont fiables, et les choix politiques ne sont pas transparents ce qui entraîne une méfiance entre les acteurs privés et l'État, et limite l'efficacité de la gestion des crises alimentaires (David-Benz et al., 2014). Ces faiblesses sont en partie compensées en partie par les systèmes internationaux d'alerte précoce (FEWS NET[4], ASP (Rembold et al., 2019), le PAM) qui ont été mis en place afin d'anticiper de possibles crises alimentaires.

Les principales cultures de l'île sont le riz, le manioc, le maïs et la pomme de terre, et d'autres produits comme le haricot et la brède. Pour les exploitations familiales, ces productions sont commercialisées à hauteur de 25 % de leur production vivrière, et représentent 45 % de la production agricole industrielle (arachide et canne à sucre) et plus de 90 % de la production de café, girofle et vanille (INSTAT, 2011). Ces productions sont vendues sur les marchés nationaux et internationaux (Losch et al., 2013).

1.2 Le Vakinankaratra

La région du Vakinankaratra est située au centre de l'île et a une superficie d'environ 19 250 km^2. L'altitude de la région varie de 240 mètres à 2570 mètres, avec la chaîne de l'Ankaratra située en son centre, et à l'Ouest, la plaine nommée Moyen Ouest où le relief disparaît pour laisser place à une pénéplaine à partir d'environ 1200 mètres (figure 12). Cette région est principalement agropastorale. Le Vakinankaratra est la seconde région la plus peuplée de l'île avec 8,1 % de la population et 117 hab/km² (INSTAT, 2019). Le climat des Hautes Terres, dont fait partie le Vakinankaratra est caractérisé par un climat tropical d'altitude, avec des pluies abondantes en été et un hiver froid,

[4]https://fews.net/fr/southern-africa/madagascar

les températures atteignant les 0°C.

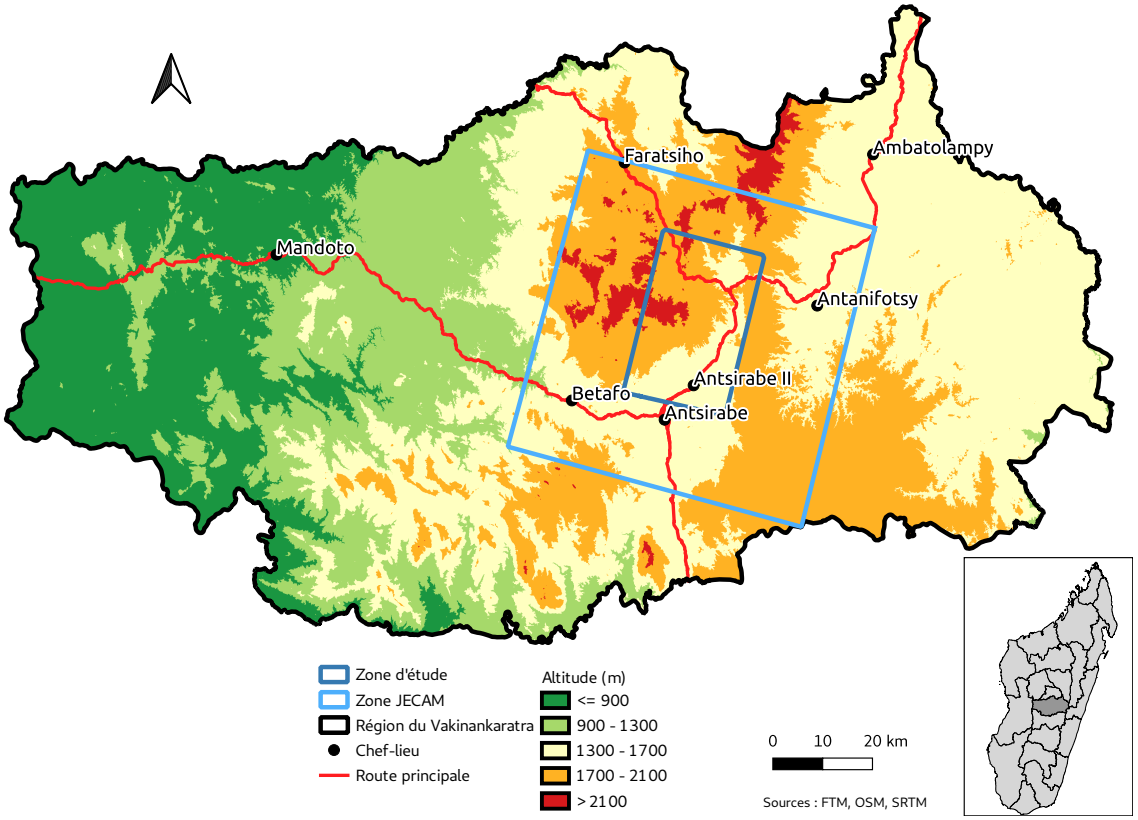


Figure 12 Carte de la région du Vakinakaratra, et localisation du site JECAM ainsi que de la zone d'étude.

Du fait de son climat tropical, la zone d'étude présente deux saisons : la saison des pluies (saison culturale principale), de novembre à avril, apportant la majorité des précipitations annuelles qui varient entre 1 200 mm et 2 000 mm selon l'altitude et l'exposition, et la saison sèche (contre-saison) de mai à octobre. En saison des pluies, lorsque des cyclones entrent sur le territoire par la côte Est, des dégâts importants peuvent être causés dans les bas-fonds par des inondations, et sur les zones pentues non boisées par une forte érosion. La couverture forestière est très faible dans le Vakinankaratra, environ 0,5 % du territoire (ONE, 2013) ; le déboisement des collines est lié à la pression foncière qui pousse les agriculteurs à cultiver sur des pentes de plus en plus importantes et à déboiser pour obtenir du bois de chauffe et de cuisson. Actuellement, les versants les plus exposés peuvent être reboisés par les agriculteurs pour stabiliser les versants et assurer des ressources énergétiques aux ménages.

D'après FewsNet (2013), les moyens d'existence au Vakinankaratra peuvent être classés en deux zones distinctes : 1) le Moyen Ouest agropastoral (riz, maïs, légumineuses), et 2) l'Ankaratra (cultures vivrières, horticulture, lait).

- dans le Moyen Ouest agropastoral, les cultures dominantes sont le riz, le maïs et les légumineuses, avec des sols majoritairement ferrallitiques et alluviaux. L'agriculture est centrée sur la production de riz et les autres productions agricoles ont pour but de garantir l'alimentation des ménages lors des périodes de soudure. La traction animale est au cœur

de l'agriculture, mais l'augmentation des vols de zébus depuis 2010 entraîne des difficultés de production pour ces ménages. La culture de légumes comme culture de rente se développe progressivement le long de la route nationale reliant Antsirabe à la côte Ouest ;

- l'Ankaratra est situé dans la zone de montagne des Hautes Terres, avec une densité de population très importante (plus de 500 hab/km²), des sols de qualité et des précipitations suffisantes. Ces conditions font de cet espace l'une des zones de production agricole les plus intensives de l'île. Un large panel de cultures est présent, dont des cultures tempérées comme les pommiers. La production de pommes de terre est développée, en faisant l'une des plus importantes du pays. L'élevage est aussi développé, principalement en stabulation, pour la production laitière. Cette zone est dominée par l'agriculture vivrière, où la riziculture est dominante. La densité de population a pour effet de faibles superficies cultivées par ménage, entre 0,1 et 0,5 ha en moyenne, les ménages les plus aisés pouvant avoir jusqu'à 5 ha de terre.

Au Vakinankaratra, la superficie économique[5] moyenne est de 0,8 ha, la superficie médiane de 0,4 ha, la part d'exploitation avec moins de 1,5 ha est de 84 %, celle de moyennes exploitations de 12,7 % et celles de plus de 4 ha représentent 3,3 % des exploitations (INSTAT, 2011). Selon le dernier recensement agricole de 2004-2005, la taille moyenne des exploitations agricoles est de 0.55 ha ce qui, ramené à la superficie moyenne par personne, correspond à 10-15 ares selon les districts. 23 % des exploitations agricoles ont moins de 25 ares et 55 % ont moins de 50 ares (Sourisseau et al., 2016).

Afin de faire face à certains aléas comme les variations du prix du riz, des politiques agricoles ont été mises en place afin d'encourager la production de pommes de terre et de riz et d'en faire une denrée d'exportation. Ces politiques ont favorisé l'utilisation d'intrants et le développement des cultures en contre-saison (durant la saison sèche) et du riz pluvial qui peut être cultivé sur tout type de sol. Le Vakinankaratra est théoriquement autosuffisant en riz, qui occupe environ 60 % des surfaces cultivées de la région. La patate douce, le manioc et la pomme de terre sont cultivés sur une superficie estimée à près de 60 000 ha en 2013/2014. Le maïs, les haricots et les arachides sont aussi largement cultivés en saison des pluies, comme le blé et l'orge en contre-saison. Le Vakinankaratra est la principale région productrice de fruits et légumes de l'île (Sourisseau et al., 2016). L'un des principaux systèmes de culture dans les Hautes Terres repose sur les associations de cultures afin d'éviter les mauvaises herbes dans les champs où presque l'intégralité du travail est manuelle. L'association de cultures peut permettre la synergie entre des plantes ayant des besoins différents, que ce soit en termes d'ensoleillement, de longueur des cycles, de système racinaire ou de besoins en nutriments (Autfray et al., 2014).

Le Vakinankaratra est desservi par des routes nationales bitumées, reliant Antsirabe, le chef-lieu régional, aux autres régions. Ces routes permettent le transport de marchandises en grande quantité. Les autres routes sont composées de pistes, majoritairement inaccessibles pendant la saison des pluies. Sur l'ensemble de la région, environ 20 % des chefs-lieux sont accessibles en

[5] La superficie économique est la superficie cultivée toute l'année en prenant en compte les différentes saisons culturales. Si un ménage cultive 20 % de ses parcelles en contre-saison, et 100 % en saison principale, alors la superficie économique sera de 120 %, soit 20 % de la superficie physique portant deux cultures dans l'année (Sourisseau et al., 2016).

camion ou 4x4 (CREMA, 2013).

Fonctionnement du maraîchage

Le maraîchage pour l'autoconsommation ne concerne que des petites surfaces de 1 à 10 ares, surtout en contre-saison. Le maraîchage durant la saison des pluies est essentiellement consacré à la vente. La production de légumes destinée à la vente est effectuée en étroite relation avec les organisations paysannes et différents acteurs de l'écoulement des denrées. Cette production « professionnelle » est, elle, essentiellement pratiquée lors de la saison des pluies. Ces exploitations varient entre 10-30 ares pour les petits paysans et de 30 ares à quelques hectares pour les grandes exploitations avec une intégration complète aux réseaux de marché. Les exploitations cultivant des légumes ne représentent que 10 % des exploitations avec 5 ares de légumes par exploitation en moyenne, les légumes étant très sensibles aux maladies. Le parcours des légumes vendus est complexe en raison du grand nombre d'intermédiaires (collecteurs, grossistes, transporteurs, détaillants ...) avec un marché principal à Antsirabe et plusieurs marchés secondaires dans la région (Sourisseau et al., 2016).

1.3 Zone d'étude

Ce travail de thèse s'est focalisé sur une sous partie du site Joint Experiment for Crop Assessment and Monitoring (JECAM) d'Antsirabe qui est situé dans les Hautes Terres du Vakinankaratra, correspondant à la zone de l'Ankaratra. Le site d'Antsirabe (19,43°S / 47.04°E) est une zone de cultures intensives et variées couvant 620 km² (20 x 31 km). Environ 6 communes composent le site et 77 Fokontany. Le site est traversé par la route nationale 7 reliant Antananarivo et Antsirabe et comporte des montagnes à l'Est et à l'Ouest Fokontany. Le site est traversé par la route nationale 7 reliant Antananarivo et Antsirabe et comporte des montagnes à l'Est et à l'Ouest (Figure 12). Le paysage est hétérogène, avec de petites parcelles, souvent séparées par de la végétation naturelle et une importante diversité agricole avec des cultures de légumes, de fruits ou de maïs et une dominance du riz irrigué. C'est le paysage agricole typique de l'Ankaratra décrit en section 1.2 Le Vakinankaratra (photographies figure 13).

Ce site est suivi depuis 2016 pour la caractérisation de l'occupation du sol par imagerie satellitaire (Lebourgeois et al., 2017), mais aussi pour l'estimation du rendement de riz pluvial et irrigué.

Pour mettre en place la méthodologie développée durant ce doctorat, nous nous sommes focalisés sur une zone d'étude plus restreinte qui correspond à l'emprise des images acquises par le satellite Venμs (section 2.1.2.2 Venμs).

Dans cette section, la situation de Madagascar, pays structuré autour de l'agriculture familiale, a été présentée ainsi que la zone d'étude qui se situe dans les Hautes Terres de la région du Vakinankaratra.



(a) Colline érodée anciennement cultivée. Des champs de maïs subsistent en bas de pente.



(b) Champ composé de pommier, de riz pluvial, de soja et d'une bordure en maïs.



(c) Mosaïque de champs de légumes, cultures pluviales et vergers. Au premier plan du maïs sèche sur pied pendant que les carottes sont cultivées.



(d) Récolte d'une rizière irriguée en faible pente. La taille du champ est représentative de la majorité des champs d'exploitations familiales.

Figure 13 Photographies de la zone d'étude, prises dans les communes d'Alakamisy, de Soanindrariny, d'Ambano et de Tritriva.

2 Données

Trois types de données ont été utilisées pour caractériser l'occupation des sols de la zone d'étude : 1) les données spatiales d'observation de la Terre, qui peuvent être divisées entre les images monodate à très haute résolution spatiale et les séries temporelles d'images à haute résolution spatiale, 2) les données de terrain, que ce soit des relevés *in situ* d'occupation et d'usage des sols ou des entretiens et enquêtes, et 3) les données thématiques, telles que l'altitude, le type de relief ou le temps de trajet d'un bâtiment à un objet, qui servent à l'application de règles spatio-temporelles sur cet espace agricole. Les caractéristiques et les prétraitements appliqués sur ces données sont décrits dans les sous-sections suivantes. Les données spatialisées sont en cordonnées géographique World Geodetic System 1984 (WGS 84) zone 38S. Le tableau 2 présente synthétiquement l'ensemble des données utilisées.

Données	**Échelle et unité spatiales**	**Fréquence et période d'acquisition**
Données satellites		
Pléiades	0,5 m x 0,5 m	Monodate - 01/02/2019
Venμs	5 m x 5 m	2 jours - 01/09/2018 - 30/06/2019
Sentinel 2	10 m x 10 m	5 jours - 01/09/2018 - 30/06/2019
Données de terrain		
Relevés *in situ* d'occupation du sol	Ponctuels et polygones	Saison culturale principale 2017-2018 et 2018-2019
28 enquêtes auprès des agriculteurs	Ponctuels	Saison culturale principale 2017-2018
Enquêtes auprès des responsables de 12 communes et Fokontany	Ponctuels	Campagne de terrain 2018-2019
Données thématiques		
Altitude	1 arc seconde x 1 arc seconde	Unique
Pente	1 arc seconde x 1 arc seconde	Unique
Type de relief	1 arc seconde x 1 arc seconde	Unique
Réseau hydrographique	1 arc seconde x 1 arc seconde	Unique
Pédo-géomorphologique	Madagascar (échelle nationale)	Unique
Communes	Madagascar	Unique
Temps de trajet entre un bâtiment et un objet	Zone d'étude, 0,5 x 0,5 m	2018-2019

Table 2 Synthèse des données utilisées.

2.1 Données satellitaires et prétraitements

La complexité spatiale et temporelle de la zone d'étude nécessite d'avoir de l'information satellite dans ces deux dimensions. Or, il n'existe pas de système d'observation de la terre qui offre une fine résolution spatiale et temporelle (Tonneau et al., 2020), ce qui nous oblige à utiliser différents types de satellites. Cette sous-section présente l'image THRS Pléiades, puis les séries temporelles des satellites Sentinel-2 et Venμs (Figure 14).

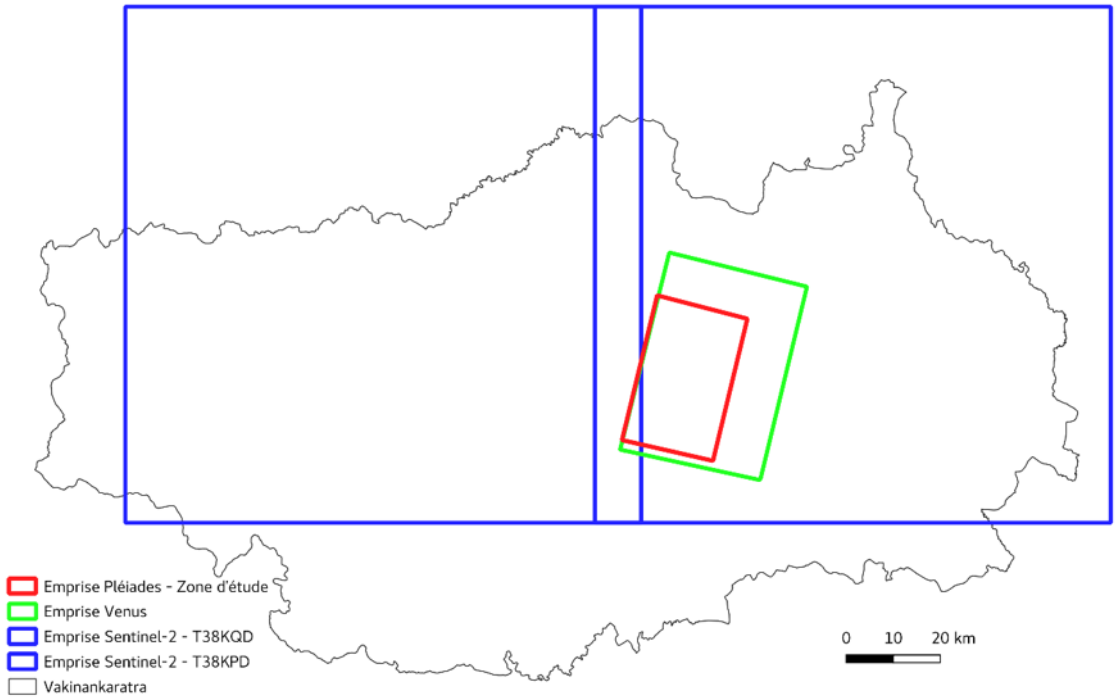


Figure 14 Fauchée des différentes images satellitaires utilisées.

2.1.1 Image mono-date annuelle à très haute résolution spatiale : Pléiades

Les données à très haute résolution spatiale servent d'appui pour les relevés d'occupation des sols sur le terrain (via une visualisation sur des tablettes de terrain adaptées) et à déterminer la structure spatiale de la zone d'étude (contours des objets à caractériser) lors des traitements. Ces images sont mono-date car elles sont disponibles uniquement via une programmation du satellite pour une période donnée. L'acquisition de l'image a été programmée au pic de la saison culturale principale avec, pour contrainte, un minimum de nuages.

Les images Pléiades livrées ont une résolution spatiale de 0,5 m en panchromatique et de 2 m en multispectral (4 bandes), avec une fauchée de 20 km au nadir (Annexe A). Pour la campagne 2018-2019, l'image Pléiades a été acquise le 01/02/2019 (Figure 15a).

2.1.2 Les séries temporelles à haute résolution spatiale

Les séries temporelles d'images satellitaires ont pour objectif de caractériser le fonctionnement des objets (phénologie, usage des sols, changement d'occupation des sols). Dans le cadre de cette étude, nous avons utilisé deux systèmes satellitaires qui fournissent des images gratuites à résolution décamétrique (entre 5m et 10 m), permettant ainsi d'augmenter les chances d'obtenir des images dégagées : Sentinel-2 et Venμs. Ces acquisitions ont été faites pendant la saison culturale principale 2018-2019.

2.1.2.1 Sentinel-2

La mission Sentinel-2 est une constellation de deux satellites lancés en 2015 et 2017 par European Space Agency (ESA) qui fournit des données optiques (Annexe A). Le choix des données Sentinel-2 a été motivé par plusieurs raisons :

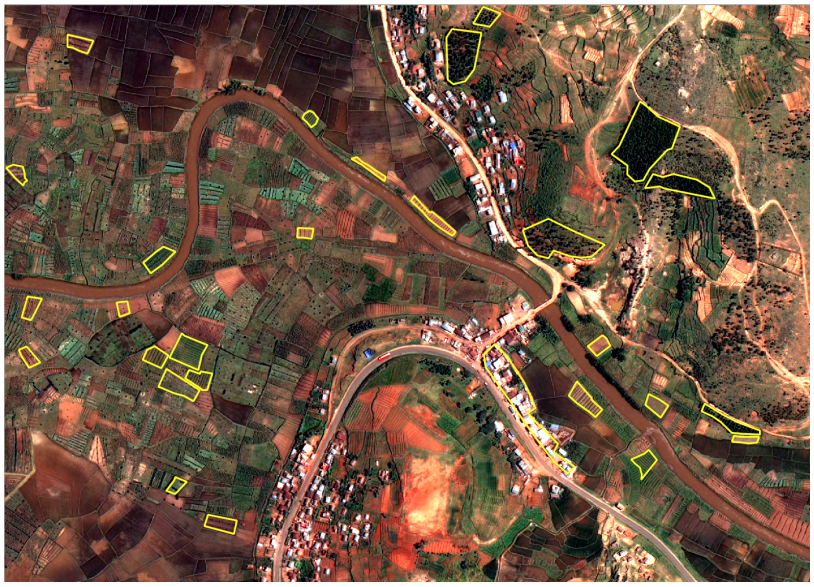
- haute fréquence d'acquisition des données avec un temps de revisite de 5 jours ;
- continuité de la mission avec la préparation de deux nouveaux satellites qui seront mis en orbite lors de la fin de vie théorique de 7 ans de Sentinel-2 A et B ;
- résolution spatiale de 10 m dans le visible et le PIR ;
- résolution spatiale de 20 m dans le moyen infrarouge (MIR)
- accessibilité des données en temps quasi réel ;
- images orthorectifiées et corrigées en réflectance de surface après correction atmosphérique, avec un masque des nuages et de leurs ombres ;
- gratuité de la donnée.

Les produits sont directement accessibles depuis le site de Theia. Elles sont fournies par le CNES au niveau L2A (Donadieu and L'Helguen, 2016), après pré-traitement avec la chaîne MAJA (Hagolle et al., 2017) afin de corriger les effets atmosphériques (Figure 15c).

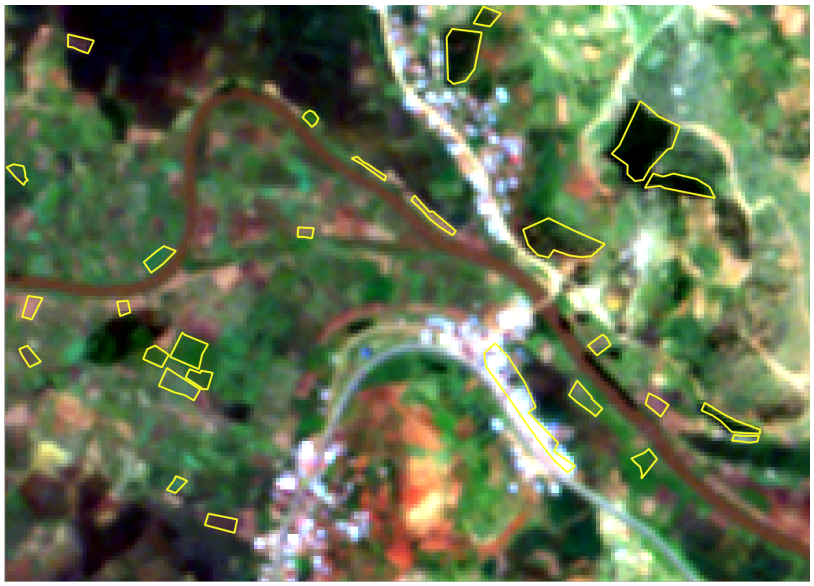
2.1.2.2 Venμs

Venμs est un satellite lancé en 2017 par le CNES et l'Isreal Space Agency (ISA). L'un des objectifs primaires de cette mission satellite à visée scientifique est de fournir des mesures des surfaces continentales afin de préparer de futurs satellites. La fauchée de ce satellite est d'environ 27 km avec une résolution spatiale d'environ 10 m, puis retraités à 5 m (Figure 15b). Les images sont fournies dans 12 bandes spectrales étroites (largeur de 16 à 40 mm sur un spectre allant de 400 nm à 920 nm) (Annexe A). Tout comme Sentinel-2, les données sont disponibles sur le site de Theia et disposent des mêmes traitements L2A que Sentinel-2. Les images acquises par Venμs sont intéressantes à utiliser, car :

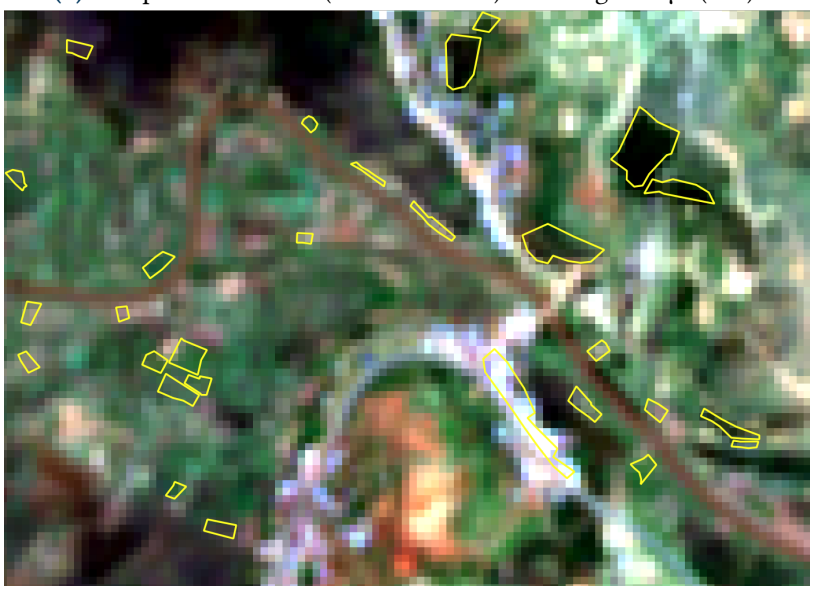
- leur résolution spatiale est deux fois plus fine que celle des images Sentinel-2 ;
- la fréquence d'acquisition est 2 jours ;



(a) Composition colorée (vraies couleurs) de l'image Pléiades (0,5 m).



(b) Composition colorée (vraies couleurs) de l'image Venμs (5 m).



(c) Composition colorée (vraies couleurs) de l'image Sentinel-2 (10 m).

Figure 15 Extraits d'images satellitaires Pléiades (acquise le 01/02/2019), Venμs (acquise le 02/02/2019) et Sentinel-2 (acquise le 03/02/2019). En jaune sont représentés les contours des relevés d'occupation et d'usage des sols effectués sur le terrain.

- elles sont prétraitées avec les mêmes algorithmes et au même niveau que les images Sentinel-2.

Les limites principales de cette mission sont liées à son caractère de démonstration scientifique, soit des acquisitions d'images sur un nombre limité de sites prédéfinis (entre 100 et 200 pour l'ensemble du globe) et de faible emprise spatiale, et pour une durée d'exploitation de mission de 2,5 ans.

L'annexe A résume l'ensemble des caractéristiques des satellites utilisés.

2.2 Les données de terrain

Les données collectées sur le terrain sont destinées à deux objectifs : 1) la création d'une base de données de vérité terrain servant à la fois à l'apprentissage du classifieur de données de télédétection et à la validation des résultats de cette classification ainsi que de l'application du module 2 : règles temporelles ; et 2) l'identification de connaissances pour modéliser les stratégies et pratiques des agriculteurs à travers entretiens et des enquêtes de terrain et une recherche bibliographique. Deux campagnes de terrain ont permis la collecte de ces données durant les saisons culturales principales 2017-2018 et 2018-2019.

2.2.1 Relevés *in-situ* d'occupation et d'usage des sols

La campagne de terrain de 2018-2019 a permis la constitution de bases de données géoréférencées de vérité terrain sur l'occupation des sols pour la saison culturale principale. Ces relevés ont été effectués au cœur de la saison des pluies, de janvier à mai 2019, pour identifier facilement les cultures en place.

Les relevés de terrain effectués lors de la saison culturale 2018-2019 s'appuient sur le guide de relevés *in situ* Joint Experiment for Crop Assessment and Monitoring (JECAM)[6], associant de bonnes pratiques de relevés et une nomenclature générique sur l'occupation des sols, dédiée à la caractérisation des espaces agricoles. Cette nomenclature est hiérarchique et agrégative avec cinq niveaux allant des grands groupes d'occupations des sols (Land cover) jusqu'aux et pratiques culturales (Cropping practices). Cette nomenclature a été conçue pour comparer différents sites d'études situés dans plusieurs endroits du monde ; elle a toutefois dû être adaptée aux spécificités locales de notre site d'étude (Bontemps et al., 2015; Lebourgeois et al., 2017).

Travaillant spécifiquement sur une zone tropicale d'agriculture familiale, la nomenclature JECAM a été adaptée afin de mieux correspondre à nos problématiques de recherche. Cette adaptation s'est faite dans deux directions : le regroupement de certaines cultures selon des critères propres à chaque groupe (les pratiques culturales par exemple, voir ci-après), et la séparation du riz en deux classes (riz pluvial et riz irrigué).

Concernant les regroupements :

[6] http://jecam.org/wp-content/uploads/2018/10/JECAM_Guidelines_for_Field_Data_Collection_v1_0.pdf

- la classe « légumes » regroupe les principales cultures maraîchères qui sont cultivées dans des conditions similaires (choux, salades, autres légumes à feuille et tige, aubergines, tomates, concombres, autres légumes et fruits, carottes, ail, oignons ou échalotes, autres légumes à racine ou bulbes) ;
- la classe « vergers » a été créée selon les mêmes critères et regroupe tous les types d'arbres fruitiers ;
- la classe « cultures potagères » regroupe les cultures principalement destinées à l'autoconsommation, qui sont cultivées sur de très petites surfaces et qui suivent des stratégies agricoles difficilement identifiables à l'échelle de l'exploitation. Les cultures de cette classe sont le manioc, les haricots, les patates douces et les autres cultures oléagineuses ou légumineuses.

Ces choix de regroupement ont été effectués sur la base de discussions avec les agriculteurs et les experts locaux, et ont également permis de pallier certains problèmes d'échantillonnage tels que les effectifs réduits de relevés d'occupation du sol pour les classes minoritaires.

La séparation entre le riz pluvial et le riz irrigué est faite en raison de l'importance de cette culture dans cette région du monde et de la signature spectrale très différente de ces deux modes de conduite, du fait de la présence d'une lame d'eau sous-jacente dans les parcelles de riz irrigué.

Les relevés de terrain ont mis en évidence deux types d'associations principales : le maïs associé à d'autres cultures, principalement des légumineuses, qui est l'association la plus courante, et le riz pluvial associé à d'autres cultures (principalement du maïs) qui est également très répandu. Les associations de culture avec du maïs ou du riz comme culture dominante étaient labellisées sur le terrain, *Maïs associé à une autre culture* et *Riz pluvial associé à une autre culture* respectivement[7].

Finalement, à son niveau le plus détaillé, notre nomenclature compte 17 classes, dans lesquelles sont répartis les 2542 relevés d'occupation et d'usage des sols et polygones photo-interprétés dont 12 classes agricoles (comprenant également une classe de *sol nu agricole*, correspondant aux parcelles rencontrées sur le terrain dans le domaine cultivé, faisant l'objet d'une préparation, mais sur lesquelles aucune culture n'était encore identifiable). L'ensemble des classes est présenté en table 3.

2.2.2 Méthode d'échantillonnage

La stratégie d'échantillonnage visait à parcourir les parties accessibles de la zone d'étude afin de prendre en compte l'importance et la diversité des espèces cultivées autour d'Ansirabe durant la saison culturale principale d'octobre à mai. Les espaces dédiés à certaines pratiques ont été ciblés de manière spécifique. Le travail de terrain a été principalement effectué par Eloise Rasoamalala en 2018 et par Valérie Andriamanga en 2019, toutes deux ingénieures agronomes et chargées des traductions malgache-français. Durant la saison des pluies, une partie des routes secondaires n'est pas praticable, car en trop mauvais état. L'accès aux zones reculées doit se faire à pied ce qui limite les surfaces prospectées. Les mesures GPS ont ainsi été contraintes par l'accessibilité, obligeant

[7] Si aucune de ces deux cultures n'était la principale, alors la culture n'a pas été relevée, sauf si les effectifs ce la classe étaient trop faibles.

Table 3 Répartition de l'ensemble des relevés de terrain par classe pour la saison culturale 2018-2019 (polygones > 200m²).

Occupation du sol (français)	Occupation du sol (anglais)	Code	Effectif par classe
Maïs	Maize	1	75
Riz irrigué	Irrigated rice	2	256
Riz pluvial	Rainfed rice	3	47
Légumes	Vegetables	6	474
Soja	Soya beans	7	76
Pomme de terre	Potatœs	10	253
Herbes et autres cultures fourragères	Grasses and other fodder crops	21	59
Verger	Fruit crops	23	154
Culture boisée de rente	Cash woody crops	24	178
Sol nu agricole	Agricultural bare soils	25	31
Cultures potagères	Garden crops	30	249
Savane	Savannah	104	319
Sols rocheux	Mineral soils	106	10
Bâti	Built-up surface	108	128
Eau	Water bodies	109	14
Maïs associé	Maize mixted with other crop	1001	146
Riz pluvial associé	Rainfed rice mixed with other crop	1002	73
Somme			2542

à rester le long des routes principales, routes goudronnées ou pistes de plaine (figure 16). À l'exception de rares excursions ciblées, les zones à plus de 15 minutes de marche n'ont pas été visitées.

Une tablette Trimble® Yuma® 2 équipée d'un GPS et du logiciel logiciel QGIS (QGIS.org, 2020) a été utilisée afin de faciliter le relevé de points d'occupation et d'usage des sols. Une image SPOT-6 (acquise le 14/02/2018)[8] et une image Pléiades (acquise le 01/02/2019) pour la saison 2018-2019 ont servi à positionner les relevés au centre de chaque parcelle (ou autres objets pour les relevés non agricoles). À chaque relevé, un formulaire de saisie permettait l'enregistrement rapide des attributs associés (date, numéro du relevé, type d'occupation/usage du sol, irrigué/pluvial) grâce à des listes déroulantes.

En accord avec les recommandations formulées dans le guide Joint Experiment for Crop Assessment and Monitoring (JECAM), les objets échantillonnés devaient avoir une aire supérieure

[8] Cette image a aussi été utilisée au début de la saison culturale 2018-2019 pour les relevés de points d'occupation des sols jusqu'au 03/02/2019, date où l'image Pléiades a été traitée pour être utilisée sur le terrain.

à 30 m * 30 m (estimation visuelle), soit neuf pixels Sentinel-2. Compte tenu de la petite taille des champs, cette recommandation a été suivie dans la mesure du possible, mais le critère de taille minimale des champs relevés est plus proche de 20 m * 20 m, soit quatre pixels Sentinel-2. Lorsque plusieurs cultures sont présentes en association, seuls les champs ayant une association de deux cultures ont été enregistrés. Les parcelles contenant plus de deux cultures n'ont pas été relevées.

Lors de ces relevés, nous avons veillé à ce que l'égalité des effectifs de chaque classe soit respectée afin de limiter les biais de classification. Cependant, face à la diversité de l'environnement agricole, à la surreprésentation de certaines cultures comme le riz irrigué dans les bas-fonds et à la spécialisation de certains espaces (production maraîchère intensive ou anciens vergers), les classes majoritaires ont été plus échantillonnées que les autres. Pour ces cultures, une attention particulière a été portée sur l'homogénéité intra parcellaire des relevés, ainsi que sur leur répartition spatiale sur l'ensemble de la zone d'étude.

En post-traitement, les contours des éléments relevés ont été digitalisés sur l'image THRS ayant servi aux relevés, en respectant une bande d'exclusion des contours de l'objet d'environ deux mètres afin d'éviter les effets de bordure (*i.e.* pixels mixtes). Lors de cette étape, d'autres polygones ont été ajoutés par photo-interprétation pour des objets facilement identifiables (bâtis, pins et autres cultures boisées de rente[9] et sol rocheux).

Durant la campagne de terrain 2018-2019, 2743 points ont été collectés sur le terrain et 126 points photo-interprétés (table 4). Les classes ayant des effectifs inférieurs à 20 ont été supprimées, tout comme les polygones de superficie inférieure à 200 m^2. Une suppression des polygones selon une aire plus élevée aurait conduit à une réduction trop importante du nombre d'échantillons. Comme présenté lors de de la section 1 Madagascar et les Hautes Terres, les champs relevés lors des campagnes sont parmi les plus grands observés.

Table 4 Nombre d'échantillons relevés par campagne de mesure, et taille des polygones. Au total, 2617 relevés ont été effectués sur le terrain, 126 ont été rajoutés par photo-interprétation et 201 polygones ont été supprimés, car jugés trop petits.

Campagne	Relevés agricoles	Relevés non-agricoles	Photo-interprétation non agricole	Polygones $\leq 200m^2$	Polygones $\leq 400m^2$	Polygones $\leq 900m^2$
2018-2019	2267	476	126	**201**	642	1497

La base de données de vérité terrain ainsi constituée permettra par la suite de créer une base de données d'apprentissage et de validation pour les étapes de classification de modélisation. L'identification des connaissances via des entretiens, des enquêtes et la littérature est présentée dans la suite de cette sous-section.

[9] Ces cultures sont principalement des plantations de pins et d'eucalyptus.

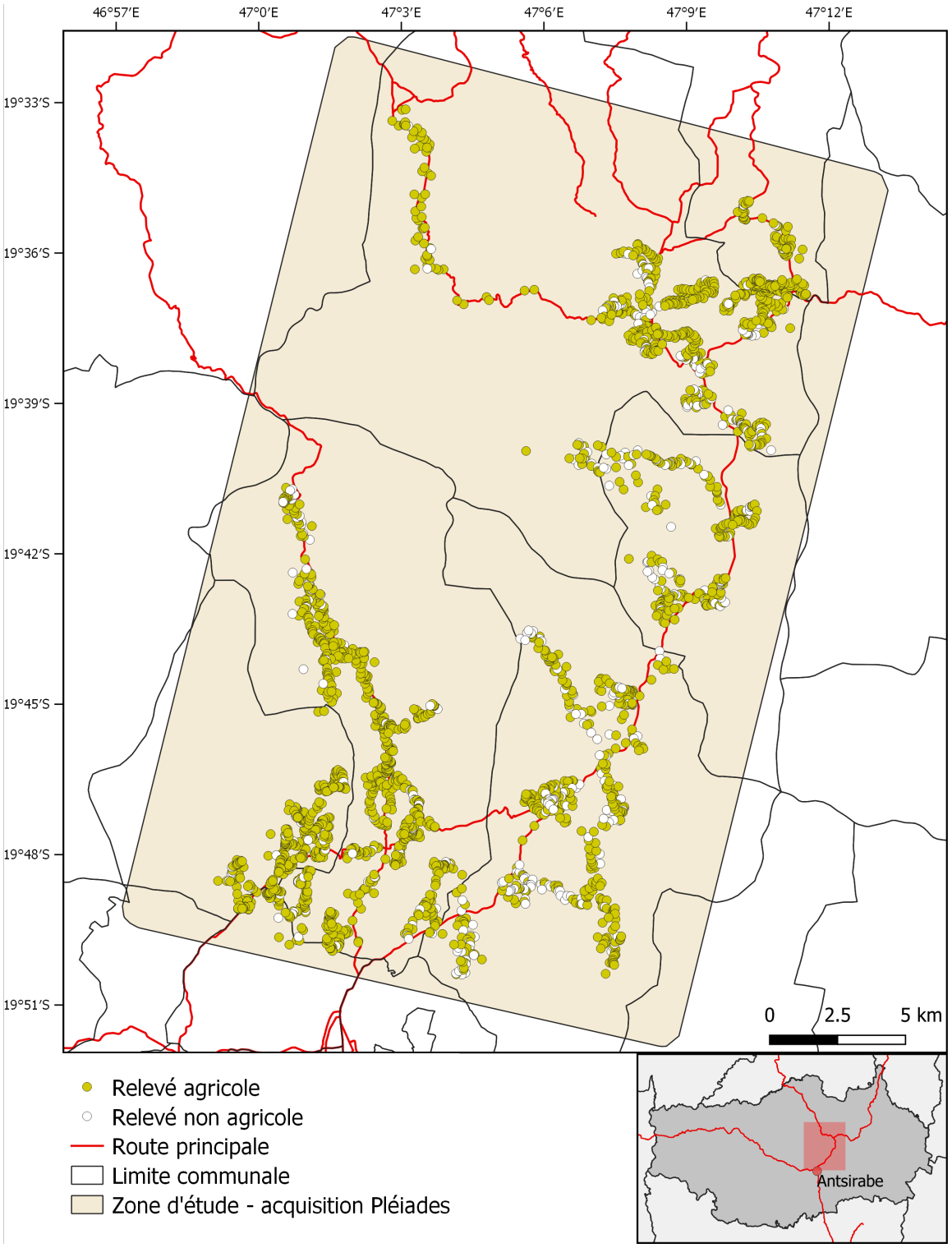


Figure 16 Cartographie des points GPS relevés entre le 31/01/2018 et le 10/04/2019. Sources : limites administratives et routes OpenStreetMap 2020.

2.2.3 Entretiens et enquêtes de terrain

Afin de comprendre l'organisation du territoire étudié et surtout les stratégies des agriculteurs vis-à-vis de leurs cultures, trois types d'entretiens et enquêtes ont été réalisés. Des entretiens avec des experts scientifiques travaillant sur place, des entretiens avec des agriculteurs, et des entretiens avec des responsables locaux. L'objectif principal de ces discussions était d'identifier des choix agricoles globaux qui déterminent une certaine organisation du territoire, puis d'en extraire des connaissances qui sont formalisées par la suite en règles spatio-temporelles.

2.2.3.1 Entretiens avec des scientifiques

Les discussions avec les scientifiques du DP SPAD ont été conduites de façon semi-dirigée afin d'obtenir des informations sur les pratiques culturales de la zone d'étude. Ces scientifiques travaillent sur des domaines divers tels que la géographie, la socio-économie ou l'agronomie. Les questions variaient selon le domaine scientifique du chercheur, son expertise locale ou bien son intérêt pour notre étude. Les thèmes abordés portaient sur le territoire d'Antsirabe, sur les principaux bassins de production, sur les contraintes existantes et sur les pratiques culturales des agriculteurs. Suite à ces entretiens, les personnes interrogées nous ont transmis un nombre important de documents scientifiques et techniques qui ont permis de construire les questionnaires à destination des agriculteurs, et de formuler certaines règles pour la modélisation. Les discussions avec les scientifiques nous ont permis d'obtenir une bibliographie technique et scientifique importante, parmi laquelle certaines références ont été utilisées pour identifier des connaissances (Annexe C).

2.2.3.2 Enquêtes avec des agriculteurs

Des entretiens individuels ont été menés auprès de 28 chefs d'exploitation lors de la saison culturale principale de 2017-2018 afin de comprendre leurs choix personnels quant à la répartition des cultures sur leurs parcelles et leurs choix d'assolement. Ces questionnaires dirigés ont été conçus à l'aide d'experts CIRAD et comportent quatre rubriques (Annexe D) :

- l'exploitation : cette rubrique comporte des questions sur le nombre de champs cultivés, le nombre de champs non mis en culture, le travail de la terre manuel ou tracté, et des demandes d'informations sur la famille de l'exploitant ;
- les cultures : dans cette seconde rubrique, étaient relevés pour chaque parcelle cultivée : la culture pendant la saison des pluies, le type de sol, la date de semis et de récolte, la localisation de la parcelle, l'irrigation et les contraintes associées (topographie, irrigation, engrais ...). Les séquences de cultures théoriques et appliquées à la parcelle ont aussi été renseignées ;
- le marché et les ventes : nous renseignions ici le devenir des productions de l'exploitation (autoconsommation ou actions d'achat/vente des semences et/ou productions). S'il y avait vente, le marché de destination était renseigné ainsi que la raison du choix du marché ;

- le lien entre l'exploitation et le Fokontany : enfin, dans cette dernière rubrique, il était demandé à l'exploitant si, de son point de vue, le Fokontany était homogène et si les cultures du Fokontany étaient les mêmes que celles cultivées par l'exploitant. Si ce n'était pas le cas, il lui était demandé qu'elles étaient les autres cultures présentes dans le Fokontany.

Ces enquêtes se sont déroulées avec l'aval des maires et chefs de Fokontany pour faire des enquêtes sur leur entité administrative, ainsi que des recommandations ou des sélections des personnes à enquêter. Les personnes répondant à ces questionnaires l'ont fait volontairement. En général une enquête durait une demi-journée, comprenant le temps de déplacement sur les lieux, l'obtention des autorisations et la visite de l'exploitation. L'utilisation des enquêtes pour la création de règles pour modéliser des processus a été consentie par les personnes enquêtées.

L'annexe D présente le questionnaire utilisé sur le terrain et l'annexe E présente la synthèse des entretiens avec les exploitants, triée par type de culture, après vérification avec les experts scientifiques.

2.2.4 Entretiens avec des responsables locaux

Lors de la seconde campagne de terrain en 2019, les entretiens ont été menés auprès de divers responsables agricoles malgaches de chaque commune et Fokontany de la zone d'étude, soit quatorze responsables ou groupes de responsables. L'objectif de ces entretiens était d'obtenir des informations plus générales, et d'identifier si certaines stratégies étaient propres à des délimitations administratives, culturelles ou topographiques. Dans cette optique, les questions étaient similaires à celles posées aux agriculteurs lors de l'année précédente, mais sur des exploitations théoriques. Un certain nombre de documents administratifs nous ont été transmis lors de ces discussions.

2.3 Données support pour l'application des connaissances

Les entretiens ont permis d'identifier des connaissances qui seront par la suite formalisées en règles (chapitre 4, section 2 De la connaissance à la probabilité d'appartenance : les règles expertes), puis utilisées dans les modules de caractérisation de l'occupation et usages des sols. Pour l'application des connaissances, certaines nécessitent des données dites « thématiques ». Pour cette étude, les données ouvertes et libres ont été privilégiées. Pour que l'application du modèle soit reproductible à d'autres systèmes agricoles de l'île, nous avons utilisé des données disponibles à l'échelle de Madagascar. Certaines données n'étant pas en accès libre ou inexistantes, elles ont été dérivées de données existantes.

Cette sous-section présente d'abord les variables dérivées du modèle numérique de terrain (MNT), la carte pédologique utilisée, le réseau routier, les données administratives et la création d'une carte du temps de trajet entre un bâtiment et un objet.

2.3.1 Topographie et relief

Un premier lot de variables thématiques ont été calculées à partir du modèle numérique de terrain (MNT) 1 Arc-Second Global[10] version 3.0 produit par la mission Shuttle Radar Topography Mission (SRTM) à une résolution d'un arc seconde (soit environ 30 mètres à l'équateur). Ce modèle numérique d'élévation est corrigé des vides (JPL, 2014). Quatre données support ont été dérivées de ce modèle numérique de terrain (MNT):

- l'altitude absolue ;
- les pentes ;
- la position topographique de chaque pixel dans l'environnement, pouvant s'apparenter à la géomorphologie, calculée selon l'approche développée par Weiss (2001) qui permet d'obtenir 10 classes topographiques. Cette topologie est décrite en Annexe G ;
- et le réseau hydrographique modélisé à l'aide du logiciel GRASS GIS (Team, 2018).

Maraîchage et relief

Le maraîchage fait partie des cultures « spécifiques » aux zones proches d'Antsirabe durant la saison culturale principale. Les cultures maraîchères représentent un investissement pour les agriculteurs par leur coût et celui des engrais, et demandent une main d'œuvre importante pour leur entretien (désherbage, irrigation, surveillance des maladies). Ces cultures ne se trouvent que dans certaines communes (annexe F) car elles demandent un savoir-faire particulier, et ce savoir se transmet de génération en génération. De plus, les cultures maraîchères nécessitent des conditions environnementales particulières.
Le relief est l'un des éléments clés des cultures maraîchères, majoritairement cultivées dans les plaines et cuvettes planes, mais aussi sur les pentes faibles des collines ou des petits cours d'eau, et dans les champs non inondés des bas-fonds. Plus rarement, ces cultures sont situées dans des fonds de vallée de montagne, ou sur les bas de pente de ces reliefs.

2.3.2 Carte pédologique

La carte pédologique de Riquier (1968) est la carte la plus précise à l'échelle de Madagascar. Cette carte permet d'identifier les principaux substrats et donc ceux plus favorables à certaines cultures.

[10] https://doi.org/10.5066/F7PR7TFT

2.3.3 Réseau routier

Madagascar ne dispose pas d'un recensement de son réseau routier en dehors des routes nationales. Pour avoir accès à des données routières géoréférencées complémentaires, la cartographie libre et collaborative est une source de données intéressante. Dans le projet OpenSteetMap, les zones vulnérables des pays « en développement » sont prioritairement cartographiées via des missions de cartographie (souvent portées par des ONG telles que CartONG[11] ou Humanitarian OSM Team[12]). Les données issues d'OpenStreetMap ont longtemps été très incomplètes sur le territoire malgache en dehors des routes principales et de la capitale Antananarivo. Une récente mission de cartographie a permis de limiter le nombre d'erreurs et surtout d'augmenter la surface cartographiée dans le pays, et partiellement sur notre zone d'étude. Malgré tout, les pistes et chemins ne sont toujours pas cartographiés, alors que ce sont les principales voies de communication.

Les données de routes ont été téléchargées au format vectoriel directement sur la plateforme OpenStreetMap[13] et les erreurs les plus grossières ont été corrigées.

Routes et agriculture

Les routes représentent les réseaux de communication les plus importants, surtout pour la vente des produits agricoles. Lorsque ces routes sont en bon état, les collecteurs viennent acheter les cultures en camion. Si les routes sont abîmées, seules les charrettes permettent le transport des marchandises. Les agriculteurs qui souhaitent vendre leurs produits sont donc très dépendants de l'état des routes, car peu d'entre eux possèdent une charrette à zébu. Lorsque l'agriculteur n'a pas de moyen de transport de marchandises, il est obligé de louer une charrette ou bien de prendre le taxi-brousse avec les produits qu'il souhaite vendre en ville.

[11] https://www.cartong.org/fr
[12] https://www.hotosm.org/
[13] https://download.geofabrik.de/

2.3.4 Données administratives

Une couche vectorielle des délimitations administratives (Fokontany et communes) a été utilisée pour l'identification des Fokontany à enquêter, ainsi que pour l'étude de la répartition spatiale des cultures. Il est à noter que le découpage de certains Fokontany n'est pas correct, car cette donnée est toujours en cours de production (communication personnelle, Defrise, 2018).

2.3.5 Bâtiments

Lors des entretiens avec les experts locaux, le **temps de trajet entre l'habitation d'un agriculteur et ses champs** a souvent été souligné comme l'un des principaux facteurs du choix du type de culture à mettre en terre par les agriculteurs (soit exprimé sous forme de distance à vol d'oiseau ou de temps dans les annexes F). En effet, les cultures ayant une forte valeur ajoutée (tel que les légumes ou les pommes de terre) sont plus susceptibles d'être volées si elles sont loin des habitations. La préférence des agriculteurs est donc d'avoir des plantations proches des habitations. Les cultures maraîchères demandent un travail au champ quotidien, ce qui nécessite de nombreux allers-retours durant la journée. Si les champs sont trop loin, alors cette culture risque d'avoir un entretien plus faible qu'un champ proche de l'habitation, mais le champ sera aussi probablement mis en culture avec des cultures pluviales ou ne sera pas cultivé.

Les données existantes ne sont pas suffisamment complètes pour pouvoir être utilisées. La couche vectorielle du bâti d'OpenStreetMap n'est pas utilisable, car trop faiblement documentée en dehors des grandes villes. Il a donc été nécessaire de créer une couche « bâti » à partir d'autres sources de données.

Pour obtenir les emplacements des bâtiments, nous avons utilisé la classification des séries temporelles d'images satellitaires afin d'en extraire la classe *bâti*, puis de calculer le temps de trajet entre un segment et un bâtiment d'habitation à partir d'un automate cellulaire. La méthode utilisée pour produire cette classification est détaillée dans l'annexe H.

Dans cette section, nous avons vu les trois types de données qui sont utilisées dans cette thèse : les données satellitaires (une image THRS et deux séries temporelles à haute résolution), les données de terrain (base de données de vérité terrain ainsi que les entretiens et enquêtes pour identifier des connaissances sur les stratégies et pratiques des agriculteurs) et les données thématiques, support pour l'application des connaissances. L'ensemble de ces données va être utilisé dans l'étape de classification et dans les modules spatiaux. Pour traiter ces données, deux principaux outils sont utilisés.

3 Outils

Deux outils originaux ont été utilisés dans la thèse : il s'agit de la chaîne de traitement de séries temporelles d'images de télédétection Moringa, et de la plateforme de modélisation des dynamiques spatiales Ocelet Modelling Platform (OMP). En complément de ces outils, l'Orfeo ToolBox (OTB) (Grizonnet et al., 2017) et des scripts Python (Foundation, 2018) ont été utilisés.

3.1 Moringa

Moringa est une chaîne de traitement développée au sein de l'UMR TETIS par Raffaele Gaetano[14] (Gaetano et al., 2019). Moringa utilise des modules de l'OTB (Grizonnet et al., 2017) orchestrés et lancés à partir de scripts Python. L'objectif de cette chaîne est de traiter les séries d'images satellite selon une approche orientée objet (Blaschke et al., 2014) basée sur une image très haute résolution spatiale (THRS). Cette chaîne est bien adaptée à la cartographie des espaces complexes tels que les zones rurales d'agriculture familiale. Moringa a déjà été utilisée en zone tropicale pour cartographier l'occupation des sols de l'île de La Réunion (Dupuy et al., 2020b) ou la zone périurbaine d'Antananarivo, la capitale de Madagascar (Dupuy et al., 2020a). L'utilisation de cette chaîne de traitement permet d'automatiser une partie du traitement des données issues de la télédétection avec une méthode déjà éprouvée sur le site JECAM de Madagascar par Lebourgeois et al. (2017). Dans nos travaux, la chaîne Moringa est utilisée pour le traitement des séries temporelles et pour le calcul des statistiques zonales (voir Chapitre 4, section 1.1.2 Séries temporelles d'images satellitaires). L'étape de classification en elle-même, à partir de la méthode ensembliste *Random Forest*, est réalisée à part, à l'aide d'un script Python dédié. La classification permet l'obtention, pour chaque objet, d'un vecteur de probabilité d'appartenance à chaque classe de la nomenclature.

3.2 Ocelet

Ocelet est une plateforme de modélisation et un langage métier permettant de réaliser des simulations numériques de dynamiques spatiales dans les territoires. Un territoire est modélisé comme un système où les entités présentes dans le territoire sont en interaction avec les autres. Le système et les entités qui le composent changent au cours du temps sous l'effet de ces interactions. Le langage propose les principaux concepts suivants pour permettre aux modélisateurs de représenter le système : entité, relation et scénario. Le formalisme d'Ocelet (Degenne et al., 2009, 2010; Degenne, 2012; Degenne and Lo Seen, 2016) se base sur le concept de graphe d'interaction, où chaque élément du système (entité) est représenté par un sommet du graphe, et où les arêtes du graphe entre les sommets portent les fonctions décrivant les interactions entre les entités (fonctions d'interaction). Chaque entité dispose de différentes caractéristiques, nommées propriétés, qui sont variables dans le temps. L'une de ces propriétés contient la représentation spa-

[14] Moringa est disponible à l'adresse https://gitlab.irstea.fr/raffaele.gaetano/moringa/

tiale de l'entité. L'utilisation conjointe des principales formes de représentation spatiale que sont les vecteurs (géométrie) et les rasters (matrices) est particulièrement aisée avec cette plateforme (Castets, 2015).

Trois niveaux de spécification de modèle sont définis par (Degenne, 2012) :

- « Le niveau des individus », où sont décrits les types d'**entités** du modèle, leurs propriétés définissant leur état, et les règles définissant leur comportement ;
- « Le niveau des interactions », où sont régis les échanges entre les entités qui sont définies via le concept de **graphe d'interaction** ;
- « Le niveau du système et de sa dynamique », qui permet d'organiser le déroulement d'un **scénario** de modélisation des dynamiques spatio-temporelles d'un système défini.

Un graphe d'interaction permet de connecter les entités susceptibles d'interagir entre elles. Un tel type de graphe en Ocelet est nommé **relation**. Il revient au modélisateur de définir les interactions (fonctions) pour décrire la manière dont les entités vont interagir et modifier ainsi la valeur de leurs propriétés.

Une relation définit ainsi la nature des actions possibles entre une entité et une autre entité distincte, ou un ensemble d'autres entités lorsqu'elles interagissent. Les interactions sont appliquées simultanément sur l'ensemble des arêtes et les valeurs des propriétés affectées des entités sont changées à la fin du parcours du graphe. Les valeurs pouvant modifier une propriété sont nommées « valeurs candidates » et conservées jusqu'à l'affectation finale, en utilisant si besoin une fonction d'agrégation.

Avec ces concepts dédiés, Ocelet permet de décrire dans l'espace et dans le temps à différentes échelles une grande diversité de situations rencontrées dans le territoire. Cette description exprime et combine généralement des connaissances de disciplines différentes (ex. agriculture, géographie, socio-économie). Les modules décrits en chapitre 2, section 3.2 pourront ainsi être mis en oeuvre utilisant Ocelet.

Une fois les données traitées à l'aide de la chaîne de traitement Moringa et avec la plateforme de modélisation Ocelet, les résultats doivent être analysés et évalués à l'aide de métriques.

4 Méthodes d'évaluation

Afin d'évaluer les résultats des traitements des données de télédétection et du modèle conceptuel, un ensemble de métriques est utilisé à la fois pour estimer les objets dont les classes sont correctement prédites et pour comprendre l'origine des erreurs de classification.

La base de données spatiale de polygones géoréférencés (section 2.2) est utilisée comme vérité terrain pour l'évaluation des attributions de classes. C'est l'**aire** de ces polygones qui est utilisée pour calculer les métriques d'évaluation.

L'évaluation d'une classification, déjà introduite dans la section 2.1.2.1, a pour but d'identifier le rapport entre les vrais positifs (éléments correctement attribués à leur classe), les faux positifs (éléments attribués à la classe analysée alors qu'ils correspondent à une autre), les faux négatifs (éléments de la classe analysée omis pour cette classe, c'est-à-dire attribués à une autre classe) et les vrais négatifs (éléments correspondant à une autre classe non associés à la classe analysée). L'ensemble de ces éléments est résumé dans une **matrice de confusion**. L'intérêt de chacun de ces éléments dépend de l'objectif de la classification. Pour la caractérisation de l'occupation des sols, c'est principalement l'analyse des vrais positifs et des faux positifs qui est importante. Les autres éléments permettent de comprendre les confusions entre les classes et l'origine des erreurs observées.

Les principales métriques utilisées pour évaluer ces éléments sont la précision globale, la précision intra-classe, le rappel de classe (Story, 1986), la F-mesure (Labatut and Cherifi, 2011) et le cœfficient Kappa (Cohen, 1960). Ces métriques sont largement utilisées en télédétection (Thenkabail, 2015).

La précision globale (OA pour Overall Accuracy) permet d'évaluer la fraction des éléments qui sont correctement classés :

$$\text{OA} = \frac{\sum_{i=1}^{n} \text{vrais positifs}}{N} \tag{1}$$

où N est la somme des échantillons de vérité terrain ;
n est le nombre de classe.

Le **rappel** (ou sensibilité) permet de calculer le taux d'éléments positifs correctement prédits par classe :

$$\text{Rappel} = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux négatifs}} \tag{2}$$

La **précision** est l'évaluation du taux de vrais positifs dans la classe prédite :

$$\text{Précision} = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux positifs}} \tag{3}$$

La **F-mesure** est la moyenne harmonique de la précision et du rappel :

$$\text{F-mesure} = 2 \cdot \frac{\text{rappel} \cdot \text{précision}}{\text{rappel} + \text{précision}} \tag{4}$$

Dans le cadre d'une évaluation multiclasses comme dans notre cas, le rappel, la précision et la F-mesure peuvent être calculés pour l'ensemble du classifieur, tout en donnant un poids égal à

chaque classe :

$$\text{Métrique globale } = \frac{\sum_{i=1}^{n} \text{métrique}}{N} \tag{5}$$

Pour prendre en compte le déséquilibre entre les classes, il est possible de pondérer ces métriques par la présence de la classe dans le jeu de données d'évaluation :

$$\text{Métrique pondérée } = \frac{\sum_{i=1}^{n} \omega_n \text{vrais positifs}}{\sum_{i=1}^{n} \omega_n} \tag{6}$$

où ω_n est le poids de la classe n ; cela correspond au nombre d'éléments dans la classe n;

Le Kappa de Cohen permet de mesurer la différence entre l'accord observé et la probabilité d'un accord uniquement dû au hasard. Si $\kappa = 1$, le classifieur et les données terrain sont en accord. Si $\kappa <= 0$, soit ils sont en désaccord, soit le résultat est uniquement dû au hasard :

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{7}$$

où $\Pr(a)$ est l'accord relatif entre les données de vérité terrain et les données prédites (soit la précision globale) ;
$\Pr(e)$ est la probabilité d'un accord aléatoire.

En complément de la mesure de précision globale, Pontius and Millones (2011) proposent deux indicateurs pour décomposer les causes de désaccord global : la mesure de désaccord de quantité Q et la mesure de désaccord de répartition[15] A (Pontius and Millones, 2011; Warrens, 2015). Ces indicateurs permettent aussi de mesurer l'accord entre la carte de référence et la carte modélisée, et d'obtenir des pistes de compréhension des sources de désaccords. Si la quantité de désaccord est importante, alors il y a une erreur dans les totaux des classes. Si l'accord de répartition est important, alors il y a des erreurs dans l'allocation spatiale des classes.

La mesure de quantité de désaccord d'une classe est donnée par :

$$q_n = |(p_{n+} - p_{+n})| \tag{8}$$

où n est la classe ;
p_{n+} est la somme des lignes ;
Et p_{+n} est la somme des colonnes.

La mesure de désaccord de répartition permet d'identifier les classes qui ont des différences

[15] *quantity disagreement* et *allocation disagreement*

importantes dans leur répartition spatiale. La mesure de désaccord de répartition est donnée par :

$$a_n = 2 \cdot \min(p_{n+}, p_{+n}) - 2p_{nn} \tag{9}$$

La quantité de désaccord global correspond à la somme totale des différences entre les classes de référence et les classes prédites selon la correspondance moins que parfaite dans les totaux des catégories p_{n+} et p_{+n} (Pontius and Millones, 2011). Dans tous les cas, $q_n \leq Q$, Q est donc la limite supérieure de quantité de désaccord total, l'analyse des q_n permet de comprendre quelle est la classe qui contribue le plus à l'erreur. La quantité de désaccord global est donnée par :

$$Q = \frac{1}{2}\sum_{i=1}^{n}|(p_{n+} - p_{+n})| \tag{10}$$

La mesure de désaccord de répartition global mesure la somme totale des différences entre les classes de référence et les classes prédites due à la correspondance moins que maximale dans la répartition spatiale des catégories compte tenu des totaux des catégories des classes de référence (Pontius and Millones, 2011). À la différence des quantités de désaccord, a_n peut être supérieur à A. Le désaccord de répartition global est donné par :

$$A = \left[\sum_{i=1}^{n} \min(p_{n+}, p_{+n}) - OA\right] = \frac{\sum_{i=1}^{n} a_n}{2} \tag{11}$$

Dans le cas où $p_{n+} = p_{+n}$ pour toutes les classes, il n'y a pas de désaccord de quantité et l'erreur est uniquement liée au désaccord de répartition et inversement si $pnn = \min(p_{n+}, p_{+n})$. Une valeur de désaccord est considérée comme conséquente si elle dépasse 0,1 (Warrens, 2015).

Pour comparer les résultats entre la classification des données de télédétection et l'application du modèle conceptuel, le test du $\chi 2$ de McNemar est utilisé McNemar (1947). Ce test permet de déterminer s'il y a une différence statistiquement significative entre deux résultats de classification. Pour que ce test soit valide, il faut que la somme des paires discordantes soit ≥ 25, sinon une correction de continuité doit être appliquée (Edwards, 1948).

Le test du $\chi 2$ de McNemar est donné par :

$$\chi 2 = \frac{(b-c)^2}{b+c} \tag{12}$$

où b est la proportion de valeurs étant correctement labellisées par le premier classifieur mais incorrectement labellisées par le second classifieur ;
c est la proportion de valeurs étant correctement labellisées par le second classifieur mais incorrectement labellisées par le premier classifieur.

La correction de continuité du $\chi 2$ est donnée par :

$$\chi 2 = \frac{(|b-c|-1)^2}{b+c} \tag{13}$$

Une évaluation thématique des résultats est aussi effectuée pour aller au-delà de l'information fournie par la base de données terrain et ainsi obtenir une compréhension spatiale de l'application du modèle conceptuel.

POINTS CLÉS DU CHAPITRE 3

- **CONTEXTE MALGACHE** :
  - Madagascar est un pays principalement agricole avec un taux de pauvreté important ainsi qu'une vulnérabilité forte vis-à-vis du changement climatique. Le mode d'agriculture dominant dans le pays est l'**agriculture familiale** ;
  - les Hautes Terres du Vakinankaratra sont une zone de production agricole intensive et diversifiée (riz irrigué, cultures pluviales, maraîchage), caractérisée par des **exploitations de petite taille** ;
- **DONNÉES** :
  - **3 sources de données de télédétection** : une image Pléiades THRS (0,5*0,5 m), une série temporelle d'images venμs (5*5 m) et une série temporelle d'images Sentinel-2 (10*10 m) ;
  - une **base de données de vérité terrain** comportant 2542 polygones répartis en 17 classes agricoles et non agricoles ;
  - des enquêtes auprès des agriculteurs et des responsables locaux malgaches, ainsi que des entretiens avec des scientifiques pour identifier les **connaissances à formaliser** pour modéliser les stratégies et pratiques des agriculteurs ;
  - des **données support pour l'application des connaissances** dans les modules : variables dérivées du modèle numérique de terrain (MNT), carte pédologique, réseau routier de la zone d'étude, limites administratives et une carte du temps de trajet entre un bâtiment et un segment ;
- **OUTILS ET ÉVALUATION** :
  - la chaîne de traitement de séries temporelles **Moringa** pour les données de télédétection ;
  - la plateforme de modélisation **Ocelet** pour l'implémentation des modules spatiaux ;
  - un ensemble de métriques d'évaluation quantitatives pour évaluer les résultats de l'application du modèle conceptuel et pour comparer les résultats entre eux.

Chapitre 4

Application du modèle conceptuel : Méthode d'application des modules à la région des Hautes Terres

Table des matières

1. Télédétection
   1.1. Traitement des images satellite
   1.2. Classification des données de télédétection
2. De la connaissance à la probabilité d'appartenance : les règles expertes
   2.1. Sélection des connaissances à formaliser
   2.2. Formalisation des connaissances en règles
   2.3. Attribution de probabilités aux règles (issues de la connaissance)
3. Application des modules
   3.1. Préparation de données support pour l'application des connaissances
   3.2. Sélection des segments
   3.3. Calcul des probabilités de présence de culture par segment
   3.4. Application du module 4 : modélisation de la répartition spatiale des classes
4. Synthèse des expérimentations menées
   4.1. Analyse de la « cohérence agricole » des cartes
   4.2. Analyse de la classification des données de télédétection
   4.3. Résultats de la combinaison des $VPAC_{RF}$ et $VPAC_{RS}$

Ce chapitre 4 présente les méthodes mises en oeuvre pour appliquer le modèle conceptuel au cas d'étude des Hautes Terres de Madagascar. Dans ce chapitre sont abordés le traitement des données de télédétection servant à obtenir des probabilités liées à l'observation d'une classe dans un segment, la formalisation des connaissances en règles expertes, la préparation des données support pour l'application des connaissances, les méthodes mises en œuvre pour utiliser les connaissances formalisées dans les deux modules spatiaux via à la plateforme Ocelet (module 2 : règles spatiales et module 4 : modélisation de la répartition spatiale des classes), et enfin, les méthodes d'évaluation des résultats.

1 Télédétection

Dans cette section est présentée la chaîne de traitement des images de télédétection utilisée pour obtenir une première cartographie de l'occupation et de l'usage des sols . La chaîne de traitement, basée sur la chaîne de traitement Moringa, est composée de quatre blocs (Figure 17). Trois blocs correspondant au traitement de l'image THRS Pléiades (**1**) (pré-traitements, segmentation de l'image, choix des indices de textures à extraire), au pré-traitement des séries temporelles Sentinel-2 et Venμs et au calcul des indices spectraux (**2**), et au calcul des statistiques zonales (**3**) sont présentés en sous-section 1.1. Le quatrième bloc sur la partition de la base de données de vérité terrain et la classification des données de télédétection à l'aide d'un algorithme de forêts aléatoires est présenté en sous-section 1.2. En sortie de la chaîne de traitement, un $VPAC_{RF}$ (vecteur de probabilité d'appartenance aux classes issu de l'algorithme de forêt aléatoire) est créé pour chaque segment, synthétisant les informations des images satellitaires.

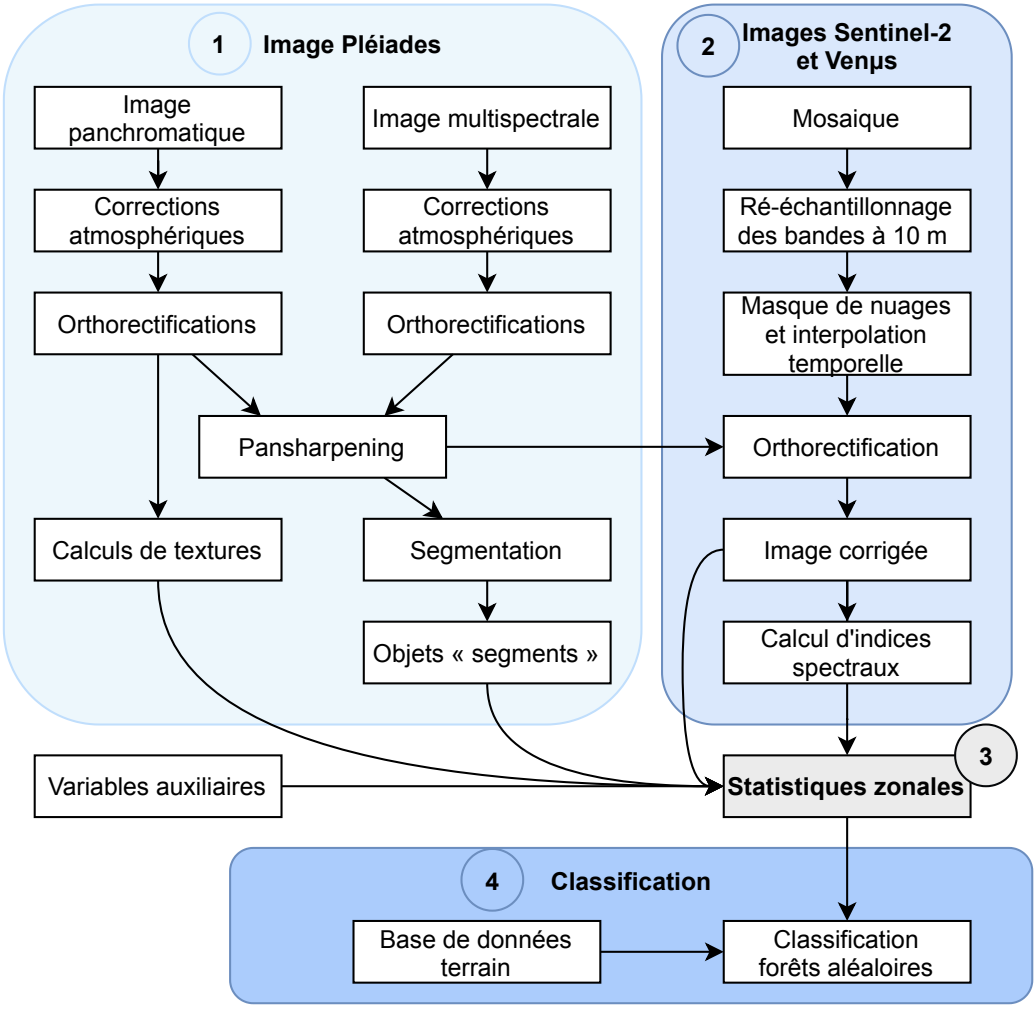


Figure 17 Schéma des étapes du traitement des données de télédétection.

1.1 Traitement des images satellite

1.1.1 Image à très haute résolution spatiale

L'image THRS Pléiades (présentée chapitre 3, section 2.1 Données satellitaires et prétraitements) est d'abord pré-traitée, puis les contours des éléments à caractériser sont extraits par une segmentation automatique de l'image, et les indices de textures sont calculés.

1.1.1.1 Pré-traitement

L'image Pléiades a été pré-traitée avec le logiciel open source OTB (Grizonnet et al., 2017). Les comptes numériques de la panchromatique et des bandes multispectrales ont d'abord été convertis en réflectance top of atmosphere (TOA). Les images ont ensuite été orthorectifiées à l'aide du MNT SRTM de l'USGS. Afin d'assurer une superposition optimale des images et la continuité des relevés sur le terrain, les images ont été géométriquement corrigées une seconde fois à l'aide d'une image THRS acquise lors de la saison culturale précédente. Ensuite, les images multispectrales et panchromatiques sont fusionnées par « pansharpening », processus qui consiste à ré-échantillonner l'image multispectrale à la résolution de l'image panchromatique. Le ré-échantillonnage est effectué par interpolation bicubique, et un algorithme de fusion bayésienne (Fasbender et al., 2008) est utilisé pour obtenir une image multispectrale à 0,5 m de résolution.

1.1.1.2 Segmentation

La segmentation est un processus qui consiste à partitionner une image en segments selon des critères définis. Un segment représente un ensemble de pixels contigus de l'image. La segmentation de l'image multispectrale à 0,5 m de résolution a été effectuée selon l'algorithme « Large Scale Generic Region Merging" » de l'OTB (Lassalle, 2015). Le critère de Baatz and Schäpe (2000), qui est basé sur une approche de fusion itérative des régions (Hong and Rosenfeld, 1984), a été utilisé.

Dans cette approche, chaque pixel est initialement considéré comme un segment. De manière itérative, chaque paire de segments peut être fusionnée selon un critère d'homogénéité, décrivant la similarité entre deux segments. Lorsque le seuil d'homogénéité est atteint ou qu'un segment a atteint la taille maximale définie par l'utilisateur (paramètre d'échelle), la fusion n'est plus possible. Le critère de Baatz and Schäpe (2000) est utilisé comme critère d'homogénéité combinant à la fois l'information spectrale et l'information spatiale. L'attribut spectral correspond à l'écart type des intensités spectrales des pixels contenus dans le segment et l'information spatiale correspond à trois attributs géométriques d'un segment qui sont sa surface, son périmètre et son enveloppe. Cette méthode a été choisie, car elle donne de bons résultats pour les images THRS (Schiewe, 2002; Meinel and Neubert, 2004) et a pour avantage de permettre la création de segments de taille hétérogène.

Le choix des paramètres de segmentation a été fait de manière subjective en évaluant visuellement les résultats de cinquante combinaisons de paramètres sur deux zones distinctes. La première zone est la cuvette d'Ambohibary, qui est un espace agricole composé de grandes parcelles de riz irrigué et de petites parcelles de légumes dans le bas-fond. Un mélange de cultures pluviales et de savanes est présent sur les pentes bordant la dépression. La seconde zone est dans la commune d'Alakamisy, composée de petites parcelles, soit de légumes, soit de cultures pluviales sur pentes douces et plateaux. Les critères choisis sont **0,3** pour l'attribut spectral, **0,7** pour l'attribut spatial et **250** pour le paramètre d'échelle.

Il est difficile d'évaluer de manière quantitative et comparative différentes segmentations. En effet, il faudrait une segmentation de référence pour l'image complète, ainsi que suffisamment de temps de calcul pour produire toutes les segmentations. Dans notre cas, la vérité terrain ne correspond qu'à 3,39 % de la surface de l'image Pléiades. Afin d'avoir un aperçu quantitatif de la segmentation, deux indicateurs ont été calculés à partir de la base de données de vérité terrain : le nombre moyen de segments par polygone et par classe, ainsi que la précision (équation 3) et la F-mesure (équation 4). Pour calculer ces indicateurs, tous les segments qui sont totalement ou partiellement à l'intérieur d'un polygone de vérité terrain sont utilisés.

Le tableau 5 présente, par classe, le nombre de segments moyen contenu par polygone de vérité terrain. Les classes non agricoles (*savane*, *sols rocheux*, *bâti* et *eau*) sont sur-segmentées, car les polygones de vérité terrain sont majoritairement de grande taille et composés d'éléments hétérogènes (terre, tôle, paille pour le bâti par exemple). Pour les classes agricoles, la segmentation est acceptable compte tenu du nombre moyen de segments pour une parcelle. Bien que ces classes soient légèrement sur-segmentées, notamment dans le cas du maïs du fait de la plus grande taille des parcelles qui ont été relevées, et dans le cas des cultures boisées de rente (pins et eucalyptus) pour lesquelles des segments sont créés à l'échelle des groupements d'arbres.

En se basant sur cette évaluation et en prenant en compte que l'objectif de cette thèse est la caractérisation de l'espace agricole, les critères de segmentation peuvent être considérés comme cohérents pour les classes agricoles, et ce, au détriment des autres classes qui sont sursegmentées.

Le tableau 6 présente les résultats globaux de la précision et la F-mesure pour l'évaluation de la segmentation. Pour évaluer la segmentation, le rappel représente la part de segments intersectant entièrement ou en partie la base de données de vérité terrain. Un rappel de 1 signifie que tous les segments présents dans les polygones de vérité terrain sont pris en compte. Un rappel de 0,99 signifie que les segments qui ont plus de 99 % de leur surface en dehors du polygone de vérité terrain ne sont pas pris en compte.
De manière théorique, un rappel inférieur à 1 peut signifier une inexactitude de la base de données de vérité terrain. C'est-à-dire que certains polygones ne sont pas correctement délimités. Le tableau 5 présente les résultats de la précision par classe. Ces indicateurs ont été calculés en considérant comme correcte l'aire du segment généré qui se trouve à l'intérieur du polygone de vérité terrain et comme incorrecte l'aire à l'extérieur de ce polygone. Cela permet d'évaluer la précision du contour de l'objet à délimiter (Levinshtein et al., 2009).

Table 5 Évaluation de la segmentation avec comme paramètre d'échelle **250**, comme paramètre spectral **0,3** et comme paramètre spatial **0,7**. La précision est calculée pour l'ensemble des segments qui intersectent un polygone de vérité terrain (rappel = 1).

Occupation du sol	Nombre moyen de segments par classe	Précision
Maïs	15.57	0.70
Riz irrigué	5.72	0.35
Riz pluvial	5.8	0.36
Légumes	2.41	0.28
Soja	5.71	0.44
Pomme de terre	2.64	0.33
Herbes et autres cultures fourragères	13.97	0.66
Verger	6.81	0.31
Culture boisée de rente	23.09	0.52
Sol nu agricole	4.77	0.52
Cultures potagères	3.22	0.36
Savane	13.03	0.58
Sols rocheux	12.2	0.45
Bâti	23.5	0.50
Eau	14.64	0.60
Maïs associé	5.83	0.39
Riz pluvial associé	5.14	0.36

Si l'on considère l'ensemble des segments qui intersectent un polygone (rappel = 1), nous obtenons une précision de **0,43**. Ce résultat montre que la segmentation ne correspond pas exactement aux polygones de référence, un rappel de 0,43 signifie que 57 % de la surface des segments qui intersectent les polygones de vérité terrain sont à l'extérieur de ces polygones. Si l'on ne prend en compte que les segments qui intersectent les polygones à plus de 90 % de leur surface (rappel = 0,90), la précision passe de **0,43** à **0,60** et la F-mesure de **0,60** à **0,72**. Un rappel de 0,90 peut être considéré comme acceptable, car il permet d'éliminer les segments en bordure des polygones qui peuvent être liés à une mauvaise délimitation d'un polygone de vérité terrain, ou à un problème d'orthorectification de l'image, en particulier sur des zones de pentes, ou encore à la présence d'objets en bordure tels que la diguette d'un champ de riz irrigué ou la présence d'un arbre.

En dehors de ces erreurs acceptables, ces métriques nous indiquent que la segmentation n'a pas permis de délimiter correctement les contours des polygones de vérité terrain. Les précisions les plus faibles étant pour les classes agricoles, excepté la classe *maïs* (avec un rappel de 1) (tableau 5).

Table 6 Évaluation de la segmentation en terme de précision et de F-mesure, en fonction de la valeur du rappel choisi.

Rappel	Précision	F-mesure
1	0.43	0.60
0.99	0.49	0.65
0.95	0.55	0.69
0.90	0.60	0.72

1.1.1.3 Textures

Les indices de texture sont une description continue de l'organisation spatiale des valeurs de réflectance. Les valeurs de texture sont indépendantes de la radiométrie, ce qui permet d'ajouter des variables pour la classification (Hall-Beyer, 2017). L'utilisation des indices de texture peut permettre de pallier le manque de résolution spectrale (Pacifici et al., 2009) et ainsi d'améliorer la précision de la classification (Pacifici et al., 2009; Regniers, 2014). Haralick (1979) propose un ensemble de 14 textures statistiques afin de résumer l'information texturale pour le traitement des images de télédétection.

Le calcul d'indices de texture permet à la fois d'identifier des groupes de pixels homogènes dans une fenêtre de n*n pixels (textures de premier ordre), et la variation spatiale des pixels au sein de cette fenêtre (textures de second ordre). Les indices de texture de premier ordre ont comme objectif d'analyser l'homogénéité/hétérogénéité dans leur fenêtre de calcul. Ces indices peuvent aider à la discrimination de différents types d'occupation/usage du sol (par exemple, une parcelle de riz sera plus homogène qu'un verger dont les alternances entre couronnes d'arbres, sol, parties ensoleillées et ombragées engendrent une forte hétérogénéité du signal radiométrique). Ces indices de texture sont calculés à partir de l'ensemble des valeurs des pixels de la fenêtre sans prendre en compte leur répartition et relation spatiale. Les indices de texture de second ordre sont basés sur les relations spatiales entre les pixels dans une fenêtre donnée, correspondant à une matrice de co-occurences des niveaux de gris (GLCM). C'est avec ces indices que l'on peut observer les motifs anisotropes.

Les indices de textures permettent d'analyser la distribution statistique des niveaux de gris d'une image, selon trois paramètres :

- la taille de la fenêtre d'analyse, soit l'ensemble des pixels considérés dans un voisinage. Une fenêtre large produit des textures stables, mais avec des effets de bord, alors qu'une fenêtre réduite minimise ces effets au détriment de la stabilité(Fern and Warner, 2002). Il faut donc une fenêtre suffisamment large pour correspondre aux variabilités de taille et d'espacement des éléments de l'occupation du sol (rangées de légumes par exemple), tout en ne recoupant pas une classe différente ;
- la distance d'analyse entre deux pixels d'un même couple. Si possible, elle doit correspondre à une corrélation élevée entre un couple de pixels. Selon Regniers (2014), les courtes distances sont plus significatives ;
- et la direction de l'analyse, qui permet d'extraire les textures anisotropes, dépendantes de l'orientation (cultures en rangs). Si la texture est isotrope, c'est-à-dire homogène ou aléatoire, l'orientation a un faible impact.

Les textures sont calculées à partir de la bande panchromatique de l'image THRS Pléiades. Deux fenêtres de textures ont été choisies afin de caractériser l'occupation des sols agricoles. Pour les petites parcelles, une taille de fenêtre réduite a été choisie afin d'avoir une transition nette entre deux parcelles ayant une occupation du sol différente. Pour détecter les surfaces composées de parcelles contiguës d'une même culture, telles que les zones maraîchères ou les bas-fonds en riziculture, une fenêtre plus large est sélectionnée (Lefebvre, 2011).

Pour notre zone d'étude, aucune direction n'a été privilégiée, car nous étudions tous les types d'occupation et d'usage des sols sur un territoire avec une structure spatiale hétérogène. Pour pallier ce problème, quatre orientations ont été calculées (0°, 45°, 90°et 135°) afin d'avoir un résumé des différentes orientations (Haralick, 1979). La moyenne de ces résultats correspond à un descripteur invariant à la rotation (Kim et al., 2009), qui permet de caractériser correctement les objets ayant des textures isotropes (Regniers, 2014).

La taille de la fenêtre optimale est obtenue à l'aide de semi-variogrammes qui permettent d'évaluer la corrélation spatiale entre des paires de pixels et de détecter à partir de quelle distance deux pixels sont considérés comme différents, signifiant que la taille de fenêtre est optimale (Regniers, 2014). Pour l'intégralité des polygones issus de la campagne de terrain effectuée en 2019, les variogrammes ont été calculés et classés, et deux fenêtres de calcul de texture ont été choisies : une fenêtre de **15*15 pixels** et une fenêtre de **29*29 pixels**[1].

Les indices de textures choisis sont ceux préconisés par Hall-Beyer (2017). Deux textures internes, l'indice de matrice de co-occurences des niveaux de gris (GLCM) **corrélation** et l'indice de GLCM **moyenne**[2], servent de descripteurs pour l'identification des variations de valeurs au sein d'une image. L'indice de **contraste**[3] est utilisé comme texture de bordure afin de décrire et détecter les limites de parcelle. Soit au total six indices de texture, trois par fenêtre.

1.1.2 Séries temporelles d'images satellitaires

La chaîne de traitement d'images satellitaires Moringa (Gaetano et al., 2019) a été utilisée pour pré-traiter les séries temporelles Sentinel-2 et Venμs (chapitre 3, section 2.1 Données satellitaires et prétraitements), calculer les variables de réflectance et les indices spectraux, et calculer les statistiques zonales (extraction des valeurs moyennes de chaque variable par objet).

1.1.2.1 Pré-traitements

Le pré-traitement des séries temporelles se décompose en 4 étapes :

1. la création d'une mosaïque de tuiles Sentinel-2 acquises à des dates proches (3 jours maximum d'écart) ;
2. le ré-échantillonnage à 10 m des bandes spectrales de Sentinel-2 ayant une résolution spatiale de 20 m (Annexe A) (soit six bandes allant du PIR au MIR) ;
3. la construction d'une image composite à intervalle régulier : tous les 10 jours pour Sentinel-2 et tous les 5 jours pour Venμs. L'image est construite à partir des masques de nuage de chaque date et des pixels non masqués les plus proches dans le temps. La valeur d'un pixel masqué est interpolée à la donnée la plus proche temporellement ;

[1] Soit un rayon de 7 pixels pour la première fenêtre et un rayon et 14 pixels pour la seconde fenêtre.
[2] « Guideline 1: When using texture for classification, use GLCM Mean as a first choice and GLCM Cor as a possible addition. » Hall-Beyer (2017)
[3] « Guideline 2: If a class is characterised by edges within its patch boundaries, choose Con. If no class has this characteristic, then an edge texture unlikely to be helpful for classification » Hall-Beyer (2017)

4. l'orthorectification des images composites sur la base de l'image THRS Pléiades et du MNT de l'USGS.

À la suite de ces pré-traitements, 31 images sont obtenues pour Sentinel-2 et 61 pour Venμs pour la période de suivi (de septembre 2018 à juin 2019). Aucune image n'a été exclue lors de la création d'images composites.

1.1.2.2 Extraction des variables

La caractérisation d'un système à partir d'images satellitaires peut se baser à la fois sur les bandes spectrales et des combinaisons de plusieurs bandes nommées **indices spectraux**. Dans notre cas d'application, nous avons utilisé des indices de végétation pour faciliter la détermination de la végétation en période de croissance (Colwell, 1974). Les indices les plus communs pour l'analyse de la végétation ont été calculés pour les images Sentinel-2 et Venμs. Pour ces deux satellites, les indices de végétation NDVI, NDWI, NDRE et un indice de brillance ont été calculés[4] :

- l'indice de végétation par différence normalisée (en anglais : Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974)) combine les bandes spectrales du rouge et du proche infrarouge. Ces bandes sont sensibles respectivement à la présence de pigments chlorophylliens et à la densité de végétation. Cet indice permet de discriminer les surfaces couvertes par des végétaux par rapport à d'autres types de surface non végétalisées (eau, bâti, sols rocheux ...), ainsi qu'à aider à la discrimination des différents types de végétation (naturelle ou cultivée) en fonction de leur développement végétatif, du stade de croissance, de leur état ou de réponses spectrales spécifiques ;
- l'indice hydrique par différence normalisée (en anglais : Normalized Difference Water Index (NDWI) (Gao, 1996)) est un indice sensible à la présence d'eau dans les feuilles, qui est utilisé pour estimer la teneur en eau des végétaux, mais également dans les applications de cartographie ou de gestion des zones irriguées, ou le suivi des sécheresses. Il utilise les bandes spectrales du vert et du PIR ;
- l'indice Normalized Difference Red-Edge (NDRE) (Barnes et al., 2000) a un intérêt similaire à celui du NDVI, à la différence qu'il est calculé avec les bandes spectrales du PIR et du *red-edge* ;
- l'indice de brillance des sols (en anglais : Brightness Index (BRI)) basé sur l'équation de (Escadafal et al., 1989) est une combinaison de l'ensemble des bandes disponibles dans une image satellite. Il est sensible à la brillance du sol qui est elle-même liée à l'humidité des sols.

Pour Sentinel-2, deux indices supplémentaires ont été calculés à partir des bandes spectrales dans le MIR (bande 11, Annexe A) :

- le modified Normalized Difference Vegetation Index (mNDVI) (Jurgens, 1997) est une amélioration du NDVI construite pour réduire l'influence des bruits liés aux conditions atmo-

[4] les formules des indices pour Sentinel-2 et Venμs sont données en annexe B.

sphériques et à la nature du sol. Dans cet indice, le MIR remplace le rouge (R).

- Le modified Normalized Difference Water Index (MNDWI) (Xu, 2006) est un indice qui permet de limiter le bruit causé par les bâtiments, la végétation et le sol afin de mieux détecter les zones en eau. Dans cet indice, le MIR remplace le PIR.

Pour caractériser le système des Hautes Terres, 10 bandes spectrales de Sentinel-2 et 11 bandes spectrales de Venμs ont été utilisées, soit un total de 310 variables de réflectance pour Sentinel-2 et de 671 variables de réflectance pour Venμs[5] pour la période considérée (31 dates). Six indices spectraux ont été calculés pour Sentinel-2, soit 186 indices pour la période, et quatre indices spectraux ont été calculés pour Venμs, soit 244 indices pour la période. Au total, 1 411 variables (réflectances et indices spectraux) issues des séries temporelles d'images satellitaires sont disponibles.

1.1.3 Statistiques zonales

Le calcul des statistiques zonales a pour but de résumer l'information d'un ensemble de pixels contenus dans un segment. Le nombre de bandes spectrales et d'indices étant élevé, seule la moyenne des valeurs de ces variables est calculée pour chaque segment. Les indices de textures présentés dans la section 1.1.1.3 Textures, l'altitude et la pente du MNT (chapitre 3, section 2.3 Données support pour l'application des connaissances) sont aussi moyennés par segment.

Ces statistiques zonales sont utilisées comme variables d'entrée de la classification supervisée des segments qui est présentée dans la section suivante.

1.2 Classification des données de télédétection

La classification des données de télédétection se fait de façon supervisée, sur la base d'une base de données de référence servant à entraîner l'algorithme de classification de forêts aléatoires. Au préalable, la base de données de vérité terrain a été divisée entre un jeu d'apprentissage servant à entraîner les modèles et un jeu de validation qui est utilisé pour évaluer la classification des données de télédétection. Ce même jeu de données est utilisé pour évaluer le résultat de l'application du modèle.

1.2.1 Partitionnement de la base de données de vérité terrain

La base de données de vérité terrain sur l'occupation/usage des sols a été séparée en deux jeux indépendants et en respectant la proportionnalité de chaque classe (présenté chapitre 3, section 2.2 Les données de terrain). Le jeu de données d'apprentissage qui sert à entraîner le modèle de forêt aléatoire (random forest) correspond à 70 % du jeu de données complet. Les 30 % restants composent le jeu de validation.

Pour augmenter la taille du jeu d'apprentissage, chaque polygone est intersecté avec la segmentation issue de l'image THRS. Puis, les statistiques zonales sont calculées pour chaque segment

[5] Le détail des bandes spectrales utilisées est visible en annexe A

intersecté[6]. La même opération est effectuée sur le jeu de validation afin d'obtenir un ensemble cohérent entre l'apprentissage et la validation.

1.2.2 Modèle de forêts aléatoires (*random forest*)

Pour classifier les données de télédétection et intégrer les résultats de cette classification dans le modèle, il est nécessaire d'utiliser un classifieur permettant d'obtenir les probabilités d'appartenance à chaque classe pour chaque segment.

L'algorithme de classification choisi est la méthode ensembliste Random Forest (Breiman, 2001), qui est largement utilisée dans le domaine de la télédétection (Debats et al., 2016; Lebourgeois et al., 2017; Song et al., 2017; Lelong et al., 2020) car il permet d'obtenir de meilleurs résultats que les autres classifieurs (Fernández-Delgado et al., 2014; Inglada et al., 2015).

L'algorithme de forêts aléatoires (*random forest* en anglais) est un méta estimateur qui fait correspondre un ensemble d'arbres de décision à différents sous-ensembles aléatoires d'un jeu de données d'apprentissage. L'algorithme utilise la moyenne de ces résultats pour améliorer la précision du classifieur et éviter le sur-apprentissage (Louppe, 2014). Chaque arbre de décision génère une classification, puis la classe finale est déterminée par un vote à la majorité sur l'ensemble de ces arbres. Le résultat de l'agrégation des votes (probabilités d'appartenance) de chaque arbre pour toutes les classes est conservé dans un $VPAC_{RF}$ (vecteur de probabilité d'appartenance aux classes issu de l'algorithme de forêt aléatoire). Un $VPAC_{RF}$ correspond à la synthèse de la classification de l'information satellitaire.

L'intérêt de cet algorithme est qu'il est basé sur les arbres de décisions qui ne sont pas des classifieurs paramétriques. L'algorithme peut ainsi modéliser arbitrairement et sans hypothèse de départ des relations complexes entre les données d'entrées et les résultats, et sélectionner les variables les plus pertinentes, limitant l'effet du bruit statistique causé par des variables inutiles[7].

La paramétrisation du modèle de forêts aléatoires a été effectuée en faisant une recherche sur une grille de six paramètres définis selon 300 combinaisons aléatoires (tableau 7). La combinaison obtenant le plus faible taux d'erreur sur le jeu d'apprentissage et sur le jeu de test a été sélectionnée.

[6] pour que le segment soit conservé et les statistiques zonales calculées, il doit être supérieur à une aire de 10 m²

[7] Dans un modèle de forêts aléatoires, le bootstrapping est une méthode de réplication multiple des données à partir des données d'apprentissage, selon les techniques de ré-échantillonnage. Lorsque celui-ci n'est pas utilisé, cela signifie qu'il n'y a pas de variation aléatoire entre les arbres, mais que la subdivision aléatoire des variables à chaque noeud est toujours aléatoire. Dans ce cas, le jeu d'apprentissage complet est toujours utilisé sans remplacement.

Table 7 Tableau récapitulatif du paramétrage de l'algorithme de forêts aléatoires et des paramètres sélectionnés.

Paramètres	Grille de test	Paramètres sélectionnés
Nombre d'arbres	200 à 2000, par pas de 10	400
Nombres de variables à prendre en compte à chaque séparation	automatique ou $\sqrt{\text{nb variables}}$	automatique
Profondeur maximum des arbres	10 à 110, par pas de 10	110
Nombre minimum d'échantillons pour séparer un noeud	[2, 5, 10, 15, 20]	5
Nombre minimum d'échantillons requis à chaque branche	[1, 2, 4, 6, 8, 10]	1
Méthode de sélection des échantillons pour l'entraînement de chaque arbre	Avec ou sans *bootstrap*	Sans *bootstrap*

À l'aide des variables extraites des séries temporelles d'images satellitaires et des données thématiques, chaque segment de l'image THRS. Les résultats de cette classification, soit la synthèse des données de télédétection, sont résumés dans un $VPAC_{RF}$ pour chaque segment.

2 De la connaissance à la probabilité d'appartenance : les règles expertes

Les modules 2 : règles spatiales et 4 : modélisation de la répartition spatiale des classes s'appuient sur la formalisation des connaissances en règles expertes. Dans le cas de l'application du modèle conceptuel à la région des Hautes Terres du Vakinankaratra, les connaissances sur les stratégies et les choix des agriculteurs pour la mise en culture des terres ont été identifiées à partir des entretiens, des enquêtes et de la littérature (chapitre 3, section 2.2.3 Entretiens et enquêtes de terrain). La première sous-section présente la sélection des connaissances à formaliser à travers les ouvrages techniques et scientifiques et les enquêtes. Dans la deuxième section, nous expliquons comment nous avons procédé pour formaliser les connaissances à l'aide de règles de production (*i.e.* si ... alors ...). La dernière section est dédiée à l'estimation des probabilités à attribuer aux règles expertes pour leur utilisation dans les modules spatiaux.

2.1 Sélection des connaissances à formaliser

L'objectif de cette thèse n'étant pas d'identifier *ex nihilo* les connaissances des agriculteurs, nous posons l'hypothèse que les connaissances identifiées à partir de la littérature scientifique et des rapports techniques sont fiables. La même logique est employée pour les entretiens avec les responsables locaux. Ces personnes, bien que non scientifiques et pas forcément expertes, ont été choisies pour leurs fonctions au sein de la commune ou du Fokontany, leur connaissance de l'agriculture et du fonctionnement agricole de la commune et leur ancienneté.

L'identification des connaissances dans la littérature scientifique et dans les rapports techniques est assez aisée. Les connaissances y sont généralement exprimées de manière claire et de façon quantitative. Ces rapports et la littérature scientifique fournissent aussi les détails et le contexte de la collecte des données et de leur analyse.

La pomme de terre dans un rapport technique

Dans le rapport de Bélières (2020a) sur l'étude de la chaîne de valeur pomme de terre, les informations sont nombreuses et quantifiées.
Le rapport fournit une estimation de l'importance d'une connaissance dans les décisions ou les stratégies des agriculteurs :
« l'importance de la production de pommes de terre est en relation avec l'altitude » (p. 21)
Qui est par la suite étayée par des informations quantitatives : « Riz pluvial et pomme de terre sont plus ou moins cultivés selon les zones avec une évolution opposée : la pomme de terre est très cultivées dans les zones de très haute altitude *[ndlr. > 1700 m]* avec 94 % des EA *[ndlr. exploitations agricoles]* qui en ont dans leur assolement annuel, soit plus que le riz irrigué, puis la part diminue avec 54 % dans la zone de haute altitude *[ndlr. 1300 – 1700 m]* et enfin 19 % dans la zone de moyenne altitude *[ndlr. 900 – 1300 m]* ; pour le riz pluvial il y a peu de producteurs (18 %) en zone de très haute altitude à cause des températures froides et un taux qui progresse vers les températures plus chaudes : 39 % pour la zone de Haute Altitude et 74 % dans le Moyen Ouest du Vakinankaratra. » (p.35).

Cet exemple est réutilisé ci-après afin de mieux comprendre les étapes de la formalisation des règles.

À la différence des écrits scientifiques et techniques, les connaissances exprimées lors des entretiens le sont dans le langage naturel à l'aide de termes approximatifs et relatifs tels que « loin », « proche », « important » ou « sol de couleur noire ». La principale raison de l'expression qualitative des connaissances est que les responsables locaux n'ont pas d'instruments de mesure et n'ont généralement pas mené d'études sur ces questions. Les informations sont rarement quantitatives, excepté lorsque cela était demandé de manière explicite. Leur vraisemblance est faible, car les personnes interrogées préfèrent parfois fournir une réponse hasardeuse plutôt que de dire qu'elles ne savent pas. Les valeurs quantitatives exprimées n'ont dans ces cas-là pas été utilisées pour l'identification de connaissances utiles au module 2 : règles expertes. Par contre, les estima-

tions des proportions de chaque culture dans la commune ont été mobilisées dans le module 4 : modélisation de la répartition spatiale des classes.

Les connaissances exprimées par les responsables ou les agriculteurs sont variées :

« Les légumes représentent un tiers des cultures et peuvent être sur deux types d'emplacements. Soit ils sont proches des maisons et ce sont de petites parcelles, soit ce sont de grandes parcelles plus loin, à environ 500 mètres. »[8]

« Les trop fortes pentes ne sont pas cultivées, mais il n'y a pas de vraie limite. »[9]

« Si les bas-fonds sont trop loin, nous ne les cultivons pas, ce sont nos enfants qui le font. Si personne ne peut les cultiver, alors ils ne sont pas mis en culture. »[10]

« Il y a des vergers surtout derrière les collines. »[12]

L'ensemble de ces connaissances collectées au travers des enquêtes auprès des agriculteurs et des entretiens avec les responsables locaux sont disponibles dans les annexes E et F.

Pour que l'application du modèle ne soit pas trop spécifique à notre zone d'étude, seules les connaissances générales et valides pour la majorité de la zone d'étude ont été conservées. Les spécificités de certains Fokontany sont donc écartées.

La sélection des connaissances s'est tout d'abord basée sur la base de la littérature scientifique et technique. Le choix des connaissances exprimées par les responsables locaux s'est révélé plus compliqué, car il était nécessaire de prendre en compte la compétence des personnes interrogées. Le niveau de compétence a été évalué a posteriori en s'appuyant sur les avis de Eloise Rasoamalala et de Valérie Andriamanga (ingénieurs agronomes malgaches, ayant participé et traduit les entretiens). Si les connaissances exprimées divergeaient de celles identifiées de la littérature, alors celles-ci n'étaient pas sélectionnées, l'intérêt des enquêtes étant d'identifier des connaissances non présentes dans la littérature. L'estimation de la répartition des cultures par les responsables locaux a été particulièrement utile pour l'application du module 4 : modélisation de la répartition spatiale des classes.

[8] Entretien avec le chef du Fokontany d'Ambano le 29/04/2019.

[9] Entretien avec le maire d'Ambohimiarivo le 30/04/2019.

[10] Entretien n°19 avec un agriculteur dans le Fokontany d'Ambatolahy le 16/03/2018. Pour comprendre cette citation, il est important de la replacer dans le contexte malgache. Le riz est la source principale de l'alimentation, tandis que les légumes sont des cultures commerciales très intéressantes à vendre pour les agriculteurs. Les bas-fonds permettent, selon leurs situations, de cultiver du riz irrigué ou des légumes lors de la saison des pluies et de cultiver des légumes en contre-saison[11]. Le fait de choisir de ne pas cultiver un bas-fond est donc un cas particulier.

[12] Entretien n°27 avec un agriculteur dans le Fokontany d'Ambatolampy, le 22/03/2018.

2.2 Formalisation des connaissances en règles

Pour utiliser des connaissances dans le modèle, celles-ci sont représentées sous forme de règles de production (Chomsky, 1959). Cette forme **SI condition ALORS action** permet de produire de nouvelles informations à partir de connaissances. Dans notre cas, la nouvelle information produite est la possibilité[13] de trouver une occupation ou un usage des sols pour un segment.

Lorsque les connaissances ne sont pas exprimées de manière quantitative (utilisation des termes « loin », « proche », « important » ou « sol noir »), il est impératif de les relier à des valeurs quantitatives pour pouvoir les formaliser en règles.

Pour les variables quantitatives (*i.e.* distance, altitude, pente, et temps de trajet), les termes tels que « loin », « proche » ou « haut » sont associés à une valeur quantitative. Pour créer ce lien, nous demandions à la personne interrogée de nous montrer à quoi correspondait le terme utilisé. Par exemple, pour le terme « loin », nous demandions à la personne de nous montrer dans le paysage environnant, ou sur une carte, où se situait un champ considéré comme « loin ». Ensuite, nous reportions sur une carte l'emplacement du point « lointain ». Cette méthode nous a permis d'estimer une gamme de distance et de l'associer à chaque terme utilisé.

Pour les variables qualitatives (*i.e.* type de sol ou géomorphologie) pouvant être exprimées sous la forme « sol de couleur noire » ou « colline », nous avons directement rattaché le terme utilisé à la classe associée (ex. le terme « colline » peut être associé à la classe de relief *6*, *8* ou *9* selon le contexte dans lequel il est exprimé (Annexe G pour les classes de relief)). Pour les termes liés à la pédologie, la carte pédologique (Riquier, 1968) a servi de référence avec Raunet (1997) et les scientifiques ont permis d'apporter des compléments d'information.

Si l'information donnée lors d'un entretien n'est pas assez précise et qu'elle ne peut être rattachée à une valeur, alors elle n'est pas conservée.

[13] Le terme *possibilité* est volontairement employé, car il correspond à des emplacements optimaux, mais le choix de mise en culture de la terre dépend des agriculteurs.

De la connaissance issue du rapport technique à la règle

Dans le rapport de Bélières (2020a) sur l'étude de la chaîne de valeur pomme de terre, l'altitude a été identifiée comme une variable importante pour la culture de la pomme de terre. Dans ce rapport, trois gammes d'altitudes sont analysées : la très haute altitude (> 1700 m), la haute altitude (1300 – 1700 m) et la moyenne altitude (900 – 1300 m). La part des exploitations agricoles qui cultivent la pomme de terre est aussi exprimée par gamme d'altitude. Cette proportion exprime un ordre d'idée quant à l'importance de la culture de pomme de terre dans ces altitudes. En revanche, le rapport ne spécifie pas la différence entre les cultures de saison culturale principale et les cultures de contre-saison. De manière approximative et subjective, des classes quantitatives sont associées à chacune des altitudes en fonction de la part des ménages.

Une retranscription de cette connaissance en règle permet d'en exprimer trois :
Règle 1 : **Si** (la valeur d'altitude du segment est $\geq$ 900 m et < 1300 m), **alors** (la probabilité d'appartenance à la classe *pomme de terre* est moyenne)
Règle 2 : **Si** (la valeur d'altitude du segment est $\geq$ 1300 m et < 1700 m), **alors** (la probabilité d'appartenance à la classe *pomme de terre* est élevée)
Règle 3 : **Si** (la valeur d'altitude du segment est $\geq$ 1700 m), **alors** (la probabilité d'appartenance à la classe *pomme de terre* est très élevée)

2.3 Attribution de probabilités aux règles (issues de la connaissance)

Les connaissances sélectionnées conduisent à la création de trois types de règles : 1) les règles qui permettent d'**attribuer une probabilité d'appartenance à une classe donnée** (ces règles pouvant être pondérées afin de prendre en compte leur importance, puis agrégées pour l'obtention d'une probabilité finale d'appartenance), 2) les **règles binaires ou exclusives**, qui ont pour but d'éliminer la possibilité d'affectation à une classe ou d'affecter le segment de façon automatique à une classe (ces règles sont utilisées pour déterminer les cultures associées et identifier les espaces qui ne sont pas cultivables) et 3) les **règles de répartition spatiale théorique des types de culture** à l'intérieur d'une entité géographique donnée (ici le Fokontany). L'application des règles a ensuite pour objectif de modifier le vecteur de probabilité d'appartenance aux différentes classes de la nomenclature.

Dans la section précédente, nous avons vu que les connaissances pouvaient être exprimées de manière quantitative (dans les rapports techniques et les articles scientifiques) ou de façon approximative (lors des entretiens et enquêtes). Il est ensuite nécessaire d'estimer l'importance de chaque connaissance dans les décisions et stratégies des agriculteurs. Cette importance est traduite par l'attribution d'une probabilité aux règles qui découlent de la connaissance.

Pour traduire l'importance d'une connaissance (*i.e.* l'action de la règle de production), nous

proposons un nuancier de six niveaux d'importance, allant de *très faible* à *extrême*[14]. Notre choix a été de quantifier l'importance de manière linéaire, *très faible* correspondant à l'importance la plus faible (soit une probabilité de 1/6) et *extrême* à l'importance la plus forte (soit une probabilité de 6/6). Le choix de la gamme d'importance est basé sur la récurrence d'une connaissance dans les discours des responsables locaux, l'estimation de cette importance d'après ces responsables locaux (*i.e.* pour la culture de pomme de terre, l'altitude a une plus forte influence que l'exposition du champ), d'après les rapports techniques, la littérature scientifique et sur la qualité des données support pour l'application des règles. Lorsqu'une règle n'est pas exprimée, son importance est nulle (soit une valeur de 0).

Dans notre modèle, c'est l'ensemble des règles d'une classe qui permet de déterminer si une culture peut être cultivée ou non, en faisant varier les probabilités d'appartenance à la culture donnée. Chaque règle entraînant une modification de la probabilité d'appartenance à la classe/culture, il est nécessaire d'agréger l'ensemble des règles et probabilités associées à une même classe.

Les règles d'une même classe sont agrégées avec un opérateur de somme pour fournir la valeur d'appartenance finale à la classe étudiée. L'ensemble des probabilités d'appartenance d'un segment aux différentes classes est regroupé dans un $VPAC_{RS}$ (vecteur de probabilité d'appartenance aux classes issu du module des règles spatiales). Ces étapes d'agrégation sont détaillées dans la section 3.3.1 Standardisation des résultats et création des $VPAC_{RS}$.

Dans le cas où une règle est binaire et donc exclusive (*i.e.* si le segment ne respecte pas la règle, il est exclu d'une classe ou bien affecté automatiquement à une autre), la règle sera formalisée ainsi :

Si la pente est $\geq$ à 35 %, **alors** le segment **doit** être affecté à une classe non agricole.

Ainsi, le segment ne pourra avoir qu'une probabilité binaire entre la classe agricole et la classe non agricole. Ensuite, ces probabilités peuvent être affinées en fonction des sous-ensembles de règles agricoles si le segment est considéré comme appartenant à une classe agricole.

Dans le cas des règles de répartitions spatiales, les connaissances sont exprimées sous la forme d'un pourcentage d'aire de chaque classe présente dans le Fokontany.

[14] Les niveaux d'importance, ou probabilités, sont *très faible*, *faible*, *moyenne*, *élevée*, *très élevée* et *extrême*.

Estimation des répartitions spatiales théoriques par culture

La répartition des cultures par entité administrative est difficile à estimer, car il existe peu de sources de données agricoles sur Madagascar. La base de données *microdata at FAO*[a] ne contient aucune donnée sur Madagascar. La base de données *CountryStat* de la FAO[b] ne recense que des données à une échelle régionale ou à l'échelle administrative des districts mais seulement pour la période 1991-1999. Ne pouvant utiliser des statistiques vérifiées, nous nous sommes basés sur les recensements agricoles des monographies des communes et sur nos questionnaires auprès des maires ou des responsables agricoles locaux. Les répartitions des cultures sont généralement estimées au niveau communal pour les cultures les plus importantes. Les monographies ne font pas de différence entre les différentes saisons culturales. Ces données sont donc des estimations non vérifiées, qui n'ont pas été utilisées.

La qualité des monographies et des enquêtes étant variable, nous avons choisi de regrouper les Fokontany en groupes lorsque les répartitions théoriques n'étaient pas assez précises. Au total, douze répartitions théoriques ont été pensées. Certaines correspondent à un Fokontany, d'autres à un groupement de Fokontany homogène dans une commune, ou à une commune entière. Une répartition théorique est utilisée pour les communes ou les Fokontany dont les données n'étaient pas disponibles ou pas suffisamment fiables. Les répartitions de culture collectées auprès des experts locaux sont disponibles en Annexe F.

[a] https://microdata.fao.org/index.php/catalog/
[b] http://madagascar.countrystat.org

Dans cette partie, nous avons décrit la méthode de sélection des connaissances à formaliser en règles expertes selon la logique des systèmes experts, ainsi que la méthode d'attribution de probabilités aux règles (de très faible à extrême). Cette pondération permet de prendre en compte l'importance de chaque règle dans l'attribution d'une probabilité d'appartenance à une classe donnée. À l'issue de l'application du module 2 : règles spatiales, les probabilités finales d'appartenance aux classes constituent le $VPAC_{RS}$. Les règles mobilisées dans le module 4 : modélisation de la répartition spatiale des classes, correspondent quant à elles à des pourcentages d'aire de chaque classe.

3 Application des modules

Cette section présente la préparation des données support pour l'application des connaissances, puis la méthode d'application à la région des Hautes Terres Malgaches des deux modules spatiaux de notre modèle conceptuel : module 2 : **règles spatiales** et le module 4 : **modélisation de la répartition spatiale des classes**.

3.1 Préparation de données support pour l'application des connaissances

Certaines règles identifiées lors de la formalisation des connaissances s'appuient sur des données support présentées en section 2.3 Données support pour l'application des connaissances du chapitre 3. Le bon fonctionnement des processus de modélisation dépend de la qualité de ces données. Cette partie est dédiée à la façon dont les données et variables support à l'application des règles sont préparées pour les intégrer à chaque segment.

Pour appliquer les processus de modélisation, les données support pour l'application des connaissances mobilisées par les modules doivent être intégrées comme propriétés des segments. Ces propriétés s'ajoutent ainsi pour chaque segment à celles issues des données et classification de télédétection (*i.e.* l'ensemble des statistiques zonales ainsi que le $VPAC_{RF}$).

Deux grands types de données sont utilisées comme support pour l'application des connaissances : les grilles, principalement dérivées du MNT, et les données vectorielles (tableau 8).

Table 8 Données utilisées dans le modèle.

Données	**Type**
Altitude	Grille quantitative continue
Pente	Grille quantitative continue
Exposition	Grille quantitative continue
Temps de trajet depuis un bâtiment	Grille quantitative continue
Route nationale 7	Géométrie linéaire
Autres routes	Géométrie linéaire
Cours d'eau	Géométrie linéaire
Bâti (issu de la classification des données de télédétection)	Géométrie polygonale nominale ponctuelle
Type de relief (géomorphologie)	Géométrie polygonale nominale continue
Pédologie	Géométrie polygonale nominale continue
Communes	Géométrie polygonale nominale continue

Pour affecter à un segment une propriété issue d'une donnée de type grille quantitative continue, la méthode utilisée est similaire à celle des statistiques zonales (section 1.1.3 Statistiques zonales). Elle consiste à attribuer au segment la valeur moyenne des cellules qui le composent. Pour qu'une cellule soit attribuée à un segment, il faut qu'au moins la moitié de sa surface soit incluse dans le segment.

Afin d'attribuer à chaque segment sa distance avec des données ayant des géométries linéaires ou ponctuelles, la distance euclidienne est calculée entre la bordure du segment et la bordure de l'élément. Puis, parmi les valeurs candidates, une fonction d'agrégation minimale permet d'affecter la distance la plus faible au segment. Cette méthode est appliquée aux routes, les cours d'eau, ainsi que pour les segments de la classe *bâti*.

Pour affecter à un segment une propriété issue d'une donnée de type géométrie polygonale nominale continue, correspondant ici à des surfaces vectorielles, l'aire de chaque classe du vecteur est calculée pour le segment. La classe ayant l'aire la plus importante est retenue pour être affectée aux propriétés du segment. Cette méthode est utilisée pour le type de relief, la pédologie et les communes.

Cette étape permet de regrouper toutes les données et variables nécessaires pour l'application des connaissances dans les deux modules spatiaux : le module 2 : règles spatiales et le module 4 : modélisation de la répartition spatiale des classes.

La première section de ce chapitre a présenté le traitement des données de télédétection qui a résulté en la création d'une segmentation qui représente la structure de l'espace étudié et le calcul d'un $VPAC_{RF}$ (vecteur de probabilité d'appartenance aux classes issu de l'algorithme de forêt aléatoire) à partir de la classification des données de télédétection (notée **Classe RF**). La seconde section a présenté la formalisation des connaissances en règles expertes et la façon dont celles-ci sont agrégées pour estimer la probabilité d'observer une classe sur un segment. Une fois que les segments contiennent les données/variables support pour l'application des connaissances et les $VPAC_{RF}$ (sous-section précédente), les connaissances peuvent être appliquées pour modéliser les stratégies des agriculteurs. Cette partie détaille les quatre étapes de la partie modélisation du modèle conceptuel pour la caractérisation de l'agriculture familiale dans les Hautes Terres du Vakinankaratra à l'aide de la plateforme Ocelet. La figure 18 présente les étapes de l'application du modèle avec les deux modules spatiaux tel que présenté sommairement dans le chapitre 2, section 3.2 (figure 10) :

1. sélection des segments pouvant être modifiés. Cette sélection se fait en utilisant le $VPAC_{RF}$ et l'application de règles spatiales (figure 18 : **1**) ;
2. calcul de probabilité de présence de dix classes et d'une classe générale pour les segments sélectionnés à partir de scénario de culture via l'application de règles spatiales.
   Agrégation et standardisation des probabilités de culture pour créer un $VPAC_{RS}$ pour chaque segment (figure 18 : **2**) ;
3. fusion des $VPAC_{RF}$ et des $VPAC_{RS}$ pour créer un $VPAC_{Fusion}$ et déterminer la classe finale 1 (figure 18 : **3**) ;
4. application du module 4 : modélisation de la répartition spatiale des classes pour prendre en compte la répartition théorique des classes par unité administrative de Fokontany. Nous obtenons à l'issue de l'application de ce module, une **classe de répartition statistique** (figure 18 : **4**).

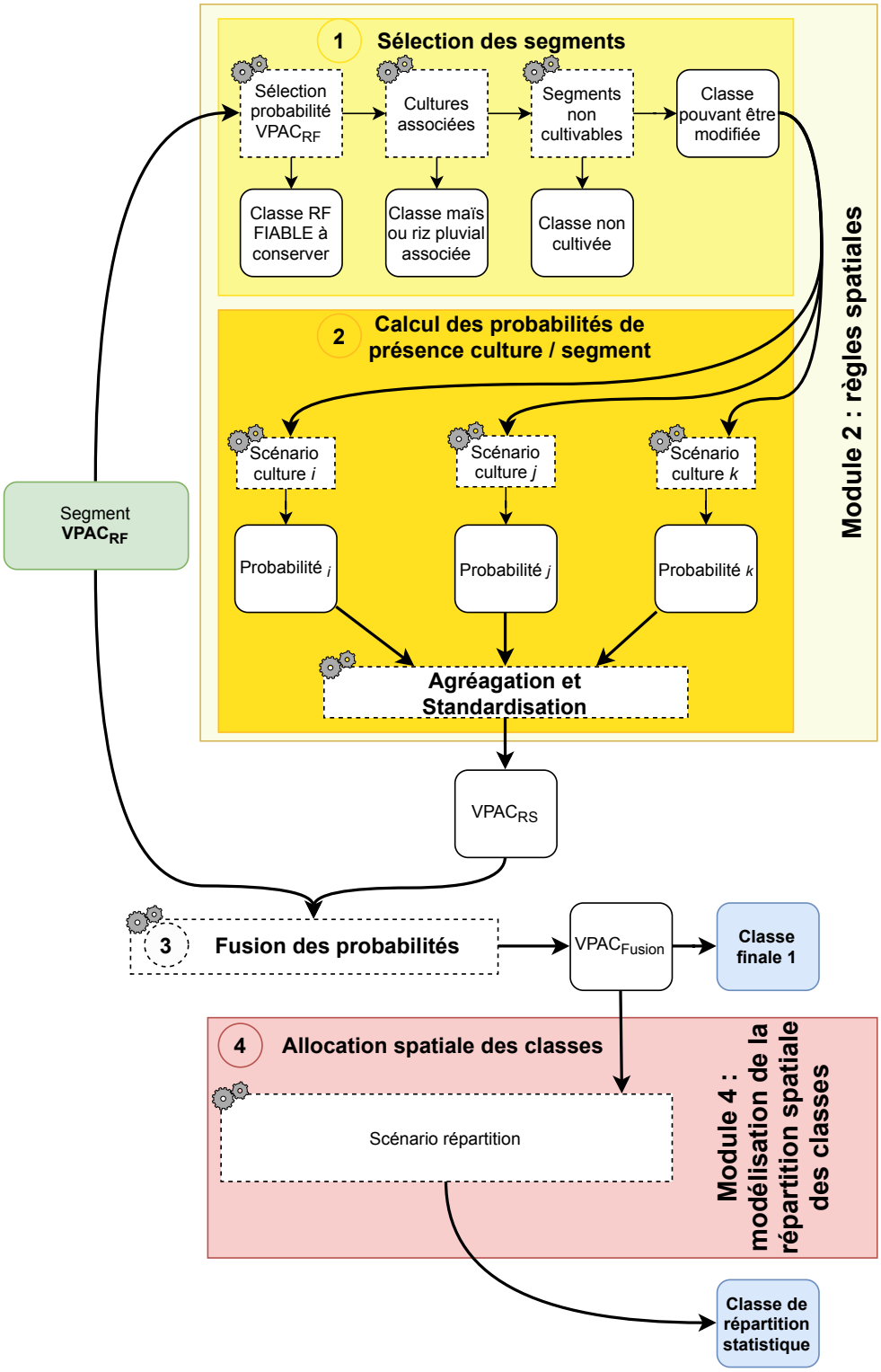


Figure 18 Schéma des étapes de l'application du modèle conceptuel.

3.2 Sélection des segments

En se basant sur des règles simples, les segments ne nécessitant pas de reclassification sont isolés selon 3 étapes de sélection : 1) la sélection des segments[15], 2) la sélection des segments pouvant correspondre à une culture associée, et 3) la sélection des segments pouvant correspondre à une classe non cultivée. Les segments restants sont ceux nécessitant une reclassification par l'étape des scénarios de cultures.

3.2.1 Sélection selon la fiabilité des $VPAC_{RF}$

Afin d'identifier les segments dont le $VPAC_{RF}$ est suffisamment fiable pour que la classe RF qui leur a été attribuée soit conservée, une sélection est effectuée en deux temps (figure 19) :

1. un « **seuil de probabilité minimum** » compris entre 0 et 1 correspond à la valeur au-delà de laquelle le $VPAC_{RF}$ et la classe RF (*i.e.* la classe avec la probabilité la plus élevée) du segment doivent être conservés. Un seuil proche de 1 aura pour conséquence un grand nombre de segments à modifier, car seuls les segments avec de fortes probabilités seront conservés. Au contraire, plus le seuil est proche de 0 et moins de segments pourront être modifiés.
   Si la probabilité d'au moins une classe dépasse ce seuil, alors le $VPAC_{RF}$ et la classe RF sont conservés. Sinon la seconde étape est appliquée ;
2. une « **différence de probabilité minimum** », ΔRF, compris entre 0 et 1 correspond à la différence minimale entre les deux probabilités les plus élevées du $VPAC_{RF}$. Un ΔRF élevé (ex. 0,8 ou 0,9) permettra que conserver la classe RF des segments qui ont un écart important entre la probabilité la plus élevée et la seconde probabilité la plus élevée, alors qu'un ΔRF faible (ex. 0,1 ou 0,2) correspond à un petit écart entre les deux probabilités les plus élevées, donc un grand nombre de segments conserveront leur la classe RF.

Cette première étape est fondamentale, car c'est elle qui détermine quels sont les segments qui vont être traités par les modules de modélisation. Le seuil ΔRF est l'un des paramètres principaux du modèle, car c'est lui qui a le contrôle le plus important sur les segments à modifier.

Pour appliquer le modèle conceptuel, 10 combinaisons de 0,1 à 1 ont été testées à la fois pour le « seuil de probabilité minimum » et pour la « différence de probabilité minimum », ΔRF,

[15] Sur la base des valeurs du $VPAC_{RF}$ de probabilité d'appartenance aux classes

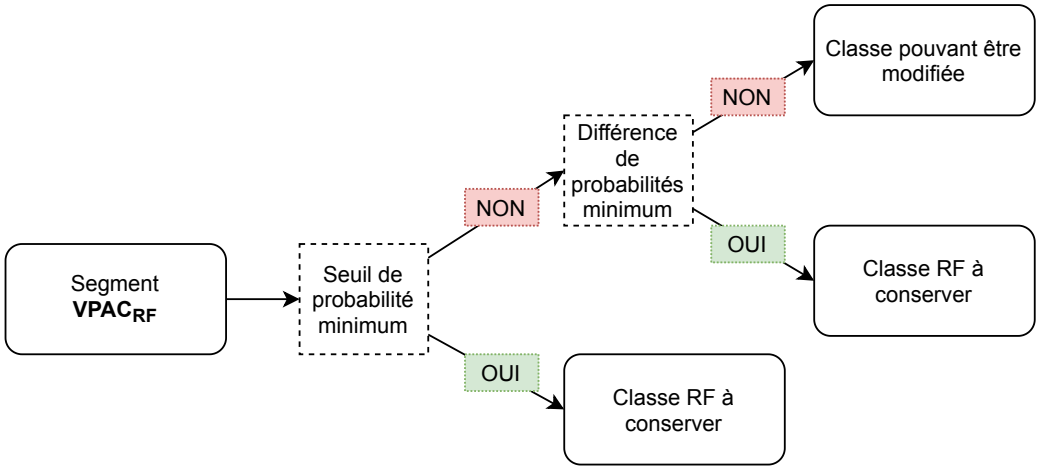


Figure 19 Sélection des segments traités par les modules de modélisation, basée sur le $VPAC_{RF}$.

3.2.2 Sélection des segments de cultures associées à partir des $VPAC_{RF}$

Une seconde étape de sélection vise à identifier et traiter les segments présentant une confusion entre les cultures pluviales *pures* et les cultures pluviales *associées* lors de la classification des données de télédétection. Les cultures associées sont au nombre de 2 dans notre nomenclature : la classe *maïs associé* (la culture associée au maïs peut être du soja ou du riz pluvial) et la classe *riz pluvial associé* (le riz pluvial lorsqu'il est associé l'est toujours avec du maïs). Cette sélection est effectuée en se basant sur les probabilités d'appartenance aux différentes classes du $VPAC_{RF}$, la figure 20 présente les différentes étapes appliquées qui sont les suivantes :

1. Une première étape **compare individuellement les probabilités de deux classes pouvant être associées** à un seuil de probabilité d'appartenance (seuil1) de **0,4**. Les classes concernées sont le *maïs*, le *soja* ou le *riz pluvial*. Si la probabilité de chacune des classes est supérieure ou égale au seuil, alors le segment est affecté à une classe de *culture associée*. Pour la combinaison *maïs-riz pluvial*, nous cherchons la culture majoritaire pour l'affecter à la classe associée. En cas d'égalité, c'est la classe *riz pluvial* qui est considérée comme la culture dominante, car cette classe occupe la plus grande surface dans la parcelle.[16] Si ces premiers critères de sélection ne sont pas validés, alors une seconde sélection est testée ;

2. la seconde sélection est appliquée uniquement pour les classes *maïs* et *riz pluvial* qui sont les associations les plus fréquentes. Cette sélection est exécutée en trois temps :
   2.1. comme pour l'étape précédente, la classe *maïs* et la classe *riz pluvial* sont chacune testées par rapport à un seuil de probabilité d'appartenance (seuil2) de **0,6**, mais doivent être inférieures ou égales à celui-ci ;

   2.2. ensuite, les probabilités du $VPAC_{RF}$ des classes *maïs*, *riz pluvial* et *maïs associé à une autre culture* sont additionnées. Il en est de même pour les probabilités des classes *maïs*, *riz pluvial* et *riz pluvial associé à une autre culture*. Si ce cumul est supérieur à **0,8**

[16] Lors de l'association entre du riz pluvial et du maïs, soit le maïs sert à entourer un assolement, soit les rangées de maïs et de riz pluvial sont alternées avec une fréquence allant de 3 à 5 rangées de riz pluvial pour une rangée de maïs.

(seuil3), alors le segment peut correspondre à une culture associée. Pour déterminer cela, une dernière vérification est effectuée ;

2.3. pour que la classe du segment soit changée en classe *maïs associé à une autre culture* il faut que la probabilité de la classe *maïs* soit inférieure à la probabilité de la classe *riz pluvial* et que la probabilité de la classe *maïs associé à une autre culture* soit inférieure à la classe *riz pluvial associé à une autre culture.*
Pour que la classe du segment soit changée en classe *riz pluvial associé à une autre culture* il faut que la probabilité de la classe *riz pluvial* soit inférieure ou égale à la probabilité de la classe *maïs* et que la probabilité de la classe *riz pluvial associé à une autre culture* soit inférieure ou égale à la classe *maïs associé à une autre culture.*
Si aucune de ces conditions n'est remplie, alors la classe du segment n'est pas changée.

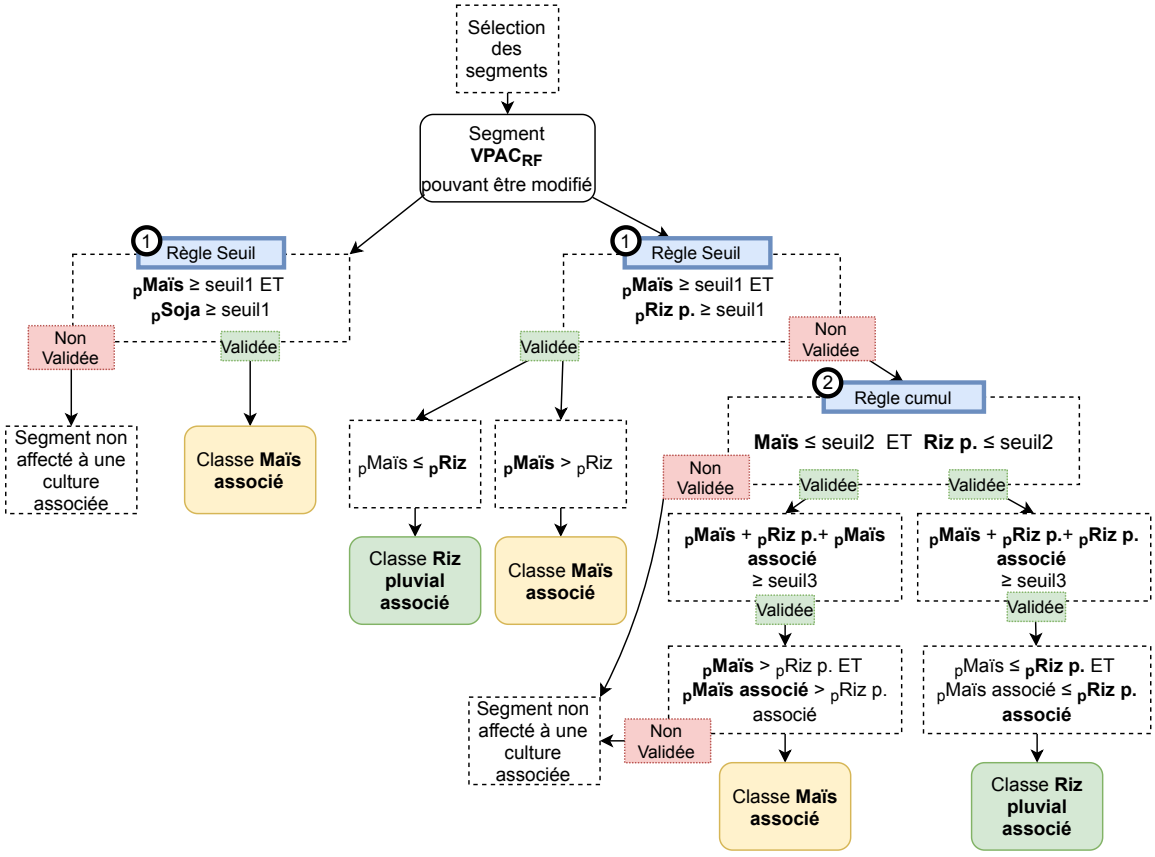


Figure 20 Sélection des segments pouvant être affectés aux classes d'associations de cultures. pCulture signifie la probabilité de la culture. Riz p. est l'acronyme de riz pluvial. Dans notre cas seuil1 = 0,4, seuil2 = 0,6 et seuil3 = 0,8.

À l'issue de cette étape sont donc identifiés les segments dont la classe peut être modifiée pour une des deux classes de culture associée à partir de l'analyse des $VPAC_{RF}$.

3.2.3 Sélection des segments non-agricoles

La dernière étape consiste à identifier les segments ne pouvant appartenir à une classe agricole à partir de règles expertes binaires. Une fois identifiés, les segments ne sont plus considérés comme pouvant être modifiés par les étapes du modèle. Une étiquette *Autre*, ou classe temporaire, est affectée aux segments. Une classe de la nomenclature sera par la suite attribuée à ces segments à partir de la probabilité de la classe la plus élevée. Pour identifier les segments non agricoles, trois règles expertes sont utilisées en s'appuyant sur les données support pour l'application des connaissances.

Les règles ont été établies selon la méthode de formalisation des connaissances présentée en section 2 De la connaissance à la probabilité d'appartenance : les règles expertes . Les variables d'exclusion sont la **pente**, l'**altitude** et le **temps de trajet à pied entre un bâtiment et un segment** (tableau : 9). Tous les segments ayant des valeurs supérieures aux valeurs fixées pour l'une des variables d'exclusion sont étiquetés comme *Autre*.

Table 9 Variables et seuils utilisés pour sélectionner les segments non-agricoles.

Variables d'exclusion	Seuils
Altitude	$\geq$ 2200 mètres
Pente	$\geq$ 35 degrés
Temps de trajet à pied depuis un bâtiment	$\geq$ 30 minutes

Dans cette sous-section nous avons vu les méthodes de sélection : des segments dont la classe RF peut être conservée du fait d'une fiabilité des valeurs des $VPAC_{RF}$, des segments pouvant correspondre à une classe de *culture associée* toujours à partir des $VPAC_{RF}$ et des segments non agricoles à partir de règles expertes binaires. Tous ces segments n'entreront pas dans les autres étapes de modélisation. Les segments restants, eux, sont considérés comme des segments dont la classe pourra être modifiée dans la suite du module.

3.3 Calcul des probabilités de présence de culture par segment

L'objectif du module des règles spatiales est d'identifier des contextes spatiaux favorables ou défavorables à chacun des systèmes de culture étudiés. C'est-à-dire de **modéliser la possibilité de retrouver chaque type de culture dans un segment**. Les règles spatiales sont basées sur des connaissances formalisées qui permettent de mettre en lien les caractéristiques biophysiques des objets, les stratégies agronomiques des agriculteurs et les contextes nécessaires à l'implantation de chaque type de culture. Pour cela, plusieurs scénarios sont conçus et chacun est composé d'un ensemble de règles spatiales qui est appliqué à tous les segments dont la classe peut être changée.

Ce module contient un scénario de modélisation pour chaque classe de culture et un scénario qui regroupe les espaces non cultivés (les classes *culture boisée de rente*, *sol nu agricole* et *herbes et autres cultures fourragères*), soit **onze scénarios**. Cette section détaille une partie des éléments développés dans la section 2 De la connaissance à la probabilité d'appartenance : les règles expertes

pour comprendre la création et l'application des scénarios.

Tous les scénarios partagent une structure commune composée de règles pouvant être prises en compte. Chaque règle est associée à une probabilité allant de *très faible* à *extrême* pour représenter l'importance de chaque règle dans l'attribution d'une probabilité d'appartenance au segment, comme défini dans la section 2 De la connaissance à la probabilité d'appartenance : les règles expertes .

Les connaissances, formalisées en plusieurs règles, se différencient en dix catégories. Les connaissances liées à l'influence de facteurs biophysiques sur le segment (composé de la pédologie, de l'altitude, de la pente, de l'exposition, de la géomorphologie) peuvent être distinguées de celles liées aux influences facteurs « socio-économiques » sur les segments (distance à un cours d'eau, temps de trajet entre un bâtiment et un segment, distance visuelle, distance aux routes secondaires et distance à la route nationale). Un scénario peut comporter une ou plusieurs de ces connaissances formalisées. Le nombre maximum de connaissances pouvant être mobilisées est au nombre de huit (classe *légumes*). Le tableau 10 présente la correspondance entre les classes et les scénarios utilisés dans le module règles spatiales. L'Annexe J présente toutes les règles appliquées dans le modèle et le type de valeur attribué à chaque règle.

Des connaissances formalisées aux scénarios

Les connaissances mobilisées pour déterminer l'emplacement possible d'une culture peuvent être facilement identifiables (nombreux rapports techniques sur une culture, contraintes biophysiques importantes), ou au contraire ne dépendre que des choix des agriculteurs (culture pouvant être cultivée en tous lieux).

À titre d'exemple, la classe *cultures potagères* (cultures de taro, patate douce, manioc) rassemble des cultures qui sont présentes dans toutes les exploitations agricoles[a]. Même si les parcelles que possèdent un agriculteur ne sont pas favorables à la mise en culture de ces plantes, une parcelle sera nécessairement consacrée à ces cultures. Pour la classe *cultures potagères*, il est donc difficile d'identifier des connaissances qui permettent de déterminer un emplacement favorable pour ces cultures. La seule connaissance qui peut être identifiée comme générale pour le choix de l'emplacement de la classe *cultures potagères* est la pente. Le **scénario *cultures potagères* est donc composé d'une connaissance : la pente**.

Pour la classe *pomme de terre*, dont les connaissances ont été identifiées, entre autres, à l'aide du rapport de Bélières (2020a) et des enquêtes auprès des responsables locaux (Annexe F), son **scénario est composé de cinq connaissances : l'altitude, le temps de trajet entre un bâtiment et un segment, le type de relief, la pente et la pédologie**.

[a] Ces cultures sont choisies, car elles ont un cycle plus long que les cultures majoritaires les plus consommées (riz, maïs) et sont consommées en période de soudure (*i.e.* « La soudure est la période de l'année précédant les premières récoltes, et où le grain de la récolte précédente peut venir à manquer » Sourisseau et al. (2016), p.112).

Lors du déroulement du scénario, une ou plusieurs connaissances (*i.e.* règles) peuvent être validées par le segment. Une connaissance est considérée comme validée quand la probabilité (*i.e.*

l'importance de la règle) affecté à la règle correspond aux probabilités *moyen*, *élevé*, *très élevé* et *extrême*. Une fois que toutes les connaissances ont été parcourues, les sommes des probabilités des règles et la somme des connaissances validées sont calculées.

Exemples de scénarios pour les classes *pomme de terre* et *cultures potagères* d'un même segment

Le scénario de la classe *pomme de terre* est composé de cinq connaissances : l'altitude, le temps de trajet entre un bâtiment et un segment, le type de relief, la pente et la pédologie (Annexe J). Le déroulement du scénario pour le traitement d'un segment va être :

Connaissance 1 : l'altitude du segment est de 1782 m, **alors** la probabilité d'appartenance à la classe *pomme de terre* est **très élevée** (valeur : 5/6) → **validée**.
Connaissance 2 : le temps de trajet entre un bâtiment et ce segment est de 14 minutes, **alors** la probabilité d'appartenance à la classe *pomme de terre* est **faible** (valeur : 2/6) → **non validée**.
Connaissance 3 : la classe de relief du segment est 4 (« sols à mull »), **alors** la probabilité d'appartenance à la classe *pomme de terre* est **moyenne** (valeur : 3/6) → **validée**.
Connaissance 4 : la pente du segment est de 2 %, **alors** la probabilité d'appartenance à la classe *pomme de terre* est **moyenne** (valeur : 3/6) → **validée**.
Connaissance 5 : la classe de pédologie du segment est 6 (pentes de colline ou de cours d'eau incisé), **alors** la probabilité d'appartenance à la classe *pomme de terre* est **moyenne** (valeur : 3/6) → **validée**.

Pour le scénario *pomme de terre*, la **somme des probabilités des règles est de 16/30** (environ 0,53) et la **somme des connaissances validées est de 4**.

Pour le scénario de la classe *cultures potagères*, composée d'une connaissance, la pente, le déroulement du scénario pour le traitement d'un segment est le suivant :

Connaissance 1 : la pente du segment est de 2 %, **alors** la probabilité d'appartenance à la classe *culture potagère* est **très faible** (valeur : 4/6) → **validée**.
Pour le scénario *cultures potagères*, la **somme des probabilités des règles est de 4/5** (environ 0,67) et la **somme des connaissances validées est de 1**.

Table 10 Correspondance entre la nomenclature et les connaissances utilisées pour la création des onze scénarios. La catégorie « Autre » correspond aux classes *culture boisée de rente*, *savane*, *sols rocheux*, *bâti*, *eau*, *herbes et autres cultures fourragères* et *sol nu agricole*. Le nombre maximum de connaissances mobilisées est de 8 (classe *légumes*).

	Connaissances utilisées dans chaque scénario										
	Pédologie	**Altitude**	**Pente**	**Exposition**	**Relief**	**Distance aux cours d'eau**	**Temps de trajet**	**Distance visuelle**	**Distance aux routes secondaires**	**Distance à la route nationale**	**Somme connaissances mobilisées**
Maïs	x		x		x		x	x		x	6
Riz irrigué		x	x		x	x	x				5
Riz pluvial	x	x	x		x		x	x		x	7
Légumes	x		x		x	x	x	x	x	x	8
Soja	x		x		x		x	x			5
Pomme de terre	x	x	x		x		x				5
Verger	x	x	x	x	x		x				6
Cultures potagères			x								1
Maïs associé	x		x		x		x	x		x	6
Riz pluvial associé	x	x	x		x		x	x		x	7
Autre		x	x		x		x				4

3.3.1 Standardisation des résultats et création des $VPAC_{RS}$

Une fois qu'un segment a été traité par tous les scénarios, les résultats de l'application des règles pour chaque classe (somme des probabilités des règles et somme des connaissances validées) doivent être standardisés afin de former un $VPAC_{RS}$ (vecteur de probabilité d'appartenance aux classes issu du module des règles spatiales), dont la somme des valeurs est égale à 1.

La standardisation est calculée à la fois par rapport à la somme des probabilités des règles et à la somme des connaissances validées. Si un segment correspond à un emplacement théoriquement idéal pour une culture (*i.e.* correspondant aux probabilités les plus élevés pour les règles mobilisés dans le scénario), alors la somme des probabilités des règles sera égale à 1. Pour éviter que les scénarios avec peu de connaissances mobilisées dépendent uniquement de la somme des probabilités des règles, cette somme est multipliée par la fraction de connaissance utilisée dans le scénario. La fraction de connaissances est calculée à partir du maximum de connaissances mobilisables dans scénario, soit huit (scénario *légumes*). La standardisation est donc calculée à la fois par rapport à la somme des probabilités des règles et à la fraction des connaissances validées de manière pondérée (équation 14) :

$$\text{Score cult} = \frac{\text{somme des probabilités des règles}}{\text{somme maximale théorique}} \cdot \left(\frac{\text{nombre de connaissances mobilisées}}{\text{maximum de connaissances mobilisables}} \cdot coef \right) \tag{14}$$

où $coef$ est le coefficient de pondération.

Pour former un $VPAC_{RS}$ dont la somme est égale à 1, les **scores cult** sont standardisés entre 0 et 1.

Dans le cas du scénario « Autre », la même valeur est attribuée aux classes concernées.

Une fois l'ensemble des scénarios exécutés pour estimer la possibilité de trouver des classes dans les segments et les $VPAC_{RS}$ constitués, nous pouvons expliquer comment un $VPAC_{Fusion}$ (vecteur de probabilité d'appartenance aux classes finales) est construit et comment la classe finale d'un segment est déterminée.

Résultats des scénarios de *pomme de terre* et de *cultures potagères* pour le même segment

Dans le précédent encart, nous avons présenté une application du scénario *pomme de terre* et du scénario *cultures potagères* à un segment. Ici est présenté le résultat de chaque scénario et d'un $VPAC_{RS}$ qui correspond à deux classes et un $coef$ de score cult de 1 (équation 14).

Pour le scénario *pomme de terre*, la somme des probabilités des règles pour ce scénario est de 16/30 et la somme des connaissances validées est de 4, soit un résultat de **0,27**, calculé ainsi :

$$\text{score } \textit{pomme de terre} = \frac{16}{30} \cdot \left(\frac{4}{8} \cdot 1\right) \tag{15}$$

Pour le scénario *cultures potagères*, la somme des probabilités des règles pour ce scénario est de 4/6 et la somme des connaissances validées est de 1, soit un résultat de **0,08**, calculé ainsi :

.

$$\text{score } \textit{cultures potagères} = \frac{4}{6} \cdot \left(\frac{1}{8} \cdot 1\right) \tag{16}$$

	PdT		Cult. potagères
Avant standardisation :	$[0,27$	;	$0,08]$
$VPAC_{RS}$ standardisé :	$[0,77$	;	$0,23]$

3.3.2 Fusion des probabilités de la télédétection du module 2 : règles spatiales

L'attribution de la classe finale au segment va se faire sur la combinaison des deux VPAC ($VPAC_{RF}$ et $VPAC_{RS}$). Pour cela, un $VPAC_{Fusion}$ est calculé et la classe finale d'un segment donné est la classe majoritaire du $VPAC_{Fusion}$ de ce segment.

Pour déterminer quelle est la classe finale à attribuer au segment, deux méthodes ont été testées pour calculer le $VPAC_{Fusion}$:

1. la **moyenne pondérée**. Dans la suite du manuscrit, cette méthode est nommée **méthode de la moyenne pondérée** ;
2. la **probabilité maximale**. Dans la suite du manuscrit, cette méthode est nommée **méthode de la probabilité maximale**.

Après cette sélection, le $VPAC_{Fusion}$ est standardisé entre 0 et 1, puis la classe qui a la plus haute probabilité est déterminée comme étant la **classe finale**.

Pour les segments qui n'ont pas été sélectionnés pour être traités par le module 2 : règles spatiales et qui conservent leur classe RF, le $VPAC_{Fusion}$ correspond au $VPAC_{RF}$.

Ce $VPAC_{Fusion}$ sera par la suite utilisé pour l'application du module 4 : modélisation de la répartition spatiale des classes.

3.4 Application du module 4 : modélisation de la répartition spatiale des classes

Cette section présente l'application aux Hautes Terres malgaches le module 4 : modélisation de la répartition spatiale des classes. Dans la section 2 De la connaissance à la probabilité d'appartenance : les règles expertes , nous avons présenté la façon dont les connaissances sont formalisées pour **obtenir une répartition théorique des cultures** (transposée en classes) à l'échelle d'un Fokontany. L'intérêt de ce module est de donner une cohérence spatiale à la caractérisation d'une unité administrative. Si les statistiques théoriques sont correctes, alors la répartition des cultures au sein d'un Fokontany sera cohérente à la fois par rapport à l'observation des classes et aux stratégies des agricultures, tout en respectant des proportions réalistes. Les segments peuvent ainsi être répartis en fonction de leur $VPAC_{Fusion}$ à un emplacement optimal.

Dans les paragraphes suivants, nous présentons un exemple de résultats d'enquêtes et de monographies pour la commune d'Alakamisy, au sein de laquelle ces deux sources divergent. Ces données ont servi à estimer les proportions théoriques de chaque classe.

Alalamisy est l'une des communes où les personnes enquêtées étaient particulièrement bien renseignées sur l'activité agricole de la commune[17]. Les informations qu'ils nous ont fournies sont détaillées et précises, mais restent des estimations à dire d'experts (tableau 11). À contrario, la monographie de la commune (Alakamisy, 2017) est très imprécise et l'erreur des estimations est directement appréhendable à la lecture (tableau 12). Dans cette configuration, c'est la description qui semble la plus fiable qui est choisie. Pour la commune d'Alakamisy la description fournie par les personnes enquêtées sert de base à l'estimation de la répartition de l'occupation des sols.

[17] Les deux personnes qui ont participé à l'enquête sont le conseiller au développement rural de la commune et un membre d'une association sur l'agriculture locale.

Table 11 Répartition des cultures dans les Fokontany Tsaratanana, Alakamisyet et Andohafarihy du Sud de la commune d'Alakamisy selon le responsable d'une association d'agriculteurs et selon le responsable du développement rural de la commune. Il est important de préciser que selon les personnes enquêtées, le maïs pur est très rare, il est majoritairement associé au riz pluvial ou aux pommes de terre (principalement en bordure de champ). Sources enquêtes auprès des responsables locaux, Annexe F.

Culture par ordre d'importance	Répartition	
	Selon le responsable de l'association	Selon le responsable du développement rural
Cultures principales		
Légumes et pomme de terre	30 %	40 %
Cultures secondaires		
Maïs	15 %	Non précisé
Maïs et riz pluvial	Non précisé	25 %
Riz pluvial et riz irrigué	20 %	Non précisé
Vergers	30 %	10 %
Pins et espaces non cultivés	Non précisé	25 %

Table 12 Extrait de la monographie de la commune rurale Alakamisy - Ambano Antsirabe II (Alakamisy, 2017). Dans le texte, la superficie de la commune est estimée à 42 km^2 alors qu'en la mesurant à partir de la couche vectorielle de la commune, son aire est de 68,43 km^2. Les calculs de pourcentage de surface ne sont pas compris dans la monographie. Les arbres sont recensés par pieds malgré le fait que la monographie indique leur superficie en hectares. Même si cela n'est pas précisé dans la monographie, les surfaces cultivées sont estimées à l'année, la somme peut donc être supérieure à 100 %, ce qui signifie que des parcelles sont en culture au moins pour la saison culturale principale et pour une autre saison culturale.

Culture	Surfaces cultivées (ha)	Aire totale (%) de 42 km^2	Aire totale (%) de 68 km^2
Maïs	1421	33,83	20,76
Riz irrigué	3532	84,10	51.61
Pomme de terre	2318	55,19	33,87
Blé	112	2,67	1,64
Soja	108	2,57	1,58
Orge	38	0,9	0,55
Patate douce	53	1,26	0,77
Carotte	2645	62,98	38,65
Choux	599	14,26	8,75
Pêchers	4800	/	/
Poiriers	12700	/	/
Pommiers	1052	/	/
Kaki	1650	/	/
Abricotier	195	/	/
Prunier	10750	/	/
Lacs	60 ares		/
Forêts	3	0,07	0,04

Le module 4 : modélisation de la répartition spatiale des classes a été implémenté en Ocelet comme décrit dans le schéma 11 du chapitre 2, section 3.2.6 Module 4 : Modélisation de la répartition spatiale des classes . Ce modèle est basé sur deux paramètres :

- un **seuil de fiabilité** qui permet de déterminer si le segment peut être considéré comme correctement classé par les étapes précédentes (*i.e.* la probabilité de la classe majoritaire dans le $VPAC_{Fusion}$). Dans la modélisation, ce seuil est de **0,8** ;
- et un **seuil de variabilité** des aires théoriques de chaque culture. Ce seuil permet d'adapter la proportion théorique des cultures en fonction de la quantité de segments considérés comme fiables par classe. Dans la modélisation, ce seuil est de **0,01**.

De manière synthétique[18] le module se déroule en trois étapes majeures :

1. comparaison entre les valeurs théoriques de répartition des classes et les proportions de chaque classe selon la classe fusion (issue du $VPAC_{Fusion}$). De cette comparaison nous savons quelles sont les classes théoriques déficitaires ou excédentaires ;
2. création d'une liste de segments considérés comme fiables et une autre où les segments doivent être reclassés, car en dessous du seuil de fiabilité. Pour les classes excédentaires en segments, les segments de la liste fiable avec les probabilités les plus faibles sont déplacés vers la liste où les segments doivent être reclassés, jusqu'à ce que la classe ne soit plus excédentaire ;
3. attribution d'une classe à chaque segment de la liste des segments à reclasser en fonction des classes théoriques déficitaires et des $VPAC_{Fusion}$ de chaque segment.

À la fin de ce module, chaque segment a pu se voir affecter une nouvelle classe d'occupation ou d'usage des sols et une nouvelle carte est obtenue. Cette carte a pour objectif d'être plus cohérente par une meilleure répartition spatiale des classes. Ce module est appliqué à l'ensemble des Fokontany de notre zone d'étude en posant l'hypothèse forte que la répartition des cultures est homogène sur le Fokontany (sauf lorsque plus de 50 % de l'aire du Fokontany est en dehors de notre zone d'étude). Le tableau 13 présente le nombre de Fokontany contenu dans notre zone d'étude.

Table 13 Nombre de Fokontany en fonction de leur aire dans la zone d'étude.

	Pourcentage d'aire dans la zone d'étude				
	100	99-75	75-50	< 50	Total
Nombre de Fokontany	38	11	9	19	77

Dans cette section 3 Application des modules, nous avons vu comment le module 2 : règles spatiales et le module 4 : modélisation de la répartition spatiale des classes sont implémentés avec Ocelet. Avec le module 2, nous avons obtenu un $VPAC_{RS}$ (vecteur de probabilité d'appartenance aux classes)

[18] Le détail de l'implémentation du module avec Ocelet est visible en annexe K

qui représente la possibilité qu'un agriculteur cultive l'une des classes définies dans un segment. Ce $VPAC_{RS}$ est ensuite combiné au $VPAC_{RF}$ (issu de l'algorithme de forêts aléatoires) pour déterminer un $VPAC_{Fusion}$ et une classe finale. Pour combiner les deux VPAC, deux méthodes sont proposées, une moyenne pondérée et une sélection de la probabilité maximale. Le résultat produit par le module 4 est une carte de l'occupation des sols après une répartition des segments selon leur $VPAC_{Fusion}$ et une répartition théoriques des classes d'occupation et d'usage des sols dans un Fokontany.

4 Synthèse des expérimentations menées

4.1 Analyse de la « cohérence agricole » des cartes

L'analyse des résultats a pour principal objectif d'évaluer et d'appréhender si à l'issue de l'application du modèle conceptuel, la caractérisation de la zone d'étude est « **cohérente d'un point de vue agricole** ». Cette cohérence thématique peut être discutée à la fois par l'analyse quantitative et qualitative des cartes produites. L'évaluation de la cohérence est effectuée à dire d'expert, car c'est le thématicien qui va estimer, selon ses connaissances, si une carte est cohérente par rapport aux stratégies des agriculteurs et à l'environnement. De cette explication de la cohérence, nous pouvons identifier deux questions principales pour analyser les résultats à partir des $VPAC_{RF}$ et des $VPAC_{Fusion}$:

1. **Les confusions entre les classes sont-elles thématiquement plus acceptables pour les $VPAC_{Fusion}$ que pour les $VPAC_{RF}$?**
   (*i.e.* Les confusions entre classes s'expliquent-elles plus facilement du fait de leur proximité biologique (légumes, pomme de terre) ou de stratégies communes à différentes cultures (associations, cultures pluviales) ?)
2. **La localisation des types de culture est-elle cohérente avec les stratégies identifiées via les connaissances ?**
   (*i.e.* Les zones défavorables sont elles bien dépourvues de cultures à l'issue de l'application de la modélisation ? Le domaine agricole est-il correctement identifié ?)

Les indicateurs quantitatifs mobilisés pour répondre à ces questions sont présentés dans le chapitre 3, section 4 Méthodes d'évaluation.

Pour répondre à la première question, l'analyse des résultats obtenus par les métriques d'évaluation nous permettra dans un premier temps d'analyser les résultats de la classification des données de télédétection pour identifier quelles sont les incohérences de cette classification. Cette identification se fera par l'analyse des indicateurs de précision des classes (F-mesure, précision, rappel) et par l'analyse des confusions entre classes (matrice de confusion, désaccords de quantité et de qualité). L'analyse des résultats issus de l'application du module 2 : règles spatiales et de la fusion des VPAC se fera en lien avec les incohérences identifiées à partir des $VPAC_{RF}$. Les mêmes méthodes vont être utilisées pour mieux comprendre l'impact de ce deuxième module sur les erreurs

de classification et les confusions entre classes afin d'évaluer de façon experte si les erreurs sont plus logiques sur le plan agricole (ex. confusion entre deux classes de culture ou de famille proche ou ayant des pratiques proches). Cette analyse permet ainsi de porter un regard différent sur les résultats, sans se baser uniquement sur des métriques de précision de classification. L'idée est ici d'évaluer si la carte produite à l'issue de l'application du module des règles spatiales présente des confusions plus acceptables que celles de la carte produite par traitement des données de télédétection, et ce même si les précisions de classification sont proches ou similaires.

Pour répondre à la seconde question, l'évaluation de la cohérence est surtout qualitative. Pour évaluer la cohérence de la localisation des types d'occupation et d'usage des sols, nous pouvons utiliser la métrique de désaccord de quantité (*i.e.* si une classe est sur-représentée) et celle de désaccord de qualité (si une classe n'est pas correctement localisée). Mais cette analyse est limitée, car elle n'est représentative que la base de données de vérité terrain (qui peut également présenter des déséquilibres). Il est donc nécessaire d'analyser les cartes de manière qualitative pour s'abstraire des limites de la base de données de vérité terrain. De cette façon, la carte issue de la classification des données de télédétection et celle issue des $VPAC_{Fusion}$ (selon la méthode de la moyenne pondérée) sont étudiées selon trois différentes configurations du module des règles spatiales. Les résultats produits par le module 4 : modélisation de la répartition spatiale des classes sont analysés de la même façon. Pour répondre à cette question, nous examinons donc la **cohérence spatiale** de la classification de la zone d'étude en fonction des connaissances que nous avons identifiée.

Dans notre cas d'étude, où nous travaillons principalement sur l'agriculture, nous considérons qu'une **erreur de confusion** (quantitative ou qualitative) **au sein du domaine cultivé est plus acceptable** qu'entre le domaine cultivé et le domaine non-cultivé. De même, une confusion entre des cultures pluviales est plus cohérente qu'une confusion entre des cultures pluviales et de la savane ou autre classe de végétation naturelle.

4.2 Analyse de la classification des données de télédétection

Dans le cadre du lancement des satellites Venμs et Sentinel-2 qui marquent une nette amélioration de la résolution spatiale et temporelle des satellites optiques, il est intéressant d'analyser brièvement l'intérêt de ces missions pour la caractérisation d'un espace d'agriculture familiale tropicale. L'une des contraintes de l'utilisation de ces données pour la caractérisation de l'occupation et de l'usage des sols en zone tropicale est la nébulosité, qui limite les acquisitions et la taille des parcelles, souvent inférieure à la taille d'un pixel (chapitre 1, section 2.1.3 La télédétection et l'agriculture familiale).
Nous formulons l'hypothèse que l'augmentation de la fréquence de revisite et l'amélioration de la résolution spatiale des données satellitaires améliorent les résultats de classification dans les espaces tropicaux. Pour valider ou non cette hypothèse, nous avons classifié l'occupation du sol de la zone d'étude à partir de différents jeux de données : 1)un premier jeu composé de toutes les variables disponibles (issues de Sentinel-2 et Venμs, ainsi que des variables auxiliaires que sont les indices de textures, la pente et l'altitude), 2) un jeu contenant les variables issues de Sentinel-2 et les

variables auxiliaires, 3) un jeu comportant les variables issues de Venμs et les variables auxiliaires, et 4) un dernier uniquement composé des variables auxiliaires. Au sein de ces jeux de données, l'ensemble des compositions possibles (utilisation ou non des bandes spectrales, indices...) a été évalué (avec la précision, le rappel, la F-mesure, la précision globale et le coefficient de Kappa). En complément de ces métriques, la Minimum Decrease In Accuracy (MDA) a été analysée pour voir quelle est l'influence de chaque variable sur la classification de l'ensemble des variables. Ensuite, nous analysons de façon détaillée le résultat de la classification de l'ensemble des variables de télédétection.

4.3 Résultats de la combinaison des $VPAC_{RF}$ et $VPAC_{RS}$

À l'issue de la classification des données de télédétection et de l'application du module 2: règles spatiales nous avons produit un $VPAC_{Fusion}$ (à partir du $VPAC_{RF}$ et du $VPAC_{RS}$). Le $VPAC_{Fusion}$ résulte de trois coefficient et seuils présentés dans la section 3 Application des modules (le coefficient de pondération pour la standardisation de chaque classe, le seuil de probabilité du $VPAC_{RF}$ minimum, le seuil ΔRF) et peut être calculé de deux façons (méthode de la moyenne pondérée et méthode de la probabilité maximale). Les résultats de précision globale sont présentés en fonction de la variation des paramètres ΔRF de 0,1 à 1 par pas de 0,1 pour la méthode de la probabilité maximale (soit 10 résultats). Pour la méthode de la moyenne pondérée, la même variation du seuil de ΔRF est utilisée ainsi que des valeurs de -1 ; -0,5 ; 1 ; 5 et 10 pour le coefficient de pondération (soit 50 combinaisons). Une combinaison de chaque méthode est ensuite analysée, chacune améliorant la précision de la classification des données de télédétection. Pour cette analyse, la précision, le rappel, la F-mesure, la précision globale et le coefficient de Kappa sont utilisés. L'indépendance des deux méthodes par rapport au résultat de la classification des données de télédétection est ensuite présentée à l'aide du $\chi 2$ de McNemar. Pour compléter l'analyse des résultats, d'autres combinaisons ne permettant pas d'améliorer la précision globale sont aussi présentées.

Ensuite, l'analyse spatiale des changements de classes affectées entre les $VPAC_{RF}$ (télédétection), le $VPAC_{Fusion}$ est mise en avant via l'étude de la localisation des changements et l'évolution de l'aire de chaque classe au sein de la zone d'étude.
Puis, afin d'évaluer si la fiabilité de l'affectation des classes est augmentée par l'application du module des règles spatiales, nous présentons la distribution des probabilités par classe pour l'ensemble des segments de la zone d'étude selon le $VPAC_{RF}$ (télédétection), le $VPAC_{Fusion}$ issu de la méthode de la moyenne pondérée et le $VPAC_{Fusion}$ selon la méthode de la probabilité maximale. Pour terminer l'analyse quantitative des résultats, la distribution de la densité des probabilités pour les segments bien classés et mal classés de chaque classe de la base de données de vérité terrain est présentée pour la classification des données de télédétection et pour une combinaison des $VPAC_{RF}$ (télédétection) et $VPAC_{RS}$ issue de l'application de la méthode de la moyenne pondérée.

Nous conclurons l'analyse des résultats du module 2 : règles spatiales et de la classification des

données de télédétection par une étude qualitative des cartes d'occupation et d'usage des sols produites. Pour cela, sont analysées les cartes issues de la classification des données de télédétection et de trois combinaisons différentes de l'application de la méthode de la moyenne pondérée dans le module des règles spatiales.

Nous terminerons le chapitre suivant par une analyse qualitative du résultat du module 4 : modélisation de la répartition spatiale des classes, la base de données de vérité terrain ne couvrant pas une zone suffisante pour évaluer ce module.

POINTS CLÉS DU CHAPITRE 4

- **TÉLÉDÉTECTION** :
  - l'image THRS est **segmentée** pour définir les unités spatiales élémentaires (segments) servant de structure spatiale aux étapes de caractérisation de l'occupation et de l'usage des sols, et des indices de textures sont également calculés à la partir de ces images ;
  - les séries temporelles Sentinel-2 et Venμs sont traitées avec la chaîne de traitement Moringa pour extraire des réflectances et des indices de végétation ;
  - ces données satellitaires et des données auxiliaires (indices de texture, pentes, altitudes) sont agrégées à l'échelle des segments, puis classifiées à l'aide d'un algorithme de **forêts aléatoires** ;
  - un $VPAC_{RF}$ est obtenu, synthétisant le résultat de la classification des données de télédétection. Ce $VPAC_{RF}$ **représente la probabilité d'observer les différentes classes de la nomenclature** dans un segment ;
- **FORMALISATION DES CONNAISSANCES** :
  - les **connaissances sur les stratégies agricoles sont formalisées** en règles expertes. De chaque connaissance peut découler une ou plusieurs règles pour un type de culture ;
  - selon l'importance de la connaissance dans les choix et stratégies des agriculteurs, une **probabilité est attribué à chaque règle** ;
  - pour chaque classe, les règles sont agrégées pour exprimer la **possibilité d'identifier une classe dans un segment** ;
- **MODULES SPATIAUX** :
  - à partir du **module 2 : règles spatiales** est obtenu un $VPAC_{RS}$**, représentant la possibilité** qu'un agriculteur cultive une culture donnée dans un segment ;
  - ce $VPAC_{RS}$ est combiné au $VPAC_{RF}$ pour obtenir un $VPAC_{Fusion}$ servant à déterminer une classe d'occupation et d'usage des sols. Cette classe finale est le fruit de l'**utilisation conjointe de la télédétection et de la modélisation spatiale** ;
  - pour combiner un $VPAC_{RF}$ et un $VPAC_{RS}$, deux méthodes sont proposées, une **moyenne pondérée** et une **sélection de la probabilité maximale** ;
  - le module 4 : modélisation de la répartition spatiale des classes permet de reclasser les segments les moins fiables pour que la répartition des classes corresponde à une répartition théorique à l'échelle d'un Fokontany.

Chapitre 5

Application du modèle conceptuel : Résultats de l'application aux Hautes Terres malgaches

Table des matières

1. Résultats de la classification des données de télédétection
   1.1. Impact des données satellitaires utilisées sur la précision globale
   1.2. Analyse des résultats de classification par classe
   1.3. Segmentation et classification
2. Application du module 2 : règles spatiales
   2.1. Probabilités issues du module des règles spatiales
   2.2. Choix des paramètres du module
   2.3. Vers une améliorations des résultats ?
   2.4. Incidence de l'application du modèle sur la répartition spatiale des classes
   2.5. Synthèse et résultats spatiaux qualitatifs
3. Résultats du module modélisation de la répartition spatiale des classes
   3.1. Changements généraux
   3.2. Plateaux et pentes, une place pour les cultures pluviales ?
   3.3. Évolution du Sud : les communes d'Alakamisy et d'Ambano

La présentation des résultats de cette thèse a comme objectif principal la compréhension de l'application du modèle conceptuel sur la région des Hautes Terres de Madagascar. Le but de cette approche est d'une part d'**améliorer statistiquement les résultats** de classification et de l'autre de **produire une carte plus cohérente par rapport aux stratégies des agriculteurs**. Cette cohérence est analysée à la fois quantitativement entre les $VPAC_{RF}$ et les $VPAC_{Fusion}$ (ex. les erreurs sont elles plus acceptable ?) et qualitativement à partir des cartes (ex. la localisation des types de culture est-elle cohérente avec les connaissances identifiées ?) pour différentes sorties du modèle. L'analyse quantitative montre que l'application du modèle conceptuel a un impact limité au niveau des résultats quantitatifs, tandis que l'analyse qualitative permet une compréhension spatiale de l'application du modèle conceptuel et aide donc à identifier des pistes d'améliorations méthodologiques et appliquées. Les résultats présentés sont issus de la classification des séries temporelles d'images satellite, du module 2 : règles spatiales et du module 4 : modélisation de la répartition spatiale des classes.

1 Résultats de la classification des données de télédétection

1.1 Impact des données satellitaires utilisées sur la précision globale

Le tableau 14, présente les résultats globaux des classifications effectuées avec le modèle de forêts aléatoire présenté Chapitre 4, tableau 7 entraîné/validé par la même base de données d'apprentissage et de validation sur différents jeux de données satellitaires (au nombre de 14). Ces jeux de données ont été présentés dans le chapitre 4 section 1 Télédétection, et ont pour objectif d'analyser l'impact de l'utilisation des différentes sources (S2, Venμs, variables auxiliaires) et types (bandes spectrales, indices) de données satellitaires sur les résultats de classification. Les deux meilleurs résultats de classification en termes de précision globale sont obtenus via la combinaison des bandes spectrales de Sentinel-2 et des variables auxiliaires (précision globale de 0,839) et par la combinaison des bandes spectrales et des indices de Sentinel-2 avec les variables auxiliaires (précision de 0,828). Les précisions de classification obtenues avec les données Venμs sont inférieures, mais tout de même extrêmement proches de ceux de Sentinel-2. La raison de cet écart n'est pas explicable en regardant uniquement les résultats de classification, mais a déjà été identifiée par Baba et al. (2020) dans une étude sur la détection de la neige. Nous pouvons toutefois souligner que la résolution spatiale deux fois inférieure de Venμs par rapport à Sentinel-2 ne semble pas dans notre cas apporter d'amélioration à la caractérisation d'un espace de petite agriculture tropicale.

Un nombre trop important de dimensions par rapport à la base de données d'apprentissage peut expliquer en partie le fait que l'utilisation de l'intégralité des données disponibles ne permet pas d'obtenir un meilleur résultat de classification(Ghosh et al., 2014; Belgiu and Drăguţ, 2016). Une autre explication est que certaines variables ont une corrélation trop importante entre elles alors que certaines classes ont un effectif réduit dans notre base de données de vérité terrain (Chapitre 3, figure 16). Une sélection d'indices plus adaptés pourrait aussi être une solution pour améliorer les résultats de classification des modèles de forêt aléatoire combinant Venμs et Sentinel-2 appliqués à la petite agriculture tropicale.

Dans la suite de notre travail, nous avons décidé d'utiliser la totalité des données des satellites Sentinel-2 et Venμs et des variables auxiliaires (tableau 14 : S2 BS, S2 I, Venμs BS, Venμs I, VA) malgré le fait que cette combinaison ne soit pas la plus performante (précision globale de 0,819, Kappa de 0,780). Les raisons qui expliquent ce choix sont que l'ensemble des traitements et tests ont été faits sur cette base, et que cela a également permis d'analyser la contribution des différentes variables via le calcul du Minimum Decrease In Accuracy (MDA). Lorsque toutes les variables sont utilisées, l'analyse du MDA montre que la contribution des textures est la plus importante, celles-ci arrivant en 1ère, 3ème et 2ème , 21ème et 25ème positions (indices de corrélation et de contraste pour les fenêtres 15 et 29). Les variables de bandes spectrales issues de Venμs correspondent à 13 des 25

variables les plus importantes. Les indices issus de Venμs représentent 7/25 des variables les plus importantes et seulement deux de ces variables sont issues de Sentinel-2 (deux fois la bande 11), ce qui est contre-intuitif par rapport à la précision globale.

Table 14 Comparaison des résultats pour chaque type de combinaison de données satellitaires et auxiliaires. **BS** signifie bandes spectrales, **I** signifie indices spectraux, **VA** signifie variables auxiliaires (textures, pente, altitude, exposition ; voir chapitre 3, section 2.3 Données support pour l'application des connaissances)

Variables	Moyenne pondérée			Précision globale	Kappa
	Précision	Rappel	F-mesure		
S2 BS, S2 I, Venμs BS, Venμs I, VA	0.83	0.82	0.83	0.819	0.780
S2 BS, S2 I, VA	0.83	0.83	0.82	0.828	0.788
S2 BS, S2 I	0.83	0.83	0.82	0.825	0.785
S2 BS, VA	0.84	0.84	0.83	**0.839**	**0.800**
S2 BS	0.82	0.82	0.81	0.820	0.777
S2 I, VA	0.82	0.81	0.80	0.807	0.765
S2 I	0.81	0.80	0.79	0.798	0.754
Venμs BS, Venμs I, VA	0.84	0.82	0.82	0.824	0.785
Venμs BS, Venμs I	0.83	0.81	0.81	0.814	0.775
Venμs BS, VA	0.83	0.82	0.82	0.819	0.780
Venμs BS	0.82	0.81	0.81	0.812	0.771
Venμs I, VA	0.82	0.81	0.81	0.813	0.772
Venμs I	0.82	0.81	0.81	0.807	0.766
VA	0.60	0.61	0.59	0.608	0.514

1.2 Analyse des résultats de classification par classe

Le tableau 15 présente les résultats de la classification obtenus à partir de l'ensemble des données de télédétection. Les classes peuvent être séparées en trois catégories : 1) les classes avec de bons résultats de classification (F-mesure entre 80 et 100), 2) les classes ayant des résultats moyens (F-mesure entre 80 et 50) et 3) les classes avec de mauvais résultats (F-mesure $\leq$ 50). L'analyse de la matrice de confusion permet d'apporter des éléments de compréhension des erreurs de classification (Annexe I, tableau 25).

La **première catégorie** est composée des classes d'occupation et d'usage des sols non agricoles (*eau* précision : 1, rappel : 0,97 ; *bâti* précision : 0,92, rappel : 0,99 ; *savane* précision : 0,92, rappel : 0,78), les classes regroupant la végétation arborée (*culture boisée de rente* précision : 0,76, rappel : 0,93 et *verger* précision : 0,88, rappel : 0,80), et les classes *riz irrigué* (précision : 0,95, rappel : 0,99) et *herbes et autres cultures fourragères* (précision : 0,97, rappel : 0,95). Ces classes ont dans l'ensemble un rappel et une précision satisfaisants. Par contre, les mesures de désaccords de quantité et de répartition sont les plus importants de toute la classification pour la classe *savane* (q : 5,64, a : 5,19). Cela signifie qu'une partie des erreurs de distribution spatiale des autres classes sont liées à une confusion avec la classe *savane* (la part de savane dans la répartition des classes est trop importante). La classe *culture boisée de rente* est aussi sur-représentée, mais dans une mesure moins

importante. Les classes *savane* et *culture boisées de rente* sont celles avec les confusions (matrice de confusion Annexe I, tableau 25) les plus importantes à cause de l'hétérogénéité de la classe *savane* qui peut être composée de savane herbeuse, arbustive ou arborée. De cette hétérogénéité résulte les confusions dans la classe *culture boisée de rente* qui peut être identifiée comme *savane* lorsque les arbres sont clairsemés. Dans l'ensemble, le résultat de classification de cette catégorie de classes est difficilement améliorable, car une tentative d'amélioration pourrait entraîner un risque de créer d'autres erreurs de classification.

La **deuxième catégorie** correspond aux classes *légumes* (précision : 0,67, rappel : 0,79), *maïs* (précision : 0,62, rappel : 0,89), *pomme de terre* (précision : 0,70, rappel : 0,67) et *sols rocheux* (précision : 0,99, rappel : 0,44). Cette catégorie est plus hétérogène dans la précision et le rappel. La classe *sols rocheux* a une excellente précision, mais un très mauvais rappel. Cela peut s'expliquer par le fait que cette classe ne comporte que 10 polygones de vérité terrain et qu'il y a très peu de zones rocheuses dans notre zone d'étude. Les résultats de la classe *pomme de terre* (q : 0,10, a : 1,61) et *légumes* (q : 0,73, a : 1,16) sont relativement équilibrés et leurs erreurs sont principalement liées à des confusions mutuelles, égales en termes de quantité, car les cultures de ces classes peuvent être cultivées dans des conditions similaires. La classe *légumes* est aussi confondue avec la classe *cultures potagères* qui est composée de petites parcelles. Le rappel de la classe *maïs* est très bon, mais cette classe est généralement confondue avec la classe *savane*. La raison principale est que les champs de maïs peuvent être dégradés donc laisser une grande partie du sol apparent. Le fort désaccord de quantité de cette classe est lié à sa prépondérance sur la classe *riz pluvial* (précision : 0,21, rappel : 0,18) (classé dans la 3ème catégorie du fait de précisions de classification faibles) qui est assimilé à la classe *maïs*. Ces classes (*riz pluvial* et *maïs*) sont intéressantes à améliorer, car ce sont des cultures répandues dans notre zone d'étude, représentant environ 38 % de l'aire cultivée de nos polygones de terrain. Ce sont aussi des cultures importantes pour l'alimentation et comme ressources financières pour la population malgache (Sourisseau et al., 2016).

La **troisième catégorie** est composée des classes *soja* (précision : 0,76, rappel : 0,37), *maïs associé à une autre culture* (précision : 0,50, rappel : 0,48), *cultures potagères* (précision : 0,41, rappel : 0,43), *riz pluvial associé à une autre culture* (précision : 0,38, rappel : 0,37), *riz pluvial* (précision : 0,21, rappel : 0,18) et *sol nu agricole* (précision : 0,0, rappel : 0,0). La classe *sol nu agricole* n'a jamais été correctement labellisée et représente une surface minime de l'usage des sols, elle sera donc écartée pour le reste des analyses. Le reste de cette catégorie est composé de cultures pluviales qui sont fréquemment associées entre elles (association de maïs-soja, maïs-riz pluvial) et qui sont rarement cultivées seules dans une parcelle, ce qui permet de comprendre les erreurs globales de cette catégorie. Le *riz pluvial* est principalement confondu avec le *riz pluvial associé* et le *maïs associé*, mais aussi avec la classe *savane*. Les erreurs de la classe *culture potagères* sont surtout liées à son attribution à tort à la classe *légumes*. Malgré les mauvais résultats de classification de cette catégorie, les erreurs restent cohérentes et sont principalement liées à des désaccords de localisation des cultures, mais pas de quantité (pas de sur-représentation). Les résultats de classification de cette catégorie peuvent être intéressants à améliorer, mais cette amélioration est compliquée en raison des pratiques culturales proches entre ces classes et du signal radiométrique

qui est rarement pur (pixels mixtes) du fait de l'association de cultures, ou de la très petite taille de ces parcelles.

Dans l'ensemble, le résultat de cette classification est plutôt satisfaisant avec une précision globale pondérée par aire à 83,23 %, un rappel global pondéré par aire à 81,93 % et une F-mesure pondérée par aire de 82,57 %. La précision globale et le kappa sont aussi acceptables (81,93 % et 78,05 %). Les erreurs de désaccord de quantité et de répartition ne sont pas déséquilibrées, bien que le désaccord de répartition soit proche de 10 %. Par ailleurs, la qualité de la segmentation entraîne une part d'erreur incompressible, car les objets créés ne sont pas conformes aux unités observées sur le terrain.

Table 15 Métriques d'évaluation par classe et globales pour la classification des données de télédétection, q correspond au désaccord de quantité, a au désaccord de répartition. Les résultats ont été calculés par aire et sont exprimés en pourcentage. La couleur verte correspond à la catégorie 1, la couleur jaune à la catégorie 2 et la couleur rouge à la catégorie 3. Pour une meilleure lisibilité, les résultats sont exprimés en pourcentage et q et a ont été multipliés par 100.

Classe	Précision	Rappel	F-mesure	q	a
Maïs	61,74	89,44	72,35	3,99	1,78
Riz irrigué	95,00	99,33	97,12	0,27	0,08
Riz pluvial	20,94	18,32	19,54	0,12	1,30
Légumes	66,71	79,26	72,45	0,73	1,61
Soja	75,85	37,00	49,73	1,29	0,59
Pomme de terre	70,30	67,75	69,00	0,10	1,61
Herbes et autres cultures fourragères	97,97	95,51	96,73	0,18	0,28
Verger	87,72	79,90	83,51	0,29	0,72
Culture boisée de rente	76,44	93,09	83,95	2,76	1,75
Sol nu agricole	0,00	0,00	0,00	0,30	0,02
Cultures potagères	40,93	43,25	42,06	0,09	1,84
Savane	92,18	78,78	84,95	5,64	5,19
Sols rocheux	99,22	44,77	61,70	0,24	0,00
Bâti	92,29	99,40	95,71	0,49	0,08
Eau	100,00	97,29	98,63	0,06	0,00
Maïs associé	50,17	47,54	48,82	0,09	1,63
Riz pluvial associé	37,59	37,46	37,53	0,00	1,01
Moyenne	68,47	65,17	66,78		
Moyenne pondérée par l'aire de la classe	83,23	81,93	82,57		
Précision globale	81,931				
Kappa	78,048				
Désaccord de quantité Q	8,326				
Désaccord de répartition A	9,744				

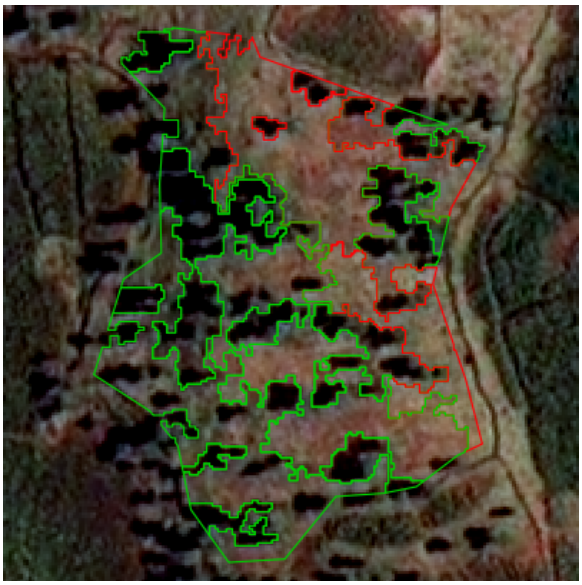


Figure 21 Forêt et savane. Polygones mal classés (en rouge) et correctement classés (en vert) à partir du traitement des données de télédétection. Les bordures délimitent la segmentation de ces polygones.

1.3 Segmentation et classification

La figure 21 permet de visualiser l'impact de la segmentation sur les résultats de classification. Les résultats de la segmentation sont présentés dans le chapitre 4, section 1.1.1.2 Segmentation et montrent que l'image est sur-segmentée. Une partie de cette sur-segmentation est normale et justifiée, par exemple dans le cas de la figure 21 où la segmentation délimite bien la différence entre la forêt et la savane. Dans ce cas, la tâche de classification est performante dans le sens où la savane herbacée est différenciée des pins. Le faux négatif créé lors de cette classification n'en est donc pas vraiment un. Cette erreur est liée à certains polygones de base de données de vérité terrain dont l'occupation du sol n'est pas pure, mais présente un mélange de savane et d'éléments boisés (principalement pour les classes *savane* et *culture boisée de rente*).

À cause de la sur-segmentation de l'image THRS, les polygones de la base de données terrain sont fragmentés et certaines parties peuvent être rattachées à des zones différentes alentour (espaces naturels, autres cultures, diguettes pour le riz irrigué). Ces segments étant souvent en bordure des polygones servant à l'évaluation, ils correspondent aux caractéristiques des autres objets délimités.

L'analyse de la classification de manière quantitative et la carte d'occupation des sols obtenue à partir du traitement des images de télédétection (figure 22) permet d'évaluer l'application du modèle conceptuel à partir des deux méthodes de fusion des $VPAC_{RF}$ et des $VPAC_{RS}$, la **moyenne pondérée** et la sélection des **probabilités maximales** (chapitre 4, Section 3.3.2 Fusion des probabilités de la télédétection du module 2 : règles spatiales).

Dans cette partie, nous avons présenté les contributions des séries temporelles Sentinel-2 et Venμs

et des données auxiliaires pour la classification de l'occupation et de l'usage des sols. À partir des deux séries temporelles et des variables auxiliaires, nous avons analysé les résultats de classification par classe.

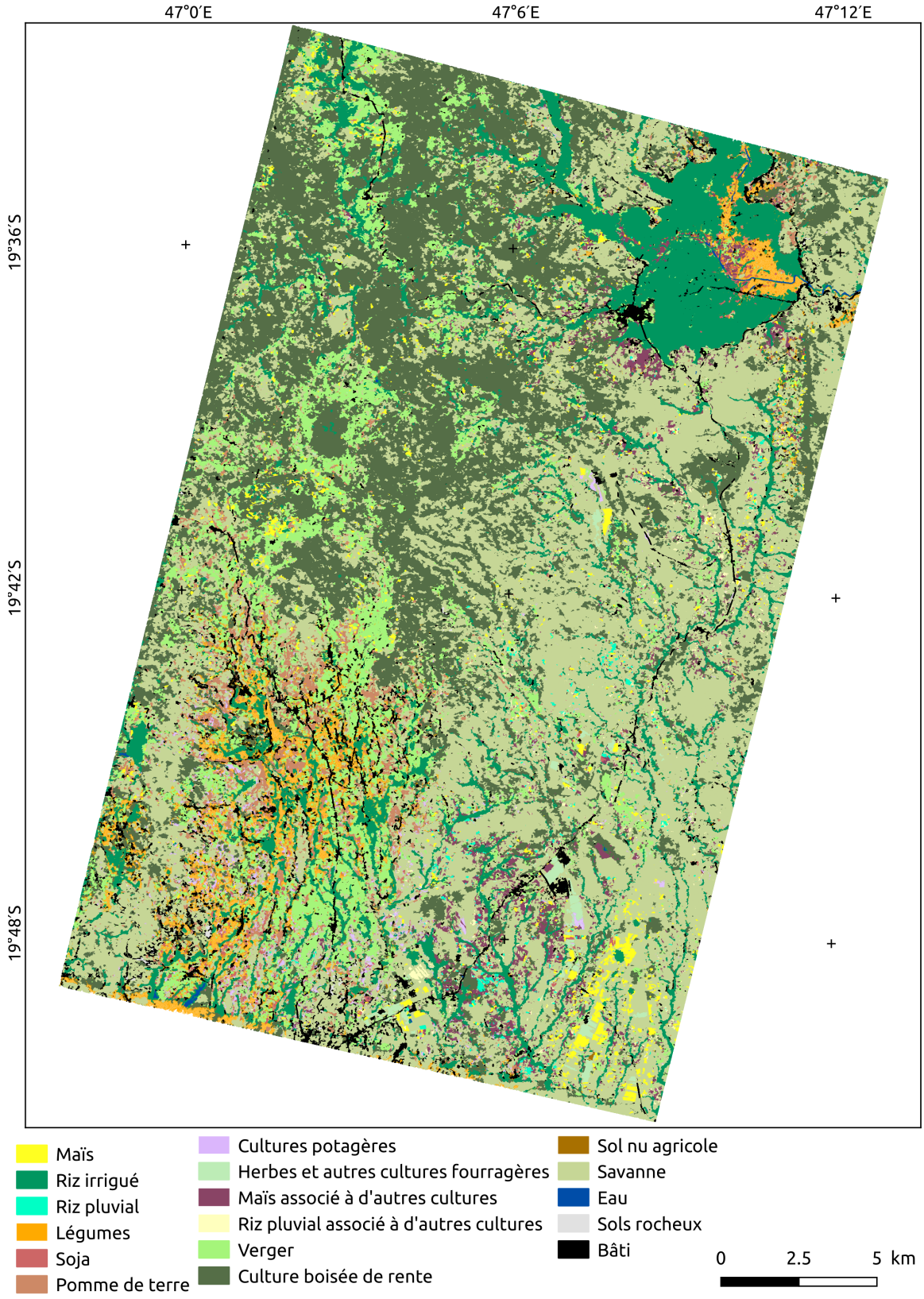


Figure 22 Classification de l'occupation des sols à partir des données de télédétection.

2 Application du module 2 : règles spatiales

2.1 Probabilités issues du module des règles spatiales

L'application du module des règles spatiales (chapitre 4, section 3.3 Calcul des probabilités de présence de culture par segment) permet d'attribuer à chaque segment un VPAC qui représente les probabilités de la possibilité de retrouver les différents types d'occupation/usage des sols de notre nomenclature. Par exemple, pour les classes de culture, cette probabilité indique si le segment est plutôt favorable (probabilité haute) ou plutôt défavorable (probabilité basse) à la mise en terre des différents types de culture. Ces résultats peuvent également être analysés pour estimer les limites de l'application des règles spatiales, notamment vis-à-vis des données de support à leur application qui sont parfois peu fiables sur la région étudiée. Des cartes de confrontation entre les classes attribuées par le module des règles spatiales (*i.e.* la classe ayant la probabilité la plus élevée dans le $VPAC_{RS}$) et celles attribuées par le traitement des données de télédétection (*i.e.* la classe ayant la probabilité la plus élevée dans le $VPAC_{RF}$) peuvent être produites, fournissant des informations sur l'accord ou les désaccords entre les règles spatiales et l'observation satellitaire.

Lorsque le module 2: règles spatiales est appliqué à tous les segments, les cartes de probabilité de la possibilité de cultiver un type de culture (figure 23) permettent de visualiser, dans une certaine mesure, l'impact des données support pour l'application des connaissances qui sont utilisées. Les probabilités issues de la modélisation sont généralement basses, car elles sont réparties entre 17 classes et sont donc difficilement comparables directement aux probabilités issues de la classification des séries temporelles (figure 23, **d**).

Le résultat de cette attribution n'est pas pertinent à analyser seul de manière quantitative, ces cartes représentent les emplacements les plus favorables pour chaque type de culture (la possibilité de cultiver une culture) et non l'emplacement des cultures (observation d'une culture). Le choix final de l'implantation d'une culture sur une parcelle repose entièrement sur l'agriculteur qui, théoriquement, prend en compte le contexte biophysique et socio-économique ici représenté. De plus, le nombre et la densité des données de vérité terrain sont trop faibles pour avoir un échantillon représentatif des choix des agriculteurs sur leurs parcelles. Pour analyser quantitativement l'impact de ce module sur la caractérisation de l'occupation et de l'usage des sols, il est nécessaire de le mettre en relation avec la classification issue du traitement des données de télédétection. Évaluer une possibilité d'occupation ou d'usage d'un sol avec la base de données de vérité terrain n'est pas pertinent. Cette base de données représente l'observation exacte de l'occupation ou de l'usage des sols alors que c'est une possibilité qui est représentée. De plus, le nombre et la densité des données de vérité terrain sont trop faibles pour avoir un échantillon représentatif des choix des agriculteurs sur leurs parcelles. Pour analyser quantitativement l'impact de ce module sur la caractérisation de l'occupation et de l'usage des sols, il est nécessaire de le mettre en relation avec la classification issues du traitement des données de télédétection.

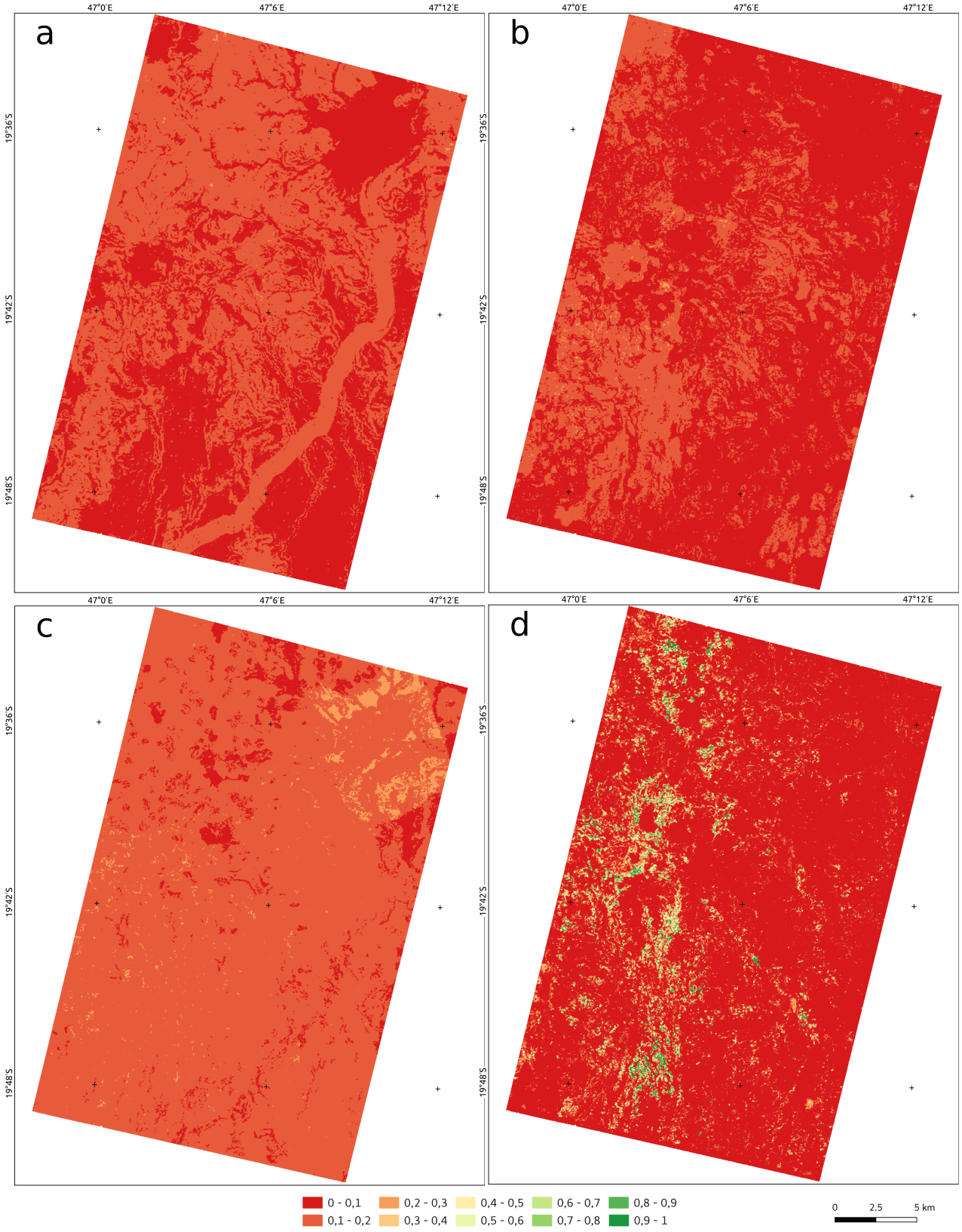


Figure 23 Probabilités d'appartenance des segments à différentes cultures issues du module règles spatiales pour les classes : (**a**) *maïs*, (**b**) *verger*, (**c**) *pomme de terre*, et probabilités issues de la classification des données de télédétection pour la classe (**d**) *verger*. L'impact de certaines règles d'importance est visible sur ces cartes (par exemple, sur la carte (a), l'importance de la route nationale 7 qui crée un corridor nord-sud ayant une probabilité d'appartenance plus élevée).

2.2 Choix des paramètres du module

Pour déterminer quels sont les meilleurs paramètres pour appliquer le module des règles spatiales et corriger la classification des données de télédétection, un ensemble de paramètres a été analysé selon différentes combinaisons. Ces paramètres sont :

- le coefficient de pondération pour la standardisation de chaque classe dans le module des règles spatiales ($coef\ classe$) (équation 14) ;
- le seuil de probabilité minimum d'un segment, c'est-à-dire la valeur de probabilité en deçà de laquelle la classe d'un segment ne sera pas modifiée par la modélisation (chapitre 4, section 3.2.1 Sélection selon la fiabilité des $VPAC_{RF}$) ;
- le seuil ΔRF, soit la différence minimale entre les deux probabilités les plus élevées du $VPAC_{RF}$. Si le seuil est $\geq \Delta RF$, alors la classe de ce segment ne peut être modifiée ;
- et pour la partie dans laquelle les $VPAC_{RS}$ et $VPAC_{RF}$ sont combinés en un $VPAC_{Fusion}$ par la moyenne pondérée, la valeur de pondération du $VPAC_{RS}$ ($coef\ wRS$) (chapitre 4, section 3.3.2 Fusion des probabilités de la télédétection du module 2 : règles spatiales).

En fonction de ces paramètres, le meilleur résultat de classification a été sélectionné en se basant sur la précision globale (tableau 16) et sur le Kappa.

L'analyse des résultats montre que le paramètre de pondération pour la standardisation des règles n'a pas, ou peu, d'influence sur les résultats de classifications. Aucun changement de la précision globale n'a pu être observé lors de l'analyse des différentes combinaisons. Le paramètre de seuil de probabilité minimum a un impact limité lorsque le seuil ΔRF est bas. Les paramètres qui influencent le plus le résultat de classification sont le seuil ΔRF et pour l'utilisation et la valeur de pondération ($coef\ wRS$) lorsque la méthode de la moyenne pondérée est utilisée pour combiner les $VPAC_{RS}$ et $VPAC_{RF}$ en un $VPAC_{Fusion}$.

Le tableau 16 présente les résultats de la précision globale à l'issue de l'application du module 2: règles spatiales (combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ en un $VPAC_{Fusion}$ incluse) pour une variation du seuil de ΔRF et du $coef\ wRS$. Le paramètre qui a la plus forte influence sur les résultats est le ΔRF. Pour le meilleur résultat, la faible valeur de ΔRF (**0,2**) signifie que peu de segments entrent dans le module de règles spatiales pour être modifiés. Pour la méthode de la moyenne pondérée, le paramètre qui apporte le plus de constance dans les résultats est une pondération égale entre le $VPAC_{RF}$ et le $VPAC_{RS}$. En dehors de ces deux valeurs de paramètres, aucun autre résultat n'est supérieur à la précision globale de la classification issue du traitement des données de télédétection.

Bien que n'améliorant pas quantitativement les résultats, les autres combinaisons entre le $VPAC_{RF}$ et le $VPAC_{RS}$ à l'aide de la méthode de la moyenne pondérée sont intéressantes à analyser de manière qualitative. En effet, ces combinaisons laissent une plus grande influence au module sur les $VPAC_{RF}$. De plus, l'augmentation de la valeur du seuil ΔRF permet la modification d'un plus grand nombre de segments.

Au cours de la présentation des résultats de l'application du module avec la méthode de la

moyenne pondérée (paramètres $\Delta RF = 0,5$ et $coef\,wRS = 1$), et de la méthode de la probabilité maximale (paramètre $\Delta RF = 0,5$), d'autres résultats de combinaison vont permettre d'illustrer les biais et avantages du module.

Table 16 Résultats de la précision globale (%) à l'issue de l'application du module des règles spatiales (combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ en un $VPAC_{Fusion}$ incluse) en fonction de la variation des paramètres ΔRF et de la pondération (équation 14 pour la méthode de la moyenne pondérée) avec un seuil de fiabilité de 1 et un $coef\ classe$ de 0,5. La précision globale de la classification des données de télédétection est de 81,93 %.

		Moyenne pondérée (valeur de pondération)					**Probabilité**
		-1	**-0,5**	**1**	**5**	**10**	**maximale**
Seuil ΔRF	**0,1**	81,93	81,93	81,93	81,93	81,93	81,93
	0,2	81,83	81,83	**82,06**	81,84	81,84	**81,97**
	0,3	79,88	77,92	81,65	79,90	79,67	81,95
	0,4	73,16	81,82	81,71	78,02	77,13	81,93
	0,5	75,58	81,51	81,69	75,89	73,61	81,88
	0,6	72,28	79,46	81,60	73,10	66,70	81,79
	0,7	70,02	77,56	81,42	72,05	60,91	81,59
	0,8	67,52	76,29	81,21	70,73	54,54	80,34
	0,9	63,66	75,41	80,91	69,77	45,77	79,46
	1	61,87	74,67	80,75	67,40	29,80	78,73

2.3 Vers une améliorations des résultats ?

2.3.1 Analyse quantitative des résultats

Le tableau 17 présente les résultats de chaque classe, selon la méthode de la moyenne pondérée (Moy.P.) et, selon la méthode de la probabilité maximale (P.Max.) pour déterminer la classe finale lorsque les $VPAC_{RF}$ et $VPAC_{RS}$ sont combinés. Le tableau 18 montre l'évaluation globale pour la classification des données satellitaires, pour la méthode de la moyenne pondérée et pour la méthode de la probabilité maximale.

Les valeurs des métriques d'évaluation obtenues par l'application du module 2 : règles spatiales, en combinant les $VPAC_{RF}$ et $VPAC_{RS}$ en un $VPAC_{Fusion}$ par l'une des deux méthodes (Moy.P. et P.Max.) par rapport à la classification des séries temporelles d'images satellite (tableau 15), montrent une très légère amélioration pour ces deux méthodes avec un kappa et une précision globale plus élevés pour la méthode de la moyenne pondérée. Par contre, si l'on regarde uniquement la moyenne de la précision, du rappel et de la F-mesure, la précision diminue tandis que le rappel augmente légèrement. Cette moyenne est intéressante, car elle ne prend pas en compte l'aire de chaque classe. Le résultat n'est donc pas augmenté artificiellement par les classes qui ont une aire plus importante que les autres dans la base de données de vérité terrain (classes non agricoles de la **catégorie 1**). Comme pour les résultats de la classification des données de télédétection, la moyenne pondérée est bien supérieure à la moyenne et ce du fait du déséquilibre des classes.

Avec la méthode de la moyenne pondérée, seules les classes de *maïs associé à une autre culture* et de *riz pluvial associé* présentent une diminution de leurs résultats à la fois pour le rappel (-2,18 et -0,37 respectivement) et pour la précision (-0,7 et -3,89 respectivement) de ces classes. Par contre, pour la méthode de la probabilité maximale, les classes dont la F-mesure diminue sont *maïs associé* (-0,62), *légumes* (-0,49), *maïs* (-0,27) et *riz irrigué* (-0,06) et les classes dont la F-mesure augmente sont les classes *riz pluvial associé à une autre culture* (+2,66), *pomme de terre* (+0,37), *herbes et autres cultures fourragères* (+0,30), *soja* (+0,28), *savane* (+0,18) et *culture boisée de rente* (+0,11).
L'effet des deux méthodes est visible avec ces résultats : la méthode de la moyenne pondérée est plutôt homogène avec une faible pondération, alors que la méthode de la probabilité maximale a tendance à favoriser les classes dominantes et avec un résultat global de classification plutôt bon (la classe *riz pluvial associé* étant une exception particulière).

Les désaccords de quantité (q) et de répartition (a) diminuent de manière générale avec peu de différence de répartition par rapport à la classification des données de télédétection. Cette similarité avec la précédente classification montre que les changements sont faibles et surtout qu'il n'y a pas d'importants changements au sein d'une classe. Les segments qui changent de classe sont donc marginaux.

Autres combinaisons de paramètres : Résultats

Parmi les différentes combinaisons qui ont été effectuées (même au-delà des paramètres présentés dans cette section), la précision globale n'a presque jamais été améliorée, mais l'influence du module sur la répartition des classes est très variable. Ces changements étaient surtout observés entre les différentes classes de cultures pluviales (*riz pluvial*, *riz pluvial associé*, *maïs* et *maïs associé*), entre la classe *savane* et les classes de cultures pluviales, et entre la classe *légumes* et la classe *pomme de terre*.

Par exemple, pour l'utilisation de la méthode de la moyenne pondérée avec les paramètres $\Delta RF = 0,4$ et $coef\ wRS = 1$, une augmentation de la F-mesure du riz pluvial entraîne une diminution de la F-mesure de la classe riz pluvial associé. Ces dynamiques peuvent être observées avec d'autres paramètres, surtout ceux permettant un changement plus radical des probabilités : soit avec un seuil ΔRF haut, soit avec une forte pondération de $coef\ wRS$. Pour plusieurs classes, on peut observer des changements entre leur précision et leur rappel, tout en conservant une F-mesure similaire.
Dans le cas la classe *savane*, les confusions avec les classes de cultures associées, *pomme de terre* et *soja* diminuent, mais ces confusions augmentent pour les classes *verger*, *herbes et autres cultures fourragères* et *cultures potagères* (de façon marginale) avec plus de faux positifs. Le nombre de faux négatifs entre les classes *pomme de terre*, *maïs* et*cultures potagères* et la savane sont aussi moins importants sauf pour la classe *maïs associé*.
En explorant ces autres combinaisons, on remarque que la répartition des probabilités entre les classes n'est pas homogène au sein d'un $VPAC_{RS}$ et que certaines classes sont presque systématiquement considérées comme favorables dans tous les contextes. On remarque aussi que **des classes sont dominantes au sein d'un groupe** (c'est-à-dire pour un ensemble de classes qui ont des conditions de mise en terre proche). C'est par exemple le cas pour la classe **pomme de terre** qui est dominante dans le groupe : *légumes-pommes de terre*, et pour la classe **maïs associé à une autre culture** pour le groupe : *maïs*, *riz pluvial*, *maïs associé* et *riz pluvial associé*. La principale raison de la domination de certaines classes sur les autres est le déséquilibre des connaissances appliquées dans le module 2 : règles spatiales. Les règles mobilisées sont proches pour les cultures pour les cultures pluviales ainsi qu'entre les classes *légumes* et *pomme de terre*. D'un point de vue thématique cette confusion est intelligible, car les stratégies pour la mise en terre de ces cultures sont similaires.

Cet équilibre difficile entre des classes qui répondent à un ensemble de règles communes pose la question de la **finesse à laquelle les règles expertes peuvent être appliquées à système complexe**.

Table 17 Métriques d'évaluation par classe et globales pour la classification des données de télédétection combinées avec le module des règles spatiales à l'aide de la méthode moyenne pondérée (Moy.P.) (paramètres $\Delta RF = 0,5$ et $coef\ wRS = 1$) et de la méthode de la sélection de la probabilité maximale (P. Max.) (paramètre $\Delta RF = 0,2$). q correspond au désaccord de quantité et a au désaccord de répartition. Les trois catégories définies dans la classification des données de télédétection sont représentées par des couleurs : ma couleur verte correspond à la catégorie 1 (F-mesure 1 - 0,8), la couleur jaune à la catégorie 2 F-mesure 0,8 - 0,5) et la couleur rouge à la catégorie 3 (F-mesure 0,5 - 0). Pour une meilleure lisibilité, les résultats sont exprimés en pourcentage et q et a ont été multipliés par 100.

Classe	Méthode	Précision	Rappel	F-mesure	q	a
Maïs	Moy.P	60,77	89,44	72,37	3,98	1,78
	P. Max.	60,36	89,44	72,08	4,07	1,78
Riz irrigué	Moy.P	95,23	99,33	97,24	0,26	0,08
	P. Max.	94,88	99,33	97,06	0,28	0,08
Riz pluvial	Moy.P	21,61	18,32	19,83	0,14	1,25
	P. Max.	20,94	18,32	19,54	0,12	1,30
Légumes	Moy.P	67,24	79,26	72,76	0,69	1,61
	P. Max.	65,88	79,26	71,96	0,79	1,61
Soja	Moy.P	74,18	37,76	50,04	1,23	0,66
	P. Max.	75,99	37,27	50,01	1,28	0,59
Pomme de terre	Moy.P	67,43	71,32	69,32	0,16	1,61
	P. Max.	70,46	68,31	69,37	0,09	1,61
Herbes et autres cultures fourragères	Moy.P	98,33	95,51	96,90	0,20	0,23
	P. Max.	98,71	95,39	97,02	0,24	0,18
Verger	Moy.P	88,01	81,26	84,50	0,25	0,71
	P. Max.	85,47	79,98	82,63	0,21	0,88
Culture boisée de rente	Moy.P	76,63	93,09	84,06	2,72	1,75
	P. Max.	76,71	92,97	84,06	2,68	1,78
Sol nu agricole	Moy.P	0,00	0,00	0,00	0,30	0,02
	P. Max.	0,00	0,00	0,00	0,30	0,02
Cultures potagères	Moy.P	44,20	43,25	43,72	0,04	1,77
	P. Max.	40,83	44,16	42,43	0,13	1,81
Savane	Moy.P	92,32	78,78	85,01	5,69	5,09
	P. Max.	92,69	78,70	85,13	5,86	4,82
Sols rocheux	Moy.P	99,22	44,77	61,70	0,24	0,00
	P. Max.	99,22	44,77	61,70	0,24	0,00
Bâti	Moy.P	92,49	99,40	95,82	0,47	0,08
	P. Max.	92,60	99,40	95,88	0,47	0,08
Eau	Moy.P	100,00	97,29	98,63	0,06	0,00
	P. Max.	100,00	97,29	98,63	0,06	0,00
Maïs associé	Moy.P	47,99	46,83	47,41	0,04	1,76
	P. Max.	48,76	47,66	48,20	0,04	1,73
Riz pluvial associé	Moy.P	37,22	33,57	35,30	0,08	0,92
	P. Max.	39,61	40,78	40,19	0,02	0,96
Moyenne	Moy.P	68,40	65,25	66,79		
	P. Max.	68,42	65,47	66,91		
Moyenne pondérée	Moy.P	83,29	82,06	82,67		
	P. Max.	83,39	81,95	82,67		

Table 18 Métriques d'évaluation globales pour la classification des données de télédétection combinées avec le module des règles spatiales à l'aide de la méthode moyenne pondérée (Moy.P.) (paramètres $\Delta RF = 0,5$ et $coef\ wRS = 1$) et de la méthode de la sélection de la probabilité maximale (P. Max.) (paramètre $\Delta RF = 0,2$). Les résultats sont exprimés en pourcentage.

	Précision globale	**Kappa**	**Désaccord de quantité Q**	**Désaccord de répartition A**
R.F.	81,931	78,048	8,326	9,744
Méthode Moy. P.	82,057	78,207	8,288	9,655
Méthode P. Max.	81,954	78,095	8,438	9,608

2.3.2 Indépendance des méthodes

Pour comparer les résultats de la classification des données de télédétection et ceux des reclassifications issues du module des règles spatiales obtenues en combinant les $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la moyenne pondérée et par celle de la probabilité maximale, le test $\chi 2$ de McNemar a été utilisé (équation 12). Pour les combinaisons, RF / P. Max. et Moy. P. / P. Max, le nombre de paires discordantes est respectivement de 15 et de 18, et la correction de continuité est donc appliquée (équation 13). Cela n'a pas été nécessaire pour la combinaison RF / Moy. P., le nombre de paires discordantes étant de 29.

On observe dans le tableau 19 que la relation entre les résultats de la classification des séries temporelles et ceux issus de la méthode de la probabilité maximale n'est pas significative. Ces deux résultats ne peuvent donc être considérés comme indépendants. La raison de cette dépendance se trouve dans l'essence même de la méthode appliquée. Celle-ci favorise les résultats issus du traitement des données de télédétection ($VPAC_{RF}$) qui ont des valeurs de probabilité beaucoup plus importantes que celles issues du module des règles spatiales ($VPAC_{RS}$) (exemple spatialisé des résultats du module de règles spatiales, figure 23).

La différence entre les résultats de la méthode de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la moyenne pondérée et ceux de la classification des données de télédétection est significative avec un risque d'erreur de 5 %. L'indépendance des résultats entre les deux méthodes de combinaison des VPAC est faiblement significative (p=0,099) car le $VPAC_{RF}$ et le $VPAC_{RS}$ sont identiques.

Table 19 Résultats du test du $\chi 2$ de McNemar avec un degré de liberté de 1.

Classifieurs	**$\chi 2$ de McNemar**	**p-value**
RF / Moyenne pondérée	4,1724	0,04109
Moyenne pondérée / Probabilité maximale	2,722	0,09896
RF / Probabilité maximale	0,2667	0,60560

2.4 Incidence de l'application du modèle sur la répartition spatiale des classes

Les changements de classe d'occupation des sols sont à l'image des résultats quantitatifs de l'évaluation des méthodes de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ dans le module 2 : règles spatiales (tableau 17 et tableau 18), avec des changements localisés et légèrement plus importants lorsque la méthode de la moyenne pondérée est utilisée (figure 24). Les cartes de la figure 24, représentent les segments ayant changé de classe après la fusion des $VPAC_{RF}$ et $VPAC_{RS}$ via les deux méthodes. Le nombre de segments ayant changé est plus important pour la méthode de la moyenne pondérée. De plus, certains changements ont lieu à des localisations similaires (à l'ouest, au nord-est), car les probabilités maximales des $VPAC_{RF}$ de ces zones sont faibles et des règles spatiales sont appliquées dans ces zones. D'autres changements de classe n'ont pas lieu aux mêmes localisations et correspondent surtout à un changement de classe des segments labellisés comme *savane* ou *verger* vers des classes de cultures pluviales ou vers les classes *pomme de terre* et *légumes*. Une partie des changements (nord-ouest et centre) est liée à la mauvaise qualité des données support pour l'application des connaissances (ou à l'absence de données dans ces zones).

La part d'aire qui a changé de classe entre la classification des données de télédétection et les méthodes de fusion des VPAC, représente 3,42 % de la zone d'étude pour la méthode de la moyenne pondérée et 3,27 % pour la méthode de la probabilité maximale. La possibilité de changement est aussi réduite par le seuil ΔRF de 0,2 qui limite le nombre de segments pouvant changer de classe. Les deux méthodes n'ont pas le même effet sur la répartition des aires par classe (tableau 20).

L'application des deux méthodes a un effet opposé sur l'aire des classes *maïs*, *culture boisée de rente*, *savane*, *bâti* et *maïs associé à une autre culture*. L'utilisation de la méthode de moyenne pondérée a tendance à augmenter la proportion des cultures dominantes au sein d'un groupe comme le montrent les résultats de classification (section 2.3 Vers une améliorations des résultats ?).

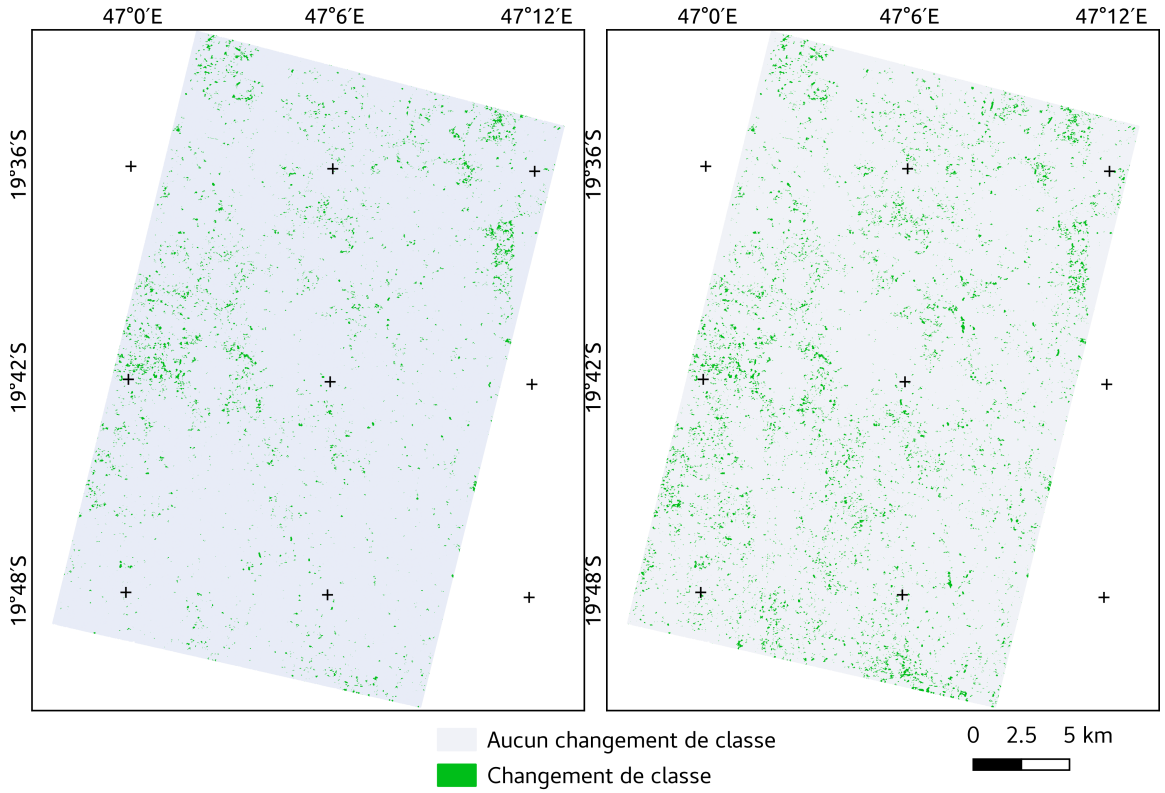


Figure 24 Cartes des segments ayant changé de classe après l'application du module des règles spatiales. À gauche, selon la méthode de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la probabilité maximale et à droite, par la moyenne pondérée.

Table 20 Répartition des aires de chaque classe en pourcentage, selon la classification des données de télédétection et l'application du module des règles spatiales en combinant les $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la probabilité maximale ($\Delta RF = 0,2$) et par la méthode de la moyenne pondérée ($\Delta RF = 0,2$ et $coef\ wRS = 1$).

	Pourcentage d'aire par classe		
Occupation du sol	**Télédétection**	**Méthode de la probabilité maximale**	**Méthode de la moyenne pondérée**
Maïs	1,65	1,48	1,68
Riz irrigué	10,07	10,15	10,12
Riz pluvial	0,26	0,25	0,24
Légumes	7,57	7,33	7,19
Soja	0,23	0,41	0,40
Pomme de terre	1,90	2,56	2,93
Herbes et autres cultures fourragères	0,42	0,41	0,39
Verger	7,81	7,75	7.92
Culture boisée de rente	25,05	25,09	25.04
Sol nu agricole	0,02	0,02	0.02
Cultures potagères	2,93	2,68	2.12
Savane	34,68	34,68	34.35
Sols rocheux	0,04	0,04	0.04
Bâti	2,90	2,91	2,87
Eau	0,06	0,06	0,06
Maïs associé	3,58	3,39	3,82
Riz pluvial associé	0,83	0,79	0,78
Pourcentage d'aire ayant changé de classe entre la carte de référence (télédétection) et l'application des méthodes			
	Référence	3,27	3,42

Autres combinaisons de paramètres : Aires

L'étude des résultats d'autres combinaisons de paramètres du module permet de mieux comprendre son impact sur la modification des classes finales dans le $VPAC_{Fusion}$. Certains changements sont amplifiés par les paramètres extrêmes du module et une partie des variations peut être analysée[a].

Paramètres $\Delta RF = 0,4$ et $coef\ wRS = 1$:
Les changements d'aires avec cette configuration ne suivent pas nécessairement la dynamique observée du modèle avec les paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$ (tableau 20. Certaines tendances sont confirmées et d'autres sont inversées. La dynamique qui concerne les cultures qui dominent certains groupes (*pommes de terre* (+4,19) et le *maïs associé* (+17,37)) est confirmée et amplifiée. Par contre, les classes *maïs* (-9,07), *riz irrigué* (-1,44), *verger* (-3,82), *culture boisée de rente* (+3,00), *sols rocheux* (+0,3) et *eau* (+1,78) ont une tendance inverse. La classe *savane* reste stable par rapport à la classification des données de télédétection. Au total, **8,63 %** de la zone d'étude se voit attribuer une nouvelle classe.

Paramètres $\Delta RF = 0,5$ et $coef\ wRS = 5$:
Pour ces paramètres, seuls le *maïs* (-37,37) et le *sol rocheux* (-22,98) ont une tendance inverse à celle observée avec les paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$. Dans cette configuration, la classe *savane* perd 31,44 points de pourcentage par rapport à la carte de référence alors qu'elle n'avait perdu que 0,95 point avec les paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$. Par contre, l'augmentation de précision de la classe *maïs associé à une autre culture* est beaucoup moins importante que pour les deux autres configurations. L'une des raisons de ce changement est l'absorption de la classe *soja* par la classe *maïs* (répondant à des règles similaires) et dans une moindre mesure avec la classe *pomme de terre*. L'aire totale qui a une nouvelle classe d'occupation des sols correspond à 33,89 % de la zone d'étude.

Les différentes dynamiques observées entre ces trois configurations du module montrent l'influence du seuil ΔRF qui permet une modification d'un nombre bien plus important de segments. L'augmentation du poids des règles spatiales dans la constitution du $VPAC_{Fusion}$ souligne le déséquilibre entre les règles expertes pour chaque classe. La très forte augmentation de l'aire et la localisation de la classe *pomme de terre* montre que les règles de cette classe sont déséquilibrées par rapport aux règles des autres classes. Les métriques d'évaluation vont dans ce sens avec une précision de 19,94 % et un désaccord de quantité de 9,8[b] pour les paramètres $\Delta RF = 0,5$ et $coef\ wRS = 5$, ce qui, avec la classe *savane* (q=8,5[b]), montre que ces deux classes sont sur-représentées.
L'analyse de l'évolution des aires reclassées et des résultats de la classification indiquent que ce sont principalement les erreurs de quantité (Q) qui sont à l'origine du mauvais résultat de classification. Les résultats montrent que ce sont les classes *pomme de terre* et *savane* qui sont à l'origine de la majorité des erreurs. Les erreurs de répartition (A) sont surtout issues de l'absorption des classes correspondant aux cultures pluviales par la classe *savane*.

[a] Les changements d'aires en points de pourcentage sont calculés par rapport à l'aire des classes issues de la classification des données de télédétection.

[b] Pour une meilleure lisibilité et une cohérence avec les autres résultats présentés, ce résultat a été multiplié par 100.

2.4.1 Incidence sur les probabilités des segments

L'un des objectifs du développement de cette méthode est de produire des cartes d'occupation et d'usage des sols plus cohérentes (chapitre 4, sous-section 4.1 Analyse de la « cohérence agricole » des cartes), objectif atteint pour l'instant de manière limitée. Le second objectif visant à produire des cartes plus fiables est analysé dans les paragraphes suivants. Pour effectuer ces comparaisons, nous utilisons la probabilité de la classe affectée au segment, c'est-à-dire la probabilité la plus élevée de ce segment.

Les résultats de la méthode de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la probabilité maximale n'étant pas indépendants des résultats de la classification des données de télédétection, la distribution statistique des probabilités maximales (*i.e.* probabilités des classes affectées à chaque segment) est très similaire à celle de la classification des données de télédétection ($VPAC_{RF}$). Par contre, l'application de la méthode des moyennes pondérées permet d'observer des résultats différents. La distribution des probabilités indépendamment de la classe est similaire entre les trois résultats. Une légère diminution pour les $VPAC_{Fusion}$ issus de la moyenne pondérée peut être notée (figure 25), principalement du fait de la distribution des probabilités qui est plus répartie entre les classes du $VPAC_{RS}$ par rapport au $VPAC_{RF}$. Le calcul de la moyenne pondérée entre ces deux VPAC entraîne nécessairement une diminution de la limite haute des probabilités des segments concernés par la reclassification. De manière opposée et logique, la méthode des probabilités maximales augmente la fiabilité de la majorité des classes d'occupation des sols.

2.4.1.1 Distribution des probabilités par classe

Pour les $VPAC_{Fusion}$ issus de la méthode de la moyenne pondérée, seules quatre classes de cultures ont une augmentation de leurs probabilités médiane et moyenne, ainsi qu'une réduction de la dispersion des probabilités (*riz pluvial*, *légumes*, *cultures potagère* et *herbes et autres cultures fourragère*). Ce changement peut aussi être observé pour les classes non agricoles (figure 25). Les autres classes agricoles voient la moyenne, médiane et dispersion de leurs probabilités augmenter à l'exception de la classe *riz pluvial associé à une autre culture* dont l'écart type diminue. Avec la méthode de la moyenne pondérée, la fiabilité des cultures qui avaient une précision de classification issue des données de télédétection déjà bas n'est pas augmentée. Cela signifie que les cartes produites avec ces paramètres ne sont pas plus fiables.

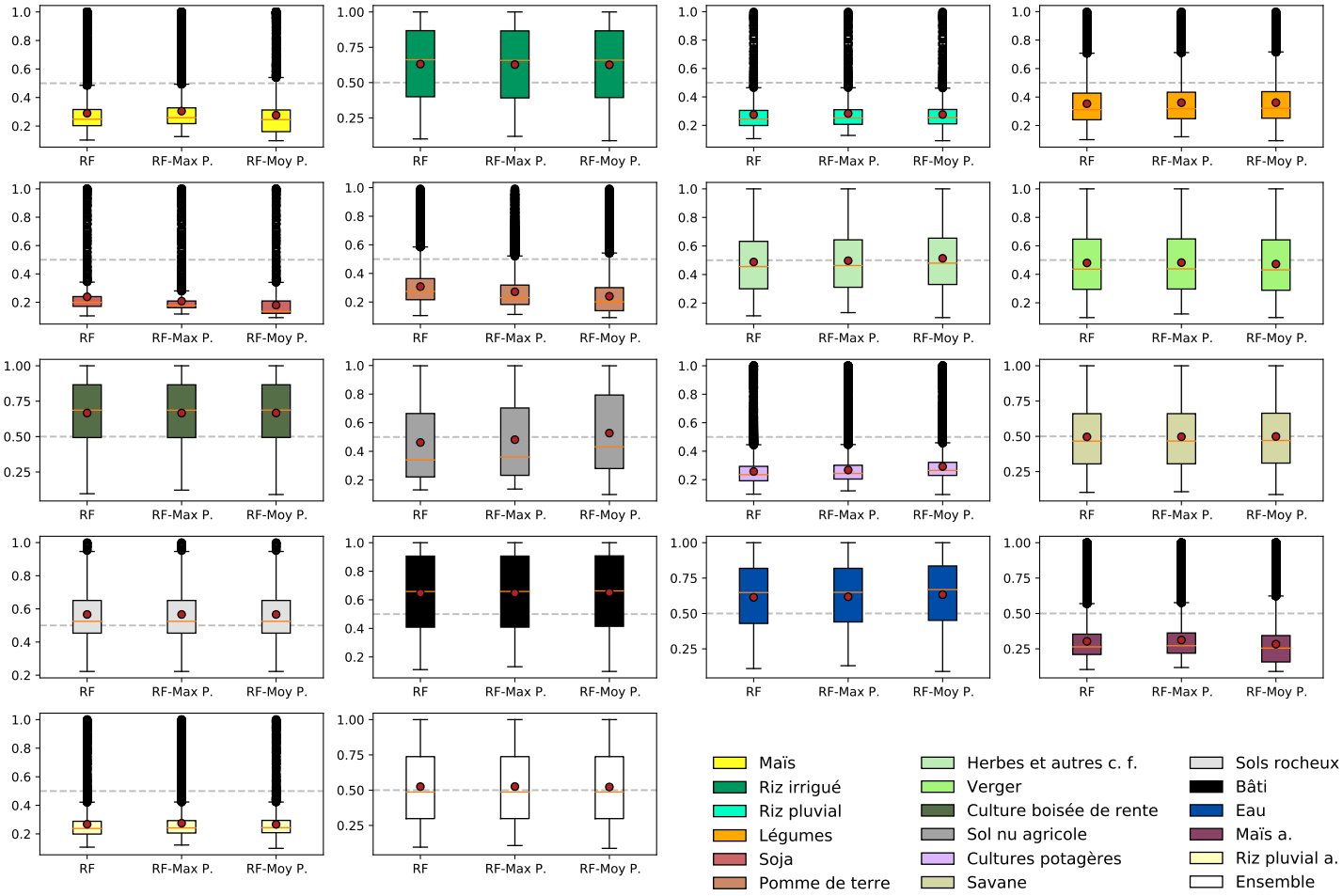


Figure 25 Distribution des probabilités par classe et pour l'ensemble des segments. **RF** correspond à classification des données de télédétection. **Max P.** et **Moy P.** correspondent respectivement aux classifications issues de l'application du module des règles spatiales et de la combinaison des des $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la probabilité maximale et de la moyenne pondérée.

2.4.1.2 Évolutions des probabilités des segments de la BD de vérité terrain

Dans un VPAC, une haute probabilité pour la classe affectée au segment signifie que le classifieur a peu de chance de se tromper. Lorsque la probabilité de la classe affectée au segment est basse, cela signifie que l'attribution de la classe est en partie liée au hasard. À partir de la base de données de vérité terrain, nous avons comparé les distributions des probabilités des classes affectées aux segments avant (probabilités issues du $VPAC_{RF}$) et après application du module des règles spatiales (probabilités issues du $VPAC_{Fusion}$ obtenu par la méthode de la moyenne pondérée). Ces résultats apportent un complément d'information aux résultats présentés dans les sections précédentes. Sur l'ensemble de la base de données, seulement une minorité de segments a changé de classe après application du module 2 : règles spatiales, l'analyse est donc limitée aux variations de probabilités de ces quelques segments.

La figure 26 représente les densités[1] des segments bien classés et mal classés pour la classi-

[1] La densité est la proportion d'observations par unité de mesure dans un intervalle.

fication des données de télédétection (bleu et orange respectivement), et pour le $VPAC_{Fusion}$ obtenu par la méthode de la moyenne pondérée (vert et rouge respectivement). Dans le cas où des segments bien classés et mal classés sont superposés et que la densité est élevée, cela signifie qu'il y a un nombre important de segments qui ont une forte chance d'être respectivement bien ou mal classés.

Les classes ayant des précisions de classification élevés (tableau 15 et tableau 17) comme les classes *riz irrigué* (n=2), *vergers* (n=23), *culture boisée de rente* (n=24), ou les classes non agricoles (n de 104 à 109) ont une densité de répartition élevée pour des valeurs probabilité proches de 1 pour les segments correctement classés. Pour ces classes, les segments mal classés ont une probabilité faible (proche de 0,25). Pour la classe *verger* (n=23), l'application de la méthode de la moyenne pondérée avec les paramètres $\Delta RF = 0,2$ et $coef\,wRS = 1$ permet une légère augmentation des probabilités pour les segments correctement classés, les rendant donc plus fiables. Les segments mal classés correspondent aussi à des probabilités plus faibles qu'avant l'application du module.

Pour les classes correspondant aux cultures pluviales, une diminution des probabilités est mesurée (n =1, 2, 7, 30, 1001, 1002 pour correspondance voir tableau 4). Cependant, peu de changements sont observables pour les classes de cultures associées. Les segments correctement classés perdent en fiabilité pour la classe *riz pluvial associé à une autre culture* (n=1002), phénomène aussi observé pour la classe *maïs associé à une autre culture* (n=1001). Pour toutes ces classes à l'exception de la classe *maïs* (n=1), les densités des segments correctement et mal classés ont des pics proches, là où les probabilités sont basses. Cela est aussi observable pour les cultures irriguées (n=6, *légumes* et n=10, *pomme de terre*) mais avec une densité pour les segments bien classés correspondant à des probabilités plus élevées. La classe *verger* (n=23) et la classe *savane* (n=104) ont quant à elles un nombre important de segments mal classés ayant des probabilités supérieures à 0,5.

La superposition des densités des segments bien et mal classés est représentative des classes qui ont des précisions de classification basses, avec de fortes confusions comme analysé dans la section 1.2. Cette superposition peut signifier deux choses :

- soit le segment ne correspond pas à une classe de la nomenclature et la base de données de vérité terrain comporte des erreurs ;
- soit le segment est sur une zone complexe et ses propriétés ne correspondent pas à une observation physique (télédétection) pure ni à un ensemble de règles spécifiques à une classe. Le segment correspond alors probablement à une occupation du sol mixte, c'est-à-dire qu'il ne correspond pas à une unité (parcelle, arbre, espace continu homogène).

Les résultats de la méthode des combinaisons maximales ne sont pas présentés visuellement, car cette méthode n'est pas indépendante de la classification des données de télédétection et peu de segments ont vu leur classe changer. La distribution des probabilités des segments de la base de données de vérité terrain est peu différente de celle de la classification des données de télédétection. Dans l'ensemble, les mêmes constats peuvent être dressés que pour méthode de la moyenne pondérée : la fiabilité des segments n'est pas améliorée.

Autres combinaisons de paramètres : Fiabilité

Paramètres $\Delta RF = 0,4$ **et** $coef\ wRS = 1$ **:**, méthode de la moyenne pondérée
L'augmentation du seuil $\Delta RF = 0,4$ permet de modifier plus de segments dans le module 2 : règles spatiales, et donc de mieux comprendre l'impact du module sur la fiabilité des segments. De manière générale, à l'issue de l'application du module des règles spatiales, les segments mal classés ont des probabilités plus basses que les ceux issus de la classification des données de . Ce résultat renforce l'hypothèse que les $VPAC_{RS}$ de ces segments ont sûrement une répartition proche de l'équiprobabilité. Dans ce cas, la combinaison entre $VPAC_{RS}$ et $VPAC_{RF}$ doit être améliorée pour que ceux-ci soient comparables de manière plus sûre. Ce nivellement vers le bas est surtout observable pour la majorité des classes de culture qui sont cultivées sur de petites parcelles.

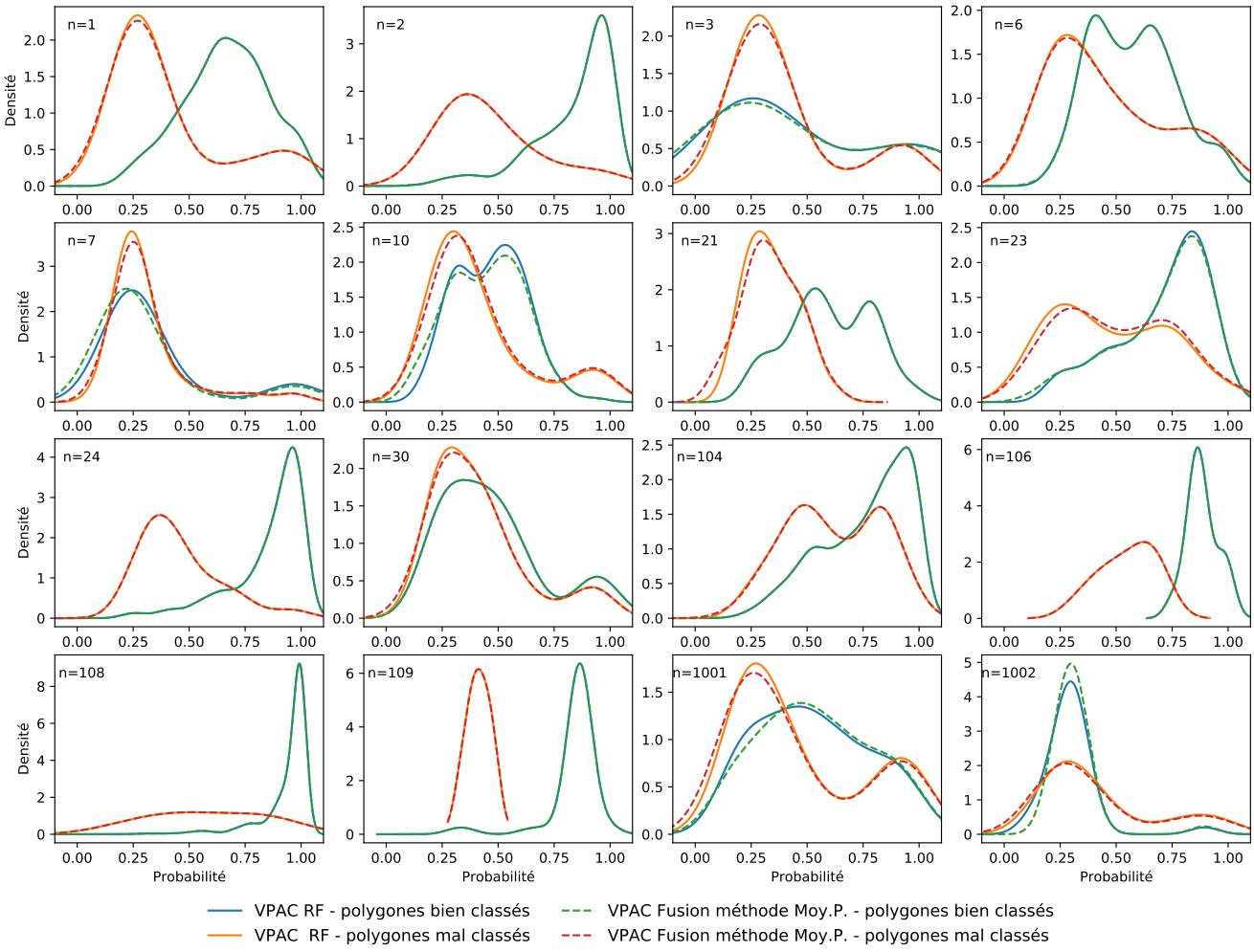


Figure 26 Distributions des probabilités de chaque classe pour les segments bien classés et mal classés. **RF** correspond à classification des données de télédétection et **Moy P.** à la classification issue de l'application du module des règles spatiales et de la méthode de combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la moyenne pondérée (paramètres $\Delta RF = 0,4$ et $coef\ wRS = 1$). La classe *sol nu agricole* n'est pas représentée car aucun segment n'a été attribué à cette classe.

2.5 Synthèse et résultats spatiaux qualitatifs

Dans les précédentes sections, l'impact de l'application du module des règles spatiales sur la classification des données de télédétection a été analysé de manière quantitative à travers les métriques d'évaluation des résultats, les changements de répartition d'aires et l'évolution de la fiabilité des segments. La section qui suit présente une synthèse de ces analyses à travers des exemples pour comprendre l'agencement spatial de ces changements. Dans cette section sont analysées qualitativement la carte issue de la classification des données de télédétection et trois cartes issues de trois configurations différentes du module 2 : répartitions spatiales et pour lesquelles la méthode d'agrégation des VPAC est la moyenne pondérée. L'analyse qualitative des résultats se fait aussi en comparaison à l'image THRS de référence Pléiades (d'une résolution de 0,5 m * 0,5 m).

2.5.1 Carte d'occupation et d'usage des sols produite avec les $VPAC_{Fusion}$

La figure 27 illustre la relation entre le type d'occupation du sol (obtenu après application du module des règles spatiales avec une combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la moyenne pondérée) et la spatialisation des probabilités les plus élevées par segment (*i.e.* spatialisation des probabilités des classes affectées aux segments). Les cultures pluviales et les parcelles de petite taille peuvent être visuellement associées aux probabilités basses.

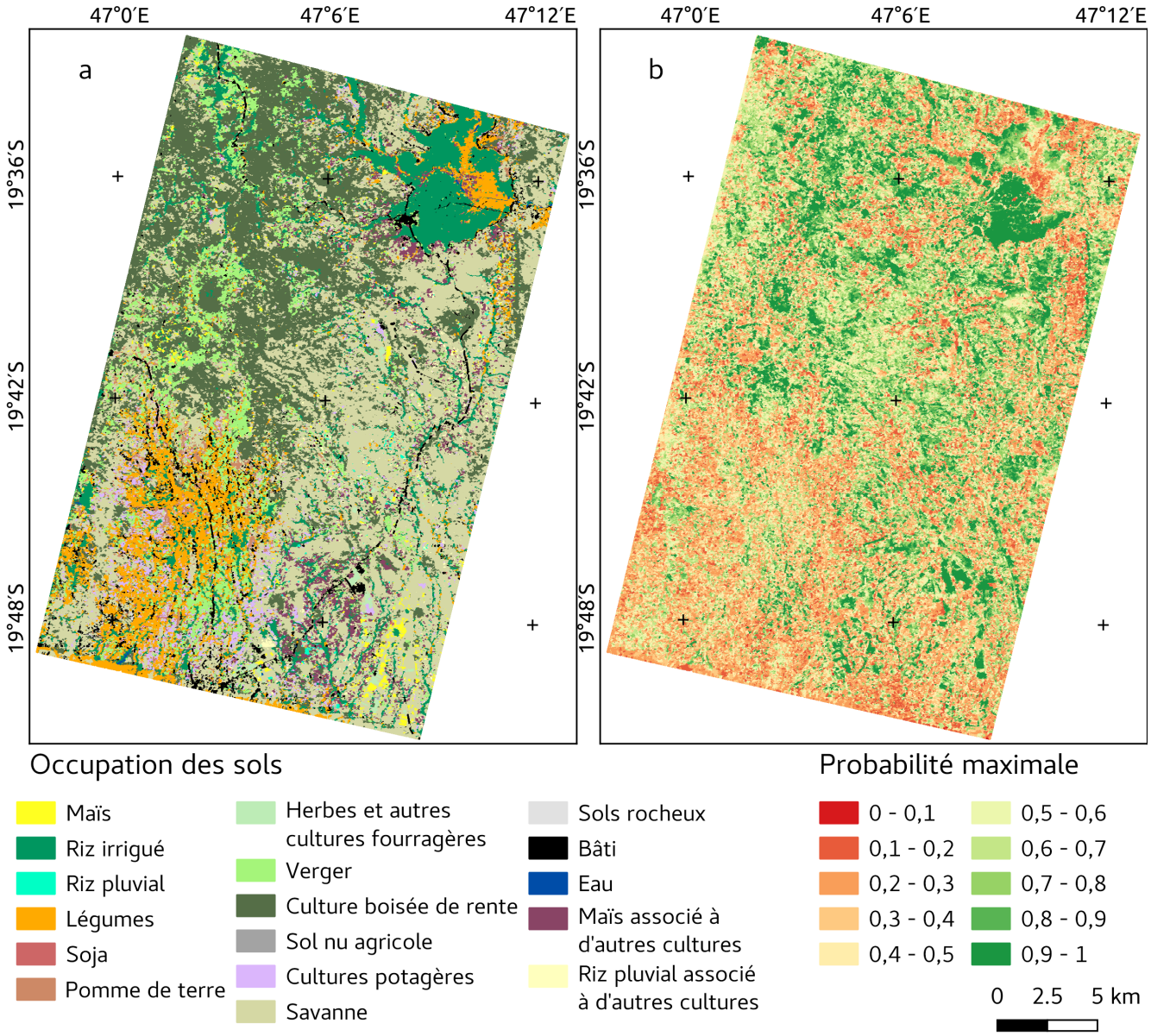


Figure 27 (**a**) Carte de l'occupation des sols obtenue après application du module des règles spatiales avec une combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la moyenne pondérée paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$) et (**b**) carte des probabilités les plus élevées pour la méthode de la moyenne pondérée (paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$).

2.5.2 Savane et cultures pluviales

La classe *savane* est sur-représentée dans la cartographie de l'occupation du sol issue de la classification des données de télédétection (figure 22), confirmant la valeur du désaccord de quantité q de 5,64[2] (tableau 15, annexe I, tableau 25). La répartition spatiale de cette classe est en partie juste, mais son hégémonie spatiale est largement sous-estimée par le désaccord de répartition a. La classe *savane* est dominante sur les reliefs où sont cultivées les cultures pluviales (centre-est de la carte). Les collines, plateaux et pentes sont les zones principalement dédiées aux cultures

[2]Pour une meilleure lisibilité, q et a ont été multipliés par 100.

pluviales, correspondant majoritairement aux classes *maïs*, *maïs associé à d'autres cultures* et *soja*. Bien que la localisation de la classe *savane* soit correcte, l'absorption des cultures pluviales par cette classe entraîne une sous-estimation des valeurs de désaccord de quantité et de répartition plus importante que ce qui est indiqué par les résultats de la classification. En effet, la valeur de rappel (78,78) reste élevée par l'utilisation de la pondération par l'aire de chaque polygone[3] (tableau 15). Cette confusion entre la savane et les cultures pluviales met aussi en avant la gestion difficile de ces champs par les agriculteurs. Ces zones ne sont généralement pas irriguées et la gestion de ces champs est généralement minimale. La densité de semis peut aussi être faible et ces champs ne sont pas prioritaires pour le désherbage contrairement aux cultures de légumes, ils peuvent donc être confondus avec la savane. Les résultats de classification montrent le fait que la classe *riz pluvial* est la plus touchée.

La figure 28 met en correspondance les cartes issues de la classification des données de télédétection (**a**) et du $VPAC_{Fusion}$ issu de la moyenne pondérée selon trois combinaisons de paramètres : (**b**)$\Delta RF = 0.2$ et $coef\ wRS = 1$, (**c**) $\Delta RF = 0.4$ et $coef\ wRS = 1$ et (**d**) $\Delta RF = 0.5$ et $coef\ wRS = 5$ pour une zone de cultures principalement pluviales et de savanes le long de route nationale 7 reliant Antsirabe à la capitale (dont le tracé est visible en noir, représenté par la classe *bâti*). Entre les cartes **a**, **b** et **c** les changements de classe sont peu importants et des permutations entre les cultures pluviales sont observables. La taille du domaine pluvial est aussi en légère augmentation dans les zones favorables à ces cultures comme le long de la route nationale, surtout pour la carte **c**. La carte **d**, bien que présentant des résultats très erronés, montre le poids du module dans le choix des classes avec une augmentation des cultures pluviales le long de la route nationale. Il est intéressant de noter l'extension de la classe *pomme de terre*[4]. Cette culture n'est presque jamais observée sur le terrain à cette localisation, mais **l'emplacement de cette classe montre les champs de culture pluviale sur les collines et plateaux**.

L'objectif d'améliorer la classification des cultures pluviales et d'éviter une trop grande confusion entre le domaine cultivé et le domaine non agricole n'est que partiellement atteint avec la méthode de la moyenne pondérée avec les paramètres $\Delta RF = 0.2$ et $coef wRS = 1$ (sous-section 2.3 Vers une améliorations des résultats ? et sous-sections suivantes). Cependant, l'utilisation du module avec une plus grande liberté dans la modification des $VPAC_{RF}$ (via un ΔRF haut et un seuil de probabilité minimum d'un segment haut) permet d'identifier les zones de cultures pluviales dans la savane. L'application du module permet l'identification de zones d'implantation favorables aux cultures pluviales à l'intérieur de la classe savane (même si la classe de culture pluviale affectée par le module à ces segments est erronée) qui sont vérifiées à partir d'une comparaison avec l'image THRS Pléiades (carte 28 **d**).

[3] La savane est la classe la plus représenté dans la base de données de vérité terrain.

[4] L'extension de la classe *pomme de terre* en particulier est liée à un impact plus important des règles spatiales pour cette classe par rapport aux autres classes (section 2.3 Vers une améliorations des résultats ? , encarts Autres combinaisons de paramètres : Résultats et Autres combinaisons de paramètres : Aires).

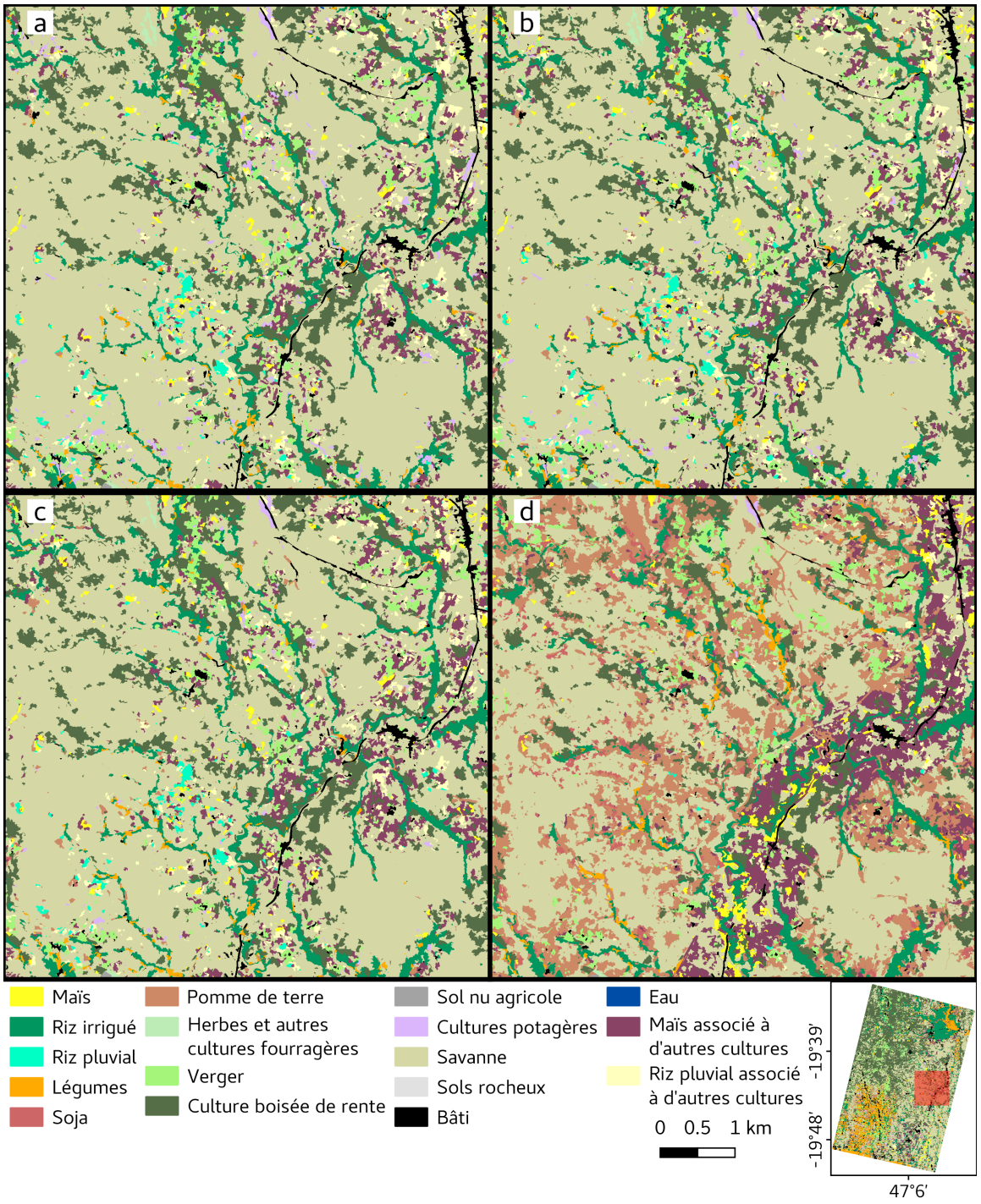


Figure 28 Cartes de l'occupation des sols selon (a) la classification des données de télédétection, l'application du module des règles spatiales avec combinaison des $VPAC_{RF}$ et $VPAC_{RS}$ par (b) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0.2$ et $coef\ wRS = 1$, (c) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0.4$ et $coef\ wRS = 1$ et (d) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0.5$ et $coef\ wRS = 5$.

2.5.3 Vergers et légumes

Pour la classification des données de télédétection, les erreurs de la classe *verger* sont de même nature que celles de la classe *savane* : une sur-représentation en termes de quantité et de nombreuses erreurs d'allocation spatiale (ces classes sont donc sur-représentés et localisées aux mauvais emplacements). Ces erreurs ne sont pas visibles au regard des résultats de classification où le désaccord de quantité est faible (q : 0,29) (figure 15). Le désaccord entre la répartition et le rappel sont plus cohérents avec les erreurs de la carte, où les confusions entre la classe *verger* et les autres classes de culture sont principalement liées au houppier d'arbres fruitiers parfois présents au sein ou au bord de parcelles dont la culture principale est une culture pluviale ou un légume. Ces houppiers couvrent la végétation des strates inférieures qui est moins visible. L'application du module avec les paramètres $\Delta RF = 0.2$ et $coef\ wRS = 1$ entraîne une **augmentation de l'aire de cette classe et ne résout pas le problème de dominance spatiale de cette classe** (cartes 22, 27, tableau 18).

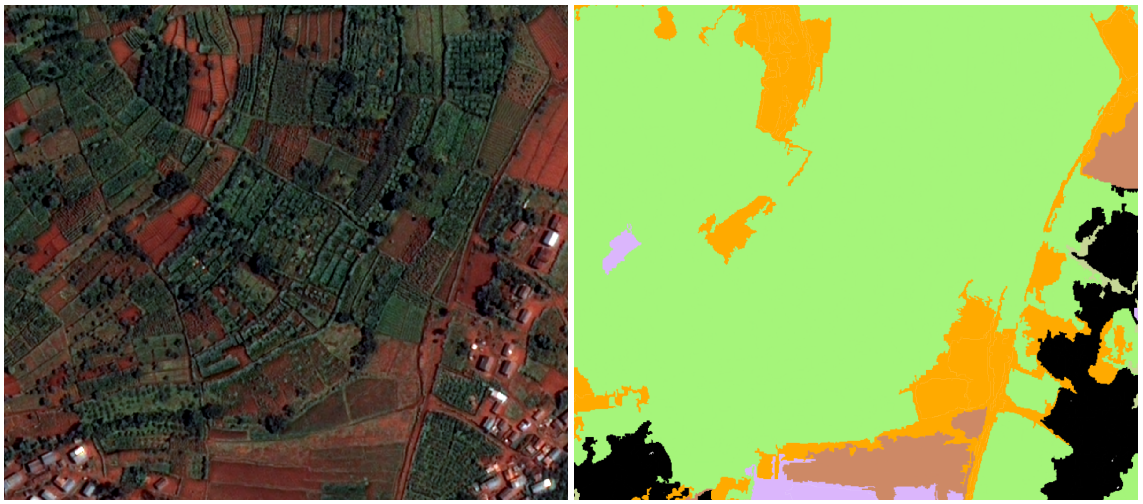
Dans la partie **nord-ouest** de la carte, les vergers sont sur-représentés par rapport aux cultures pluviales et cela n'est pas observable via les métriques utilisées pour évaluer les trois résultats (télédétection : tableau 15 et module spatial : 18), car il n'y a presque aucun point de vérité terrain sur cette zone (Chapitre 3, figure 16). Les vergers sont tout de même la culture principale de ces Fokontany à cause de leur éloignement au réseau routier ou aux principaux points de vente[5].

Dans la partie **Ouest** de la carte issue de la classification des données de télédétection (carte 22), les vergers sont aussi sur-estimés. Pour la classe *verger*, les résultats de classification semblent donc particulièrement inexacts dans les zones où cette culture est censée être cultivée. Sa quantité est sur-estimée et sa répartition spatiale n'est pas réaliste avec les observations de l'image THRS Pléiades. Cette erreur est présente malgré les différents paramétrages du module des règles spatiales. L'application du module des règles spatiales ne permet pas d'allouer une classe d'occupation ou d'usage des sols différente, le résultat n'est donc thématiquement pas plus juste.

Au **Sud**, c'est en majorité des légumes et des pommes de terre qui sont cultivés alors que l'occupation des sols est identifiée comme appartenant à la classe *verger*. Les communes d'Ambano et d'Alakamisy, situées au **Sud Ouest** de la carte, sont spécialisées dans la production de légumes et de pomme de terre. Ces cultures sont assez bien représentées, mais les vergers sont de nouveau sur-représentés. La figure 29 montre le type de confusion entre parcelles de légumes et bordures de champs pouvant contenir des arbres fruitiers. **L'application du module permet des changements de classe, surtout entre les classes *verger, légumes, pomme de terre* et *cultures potagères***. À partir d'un certain seuil, c'est la classe *pomme de terre* qui devient dominante, mais la classe *légumes* reste fiable.

Hors de ces zones, **la répartition de cette culture est juste et les résultats de la classification sont satisfaisants**.

[5]La culture majoritaire de ces Fokontany est le pommier, ce fruit étant le seul qui peut être conservé et transporté facilement durant la saison des pluies où les routes sont coupées



(a) Image Pléiades. **(b)** Résultat de la classification de l'image Pléiades.

Figure 29 Exemples d'erreurs de classification des données de télédétection dans la commune d'Alakamisy où les parcelles de *légumes* (orange) sont classées comme *verger* (couleur vert), alors que seuls quelques pieds de pommiers sont présents. Ce cas est typique et peut être corrigé par le passage de la classe *verger* à la classe *légumes* ou *pomme de terre* grâce au module 2 : règles spatiales. Plus occasionnellement, ce changement peut être en faveur des cultures potagères. Les autres classes présentes sur la carte (b) sont : *pomme de terre* (marron) et *cultures potagères* (mauve) et *bâti* (noir).

2.5.4 Cuvette d'Ambohibary

Au **Nord-Est** de la zone, la cuvette rizicole d'Ambohibary est correctement cartographiée à la fois à partir des données de télédétection et par l'application du module spatial, avec une représentation spatiale correspondant à l'image THRS Pléiades pour les classes *légumes* et *riz irrigué* (figure 30 **a**, **b** et **c**). D'après la base de données de vérité terrain et notre expertise de la zone d'étude, des cultures pluviales sont aussi présentes parmi les légumes sur une partie surélevée dans le bas-fond. Ces cultures pluviales sont correctement représenté sur la carte de la classification des données de télédétection (figure 30 **a**). Mais la présence des cultures pluviales est légèrement diminuée avec l'application du module des règles spatiales et l'utilisation de la méthode de la moyenne pondérée avec les paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$ (figure 30 **b**) et $\Delta RF = 0,4$ et $coef\ wRS = 1$ (figure 30 **c**). Les changements de classe sont en accord avec l'image THRS Pléiades pour les espaces où les parcelles sont particulièrement petites et où la segmentation ne donne pas de résultats probants. Par contre, l'utilisation de paramètres permettant un plus grand degré de liberté pour la modélisation des $VPAC_{RF}$ (figure 30 **d**, paramètres $\Delta RF = 0,5$ et $coef\ wRS = 5$), entraîne une détérioration importante de la classification de cette zone où es classes étaient correctement localisées, bien que le domaine cultivé soit bien délimité si l'on considère que la pomme de terre est représentative du domaine agricole.

Sur les pentes de la cuvette, la classe *savane* est aussi prédominante par rapport aux cultures pluviales (Nord-Est et Sud-Est de la figure 30) et les mêmes remarques peuvent être faites que pour la section 2.5.2 Savane et cultures pluviales.

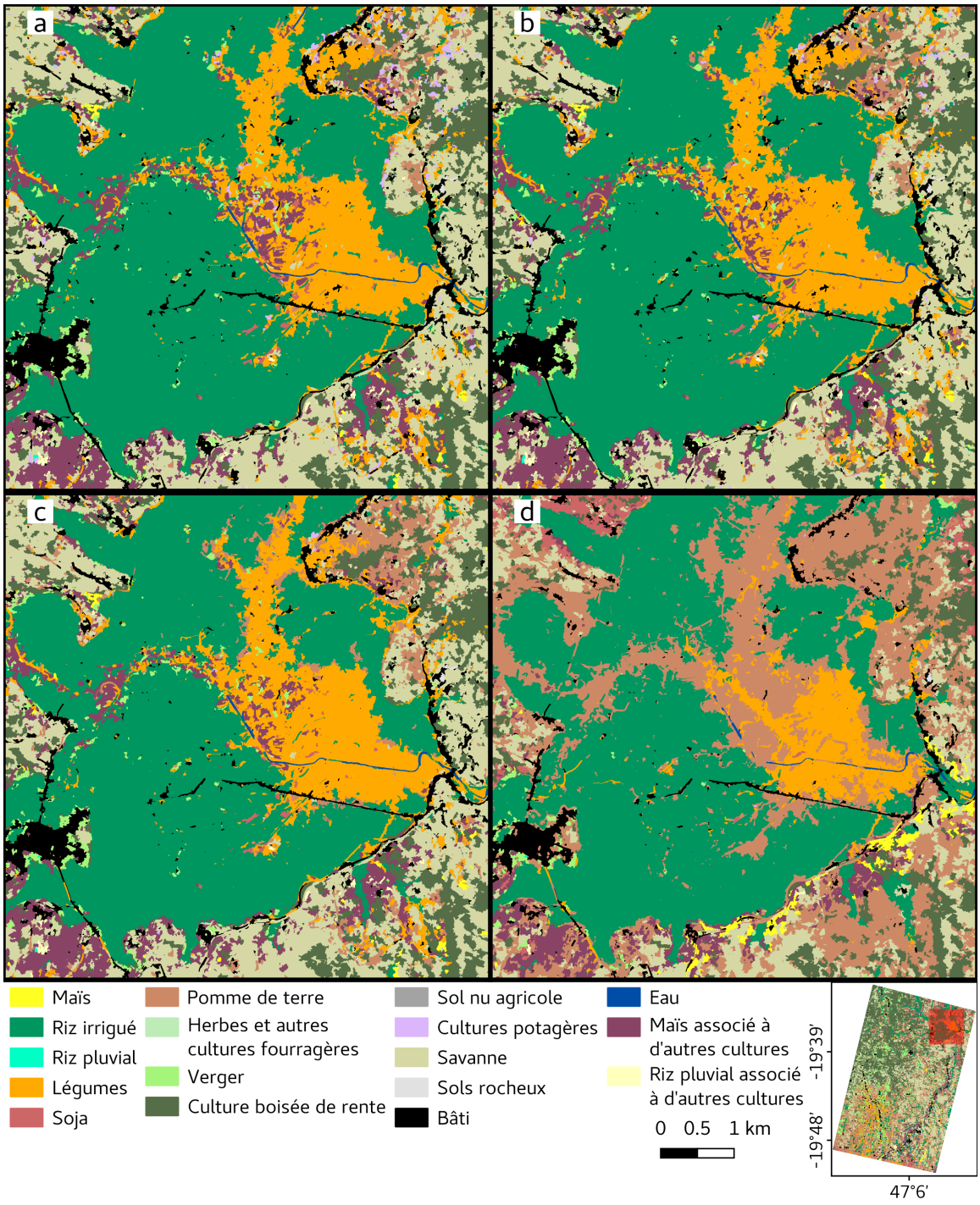


Figure 30 Cartes de l'occupation et de l'usage des sols de la cuvette d'Ambohibary, selon (a) la classification des données de télédétection, (b) l'application du module des règles spatiales en combinant les $VPAC_{RF}$ et $VPAC_{RS}$ par la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0,2$ et $coef\ wRS = 1$, (c) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0,4$ et $coef\ wRS = 1$ et (d) la méthode de la moyenne pondérée avec les paramètre $\Delta RF = 0,5$ et $coef\ wRS = 5$

2.5.5 Autres classes et localisations

Au **Sud-Est**, les parcelles de grande taille sont correctement caractérisées pour deux principales raisons. Cet espace est cultivé selon une forme entrepreneuriale et non pas familiale, les champs sont donc spatialement homogènes et les pixels sont purs (voir tableau 1, Chapitre 1). L'application du module ne change pas la classe de ces champs, sauf de manière marginale pour l'application du module avec les paramètres $\Delta RF = 0,5$ et $coef\ wRS = 5$.

La classe *bâti* est concernée par des erreurs de précision, dues à l'identification des sols nus comme du bâti. C'est notamment le cas des routes de terre. Cependant, la fiabilité de ces segments (c'est-à-dire leur probabilité de classe) est majoritairement inférieure à 0,9 et n'impacte donc que partiellement la production des données de temps de trajet aux bâtiments (chapitre 4, section 3.1 Préparation de données support pour l'application des connaissances et Annexe H). En dehors de cette erreur, faute d'une classe « *route* », la classe *bâti* correspond à la réalité. L'application du module n'a pas d'impact négatif sur ces classes, leur probabilité étant suffisamment élevée pour que la classe de ces segments ne soit pas modifiée.

2.6 Synthèse des résultats qualitatifs et quantitatifs

Dans cette section, nous avons présenté les choix de paramètres du module 2 : règles spatiales et de la fusion des $VPAC_{RF}$ et des $VPAC_{RS}$, dont une configuration pour la méthode la moyenne pondérée et une pour la méthode la probabilité maximale permettent d'améliorer les résultats de la classification des données de télédétection. Entre ces deux méthodes, seule la méthode de la moyenne pondérée montre des résultats significatifs. Ensuite, l'incidence de l'application du module avec différentes configurations a été étudiée à travers l'évolution de la probabilité des segments, les changements d'aire par classe, la distribution de la probabilité des segments et l'analyse qualitative des cartes produites. Les résultats montrent une légère amélioration par rapport à la classification des données de télédétection, moins de confusions entre le domaine cultivé et le domaine non-cultivé et une correcte identification des champs destinés à des cultures pluviales, mais sans attribuer le bon type de culture.

Afin de synthétiser les résultats présentés dans cette section, la Figure 31 présente à la fois une composition colorée de l'image Pléiades acquise au coeur de la saison culturale (Figure 31b), le résultat de la classification des données de télédétection (Figure 31a), et le résultat de la fusion de la classification des données de télédétection et du module règles spatiales à partir de la méthode de la moyenne pondérée, paramètre $\Delta RF = 0,5$ et $coef\ wRS = 5$ (Figure 31c). L'évolution de la caractérisation de l'usage des sols est particulièrement visible entre ces deux classifications vis-à-vis de l'image Pléiades. L'image Pléiades présente un paysage agricole composé à la fois de savanes et de champs de culture pluviales, ainsi que de zones de forêts éparses. Sur la Figure 31a, la classe dominante est la *savane* et la classe *culture boisée de rente* est correctement détectée. Par contre, les champs sont peu détectés et les classes attribuées sont *riz pluvial* ou *cultures potagères* ou *maïs associé à d'autres cultures*. Malgré la sous-détection des champs pluviaux, la précision globale à

l'échelle de la zone d'étude complète est élevée : 81,93 %. Ce résultat ne reflète pas les erreurs de classification tels que la confusion importante des classes de culture pluviale avec la classe *savane*. La Figure 31c présente un paramétrage qui donne un poids plus important au module des règles spatiales et permet par cela une détection des champs pluviaux dans les espaces de savane. Les espaces forestiers sont toujours bien détectés au sein de la classe *culture boisée de rente*, mais les champs sont attribués aux classes *pomme de terre*, *soja* ou *maïs associé à d'autres cultures*. Ces classes sont sur-représentées sur la vignette mais aussi sur la zone d'étude ce qui fait diminuer la précision globale à 75,89 %. En analysant ces deux classifications avec l'image Pléiades, il est possible d'affirmer que **l'application de paramètres donnant un poids important au module des règles spatiales a l'avantage de mieux détecter les champs de cultures pluviale en zone de savane, mais ne permet pas d'attribuer une classe correcte à ces champs, ce qui diminue les résultats de classification**.

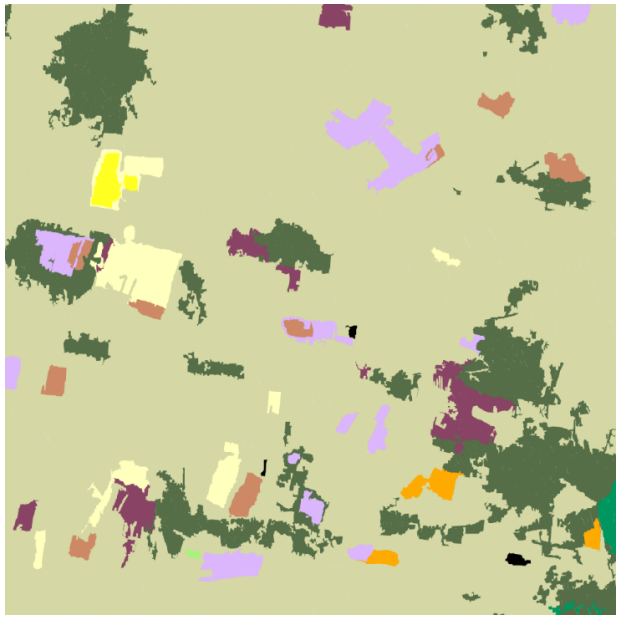
3 Résultats du module modélisation de la répartition spatiale des classes

Le module de répartition spatiale des classes utilise les $VPAC_{Fusion}$ pour distribuer les classes d'occupation des sols en fonction des statistiques agricoles des communes (chapitre 4, section 3.4 Application du module 4 : modélisation de la répartition spatiale des classes). Ce module permet néanmoins d'utiliser la connaissance que l'on a sur les habitudes culturales de certains villages. Ces données ne sont pas fiables et peu de Fokontany sont complets dans notre zone d'étude (Chapitre 4, tableau 13). L'application de ce module à la région des Hautes Terres ne peut être évaluée de manière quantitative, car la base de données terrain est trop peu représentative des répartitions au sein de chacun des Fokontany pour rendre compte des changements (la base de données représente 3,6 % de la zone d'étude). L'analyse des résultats est donc uniquement qualitative pour observer les changements effectués et leur pertinence.

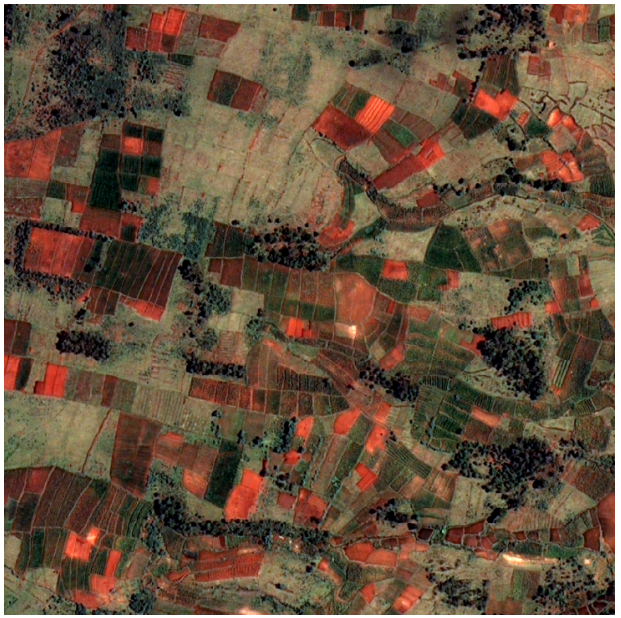
Dans la continuité des résultats présentés pour le module des règles spatiales, l'analyse porte sur la spatialisation des cultures pluviales dans les espaces dominés par la savane. L'exemple proposé dans le chapitre 4, section 3.4 Application du module 4 : modélisation de la répartition spatiale des classes de la commune d'Alakamisy est approfondi, et les statistiques qui induisent les changements de culture sont analysées.

Pour l'application de ce module, les résultats du module des règles spatiales ($VPAC_{Fusion}$ obtenus par la moyenne pondérée avec les paramètres $\Delta RF = 0,2$ et $coefwRS = 1$) sont utilisés comme point de départ. Les paramètres du module de répartition spatiale des classes sont :

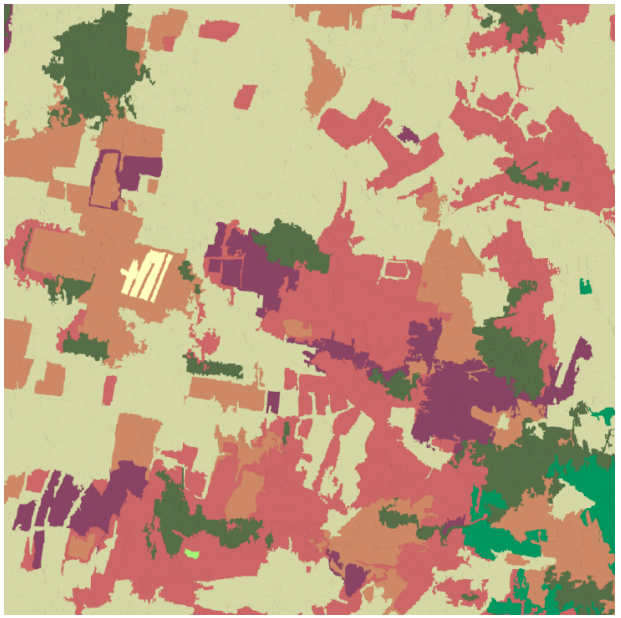
- le seuil de fiabilité des segments est fixé à **0,8** car ce seuil correspond au $3^{\text{ème}}$ quartile de la distribution des données ;
- et le seuil de variabilité entre les aires observées et théoriques est fixé à **0,05** car les statistiques ne sont pas sûres. Une hausse de ce seuil entraîne une sur-représentation des cultures qui ont une fiabilité élevée, telles que les classes non agricoles.



(a) Classification des données de télédétection. Précision globale : 81,93 %.



(b) Image Pléiades.



(c) Fusion de la classification des données de télédétection et du module règles spatiales à partir de la méthode de la moyenne pondérée, paramètre $\Delta RF = 0,5$ et $coef\ wRS = 5$. Précision globale : 75,89 %.

Figure 31 Différents résultats de classification en comparaison avec l'image Pléiades de référence (au centre). La légende des classification est disponible figure 30. L'image 31a représente la classification des données de télédétection telle que présentée dans la section 1. La classe majoritaire est la savane et peu de champs sont identifiés. L'image 31b permet de visualiser les espaces de savane (vert pâle) et les champs (marron et vert foncé). La figure 31c montre la détection des champs à l'aide de de la fusion de la classification des données de télédétection et du module des règles spatiales.

3.1 Changements généraux

La figure 32 présente la carte d'occupation et d'usage des sols issue de l'application du module de répartition spatiale des classes (**a**) ainsi que la carte des segments qui ont changé de classe (**b**). Les changements d'occupation des sols sont nombreux et sont logiques vis-à-vis de la répartition statistique utilisée. Les erreurs des données de répartition utilisées sont observables et les causes peuvent être comprises. Par exemple, la définition d'un espace agricole varie selon les Fokontany. Lorsque des pins sont proches des habitations, ils sont considérés comme faisant partie de l'espace agricole. Lorsqu'ils sont trop éloignés, ils sont considérés comme des espaces non exploités. Les Fokontany dont les statistiques sont disponibles et crédibles sont bien représentés dans leur ensemble. Par contre, lorsque les statistiques crédibles sont manquantes, cela entraîne une répartition non cohérente avec la classification des données de télédétection et avec ce qui peut être observé sur le terrain. Par exemple, au nord-est dans la cuvette d'Ambohibary, la proportion de terre considérée comme non agricole est surestimée pour certains Fokontany, transformant les rizières en classe *eau*, et sous-estimée pour d'autres, transformant la savane et les bois en cultures pluviales ou en classe *maïs associé à une autre culture*. Dans ces Fokontany, le résultat de l'application du module de répartition spatiale est donc médiocre.

L'utilisation d'une classe générale regroupant les classes de l'ensemble « Autre »[6] donne de bons résultats et permet une répartition correcte des classes qui composent cet ensemble. Les erreurs sont cohérentes (ex. classe *eau* à la place de la classe *riz irrigué*) et la répartition des classes étant basée sur les probabilités, elle respecte bien les occupations des sols possibles.

Dans les cas où les Fokontany n'ont pas de statistiques précises, les statistiques générales sont souvent inadaptées. La proportion de maïs associé à une autre culture est largement surestimée et l'obligation de répartition de certaines classes entraîne une répartition improbable.

[6] Ces classes sont : *herbes et autres cultures fourragères*, *sol nu agricole*, *culture boisée de rente*, *savane*, *sols rocheux*, *bâti* et *eau*.

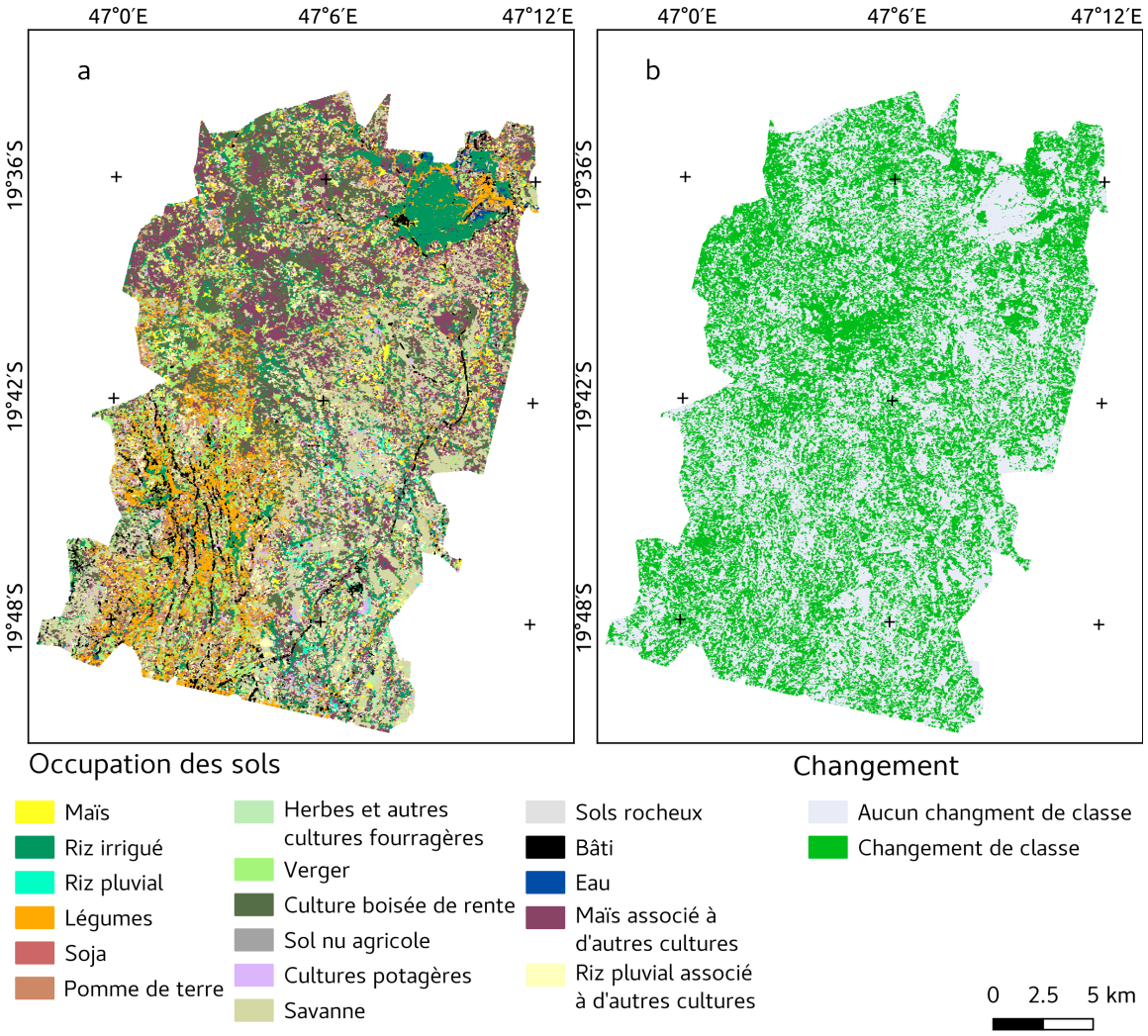


Figure 32 (a) Carte de l'occupation et de l'usage des sols après l'application du module de modélisation de la répartition spatiale des classes avec un seuil de fiabilité de 0,8 et un seuil de variabilité de 0,01. (b) Carte des segments ayant changé de classe entre l'application du module des règles spatiales (paramètres $\Delta RF = 0,2$ et $coef\ wRS = 1$) et l'application du module de modélisation de la répartition spatiale des classes avec un seuil de fiabilité de 0,8 et un seuil de variabilité de 0,01.

3.2 Plateaux et pentes, une place pour les cultures pluviales ?

Sur les pentes et plateaux dominés par la savane, les cultures pluviales et les bois, l'application du module a des effets contrastés. Dans le quart **Nord-Ouest** de la carte, la majorité de la classe *culture boisée de rente* est remplacée par la classe *maïs associé à une autre culture*. La raison de cette répartition incorrecte est le manque de données statistiques pour ces Fokontany éloignés.

Pour les pentes et plateaux proches de la route nationale où la savane est dominante, les changements sont plus cohérents. La classe *savane* est celle qui connaît le moins de changements et lorsque cela arrive, les changements de classes sont souvent de la classe *savane* vers une classe de culture pluviale.
Lorsqu'un changement de classe est effectué sur un segment initialement classé en culture plu-

viale dans le $VPAC_{Fusion}$, le changement s'effectue en général vers une autre classe de culture pluviale. La raison de ces changements s'explique par la méthode utilisée pour attribuer une classe à un segment. La probabilité de la classe de ce segment est comparée, non pas à celles des autres classes, mais à celles des autres segments (chapitre 4, section 3.3.1 Standardisation des résultats et création des $VPAC_{RS}$). La cohérence de la répartition des classes de culture pluviale ne peut donc être analysée visuellement.

Comme remarqué avec l'analyse du module des règles spatiales, la répartition des classes de cultures pluviales sur les zones auparavant labellisées comme classe *savane* est assez pertinente. La majorité des champs sont identifiés et la savane est correctement localisée. La classe *riz pluvial* sert parfois de variable d'ajustement lorsque l'occupation des sols peut être de la savane ou une culture pluviale (champ dégradé).

3.3 Évolution du Sud : les communes d'Alakamisy et d'Ambano

Le nord de la commune d'Alakamisy est enclavé et est majoritairement composé de cultures pluviales. A l'inverse, les Fokontany au sud de la commune cultivent majoritairement des légumes et des pommes de terre (chapitre 4, tableau 11). La commune voisine d'Ambano suit le même schéma.[7] Les données statistiques administratives de ces communes ne sont pas fiables mais dans ces deux communes les personnes enquêtées ont fourni des informations détaillées sur la répartition des cultures. De manière globale, la répartition des cultures obtenue est correcte et leurs emplacements sont réalistes. Les vergers sont correctement localisés, limitant le problème de confusion de ceux-ci avec les classes *légumes*, *pomme de terre* ou *cultures potagères* (figure 29). Une nette amélioration de la discrimination de ces classes peut même être soulignée. Par contre, la désagrégation des données sur les cultures que le personnel de la commune d'Alakamisy nous a fournies (chapitre 4, tableau 11) est incorrecte. Les classes *riz irrigué* et *légumes* sont sous-estimées, ce qui entraîne une légère sur-estimation de la représentation des classes *savane*, *riz pluvial* et *riz pluvial associé à une autre culture*. De manière générale, le domaine cultivé est surestimé (surtout la proportion de la classe *maïs associé à une autre culture*).

[7] Au total, ces deux Fokontany ont entre 50 et 75 % de leur aire dans notre zone d'étude. Quatre Fokontany y ont entre 75 et 99 % de leur aire et douze y sont en intégralité.

POINTS CLÉS DU CHAPITRE 5

- **TÉLÉDÉTECTION** :
  - la contribution de Sentinel-2, Venμs et des variables auxiliaires sur la précision globale a été évaluée, les deux meilleurs résultats de classification en termes de précision globale étant obtenus via la combinaison des bandes spectrales de Sentinel-2 et des variables auxiliaires (précision globale de 0,839) et par la combinaison des bandes spectrales et des indices de Sentinel-2 avec les variables auxiliaires (précision de 0,828) ;
  - les 17 classes sont réparties en 3 catégories de précision de classification : 7 classes, principalement non agricoles, avec de bons résultats (F-mesure 1 - 0,8), 4 classes avec des résultats moyens (F-mesure 0,8 - 0,5) et 6 avec de mauvais résultats (F-mesure 0,5 - 0). **Les erreurs sont liées à la fois à un désaccord de quantité et de répartition** ;
  - la **classe *savane* est la principale contributrice aux erreurs** de répartition et est surreprésentée dans la zone d'étude.

- **ASSOCIATION TÉLÉDÉTECTION - RÈGLES SPATIALES** :
  - une combinaison de paramètres pour la méthode de la moyenne pondérée et une combinaison pour la méthode de la probabilité maximale améliorent les résultats de classification des données de télédétection, mais **seule la méthode de la moyenne pondérée montre des résultats significatifs** ;
  - pour la méthode de la moyenne pondérée, la **F-mesure de toutes les classes est améliorée sauf pour les classes *maïs associé* et *riz associé***. Pour la méthode des probabilités maximales, les résultats sont plus hétérogènes entre les classes : 4 ne sont pas améliorées et 6 sont améliorées. Les autres classes sont stables ;
  - l'impact du module a été évalué selon différents paramétrages via l'analyse de **l'évolution des aires de chaque classe et des probabilités** ;
  - la **fiabilité des segments n'est pas améliorée** en appliquant le module 2 : règles spatiales, mais les classes sont réparties correctement ;
  - nous constatons une **meilleure détection des cultures pluviales en zone de savane** et une meilleure discrimination entre les classes *légumes* et *pomme de terre* par rapport à la classe *verger* (paramètres $\Delta RF = 0,5$ et $coef\ wRS = 1$) ;
  - lorsque le seuil ΔRF est diminué pour permettre à un nombre plus important de segments de voir leur classe RF modifiée par l'application du module des règles spatiales, le domaine cultivé est mieux identifié dans les zones de savane selon l'image THRS Pléiades, mais le type de culture affecté n'est pas forcément correct.

- **MODÉLISATION DE LA RÉPARTITION SPATIALE DES CLASSES** :
  - les résultats sont variables selon les données de répartition théorique des Fokontany ;
  - la classe *maïs associé* est sur-estimée dans tous les Fokontany. Cette classe est attribuée par le module au détriment de la classe *culture boisée de rente* ;
  - la **répartition des classes *cultures potagères, légumes, pomme de terre* et *verger* est réaliste** pour les Fokontany où les **connaissances sont considérées comme fiables** ;
  - dans ces Fokontany, la **délimitation du domaine agricole est acceptable** lorsque la carte est comparée à l'image THRS Pléiades.

Chapitre 6

Discussion

Table des matières

1. Bilan général de l'application des modules
   1.1. Rappel des limites existantes de la caractérisation des espaces agricoles complexes et des concepts principaux de l'approche développée
   1.2. Une application du modèle conceptuel sans temporalité
   1.3. Résultats globaux de l'application des modules spatiaux
2. Application : pistes et limites
   2.1. Une meilleure exploitation des images de télédétection
   2.2. L'aspect multiscalaire du modèle
   2.3. L'impact de la qualité des données sur les règles et les résultats
   2.4. Impact des choix pour la formalisation des connaissances et pour l'utilisation des VPAC
   2.5. Validation du modèle
3. Quels changements pour le modèle conceptuel ?
   3.1. Un modèle trop centré sur Madagascar, mais déjà pluridisciplinaire
   3.2. Penser une nouvelle approche thématique
4. Perspectives

Ce dernier chapitre présente le bilan de l'approche développée dans cette thèse à travers l'analyse de l'application du modèle conceptuel aux Hautes Terres de Madagascar. Il est structuré autour des trois objectifs fixés au début du manuscrit qui sont à la fois conceptuels et appliqués pour essayer d'améliorer la caractérisation des espaces façonnés par l'agriculture familiale en zone tropicale. Le résumé des résultats obtenus et leurs limites permettent d'identifier de nouvelles pistes pour améliorer l'approche proposée. Ces pistes sont dédiées à la fois au modèle conceptuel proposé et à son application à un espace agricole complexe. Pour terminer, deux principales perspectives de recherche sont proposées.

Les principaux objectifs présentés dans ce manuscrit pour tenter d'**améliorer la caractérisation des espaces façonnés par l'agriculture familiale en zone tropicale** sont :

1. de développer un modèle conceptuel pour combiner à la fois les approches de traitement de données de télédétection et de modélisation spatiale.
   Cet objectif a été traité dans le chapitre 2 Modèle conceptuel avec la proposition d'un modèle conceptuel mobilisant à la fois des concepts et méthodes issus du traitement des données de télédétection et de la modélisation des dynamiques spatiales ;
2. d'appliquer ce modèle conceptuel à la région des Hautes Terres de Madagascar, système agricole complexe en zone tropicale, durant la saison culturale principale.
   Cet objectif est abordé dans les chapitres 3 Zone d'étude, matériel et méthodes et 4 Méthode d'application des modules à la région des Hautes Terres ;
3. d'évaluer l'application du modèle conceptuel en étudiant les résultats de caractérisation de l'occupation des sols intermédiaires et finaux, selon différents paramétrages du modèle, pour obtenir une cartographie plus cohérente et fiable de l'occupation et de l'usage des sols.
   Le chapitre 5 Résultats d'applications présente les résultats de l'application du modèle conceptuel et propose une évaluation à la fois quantitative et qualitative des cartes produites.

La première section de ce chapitre résume les modules du modèle conceptuel et explique pourquoi l'aspect temporel n'a pas été appliqué aux Hautes Terres de Madagascar. Ensuite, les principaux résultats de l'application des modules spatiaux sont rappelés. La deuxième section est consacrée à l'analyse de l'application des modules à travers leurs aspects positifs et à identifier quelles sont les pistes d'amélioration pour l'application du modèle conceptuel. Via cette section, sont abordés l'aspect multiscalaire du modèle, l'impact des données sur la classification des données de télédétection et sur l'application des règles, les limites de notre méthode de formalisation des connaissances et de l'utilisation des VPAC (vecteur de probabilité d'appartenance aux classes) et la méthode de validation du modèle. La troisième section est dédiée à une réflexion sur le modèle conceptuel et à une adaptation possible pour le rendre générique à d'autres espaces et le rendre interdisciplinaire. La dernière section présente les perspectives de ces travaux, à la fois d'un point de vue méthodologique en limitant les contraintes externes pour permettre une exploration complète du modèle, et pour développer une approche facilement applicable au système d'agriculture familiale tropicale.

1 Bilan général de l'application des modules

Dans cette section nous rappelons d'abord les contraintes à la fois liées aux données de télédétection et à la caractérisation spatiale de l'agriculture familiale tropicale ainsi que l'idée générale développée dans cette thèse. Puis, nous expliquons pourquoi les modules temporels n'ont pas été appliqués à notre zone d'étude. Enfin, nous résumons les principaux résultats de l'application des modules spatiaux (module 2 : règles spatiales et module 4 : modélisation de la répartition spatiale des classes).

1.1 Rappel des limites existantes de la caractérisation des espaces agricoles complexes et des concepts principaux de l'approche développée

Dans le chapitre 1, nous avons résumé les difficultés rencontrées dans l'utilisation de données de télédétection pour la caractérisation de l'occupation des sols dans des espaces d'agriculture familiale tropicale. Les principales limites peuvent être différenciées entre celles liées aux données de télédétection et celles liées à l'objet d'étude. Les contraintes inhérentes à la télédétection empêchent l'application de certaines méthodes en zone tropicale et peuvent être résumées en trois points majeurs :

- la nébulosité qui limite l'acquisition d'images optiques ;
- la forte hétérogénéité de l'environnement qui, selon la résolution spatiale des images, augmente la part de pixels mixtes ou limite l'application de certaines méthodes comme la segmentation ;
- et la synchronisation des phénologies des agrosystèmes et des écosystèmes naturels sur la saison des pluies qui engendre des signaux temporels proches dans les séries temporelles d'images satellites.

L'étude de l'agriculture familiale ne fait qu'ajouter des obstacles à l'utilisation des données de télédétection. Ces limites sont :

- la petite taille des parcelles, qui est très souvent inférieure à la taille d'un pixel d'une image à haute résolution spatiale ;
- la variabilité des dynamiques spatio-temporelles pour une même culture ;
- et la diversité des pratiques culturales.

Certaines des limites liées aux données de télédétection peuvent être réduites à l'aide d'une meilleure résolution spatiale et temporelle des images. Ce type d'offre satellitaire (haute résolution spatiale et temporelle) existe maintenant, depuis le lancement de la constellation Sentinel-2 par l'ESA. Cependant, même si de nouvelles méthodes sont développées pour traiter ce type de données, les contraintes ne sont pas totalement dépassées lorsqu'il s'agit de caractériser des systèmes complexes (Lambert et al., 2018; Dupuy et al., 2020a). Les caractéristiques des systèmes

d'agriculture familiale sont quant à elles inaltérables, mais à défaut d'être caractérisées spatialement à l'aide de données de télédétection, d'autres sciences cherchent à comprendre leur fonctionnement et leurs dynamiques (Bélières et al., 2013; Alpízar et al., 2020; Funk et al., 2020).

Face à ces constats, nous avons choisi de développer une approche qui s'appuie sur la modélisation des dynamiques spatiales via l'utilisation de connaissances issues des sciences socio-économiques, de l'agronomie et de la géographie, et visant à comprendre et caractériser le fonctionnement de l'agriculture familiale pour pallier les limites de la télédétection. De manière succincte, la méthode développée vise à utiliser à la fois les méthodes issues du traitement des données de télédétection pour obtenir des informations d'observation de la Terre et les méthodes de modélisation spatio-temporelles pour caractériser les possibilités d'occupation et d'usage des sols selon les stratégies des agriculteurs. Cette thèse s'est efforcée de développer à la fois une méthodologie nouvelle et encore peu abordée dans l'utilisation conjointe de ces deux sciences pour la caractérisation de l'occupation et de l'usage des sols.

Le modèle conceptuel développé dans le chapitre 2 se base sur le traitement des données de télédétection, puis sur l'application de quatre modules, deux modules spatiaux et deux modules temporels. La conception de ces modules s'est basée à la fois sur des connaissances collectées sur le terrain (entretiens et enquêtes) et sur de la littérature technique et scientifique. L'objectif du modèle conceptuel est de caractériser l'occupation et l'usage des sols à partir de l'approche OBIA qui consiste à regrouper des pixels cohérents pour créer des segments représentant la structure spatiale de la zone d'étude, puis de les caractériser (en leur affectant une classe d'occupation des sols). Le segment, comme objet élémentaire, et le VPAC (vecteur de probabilité d'appartenance aux classes), utilisé pour synthétiser les informations issues de la télédétection ($VPAC_{RF}$) et des connaissances ($VPAC_{RS}$), sont les concepts centraux de notre approche. Le traitement des données de télédétection se base sur les segments en leur associant un ensemble de variables spectrales (réflectance spectrale, indices spectraux, indices de texture, pente et altitude). À l'aide d'une classification supervisée (algorithme de forêts aléatoires) les résultats du traitement de ces données sont synthétisés dans un $VPAC_{RF}$. La possibilité de trouver un type d'occupation ou d'usage des sols sur un segment se base sur des connaissances qui doivent être identifiées, puis formalisées en règles expertes. L'application des modules visant à améliorer la caractérisation de l'occupation des sols est basée sur des règles expertes qui traduisent des zones d'implantation favorables aux différentes cultures ou des répartitions spatiales propres (modules spatiaux), des calendriers culturaux ou dynamiques temporelles particulières (modules temporels). L'implémentation de ces modules nécessite certaines données de support à l'application des connaissances. Ces données support enrichissent le segment de nouvelles propriétés qui permettent de modéliser la possibilité d'observer un type d'occupation des sols pour chaque segment. Les quatre modules ont soit un objectif spatial, soit un objectif temporel. Ces modules sont :

- le module 1 : **modélisation du fonctionnement temporel** simule la probabilité d'observer une classe, selon l'antécédence du segment. Ce module se base sur les connaissances des rotations culturales et sur différentes classifications temporelles (pluriannuelles) où chaque segment est caractérisé par un VPAC temporel.

- le module 2 : **règles spatiales** a comme objectif de produire un $VPAC_{RS}$ pour chaque segment qui symbolise la possibilité d'observer une classe en fonction de l'emplacement du segment dans l'environnement (biophysique, socio-économique) ;
- le module 3 : **règles temporelles** utilise à la fois des données de phénologie, de précipitation et les calendriers culturaux pour différencier les types de culture entre eux et favoriser une meilleure discrimination avec la végétation naturelle ;
- et le module 4 : **modélisation de la répartition spatiale des classes** est basé sur une méthode de désagrégation de statistiques pour réallouer une classe à un segment en fonction de son VPAC. Ce module se base sur des données de proportion théorique d'occupation des sols dans une unité administrative ;

1.2 Une application du modèle conceptuel sans temporalité

Cette sous-section présente les deux raisons pour lesquelles les modules temporels n'ont pas été appliqués à notre zone d'étude.

Parmi les quatre modules que compte notre modèle conceptuel, seuls les deux modules spatiaux (2 et 4) ont été appliqués à notre zone d'étude. Les raisons pour lesquelles l'application des modules temporels n'a pas été testée sont à la fois liées au contexte de l'agriculture familiale et à la disponibilité des données soutenant l'application de ces modules temporels . Ce travail de thèse s'appuie sur des hypothèses dont certaines ont été explorées avec les modules spatiaux. Les tentatives d'application des modules temporels ont échoué pour deux principales raisons : la première contraint l'utilisation du module 3 (règles temporelles), et la seconde l'utilisation du module 1 (modélisation du fonctionnement temporel).

Concernant le module 3, certaines de nos hypothèses de départ sur les stratégies des agriculteurs dans notre zone d'étude se sont révélées fausses. Dans la région du Vakinankaratra, aucune stratégie relative à la gestion d'un champ vis-à-vis du démarrage la saison des pluies n'a pu être identifiée, ni sur le terrain ni à travers les recherches bibliographiques. Cela écarte donc l'intégration de règles temporelles liées aux précipitations dans le module 3. Une fois cette hypothèse écartée, le traitement des données de télédétection pour essayer de différencier des signaux temporels restait l'une des solutions envisageables pour ce module (par classification ou comparaison du signal temporel d'une ou plusieurs bandes spectrales ou indices d'un segment). Nous avons choisi de ne pas effectuer ces traitements, car ils ne mobilisent ni concepts, ni méthodes, ni données liées à la modélisation spatio-temporelle. Si la modélisation spatio-temporelle ne peut être utilisée, l'apport de ce module se résume à un traitement des données de télédétection, donc soumis aux contraintes de ces données dans un contexte d'agriculture familiale[1]. Ces difficultés ont déjà été soulignées et restent un enjeu pour la caractérisation de l'agriculture familiale.

[1]Dans notre cas, le signal temporel des cultures pluviales ne permet pas de différencier visuellement le type de culture et suit une trajectoire similaire au signal temporel de la savane

Bellón de la Cruz (2018) a mis en avant les difficultés d'application d'une méthode de caractérisation de système agricole à une échelle régionale à partir de classification binaire de signal temporel pour l'agriculture familiale du Burkina Faso. Vintrou (2012) a utilisé des méthodes similaires pour suivre la phénologie des cultures à l'aide d'images MODIS au Mali et a montré des incohérences dans la détection du démarrage de la saison des pluies lorsque des données sont manquantes. Pour Vintrou (2012), il est nécessaire d'utiliser des séries temporelles régulières et sur plusieurs années pour obtenir un résultat cohérent et une bonne détection des retards dans le démarrage de la saison des pluies.

Concernant le module 1 (modélisation du fonctionnement temporel), la raison pour laquelle il n'a pu être appliqué est l'échelle temporelle de l'application du modèle aux Hautes Terres qui est centrée sur la saison culturale principale. Bien que les rotations culturales et les stratégies des agricultures sur plusieurs saisons culturales et sur plusieurs années soient importantes pour comprendre l'occupation ou l'usage des sols (Annexe E et Annexe F, Jouve (2006)), la disponibilité des données (*i.e.* cartes d'occupation des sols sur les saisons précédentes) et l'allongement de la période à étudier sont des facteurs limitants. D'un point de vue théorique, l'application de ce module semble prometteuse, mais nécessite une classification de l'occupation et de l'usage des sols sur plusieurs années. Des recherches sont en cours sur ce sujet (Osman et al., 2015; Bailly et al., 2018), mais l'intégration de connaissances pour modéliser les cultures possibles n'est pas encore abordée dans la littérature. Une prise en compte d'une temporalité cohérente (successions culturales, stratégies temporelles pour gestion de l'exploitation) avec l'objet d'étude dans le modèle semble essentielle pour correctement modéliser les choix et stratégies des agriculteurs. L'étude des rotations culturales permettrait une utilisation conjointe de la télédétection et de la modélisation des dynamiques spatiales à la fois pour prédire les cultures observables et pour corriger les rotations théoriques. Cette étude pluriannuelle pourrait aider à définir, ou à améliorer, les répartitions théoriques des classes, permettant l'application d'algorithmes d'espérance-maximisation (Dempster et al., 1977).

1.3 Résultats globaux de l'application des modules spatiaux

Dans cette sous-section nous récapitulons les apports des chapitres 3 et 4 dans le développement d'une **approche générique** (malgré un besoin en expertise important) et la **nécessité de raisonner sur la fiabilité d'un segment pour déterminer la manière de combiner des probabilités de natures différentes** : celles issues d'une observation (télédétection) et celles exprimant la présence possible d'une classe à un endroit donné (règles spatiales). Ensuite, les résultats analysés dans le chapitre 5 sont résumés pour permettre une meilleure identification des **apports de l'approche et de ses limites**. Les résultats sont analysés à travers le module des règles spatiales (module 2) et le module de modélisation de la répartition spatiale des classes (module 4).

Le chapitre 3 et le chapitre 4 explicitent l'application des deux modules spatiaux aux Hautes Terres du Vakinankaratra. L'implémentation des modules a été conçue pour permettre la repro-

duction de ce travail sur une autre zone d'étude ou pour approfondir le travail mené durant cette thèse en s'appuyant sur le principe de parcimonie (les règles ne sont pas redondantes et les plus simples possible) ; et sur la reproductibilité de l'approche développée (tous les paramètres sont définis de manière explicite et la sélection des règles et des données n'est pas spécifique à la zone d'étude). Par contre, la configuration des modules demande une expertise importante pour que les processus appliqués soient cohérents avec la zone d'étude. Certains choix et partis pris peuvent être discutés et changés sans que la structure de l'approche en soit grandement modifiée.

À partir de l'approche OBIA qui vise à caractériser des segments (ensemble spatial cohérent comme un champ, une forêt, un bâtiment) et les VPAC (vecteur de probabilité d'appartenance aux classes, qui servent à caractériser un segment), la méthode développée est basée sur deux concepts : la **fiabilité** des segments et la **complémentarité entre observations** d'une occupation ou d'un usage des sols ($VPAC_{RF}$, issu de la classification des données de télédétection) et la **possibilité** de présence d'une occupation ou d'un usage des sols ($VPAC_{RS}$, issu de l'application des règles spatiales). La fiabilité d'un segment, définie à partir d'un VPAC, est un élément central pour la sélection des segments nécessitant un changement de classe et pour définir la classe à affecter à ces segments. La complémentarité entre observation et possibilité se base sur deux méthodes de combinaison du $VPAC_{RF}$ et du $VPAC_{RS}$ d'un segment pour produire un $VPAC_{Fusion}$. La première méthode est une **moyenne pondérée** qui fusionne les deux VPAC en donnant une priorité à l'observation ou à la possibilité. La classe qui obtient la probabilité la plus élevée est utilisée pour labelliser le segment. La seconde méthode sert à **sélectionner la classe qui a la plus forte probabilité** entre les deux VPAC.

Le chapitre 5 présente l'analyse des résultats de l'application des deux modules spatiaux pour évaluer si la caractérisation de la zone d'étude est cohérente d'un point de vue agricole. Cette analyse de la cohérence thématique des cartes produites vise à estimer si les confusions de classes présentes dans la cartographie obtenue à l'issue de l'application du module des règles spatiales sont plus acceptables que celles résultant de la classification des données de télédétection. Cette analyse est effectuée à la fois de manière quantitative et qualitative à partir de plusieurs configurations du module des règles spatiales (module 2). Bien que ces résultats quantitatifs (métriques d'évaluation) montrent que l'objectif n°3 n'a pu être que partiellement atteint, de nombreuses pistes d'amélioration et de recherche peuvent être identifiées.

L'application du **module des règles spatiales (module 2)** a montré une amélioration quantitative des résultats uniquement avec une combinaison de paramètres qui limitent le nombre de segments pouvant changer de classe. Parmi les deux méthodes proposées, seule la méthode de la moyenne pondérée a montré des résultats significatifs par rapport à la classification des données de télédétection. Quantitativement, les résultats de l'application de ce module présentent des changements marginaux, mais encourageants. Les résultats des classes ayant déjà une F-mesure élevée sont améliorés, mais la progression reste faible pour les classes de cultures pluviales.

L'objectif d'améliorer la fiabilité de la caractérisation de l'occupation des sols reste une piste de recherche à développer, car une réponse à cet objectif n'a pas pu être clairement identifiée. L'une

des limites pour l'analyse de l'évolution de la fiabilité des segments est la différence entre les valeurs maximales possibles des $VPAC_{RS}$ (il est très rare qu'un segment ait des propriétés compatibles avec une unique classe d'occupation du sol, donc une valeur maximale de 1 est presque impossible à atteindre) par rapport à celles des $VPAC_{RF}$ (dans la classification des données de télédétection, il est possible qu'une unique classe soit détectée et que toutes les autres possibilités soient écartées, la valeur maximale peut donc être de 1 ou proche de 1). Cette asymétrie entre les valeurs des deux VPAC est compréhensible d'un point de vue thématique, mais est techniquement limitante.
Le but d'améliorer la fiabilité des segments n'a pas été atteint avec la méthode de la moyenne pondérée, car cette méthode entraîne nécessairement une diminution globale des valeurs du VPAC par le calcul de la moyenne. Et, bien que la méthode de la probabilité maximale n'ait pas permis d'obtenir des résultats significatifs, cette méthode est intéressante pour l'amélioration de la fiabilité des segments.
L'application du module des règles spatiales avec une configuration qui permet d'améliorer les résultats de classification est finalement limitée dans notre cas d'étude, car les **changements sont peu importants pour tous les indicateurs analysés** (métriques quantitatives, fiabilités, distribution des probabilités, changements d'aires). Par contre, sur le plan qualitatif, l'analyse des confusions entre les classes présente des résultats plus cohérents lors de l'application de connaissances représentant les stratégies des agriculteurs. Les confusions entre le domaine agricole et le domaine non agricole sont moins importantes.

D'autres configurations du module des règles spatiales ne permettent pas un perfectionnement du résultat global, mais une **amélioration pour certaines classes de culture pluviale** et une **meilleure localisation du domaine pluvial**, rendant la carte produite plus **cohérente** avec les stratégies des agriculteurs. Ces configurations, qui permettent de changer la classe d'un plus grand nombre de segments, sont intéressantes, car elles aident à mieux comprendre le fonctionnement du module, ses avantages, ses limites et ont présenté des résultats inattendus. Par exemple, le **déséquilibre des règles expertes** est particulièrement visible à travers des classes qui dominent des groupes de culture (*pomme de terre* et *maïs associé*). Cette analyse permet d'affiner la réflexion autour de la formalisation des connaissances en règles expertes, surtout pour les classes de cultures qui ont des conditions de mise en culture similaires. On note avec cette configuration une amélioration de la cohérence spatiale vis-à-vis de l'image THRS Pléiades, ainsi qu'une meilleure précision des classes de culture pluviale qui permet de **mieux discriminer les champs dans les zones dominées par la savane**. De manière qualitative, la localisation des champs identifiés est correcte et en adéquation avec les connaissances. L'analyse de la cohérence thématique ouvre sur des réflexions plus ciblées autour de l'attribution des classes entre les champs de culture pluviale et les zones de savane.

Le **module de modélisation de la répartition spatiale des classes (module 4)** a été analysé uniquement de manière qualitative, car la base de données de vérité terrain n'est pas représentative de l'ensemble des Fokontany modélisés. Là où les données de répartition théorique des cultures sont disponibles, le résultat est spatialement cohérent, excepté pour certaines classes dont la sur-

face est surestimée. L'**identification des cultures pluviales dans les zones de savane** est aussi intéressant et montre une amélioration par rapport à la classification des données de télédétection qui est particulièrement mauvaise pour cette tâche. La méthode de répartition des cultures de ce module permet d'obtenir un ensemble de types de cultures disparate qui est cohérent avec les stratégies des agricultures qui pratiquent la polyculture, mais la localisation exacte des cultures ne peut être vérifiée. Le principal défaut de ce module est sa **dépendance à des données de répartition théorique précises**. L'algorithme doit être amélioré afin de mieux prendre en compte les incertitudes de répartition. Mais, même lorsque les répartitions sont estimées de manière approximative par les responsables locaux, les résultats restent qualitativement cohérents.
L'application du module des règles spatiales montre que la base de données de vérité terrain est limitante pour l'évaluation des modules. Le déséquilibre des classes et leur répartition spatiale inégale et problématique.

L'analyse de l'application des modules spatiaux (module 2 : règles spatiales et module 4 : modélisation de la répartition spatiale des classes) par les résultats obtenus souligne les limites rappelées au début de ce chapitre (données de télédétection et agriculture familiale tropicale), mais montre aussi des résultats encourageants pour une meilleure discrimination des cultures pluviales en zone de savane. Cette identification permet de lever certaines limites des données de télédétection à l'aide de connaissances formalisées dans un contexte d'agriculture familiale tropicale. Ces résultats incitent à développer l'esquisse d'approche hiérarchique, méthode utilisée pour répartir les classes de l'ensemble « Autre ». Ces résultats sont encourageants, car la cartographie, que ce soit du type de culture ou du domaine cultivé, est difficile à produire pour les systèmes familiaux complexes qui sont majoritaires sur continent africain (Xiong et al., 2017), et ce quelle que soit l'échelle, régionale (Vintrou et al., 2012a; Leroux et al., 2014) ou locale (Bontemps et al., 2015; Lebourgeois et al., 2017; Dupuy et al., 2020a). Des méthodes continuent à être développées dans ce sens, que ce soit pour la cartographie des cultures pluviales et irriguées à partir de séries temporelles pluriannuelles de LandSat-8 (Landmann et al., 2019), ou en estimant la proportion de cultures dans des fenêtres de 10 x 10 m à partir d'image WorldView (Neigh et al., 2018). Parmi les tentatives de cartographier les types de culture dans le domaine cultivé avec une nomenclature plus ou moins détaillée, peu de recherches sont spécifiquement axées sur l'agriculture familiale (Lebourgeois et al., 2017; Lambert et al., 2018; Dupuy et al., 2020a), alors que la question de la cartographie des systèmes de culture reste un enjeu majeur (Fritz et al., 2013; Begue et al., 2016) avec une difficulté particulièrement importante pour discriminer les cultures pluviales des espaces de savane herbeuse (Sweeney et al., 2015; Waldner et al., 2017). Dans ce contexte, les résultats obtenus pour améliorer les résultats de classification du domaine cultivé et surtout pour identifier correctement les cultures pluviales à une échelle fine (0,5 m) représentent une piste de recherche prometteuse.

Dans cette section nous avons dressé un bilan général de l'application des modules. Tout d'abord, nous avons rappelé les limites existantes de la caractérisation de l'agriculture familiale tropicale à partir de données de télédétection optique, puis nous avons retracé le fonctionnement du modèle con-

ceptuel à travers les concepts mobilisés (segments, connaissances formalisées, VPAC) et les quatre modules ayant des dimensions spatiales ou temporelles. Ensuite, nous avons expliqué pourquoi les modules temporels n'avaient pas été appliqués à notre zone d'étude, à la fois par le manque de stratégies temporelles identifiées et par l'échelle temporelle d'application courte du module, centrée sur la saison culturale principale (saison des pluies). Pour finir, nous avons présenté les résultats généraux des deux modules spatiaux appliqués, le module 2 : règles spatiales et le module 4 : modélisation de la répartition spatiale des classes. Ces modules montrent des changements limités dans l'amélioration générale de la classification, mais exposent des résultats intéressants pour la caractérisation des cultures pluviales et surtout pour l'identification du domaine pluvial dans les zones de savane.

2 Application : pistes et limites

La deuxième section de ce chapitre porte sur les pistes d'amélioration issues des constats de l'application du modèle conceptuel aux Hautes Terres malgaches. Tout d'abord, malgré une approche éprouvée et validée du traitement des données de télédétection, nous présentons des pistes pour mieux utiliser l'information satellitaire, particulièrement pour l'image THRS Pléiades, et pour identifier d'autres méthodes pouvant être mobilisées. Dans une deuxième partie, l'aspect multi-scalaire du modèle est abordé. Les échelles du champ (segment) et du village (Fokontany) sont correctement représentées, mais l'échelle de l'exploitation, bien qu'indirectement décrite, peut être considérée comme lacunaire. La troisième sous-section analyse l'impact de la qualité des données et des connaissances formalisées sur les résultats. Nous aborderons la forte demande en expertise et en données du modèle, la qualité de la segmentation qui influe sur les objets à caractériser, la base de données de vérité terrain qui limite certaines analyses, mais qui pourrait être utilisée pour mieux comprendre les règles qui sont validées par le modèle, et enfin les données support pour l'application des connaissances qui peuvent présenter de nouvelles contraintes à surmonter. La quatrième section est consacrée à nos choix effectués pour formaliser les connaissances en règles expertes et pour déterminer l'importance de celles-ci dans les étapes de modélisation. L'utilisation des VPAC (vecteur de probabilité d'appartenance aux classes) est aussi discutée pour identifier les pistes d'amélioration d'agrégation et de comparaison de ceux-ci. La dernière section est consacrée à la validation du modèle et à son exploration.

2.1 Une meilleure exploitation des images de télédétection

Bien que n'étant pas au coeur de notre approche, il est intéressant de présenter des pistes d'amélio-ration des méthodes de télédétection que nous avons appliquées pour déterminer les $VPAC_{RF}$. Le traitement des données de télédétection se base sur des méthodes déjà éprouvées, mais celles-ci pourraient être adaptées avec l'utilisation de méthodes plus récentes qui montrent des résultats intéressants tel que l'apprentissage profond (Benedetti et al., 2018; Interdonato et al., 2019), l'utilisation de la classification hiérarchique (Jiao et al., 2019), ou l'apprentissage à partir de jeux de données réduits (Yao et al., 2020). L'utilisation de l'approche de *dynamic time wraping*

(Sakoe and Chiba, 1978) pour les images Sentinel-2 par Belgiu and Csillik (2018) peut aussi être intéressante. Cette méthode mobilise à la fois des profils temporels de NDVI depuis les données de vérité terrain pour prendre en compte la dynamique de la végétation dans l'étape de classification.

L'image THRS Pléiades, utilisée uniquement pour la segmentation et le calcul d'indices de texture (chapitre 4, section 1.1.1.2 Segmentation et 1.1.1.3 Textures) pourrait être mieux exploitée afin de pallier certaines limites de la télédétection en zone tropicale dans un contexte d'agriculture familiale telle qu'une classification de cette image (Debats et al., 2016; Lelong et al., 2020).

2.2 L'aspect multiscalaire du modèle

Dans cette sous-section, nous présentons d'abord l'échelle du segment, qui sert à caractériser l'occupation ou l'usage des sols, puis l'échelle du Fokontany. L'échelle de l'exploitation sera abordée comme chaînon manquant des étapes de modélisation, laissant de côté certains processus importants tels que les décisions stratégiques finales des agriculteurs.

L'aspect multiscalaire de l'approche n'est mobilisé qu'avec le module 4 : modélisation de la répartition spatiale des classes. Le module 2 : règles spatiales est appliqué uniquement à l'échelle du segment. Dans un espace agricole, un segment est la représentation spatiale et générique d'un champ et hors de l'espace agricole il peut représenter différentes structures de végétation naturelle, d'espaces batis ou autres occupations des sols plus ou moins cohérentes. L'échelle du champ et l'échelle du Fokontany sont modélisées de manière explicite. L'échelle de l'exploitation, pourtant centrale, n'est décrite que de manière implicite et n'est pas directement modélisée dans le chapitre 4. L'analyse de ces trois échelles est importante pour comprendre les fonctionnements de l'application des modules ainsi que leurs manquements.

L'échelle du segment semble correctement modélisée. Basée sur l'approche OBIA, cette échelle est pertinente pour résumer l'information satellitaire et contextuelle des objets. L'un des manquements de la conceptualisation et de l'implémentation du modèle des règles spatiales (2) est la relation entre des segments voisins ou proches. La modélisation de la proximité entre plusieurs segments permettrait d'estimer une probabilité d'appartenance à une classe qui appartient à un groupe plus général (culture pluviale ou irriguée, espace non agricole) via sa proximité avec d'autres segments. Les enquêtes auprès des responsables locaux nous ont indiqué qu'il y a une reproduction des pratiques culturales entre agriculteurs proches dont certaines consistent à cultiver les mêmes cultures (Annexe F).

L'échelle du Fokontany utilisée dans le module4 : modélisation de la répartition spatiale des classes est l'une des supra-échelles évoquées dans le modèle conceptuel. Cette échelle est adaptée à notre zone d'étude qui correspond à une surface restreinte où uniquement certaines communes sont représentées dans leur intégralité. L'évaluation qualitative de la modélisation à cette échelle a montré des résultats encourageants malgré la qualité médiocre des données de répartition théorique de l'occupation des sols. L'utilisation des VPAC pour répartir les classes permet un bon compromis entre l'observation d'une classe et sa possibilité à la fois grâce aux règles de

répartition et par la fiabilité des segments qui permet une attribution spatiale où l'aléatoire est fortement limité. L'un des éléments importants à adapter pour une modélisation plus pertinente de la répartition des classes est une meilleure prise en compte de l'incertitude des probabilités théoriques (You et al., 2014). Selon la taille de la zone d'étude et la disponibilité des données, ce module peut être réutilisé à différentes échelles.

L'échelle de l'exploitation n'est pas directement modélisée et peut être considérée comme manquante, car une partie des pratiques et choix culturaux se décident au sein de l'exploitation. Les règles définies ne représentent jamais les choix des agriculteurs, mais uniquement un résumé de leurs choix possibles. Indirectement, le choix des agriculteurs dans leur exploitation est pensé dans le module 2: règles spatiales, mais uniquement au niveau du segment. La modélisation de cette échelle pourrait permettre une répartition des cultures à une échelle située entre celle des champs et celle du village en fonction des stratégies des agriculteurs (Jahel et al., 2018). Se baser sur des typologies d'exploitation peut être une solution pour modéliser les choix des agriculteurs. La modélisation d'exploitations permettrait de dépasser la dimension binaire entre observation et possibilité pour intégrer une modélisation des choix et décisions selon des comportements d'agriculteurs (Jahel, 2016; Holland et al., 2017). L'utilisation d'une méthode permettant l'identification de types d'exploitation à travers les espèces cultivées et des choix des agriculteurs tels que présenté par Chopin et al. (2015a) appliquée pour modéliser la répartition de l'occupation et de l'usage des sols à une échelle régionale (Chopin et al., 2015b) semble une piste intéressante à développer.

Une modélisation multiscalaire idéale serait la modélisation de l'attribution de l'occupation et de l'usage des sols en fonction 1) des paramètres biophysiques et socio-économiques tel que présenté dans le module des règles spatiales (module 2) et par l'utilisation des VPAC, 2) des décisions des agriculteurs selon les systèmes de culture présents, leur attitude face au risque (nouvelles cultures, investissements) et 3) d'une répartition théorique d'occupation et d'usage des sols à une petite échelle (village, commune, district). Cette modélisation se baserait sur les recherches développées par Jahel et al. (2018) et Chopin et al. (2015b), ainsi que les recherches présentées dans cette thèse. L'entrée principale du modèle serait l'utilisation de la télédétection pour une spatialisation des champs, puis l'application des modules pour produire des cartes plus cohérentes avec la réalité.

2.3 L'impact de la qualité des données sur les règles et les résultats

Le choix d'utiliser une approche basée sur des connaissances formalisées en règles expertes et sur l'utilisation de données de télédétection rend la méthode développée particulièrement dépendante aux données et à leur qualité. La problématique de la « qualité » des données de télédétection a été largement présenté dans le chapitre 1 et rappelée au début de ce chapitre. Par contre, la qualité des données utilisées pour appliquer les connaissances à notre zone d'étude n'a été discutée que ponctuellement. Cette section présente d'abord l'impact de la qualité de la segmentation, qui correspond à la création des segments qui sont caractérisés tout au long des étapes du modèle (télédétection, modules). Ensuite, est abordée de manière générale la qualité des données support

pour l'application des connaissances. Pour finir cette section, les avantages et limites de la base de données de vérité terrain sont analysés en partant du constat qu'elle ne peut être que difficilement améliorée. Néanmoins, cette base de données peut présenter des pistes d'exploration pour mieux comprendre notre objet d'étude.

Dans le chapitre 2, la possibilité d'utiliser tout type de données représentant la structure spatiale de la zone (*i.e.* parcellaires ou cadastres existants) à caractériser était évoquée. Dans un contexte théorique cette idée est pertinente, mais pour une application sur les espaces tropicaux d'agriculture familiale, il est nécessaire de produire une segmentation de la zone d'étude (Lebourgeois et al., 2017; McCarty et al., 2017). Une partie importante de la démarche développée repose sur cette segmentation. Par conséquent, l'obtention qu'une segmentation de qualité est nécessaire, surtout pour l'espace agricole. Or, lors du processus de segmentation de la zone d'étude puis de son évaluation, les résultats ont montré les limites de la segmentation qui n'est pas complètement représentative des données collectées sur le terrain (chapitre 4, section 1.1.1.2 Segmentation, tableau 5). Une plus grande attention doit être portée à ce processus et à son évaluation qui doit être plus complète pour que les segments à caractériser correspondent mieux à la réalité du terrain. L'évolution des méthodes de segmentation des images THRS montre des résultats encourageants qui peuvent être appliqués à notre cas d'étude où la qualité de la segmentation structure notre approche (Hadavand et al., 2017; Hossain and Chen, 2019; Derksen et al., 2020). Une meilleure segmentation peut aussi permettre l'utilisation de variables de forme telles que les contours des segments (géométrie, sinuosité) pour améliorer la cartographie de l'usage des sols (Goodin et al., 2015).

Les données support pour l'application des connaissances sont au cœur de la mise en place du modèle conceptuel : de ces données dépendent la bonne application des connaissances et donc la possibilité d'améliorer la caractérisation et la fiabilité de l'occupation et de l'usage des sols. Le cas d'étude des Hautes Terres malgaches montre les enjeux autour de ces données : il existe peu d'études récentes visant à collecter des informations à grande échelle, les lacunes temporelles dans les données sont importantes, la spatialisation des données est inégale à l'échelle du pays, des régions et des communes et les données disponibles sont de faible qualité[2].

Une fois qu'une connaissance est formalisée, son application dépend de la qualité des données sur lesquelles elle s'appuie. Si cette donnée est de trop mauvaise qualité, trop spécifique à une zone ou inexistante, la connaissance formalisée (*i.e.* la règle) ne peut être mise en place. L'une des solutions pour pallier ce problème est la production de données lorsque cela est possible (tel que le temps de trajet entre un segment et un bâtiment (Annexe H), ou la description de la topographie (Annexe G)), mais ce travail demande un processus supplémentaire de validation et d'estimation de l'incertitude de ces données (Irmischer and Clarke, 2018; Muddarisna et al., 2020). Dans le cadre de cette thèse, l'estimation des erreur n'a été effectuée que de manière qualitative et basées

[2]Ces constats sont issus de la recherche de données pour supporter l'application des connaissances présentée dans le chapitre 3, section 2 Données. Un aperçu des données libres et disponibles à l'échelle de Madagascar ou à l'échelle infranationale est visible sur le site d'OpenStreetMap : carte d'Antsirabe.

sur l'expertise. Lorsque les données support n'existent pas, les règles ne peuvent être appliquées alors qu'elles peuvent correspondre à des dynamiques ou stratégies qui ont un impact majeur sur les choix des agriculteurs. Les points de collectes de production agricole, l'accessibilité de certaines zones à partir des routes, et donc l'accès à la vente ou aux marchés sont des exemples de connaissances qui n'ont pu être appliquées (voir les annexes E, F et C. Malgré le fait que nous avons pu identifier les liens entre les marchés et la spécialisation vers certains types de culture dans différents Fokontany, la modélisation de l'accès aux marchés n'a pas été possible faute de données support.

La mauvaise qualité des données support pour l'application des règles spatiales peut être analysée avec les cartes de possibilités de culture (issues des $VPAC_{RS}$, chapitre 5, section 1 Résultats de la classification des données de télédétection) et en les confrontant avec la classification des données de télédétection à partir de seuils de probabilité. La mauvaise couverture spatiale de certaines données est aussi problématique, avec des connaissances qui ne sont pas appliquées aux zones moins accessibles ou avec moins de ressources pour produire ces données. L'application du module 4 : modélisation de la répartition spatiale des classes à partir de données théoriques erronées permet uniquement de donner un aperçu des résultats possibles sans pouvoir considérer que les données et connaissances utilisées sont représentatives de la réalité.

Cette approche basée sur l'expertise et les données peut, dans certains contextes, compromettre l'application du modèle et entraîne donc le risque de ne pas améliorer la caractérisation de l'occupation et de l'usage des sols. Les solutions pour surmonter ces limites sont peu nombreuses, car la production de données support de qualité est un processus coûteux. Nous pouvons tout de même espérer qu'à long terme la disponibilité des données libres et de qualité devrait augmenter, par exemple avec la cartographie participative telle qu'OpenStreetMap[3]. Une utilisation du modèle à petite échelle permettrait aussi l'utilisation de données telle que le « Spatial Production Allocation Model »[4] (You et al., 2014).

Le chapitre 3 a mis en avant les lacunes de la base de données de vérité terrain (déséquilibre entre les classes, répartition spatiale des données inégales), mais l'amélioration (en termes de quantité de relevés et de répartition entre classes et dans l'espace) de cette base de données semble difficile. Cette base de données est le fruit de plusieurs semaines de travail de terrain et dépend des contraintes structurelles de l'agriculture familiale tropicale sur notre zone d'étude (difficultés d'accès, routes coupées en saison des pluies, zones d'insécurité). L'inégalité du nombre d'échantillons entre les classes et au sein de la zone d'étude pose un problème de représentativité qui biaise à la fois l'apprentissage en télédétection et limite la validation de la méthode proposée. Cette partie présente succinctement les réflexions autour de la base de données de vérité terrain à travers l'échantillonnage des données sur le terrain, et le déséquilibre entre les classes.

La méthodologie d'échantillonnage des données de vérité terrain en bord de route et sur des espaces identifiés à l'avance pour une diversité des types d'occupation et d'usage des sols et de

[3] https://www.openstreetmap.org

[4] Le modèle SPAM (https://www.mapspam.info/) estime des données de répartition des cultures, allant de l'échelle nationale à une désagrégation à une grille de 10 x 10 km.

leur répartition est validée par l'étude de Waldner et al. (2019a). Waldner et al. (2019a) montrent que la collecte des données en bord de route diminue la précision des classifications, mais reste la meilleure stratégie d'échantillonnage des données avec une identification en amont des zones à échantillonner. Ce principe a été utilisé pour la collecte des données de terrain dans notre étude. Malgré ces méthodes, une partie des types d'occupation des sols ne peut être relevés car non exploitables avec des images à haute résolution. D'autres méthodes existent dans la littérature, mais demandent généralement plus de moyens techniques et ne permettent pas forcément de dépasser les contraintes des systèmes complexes (à l'aide de vidéos (Wu and Li, 2012) ou à l'aide de drone (Durand, 2016)).

Le déséquilibre entre les classes peut quant à lui être responsable d'une part importante de la variabilité des résultats de classification (Waldner et al., 2019b). Pour contrer cette asymétrie, les méthodes utilisées ont nécessairement un impact, soit sur les cultures majoritaires qui sont alors moins bien représentées, soit sur les cultures minoritaires Douzas et al. (2019). Pour pallier ce déséquilibre, une estimation du nombre de classes optimal par rapport à la distribution du jeu de données, puis un choix du nombre de classes en fonction des objectifs de l'étude peut être une solution. Le degré de déséquilibre entre les classes peut être un critère de choix de regroupement ou de séparation (Ortigosa-Hernández et al., 2017). Des méthodes de sur-échantillonnages des classes minoritaires peuvent aussi être utilisées (Douzas et al., 2019; Waldner et al., 2019b).

La base de données de vérité terrain est dépendante de limites qui sont donc difficiles à dépasser. Une quantification des manquements et des erreurs peut être une solution pour mieux évaluer et appréhender les enjeux de la zone étudiée, par exemple avec un index pour mesurer l'impact des espaces non échantillonnés (Wu et al., 2018).

Dans cette sous-section, nous avons analysé l'impact de la qualité des données utilisées, qu'elles soient dérivées d'autres données (segmentation, temps de trajet, topographie), utilisées pour l'application des connaissances ou pour la classification des données de télédétection et la validation des résultats. Certaines limites soulevées semblent difficiles à lever, car soient dépendantes de facteurs structuraux (agriculture familiale), ou simplement par manque de données existantes. Des solutions ont été proposées pour analyser autrement les données de la base de données de vérité terrain pour produire une segmentation plus fiable.

2.4 Impact des choix pour la formalisation des connaissances et pour l'utilisation des VPAC

Dans cette sous-section, les questionnements liés à l'aspect multiscalaire du modèle et à la qualité des données ont été abordés dans les deux sous-sections précédentes. Cette sous-section porte sur les impacts des hypothèses et des choix que nous avons formulés principalement durant le chapitre 4. La sous-section précédente abordait le concept du segment à caractériser, cette sous-section traite de la formalisation des connaissances et de l'utilisation des VPAC (vecteur de probabilité d'appartenance aux classes) pour caractériser les segments. Tout d'abord, le choix des

connaissances va être discuté, puis les choix effectués pour la formalisation des connaissances et la validation des connaissances mobilisées sont analysés. Enfin, la méthode d'agrégation des VPAC est mise en perspective avec d'autres méthodes pouvant être mobilisées.

La décision de choisir les connaissances agronomiques et socio-économiques à utiliser en se basant sur des articles scientifiques et des rapports techniques est une bonne base pour ce travail, mais peut être améliorée en utilisant des données déjà structurées dans cet objectif (bases de données d'enquêtes existantes), comme l'ont fait Vintrou et al. (2012a) ou Jahel (2016). Dans notre zone d'étude peu d'études compilent ces informations, et lorsque c'est le cas, les documents sont destinés à des diagnostiques territoriaux généraux Sourisseau et al. (2016) ou spécifiques à une thématique (Autfray, 2020; Bélières, 2020a,b; Ben Naâmane and Autfray, 2020). Ces documents sont très pointus et riches en informations, mais demandent aussi un haut niveau d'expertise thématique pour extraire les informations pertinentes à partir de ces ensembles. Des bases de données d'enquêtes construites sur des thématiques proches et sur des régions similaires (Raharison et al., 2018) peuvent être explorées pour en extraire des règles (Vintrou et al., 2012a). Les programmes de recherche tels que Secure[5], bien que non directement portées sur les questions des stratégies des agriculteurs, étudient celles-ci. Ils peuvent être sources de données et de collaborations pour identifier, structurer et décider de l'importance des connaissances.

Une validation des choix des connaissances et de leur importance de manière structurée par des experts est aussi une solution pour obtenir une sélection et une pondération des connaissances (ou graduation en fonction de l'importance) à l'aide de méthodes telles que celle utilisée par Holland et al. (2017). L'un des points négatifs de ces méthodes est qu'elles demandent à un grand nombre d'experts pour obtenir une information fiable et homogène. Dans la littérature, l'utilisation de systèmes experts est souvent associée à des méthodes de logique floue (Negoita, 1985; Zhu et al., 2001; El Hajj et al., 2009; Tsangaratos et al., 2018). L'utilisation de la logique floue pour déterminer l'importance des règles entre elles et dans le modèle semble intéressante à étudier et à mettre en place pour avoir une méthodologie structurée autour d'une approche définie dans le but de combiner des données imprécises et variables.

Dans notre travail, les connaissances sélectionnées ont pour but l'identification des stratégies agricoles, mais elles sont généralement représentatives d'un ensemble de cultures plutôt que d'une culture spécifique. En effet, les stratégies des agriculteurs (Annexe F) et les besoins agronomiques des plantes (Autfray et al., 2014) peuvent concorder pour différentes cultures. Il est donc difficile d'identifier des règles uniquement applicables à un seul type de culture. Ce constat montre la limite de l'utilisation des règles par type de culture, qui pourrait être levé en utilisant l'échelle de l'exploitation qui permettrait de représenter les choix des agriculteurs parmi un ensemble de culture similaires (cultures pluviales, légumes et pomme de terre). L'analyse et la sélection des règles les plus pertinentes et spécifiques à un type de culture ou à un ensemble de cultures correspondant à des connaissances similaires peuvent être une solution (ex. pour les cultures pluviales ou pour les légumes).Les méthodes citées précédemment peuvent être de bonnes pistes d'analyse

[5]https://www.secure.mg/le-projet-secure

et d'affinage des règles (analyse quantitative d'enquêtes, exploration de base de données).

Des méthodes pour sélectionner des connaissances et estimer leur importance ont été identifiées afin de créer un cadre méthodologique solide à la méthode que nous proposons.

La validation des règles à travers la base de données de vérité terrain est aussi une piste intéressante, mais peut présenter des limites similaires à celles liées à la construction de cette base de données. Par exemple, les polygones de vérité terrain de notre base de données ont été largement prélevés en bordure ou proche des routes ou autres voies de circulation accessibles. Uniquement ces polygones et les règles associées pourront être évalués. L'analyse des autres connaissances identifiées et règles ne correspondant pas à ce critère seront donc biaisées. Cette évaluation peut toutefois être pertinente pour identifier la redondance des règles entre les types de culture. Pour cela, un indicateur comme le TF-IDF (Salton and Buckley, 1991) peut être utilisé pour identifier les règles par rapport à leurs récurrences dans la base de données de vérité terrain.

La méthode d'association des règles, c'est-à-dire le processus de création du $VPAC_{RS}$, peut être remise en cause vis-à-vis des résultats obtenus dans cette thèse. Les $VPAC_{RF}$ et $VPAC_{RS}$ sont difficilement comparables, car ils correspondent à deux concepts différents, l'observation d'une classe d'un côté, et la possibilité qu'une classe soit présente dans un segment de l'autre. La standardisation des VPAC entre 0 et 1 (chapitre 4, section 3.3.1 Standardisation des résultats et création des $VPAC_{RS}$) peut être remise en question pour essayer de rendre comparable les VPAC et pour les associer. La normalisation de chaque probabilité (i.e. classe d'un VPAC), par exemple avec une méthode centrée-réduite, peut être envisagée. Mais cette question reste en suspens, car les connaissances peuvent correspondre à des conditions idéales pour la mise en terre d'un type de culture qui n'est pas dans la zone d'étude. Une normalisation des données risque d'entraîner une mise au même niveau de l'observation ($VPAC_{RF}$) et de la possibilité ($VPAC_{RS}$). Les méthodes d'agrégation des VPAC proposées peuvent permettre de conserver un équilibre entre ces concepts. La méthode de la moyenne pondérée permettrait ainsi de définir l'importance de chacun des VPAC lors de l'agrégation. Dans le cas de l'utilisation d'une pondération de chaque VPAC, puis de l'application de la méthode de la probabilité maximale, cette méthode permettrait de mieux analyser l'évolution de la fiabilité des segments.

Les résultats obtenus par les tests effectuées dans le cas d'étude indiquent que la méthode éprouvée n'est pas satisfaisante. Repenser et formaliser l'approche de manière plus rigoureuse est nécessaire. Les méthodes de fusion de la logique floue (Zadeh, 1978) permettent d'utiliser des opérateurs d'agrégation et de décision variées qui peuvent être intéressants à explorer (Hajj et al., 2009). Il serait intéressant pour une meilleure compréhension du modèle et des liens entre les VPAC d'analyser différentes configurations de manière exhaustive, avec différents opérateurs de combinaison de VPAC et selon différents niveaux de représentation des VPAC (normalisés ou non, standardisés).

Dans cette sous-section, nous sommes revenus sur les impacts de nos choix méthodologiques. Des méthodes robustes, utilisées dans d'autres recherches ont été préconisées pour mieux identifier les

connaissances à utiliser, déterminer leur importance et les combiner. Ces solutions peuvent être issues des sciences sociales (exploration de bases de données d'enquêtes, panel d'experts) et être combinées avec des approches mathématiques telles que la logique floue. La combinaison pertinente des VPAC entre eux reste une question ouverte et une solution satisfaisante est encore à développer.

2.5 Validation du modèle

Pour conclure cette section portant sur l'analyse de l'application du modèle conceptuel aux Hautes Terres de Madagascar à travers ses limites et les pistes de recherches identifiées pour pallier celles-ci, cette sous-section présente une courte réflexion sur la validation du modèle et son exploration.

L'évaluation de l'application du modèle conceptuel est similaire à celle utilisée pour évaluer la classification des données de télédétection. Plusieurs configurations du modèle ont été analysées pour mieux comprendre son fonctionnement, ses limites et surtout comment le rendre plus cohérent et réaliste pour la production de cartes d'occupation et d'usage des sols. Pour aller plus loin dans cette observation du modèle, une plus grande calibration de ses paramètres ainsi qu'une analyse de sensibilité (Saltelli, 2008) peuvent être envisagées. Ces analyses permettent d'évaluer la pertinence des relations et processus exprimés dans le modèle, particulièrement avec le développement de modules stochastiques (par exemple, les choix des agriculteurs). La robustesse du modèle pourrait être évaluée en se basant sur des règles spatiales et les stratégies des agriculteurs, avec une exploration des résultats pour déterminer si les règles exprimées sont pertinentes où si elles ne reproduisent pas correctement les possibilités de mise en culture (Chérel et al., 2015). Cette analyse du modèle peut être complémentaire à l'analyse menée en confrontant les règles à la base de données de vérité terrain.

Dans cette section nous avons tenté de répondre à une partie des questions soulevées par l'application du modèle conceptuel à travers ses deux modules spatiaux. Parmi les pistes et limites identifiées nous avons brièvement traité d'une meilleure exploitation des images de télédétection. L'aspect multi-scalaire du modèle, qui représente bien l'échelle du segment (champ) et celle du Fokontany, mais où l'échelle de l'exploitation agricole est manquante et doit être intégrée est également discuté. La question des typologies pour caractériser cette échelle a été soulevée et sera plus détaillée dans la section suivante. Ensuite, la qualité des données (de terrain, de télédétection, de support à l'application des règles) utilisées a été analysée. Ainsi, il en ressort que la segmentation des images THRS pourrait encore être améliorée pour mieux définir spatialement les objets d'étude. Les données support pour l'application des connaissances présentent quant à elles des qualités inégales et sont parfois inexistantes, avec peu de chance d'amélioration à court terme. La base de données de vérité terrain enfin a été mise en lien avec les différentes méthodes d'échantillonnage et des méthodes pour réduire le déséquilibre entre les classes sont proposées. La quatrième sous-section portait sur l'un des enjeux de la thèse : la sélection des connaissances et les méthodes pour déterminer leur importance, leur combinaison pour créer une probabilité de trouver une classe dans un segment (donnant un $VPAC_{RS}$), ainsi que l'agrégation de ce VPAC avec le $VPAC_{RF}$ issu du traitement des données de télédétection.

Plusieurs méthodes ont été proposées pour rendre l'approche plus robuste et homogène, mais celles-ci demandent soit un nombre important d'experts ou des ressources pour traiter un nombre important d'informations ou encore des connaissances techniques avancées. L'une des solutions techniques est l'utilisation de la logique floue. Pour l'agrégation des VPAC, cette question est toujours en suspens. La dernière partie concerne les méthodes d'évaluation du modèle à l'aide d'analyses de sensibilité.

3 Quels changements pour le modèle conceptuel ? Perspectives méthodologiques

L'analyse de l'application du modèle conceptuel aux Hautes Terres malgaches a permis de mettre en évidence des limites et des pistes d'améliorations possibles. Cette analyse permet aussi de repenser le modèle conceptuel pour qu'il soit plus adapté à l'objectif de développer un modèle conceptuel combinant à la fois les approches de traitement de données de télédétection et de modélisation spatiale. De cet objectif découle la problématique de la reproductibilité du modèle qui peut être considérée comme partiellement atteinte. Par contre, le modèle conceptuel peut être mûri pour essayer de le rendre plus cohérent, générique et interdisciplinaire. Ces réflexions structurent les deux sections suivantes.

3.1 Un modèle trop centré sur Madagascar, mais déjà pluridisciplinaire

Le modèle conceptuel a été construit en s'appuyant sur le cas d'étude des Hautes Terres de Madagascar et pourrait souffrir d'un manque de généralisation à d'autres espaces d'agriculture familiale diversifiés. Pour améliorer le modèle conceptuel dans ce sens et avoir un meilleur aperçu des systèmes de petite agriculture familiale, il est important de se pencher sur la littérature à la fois thématique et technique. Une fois des cas d'étude et leurs spécificités analysées, il sera possible de les intégrer dans le modèle conceptuel pour que celui-ci soit plus générique.

Bien que l'aspect générique du modèle peut paraître limité, les principes qui sont développés le sont ou tendent à l'être. La décomposition en quatre modules, chacun destiné à une tâche particulière, permet une modularité du modèle en fonction des données disponibles et de l'objectif de la recherche. L'association de méthodes utilisées pour le traitement des données de télédétection et de méthodes de modélisation spatio-temporelle a permis une approche méthodologique pluridisciplinaire qui vise à devenir une approche interdisciplinaire, jusque là peu abordée dans la littérature (Vannier, 2011; Jahel, 2016). Cette approche pluridisciplinaire est ancrée dans la volonté de développer une méthode à la croisée des disciplines de la télédétection et de la modélisation des dynamiques spatiales tout en mobilisant d'autres sciences pour mieux appréhender l'agriculture familiale tropicale. Pour cela, la géographie, l'agronomie, la sociologie et l'économie ont été utilisées pour comprendre et interpréter les stratégies et décisions des agriculteurs dans leur contexte. L'élargissement de cette pluridisciplinarité à d'autres sciences pour mieux comprendre les

enjeux et les besoins de ces sciences peut permettre de lever certains verrous, qu'ils soient techniques ou bien thématiques. Les approches développées dans cette thèse doivent donc devenir méthodologiquement et thématiquement plus cohérentes pour permettre une conceptualisation complète du modèle et pour proposer un modèle interdisciplinaire et générique.

3.2 Penser une nouvelle approche thématique

L'ensemble des limites présentées précédemment ainsi que les résultats poussent à remettre en cause le modèle conceptuel. Les aspects techniques et appliqués ont été présentés dans la section précédente, cette sous-section reprend certaines limites abordées telles que l'aspect thématique du modèle, les typologies d'exploitation via l'échelle de l'exploitation et l'apport d'une approche hiérarchique pour mieux caractériser l'occupation et l'usage des sols.

La conception de ce modèle est orientée par un objectif méthodologique, la combinaison de connaissances différentes et de données de télédétection pour améliorer la caractérisation de l'occupation et de l'usage de sols. L'approche thématique n'est pas directement prise en compte comme le montre l'échelle d'application des modules, segment par segment, plutôt que de chercher à caractériser un espace thématiquement glscoherent. Cet espace peut être un agro-écosystème complet et cela peut être en parti effectué dans le module 4 en utilisant une unité de regroupement thématique non administrative. Cette approche thématique permet de mobiliser des notions non utilisées dans cette thèse comme le paysage par exemple pour en extraire des unités de paysage cohérentes correspondant à des systèmes d'utilisation agricole des terres (Almeida et al., 2016; Bellón de la Cruz, 2018). L'approche développée par (Bellón de la Cruz, 2018) permet d'appréhender de grands espaces via la création d'unités paysagères cohérentes à partir d'images à haute ou moyenne résolution spatiale (Sentinel-2, MODIS), même si les résultats sont encore limités pour les espaces d'agriculture familiale sub-tropicaux avec des données à haute et moyenne résolution. Cette orientation thématique pour déterminer à quelle échelle analyser un paysage et modéliser ses dynamiques pourrait permettre d'obtenir des résultats plus cohérents.

Le développement d'une approche spatialement thématique (paysage, agro-système) peut aussi permettre de mieux intégrer la modélisation des choix et stratégies propres aux agriculteurs par l'ajout d'une échelle de modélisation explicite de l'exploitation et de typologies.

Les typologies peuvent représenter un apport à la fois méthodologique et thématique pour l'amélioration du modèle conceptuel (Matus et al., 2013). L'utilisation de typologies permettrait de modéliser un ensemble d'exploitations familiales ou de formes entrepreneuriales pour comprendre et simuler les décisions des agricultures (Raharison et al., 2018). Ces simulations s'appuieraient sur des connaissances représentant les choix des agriculteurs avec une meilleure intégration des problématiques socio-économiques. La définition de plusieurs types d'exploitations permet de modéliser les possibilités de trouver une culture à l'aide de critères non utilisés dans notre modèle conceptuel tels que la gestion de l'exploitation à travers des assolements (Jahel et al., 2018). Les processus de décision des agriculteurs développés par Jahel (2016) pourraient permettre de

présélectionner un ensemble de segments pour les attribuer à une exploitation, puis modéliser les décisions des agriculteurs en fonction de comportements et stratégies prédéfinies. En itérant ce processus à travers différentes typologies et scénarios, la stochasticité de ce processus permettrait à la fois de montrer la variabilité des cultures possibles et donc de déterminer quels sont les choix plus probables des agriculteurs à travers les systèmes de cultures ou la tradition, notions non abordées dans ce modèle conceptuel. La création de processus stochastiques est une piste intéressante à développer, à la fois pour permettre de mesurer la variabilité des processus mobilisés, mais aussi pour la mise en place de nouveaux processus, là où le modèle proposé est déterministe en termes de règles spatiales.

En plus de l'ajout de l'échelle de l'exploitation et de typologies d'agriculteurs, une approche hiérarchique des classes (nomenclature emboîtée) permettrait d'aller plus loin à la fois dans les méthodes mobilisées et dans l'approche thématique.

L'analyse des résultats a montré l'intérêt de changer l'approche vis-à-vis des classes. L'approche hiérarchique telle qu'elle peut être appliquée en télédétection (Lebourgeois et al., 2017; Dupuy et al., 2020a) semble être une solution pertinente pour aborder les processus de modélisation et la classification des données de télédétection. Dans l'application des modules, un rudiment d'approche hiérarchique a été appliqué avec la modélisation de la catégorie « Autres », puis la répartition des classes qui la compose selon les $VPAC_{RF}$ (chapitre 4, section 3.2.1 Sélection selon la fiabilité des $VPAC_{RF}$). La détection du domaine cultivé pluvial par les modules spatiaux malgré la mauvaise détection du type de culture va dans le sens de la hiérarchisation des classes. Les trop nombreuses classes et la redondance des règles pour des cultures répondant aux mêmes conditions de mise en culture montrent une limite qui pourrait être dépassée en adoptant une nouvelle approche. Le regroupement des classes en catégories cohérentes thématiquement et radiométriquement pourrait être une solution de hiérarchisation telle que celle proposée par Jouve (2006) : « l'ager (l'espace cultivé), la sylva (l'espace forestier), le saltus (l'espace pastoral), et l'hortus (les jardins ou les champs de case) ». Une nouvelle approche en caractérisant en premier lieu le domaine cultivé, puis en déterminant le type de culture peut être une perspective méthodologique et conceptuelle intéressante à développer.

L'intégration dans le modèle conceptuel de typologies d'exploitation et d'une approche hiérarchique des classes est à la fois un changement spatial et temporel (gestion des assolements dans une exploitation sur plusieurs saisons) dans la conception du modèle. Ce changement d'approche dans le modèle conceptuel permettrait de briser les cloisons définies par les quatre modules pour permettre à la fois une approche spatio-temporelle et multiscalaire.

Cette section a permis d'identifier les changements majeurs à opérer pour une amélioration du modèle conceptuel, notamment en adoptant une approche thématique (stratification) en plus de l'approche méthodologique. Pour cela, il est nécessaire de décloisonner les modules pour passer d'une approche pluridisciplinaire à une approche interdisciplinaire. Ce décloisonnement passe par une échelle d'analyse thématique prenant en compte des ensembles spatiaux cohérents à moyenne échelle (et non

uniquement l'échelle du champ (segment) et l'échelle du Fokontany (administrative). L'intégration d'une échelle de modélisation correspondant aux décisions des agriculteurs (échelle de l'exploitation) permettrait d'utiliser des typologies et d'intégrer de la stochasticité dans le modèle. Une approche hiérarchique des classes serait un complément intéressant à la vue des résultats obtenus et obtenir un modèle spatio-temporel et multiscalaire.

4 Perspectives

Deux principales faiblesses ont été identifiées dans la démarche de cette thèse : une forte dépendance aux données qui a limité l'application des modules et une volonté de développement méthodologique trop importante par rapport à l'espace où le modèle était appliqué. Pour conclure, en plus de l'ensemble des pistes de recherche proposées dans ce chapitre, nous proposons deux orientations pour continuer ces travaux et répondre à deux objectifs différents : le développement méthodologique et la cartographie opérationnelle. Ainsi, nous proposons de développer un modèle en s'appuyant sur un milieu contrôlé, et de simplifier le modèle présenter pour qu'il soit opérationnel dans un système d'agriculture familiale tropical.

Pour le développement complet d'une méthode autour d'un sujet pluridisciplinaire peu abordé dans la littérature, il est nécessaire de contrôler les paramètres qui permettent l'application des processus, ainsi que d'évaluer le bon fonctionnement du modèle. Un milieu contrôlé pour appliquer un modèle conceptuel permettrait de supprimer la variabilité des résultats liés aux données afin de développer l'approche. L'évaluation de la sensibilité du modèle aux données ou à la zone d'étude serait possible pour transférer et adapter la méthode à d'autres systèmes complexes d'agriculture familiale tropicale.

Les résultats de l'utilisation à la fois d'une classification des données de télédétection et de méthodes de modélisation spatiale ont montré un résultat satisfaisant pour la caractérisation du domaine cultivé pluvial. L'identification d'un grand nombre de types de culture avec une précision moyenne n'est pas utilisable par les scientifiques thématiciens ou pour l'application de politiques publiques. La simplification du nombre de classes pour identifier des groupes de culture de manière faible est une solution pour produire des cartes fiables et pratiques d'occupation et d'usage des sols. Adapter le modèle conceptuel pour mieux caractériser le domaine cultivé et plus particulièrement pour différencier les espaces non agricoles, tel que la savane, des cultures pluviales semble prometteur là où ces discriminations sont toujours difficiles. Une caractérisation de l'occupation ou de l'usage des sols pourrait être produite à différentes résolutions thématiques selon les besoins (groupes de cultures, cultures pluviales, cultures irriguées, pratiques culturales). Ainsi, un modèle alliant à la fois des données de télédétection et des connaissances pour une cartographie thématique simple des systèmes agricoles complexes peut être une perspective opérationnelle intéressante.

POINTS CLÉS DU CHAPITRE 6

- **APPLICATION DU MODÈLE CONCEPTUEL** :
  - **pas d'application des modules temporels**, car aucune stratégie relative à la gestion d'un champ vis-à-vis de la saison des pluies n'a pu être identifiée (module 3 : règles temporelles). L'échelle temporelle d'application du modèle (une saison culturale principale centrée sur la saison des pluies) n'est pas suffisamment étendue pour intégrer des connaissances pluri-annuelles (module 1 : modélisation du fonctionnement temporel) ;

- **PISTES D'AMÉLIORATION** :
  - une meilleure exploitation des images satellitaires (autres méthodes de classification, THRS) ;
  - une modélisation **multiscalaire** basée sur des paramètres biophysiques et socio-économiques (champ), des décisions des agriculteurs et d'une répartition théorique d'occupation et d'usage des sols à une petite échelle (village, commune, district) ;
  - pour les systèmes d'agriculture familiale tropicaux, une meilleure qualité des données support pour l'application des connaissances n'est pas envisageable à court terme, il faut donc identifier d'autres moyens d'utiliser ou évaluer ces données ;
  - les méthodes pouvant rendre l'approche plus robuste et homogène **demandent une expertise importante** ou l'utilisation de **méthodes mathématiques** tel que la logique floue ;
  - la question de l'amélioration de l'agrégation des VPAC est quant à elle toujours en suspens.

- **CHANGEMENTS DANS L'APPROCHE CONCEPTUELLE** :
  - pour proposer un modèle interdisciplinaire et générique, il faut **approfondir l'approche méthodologique et thématique** développée dans cette thèse ;
  - l'ajout d'une échelle de modélisation correspondant aux **choix des agriculteurs sur leur exploitation** permettrait d'utiliser des **typologies** et d'intégrer de la **stochasticité** dans le modèle
  - l'intégration de typologies d'exploitation et d'une **approche hiérarchique** des classes pour permettre à la fois une **approche spatio-temporelle et multiscalaire**.

- **PERSPECTIVES** :
  - il serait intéressant de développer un modèle avec un minimum de contraintes sur les données pour mieux évaluer et comprendre le fonctionnement du modèle ;
  - le développement d'un modèle plus simple pour mieux caractériser le domaine pluvial dans les systèmes tropicaux d'agriculture familiale serait intéressant d'un point de vue opérationnel.

CONCLUSION GÉNÉRALE

La cartographie de l'occupation et de l'usage des sols par classification de données de télédétection montre des performances limitées pour les systèmes agricoles complexes que sont les agricultures familiales des pays des suds. A l'ère du *big data*, les récents développements méthodologiques visant à lever ces limites portent avant tout sur les techniques d'apprentissage automatique pour mieux exploiter le potentiel des images satellitaires, notamment avec les approches basées sur de l'apprentissage profond. Cette thèse ne s'inscrit pas dans cette tendance actuelle, mais explore plutôt la possibilité de tirer parti des connaissances accumulées sur ces espaces et issues de domaines multiples – ex. agriculture, socio-économie, ou géographie. Il s'agit ainsi, en complémentarité de méthodes automatiques, d'obtenir des cartographies de l'occupation et de l'usage des sols qui soient cohérentes avec ces connaissances, et donc plus fiables. Pour cela, cette thèse a proposé deux sous objectifs : le premier correspondant au développement d'un modèle conceptuel qui allie à la fois des méthodes de traitement de données de télédétection ainsi que des méthodes de modélisation spatiale, et le second traitant de l'application de ce modèle conceptuel à un cas d'agriculture familiale, sur une zone d'étude située dans les Hautes Terres du Vakinankaratra, à Madagascar.

Le modèle conceptuel proposé se base à la fois sur l'approche orientée objet (OBIA), utilisée en télédétection pour déterminer la structure de l'espace, puis sur la caractérisation de chaque objet à l'aide de séries temporelles d'images satellitaires et de connaissances spatiales et temporelles à différentes échelles. Dans le modèle, l'observation d'une classe dans un objet est définie par l'utilisation de données d'observation de la Terre, tandis que la possibilité de trouver une classe dans un objet est représentée par la modélisation spatio-temporelle des stratégies des agriculteurs. Quatre modules sont proposés pour modéliser et définir les stratégies des agriculteurs : deux modules spatiaux et deux modules temporels.

Pour intégrer des connaissances basées sur l'expertise, il est nécessaire de les formaliser en règles, qui peuvent correspondre à des stratégies agricoles (spatiales et temporelles) différentes et qui peuvent être hiérarchisées selon leur importance. Ces connaissances formalisées sont combinées à une classification supervisée d'images de télédétection dans le but de spatialiser à l'échelle du champ les pratiques et stratégies des agriculteurs. Avec les connaissances identifiées sur notre zone d'étude, nous avons pu appliquer et évaluer les deux modules spatiaux du modèle conceptuel, l'un étant basé sur des règles spatiales identifiant des contextes favorables aux différentes cultures, et l'autre sur la modélisation de la répartition spatiale des types de cultures pour un espace administratif donné.

L'évaluation de la combinaison de la télédétection et de la modélisation spatiale s'est basée sur l'analyse de la cohérence des cartes produites. L'application de connaissances à travers ces deux modules permet une légère amélioration quantitative des résultats, mais surtout une meilleure discrimination entre les cultures pluviales et les espaces de savane ainsi que l'obtention de confu-

sions thématiquement plus cohérentes entre les classes. Les enseignements issus de l'application du modèle conceptuel nous permettent de proposer des pistes pour mieux organiser le modèle conceptuel et essayer d'aller au-delà des limites identifiées.

Basée sur un modèle simple et sans avoir un caractère générique hors de la zone d'étude, la méthode développée dans cette thèse est un début de réponse pour améliorer la caractérisation des espaces façonnés par l'agriculture familiale en zone tropicale. Bien que les résultats montrent qu'un important travail doit encore être effectué pour proposer une méthode robuste et générique, cette méthode est originale par l'intégration de connaissances agricoles via la modélisation spatiale, approche jusque là peu abordée dans la littérature. En mobilisant à la fois des connaissances en sciences sociales et agronomiques et des données de télédétection, cette approche peut être complémentaire d'autres méthodes en cours de développement basées sur des données thématiques et la modélisation mathématique des systèmes étudiés.

Si les travaux de cette thèse ne montrent pas de résultats quantitatifs probants, ils s'inscrivent dans une réflexion pour mieux comprendre l'agriculture familiale et sa représentation. Ce système, diversifié et dynamique, est la principale forme de production alimentaire mondiale et est un élément majeur pour lutter contre l'insécurité alimentaire tout en essayant de limiter les impacts environnementaux associés. En faisant de la décennie 2019 – 2028 la « Décennie des Nations Unies pour l'agriculture familiale », les Nations Unies replacent l'agriculture familiale au centre des politiques nationales et internationales, tout en mettant en exergue le besoin de mieux la comprendre et la caractériser.

Bibliographie

Adger, W. N., Huq, S., Brown, K., Conway, D., and Hulme, M. (2003). Adaptation to climate change in the developing world. *Progress in Development Studies*, 3(3):179–195, DOI: 10.1191/1464993403ps060oa.

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., and Schweik, C. M. (2002). A review and assessment of land-use change models: dynamics of space, time and human choice. Technical Report NE-297, Newtown Square, PA : U.S. Department of Agriculture, Forest Service, Northeastern Research Station.

Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E., and A. de By, R. (2018). A Cloud-Based Multi-Temporal Ensemble Classifier to Map Smallholder Farming Systems. *Remote Sensing*, 10(5):729, DOI: 10.3390/rs10050729. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Alakamisy, C. (2017). Monographie de la commune rurale Alakamisy - Ambano Antsirabe II. Technical report.

Almeida, C., Mourão, M., Dessay, N., Lacques, A.-E., Monteiro, A., Durieux, L., Venturieri, A., and Seyler, F. (2016). Typologies and Spatialization of Agricultural Production Systems in Rondônia, Brazil: Linking Land Use, Socioeconomics and Territorial Configuration. *Land*, 5(2):18, DOI: 10.3390/land5020018. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Alpízar, F., Saborío-Rodríguez, M., Martínez-Rodríguez, M. R., Viguera, B., Vignola, R., Capitán, T., and Harvey, C. A. (2020). Determinants of food insecurity among smallholder farmer households in Central America: recurrent versus extreme weather-driven events. *Regional Environmental Change*, 20(1):22, DOI: 10.1007/s10113-020-01592-y.

Altieri and Miguel A (2008). *Small farms as a planetary ecological asset: five key reasons why we should support the revitalisation of small farms in the global South.* Third World network (TWN), Penang. OCLC: 316209591.

Alvarez, S., Timler, C. J., Michalscheck, M., Paas, W., Descheemaeker, K., Tittonell, P., Andersson, J. A., and Groot, J. C. J. (2018). Capturing farm diversity with hypothesis-based typologies: An innovative methodological framework for farming system typology development. *PLOS ONE*, 13(5):e0194757, DOI: 10.1371/journal.pone.0194757.

Andriamanohy, F. R. W., Bélières, J.-F., Razafimahatratra, J. C., Mamy, H. M., and Emilson, N. F. H. (2016). Production agricole, travail salarie, et revenu dans les exploitations agricoles de la zone du lac Alaotra a Madagascar. page 25, Antananarivo, Madagascar.

Andriandrahona, F. R., Douzet, J.-M., Dusserre, J., Scopel, E., Rabenjanahary, H., Ralaiarimalala, S., and Doré, T. (2007). METHODE DE DIAGNOSTIC AGRONOMIQUE REGIONAL (APPLICATION AU CAS D'ANTSAMPANIMAHAZO). page 51.

Atzberger, C. (2013). Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing*, 5(2):949–981, DOI: 10.3390/rs5020949.

Autfray, P. (2020). Etude chaîne de valeur Maïs et Soja - Alimentation Animale dans les régions d'Analamanga, Itasy et Vakinankaratra 2019/2020. Augmentation de l'offre en maïs et soja et conditions de contractualisation avec le secteur de production d'aliments pour élevage. Volet 1 : Production. Technical report, CIRAD, Montpellier, France.

Autfray, P., Raharison, T., Razafimahatratra, M. H., Randrianjafizanaka, M., Rakotoarivelo, N. M., Moussa, N., Penot, E., and Fofifa, C. (2014). Conservation Agriculture (CA) adoption by smallholder in Lake Alaotra area in Madagascar: from CA to agro-ecological practices. *Agroecology and Sustainability of Tropical Rainfed Cropping Systems*, page 6.

Baatz, M. and Schäpe, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informations-Verarbeitung XII*, pages 12–23.

Baba, M., Gascoin, S., Hagolle, O., Bourgeois, E., Desjardins, C., and Dedieu, G. (2020). Evaluation of Methods for Mapping the Snow Cover Area at High Spatio-Temporal Resolution with VEN\micro{}S. *Remote Sensing*, 12:3058, DOI: 10.3390/rs12183058.

Bailly, S., Giordano, S., Landrieu, L., and Chehata, N. (2018). Crop-Rotation Structured Classification using Multi-Source Sentinel Images and LPIS for Crop Type Mapping. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1950–1953, Valencia. IEEE, DOI: 10.1109/IGARSS.2018.8518427.

Balstrøm, T. (2002). On identifying the most time-saving walking route in a trackless mountainous terrain. *Geografisk Tidsskrift-Danish Journal of Geography*, 102(1):51–58, DOI: 10.1080/00167223.2002.10649465.

Barcza, Z., Kern, A., Haszpra, L., and Kljun, N. (2009). Spatial representativeness of tall tower eddy covariance measurements using remote sensing and footprint analysis. *Agricultural and Forest Meteorology*, 149(5):795–807, DOI: 10.1016/j.agrformet.2008.10.021.

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., and Moran, M. S. (2000). Coincident detection of crop water stress, nitrogen status, and canopy density using ground based

multispectral data. *Proceedings of the Fifth International Conference on Precision Agriculture*, page 15.

Begue, A., Leroux, L., Lo Seen, D., Tonneau, J.-P., and Morant, P. (2016). Observation spatiale pour l'agriculture en Afrique : potentiels et défis. Technical Report 12, AFD-CIRAD.

Belgiu, M. and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204:509–523, DOI: 10.1016/j.rse.2017.10.005.

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, DOI: 10.1016/j.isprsjprs.2016.01.011.

Belhouchette, H., Louhichi, K., Therond, O., Mouratiadou, I., Wery, J., Ittersum, M. v., and Flichman, G. (2011). Assessing the impact of the Nitrate Directive on farming systems using a bio-economic modelling chain. *Agricultural Systems*, 104(2):135–145, DOI: 10.1016/j.agsy.2010.09.003.

Bellón, B., Bégué, A., Lo Seen, D., de Almeida, C., and Simões, M. (2017). A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. *Remote Sensing*, 9(6):600, DOI: 10.3390/rs9060600.

Bellón, B., Bégué, A., Lo Seen, D., Lebourgeois, V., Evangelista, B. A., Simões, M., and Demonte Ferraz, R. P. (2018). Improved regional-scale Brazilian cropping systems' mapping based on a semi-automatic object-based clustering approach. *International Journal of Applied Earth Observation and Geoinformation*, 68:127–138, DOI: 10.1016/j.jag.2018.01.019.

Bellón de la Cruz, B. (2018). *Une approche multiscalaire par télédétection pour la cartographie et la caractérisation des systèmes agricoles à l'échelle régionale.* PhD thesis, AgroParisTech.

Ben Naâmane, C. K. and Autfray, P. (2020). Gestion comparée de la fertilité au niveau des exploitations agricoles familiales du Moyen-Ouest de la région Vakinankaratra et des Hautes-Terres de la région Itasy à Madagascar. Technical report, CIRAD. Publisher: Secure.

Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R. G., and Dupuy, S. (2018). $M^3\textFusion$: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12):4939–4949, DOI: 10.1109/JSTARS.2018.2876357. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Beshers, J. M., editor (1965). *Computer methods in the analysis of large-scale social systems: proceedings of a conference held at the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University in Cambridge, Massachusetts on October 19-21, 1964.* Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, Cambridge, Mass., joint center for urban studies edition. OCLC: 166299.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., and Tiede, D. (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180–191, DOI: 10.1016/j.isprsjprs.2013.09.014.

Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Matton, N., Morin, D., Popescu, R., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Ahmad, I., Bégué, A., Wu, B., De Abelleyra, D., Diarra, A., Dupuy, S., French, A., Ul Hassan Akhtar, I., Kussul, N., Lebourgeois, V., Le Page, M., Newby, T., Savin, I., Verón, S. R., Koetz, B., and Defourny, P. (2015). Building a Data Set over 12 Globally Distributed Sites to Support the Development of Agriculture Monitoring Applications with Sentinel-2. *Remote Sensing*, 7(12):16062–16090, DOI: 10.3390/rs71215815.

Borshchev, A. and al, e. (2004). From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. In *In Proceedings of the 22nd International Conference of the System Dynamics Society, July 25 - 29*.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, COLT '92, pages 144–152, New York, NY, USA. Association for Computing Machinery, DOI: 10.1145/130385.130401.

Bousquet, F. and Le Page, C. (2004). Multi-agent simulations and ecosystem management: a review. *Ecological Modelling*, 176(3-4):313–332, DOI: 10.1016/j.ecolmodel.2004.01.011.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brunet, R. (1980). La composition des modèles dans l'analyse spatiale. *L'Espace géographique*, 9(4):253–265, DOI: 10.3406/spgeo.1980.3572. Publisher: Persée - Portail des revues scientifiques en SHS.

Buchmüller, S. and Weidmann, U. (2006). Parameters of pedestrians, pedestrian traffic and walking facilities. Technical report, ETH Zurich, DOI: 10.3929/ETHZ-B-000047950. Artwork Size: 56 p. Medium: application/pdf.

Bureau, J.-C., Guyomard, H., and Réquillart, V. (2001). On inefficiencies in the European sugar regime. *Journal of Policy Modeling*, 23(6):659–667, DOI: 10.1016/S0161-8938(01)00080-1.

Burnod, P., Rakotomalala, H., Andriamanalina, B. S., and Di Roberto, H. (2016). Composer entre la famille et le marché à Madagascar: Évolution de l'accès des jeunes agriculteurs à la terre. *Afrique contemporaine*, 259(3):23, DOI: 10.3917/afco.259.0023.

BVLac (n.d.). Fiche technique : Maïs associé aux légumineuses.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., and R. Verón, S. (2018). Remote Sensing and Cropping Practices: A Review. *Remote Sensing*, 10(2):99, DOI: 10.3390/rs10010099.

Bélières, J.-F. (2020a). Etude chaîne de valeur pomme de terre dans les régions d'Analamanga, Itasy et Vakinankaratra 2019/2020. Partie 1 : Importance de la culture de la pomme de terre pour les exploitations agricoles et rentabilité de la production de plants de semence et de consommation. Technical Report 1, Cirad, CASEF, MAEP.

Bélières, J.-F. (2020b). Étude diagnostic relative au potentiel de croissance de la chaine de valeur lait et produits dérivés (Hautes Terres - Madagascar). Technical report, Cirad, CASEF, Antananarivo.

Bélières, J.-F., Bonnal, P., Bosc, P.-M., Losch, B., Marzin, J., Sourisseau, J.-M., Baron, V., and Loyat, J. (2013). Les agricultures familiales du monde. Définitions, contributions et politiques publiques. Rapport pour l'AFD, le MAZ et le MAAF, Cirad, Montpellier, France.

Castets, M. (2015). *Pavages réguliers et modélisation des dynamiques spatiales à base de graphes d'interaction: conception, implémentation, application.* PhD thesis, Université de Montpellier.

Castillejo-González, I. L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J. M., Jurado-Expósito, M., de la Orden, M. S., and González-Audicana, M. (2009). Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture*, 68(2):207–215, DOI: 10.1016/j.compag.2009.06.004.

Chauvigne, V. (2005). Enjeux et perspectives du développement de la riziculture pluviale à Madagascar.

Chery, J.-P. and Smektala, G. (2004). D'un modèle chorématique à un modèle de simulation : gestion des ressources ligneuses en zone soudano-sahélienne. *Cahiers d'études et de recherches francophones / Agriculture*, 6(13):528–538.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2(2):137–167, DOI: 10.1016/S0019-9958(59)90362-6.

Chopin, P. and Blazy, J.-M. (2013). Assessment of regional variability in crop yields with spatial autocorrelation: Banana farms and policy implications in Martinique. *Agriculture, Ecosystems et Environment*, 181:12–21, DOI: 10.1016/j.agee.2013.09.001.

Chopin, P., Blazy, J.-M., and Doré, T. (2015a). A new method to assess farming system evolution at the landscape scale. *Agronomy for Sustainable Development*, 35(1):325–337, DOI: 10.1007/s13593-014-0250-5.

Chopin, P., Doré, T., Guindé, L., and Blazy, J.-M. (2015b). MOSAICA: A multi-scale bioeconomic model for the design and ex ante assessment of cropping system mosaics. *Agricultural Systems*, 140:26–39, DOI: 10.1016/j.agsy.2015.08.006.

Chávez, A. B., Broadbent, E. N., and Almeyda Zambrano, A. M. (2014). Smallholder policy adoption and land cover change in the southeastern Peruvian Amazon: A twenty-year perspective. *Applied Geography*, 53:223–233, DOI: 10.1016/j.apgeog.2014.06.017.

Chérel, G., Cottineau, C., and Reuillon, R. (2015). Beyond Corroboration: Strengthening Model Validation by Looking for Unexpected Patterns. *PLOS ONE*, 10(9):e0138212, DOI: 10.1371/journal.pone.0138212. Publisher: Public Library of Science.

Clavel, L., Soudais, J., Baudet, D., and Leenhardt, D. (2011). Integrating expert knowledge and quantitative information for mapping cropping systems. *Land Use Policy*, 28(1):57–65, DOI: 10.1016/j.landusepol.2010.05.001.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, DOI: 10.1177/001316446002000104. Publisher: SAGE Publications Inc.

Collin, Y. (2012). Rapport d'information sur le défi alimentaire à l'horizon 2050. Rapport d'information 504, Sénat.

Colonna, A., Di Stefano, V., Lombardo, S., Papini, L., and Rabino, G. A. (1998). Learning Urban Cellular Automata In A Real World. In Bandini, S., Serra, R., and Liverani, F. S., editors, *Cellular Automata: Research Towards Industry*, pages 165–184, London. Springer, DOI: 10.1007/978-1-4471-1281-5_16.

Colwell, J. E. (1974). Vegetation canopy reflectance. *Remote Sensing of Environment*, 3(3):175–183, DOI: 10.1016/0034-4257(74)90003-0.

Comber, A., Fisher, P., and Wadsworth, R. (2005). What is Land Cover? *Environment and Planning B: Planning and Design*, 32(2):199–209, DOI: 10.1068/b31135.

Costanza, R. (1987). Simulation Modeling on the Macintosh Using STELLA. *BioScience*, 37(2):129–132, DOI: 10.2307/1310367.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P., and van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387(6630):253–260, DOI: 10.1038/387253a0. Number: 6630 Publisher: Nature Publishing Group.

Costanza, R. and Voinov, A. (2004). *Landscape Simulation Modeling. A Spatially Explicit, Dynamic Approach.* Modeling Dynamic Systems. Springer, New York.

Cousins, B. (2013). Smallholder Irrigation Schemes, Agrarian Reform and 'Accumulation from Above and from Below' in South Africa: Irrigation Schemes, Reform and Accumulation in South Africa. *Journal of Agrarian Change*, 13(1):116–139, DOI: 10.1111/joac.12000.

CREMA (2013). Monographie région Vakinankaratra. Technical report, Centre de Recherches, d'Études et d'Appui à l'Analyse Économique de Madagascar.

Dabat, M.-H., Jenn-Treyer, O., Magnay, J., and Razafintsalama, Z. (2006). Rice Markets in Madagascar in Disarray: Policy Options for Increased Efficiency and Price Stabilization. Working paper 101, World Bank.

Dan, G. Y. (2006). *Agriculture, Rural Areas and Farmers in China*. China Intercontinental Press, Agriculture, Rural Areas and Farmers in China.

David-Benz, H., Rasolofo, P., and Andriamparany, S. (2014). Rice price management in Madagascar: Joint action of information and multi-stakeholder dialogue. *Cahiers Agricultures*, 23(4-5):295–303, DOI: 10.1684/agr.2014.0707.

Debats, S. R., Luo, D., Estes, L. D., Fuchs, T. J., and Caylor, K. K. (2016). A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sensing of Environment*, 179:210–221, DOI: 10.1016/j.rse.2016.03.010.

Defrise, L., Burnod, P., Tonneau, J. P., and Andriamanga, V. (2019). Disparition et permanence de l'agriculture urbaine à Antananarivo. *L'Espace géographique*, 48(3):263, DOI: 10.3917/eg.483.0263.

Degenne, P. (2012). *Une approche générique de modélisation spatiale et temporelle: application à la modélisation de la dynamique des paysages*. PhD thesis, Université Paris-Est.

Degenne, P., Lahcen, A. A., Curé, O., Forax, R., Parigot, D., and Seen, D. L. (2010). Modelling the environment using graphs with behaviour: do you speak Ocelet? In *iEMSs 2010*, pages 8–pages.

Degenne, P. and Lo Seen, D. (2016). Ocelet: Simulating processes of landscape changes using interaction graphs. *SoftwareX*, 5:89–95, DOI: 10.1016/j.softx.2016.05.002.

Degenne, P., Lo Seen, D., Parigot, D., Forax, R., Tran, A., Ait Lahcen, A., Curé, O., and Jeansoulin, R. (2009). Design of a Domain Specific Language for modelling processes in landscapes. *Ecological Modelling*, 220(24):3527–3535, DOI: 10.1016/j.ecolmodel.2009.06.018.

Delfosse, C. (2015). L'agriculture familiale. *Bulletin de l'association de géographes français*, 92(3):283–287, DOI: https://doi.org/10.4000/bagf.658.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, DOI: 10.1111/j.2517-6161.1977.tb01600.x.

Derksen, D., Inglada, J., and Michel, J. (2020). Geometry Aware Evaluation of Handcrafted Superpixel-Based Features and Convolutional Neural Networks for Land Cover Mapping Using Satellite Imagery. *Remote Sensing*, 12(3):513, DOI: 10.3390/rs12030513. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Dieppe, A. and World Bank (2020). *Global economic prospects: slow growth, policy challenges*. OCLC: 1111771966.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271.

Dixon, J., Gibbon, D. P., Gulliver, A., and Hall, M. (2001). *Farming systems and poverty: improving farmers' livelihoods in a changing world.* FAO ; World Bank, Rome : Washington, D.C. OCLC: ocm48933776.

Donadieu, J. and L'Helguen, C. (2016). Technical note SENTINEL-2A L2A Products description. Technical note, CNES, Direction du Centre Spatial de Toulouse.

Douzas, G., Bacao, F., Fonseca, J., and Khudinyan, M. (2019). Imbalanced Learning in Land Cover Classification: Improving Minority Classes' Prediction Accuracy Using the Geometric SMOTE Algorithm. *Remote Sensing*, 11(24):3040, DOI: 10.3390/rs11243040.

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., and Liu, S. (2012). Multiple Classifier System for Remote Sensing Image Classification: A Review. *Sensors (Basel, Switzerland)*, 12(4):4764–4792, DOI: 10.3390/s120404764.

Dupuy, S., Defrise, L., Lebourgeois, V., Gaetano, R., Burnod, P., and Tonneau, J.-P. (2020a). Analyzing Urban Agriculture's Contribution to a Southern City's Resilience through Land Cover Mapping: The Case of Antananarivo, Capital of Madagascar. *Remote Sensing*, 12(12):1962, DOI: 10.3390/rs12121962.

Dupuy, S., Gaetano, R., and Le Mézo, L. (2020b). Mapping land cover on Reunion Island in 2017 using satellite imagery and geospatial ground data. *Data in Brief*, 28:104934, DOI: 10.1016/j.dib.2019.104934.

Durand, W. (2016). Crop yield forecasting in South Africa. In *Crop Yield Forecasting: Methodological andInstitutional Aspects. Current Practices From Selected Countries (Belgium, China,Morocco, South Africa, USA) With a Focus on AMIS Crops (Maize, Rice, Soybeans and Wheat)*, pages 111–150. FAO„ Rome, food and agricultureorganization of the united nations edition.

Dury, J., Schaller, N., Garcia, F., Reynaud, A., and Bergez, J. E. (2012). Models to support cropping plan and crop rotation decisions. A review. *Agronomy for Sustainable Development*, 32(2):567–580, DOI: 10.1007/s13593-011-0037-x.

Dutton, J. M. and Starbuck, W. H. (1971). Computer Simulation Models of Human Behavior: A History of an Intellectual Technology. *IEEE Trans. Syst. Man Cybern.*, DOI: 10.1109/TSMC.1971.4308269.

Duveiller, G. and Defourny, P. (2010). A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sensing of Environment*, 114(11):2637–2650, DOI: 10.1016/j.rse.2010.06.001.

Eberlein, R. L. and Peterson, D. W. (1992). Understanding models with Vensim™. *European Journal of Operational Research*, 59(1):216–219, DOI: 10.1016/0377-2217(92)90018-5.

Edwards, A. L. (1948). Note on the "correction for continuity" in testing the significance of the difference between correlated proportions. *Psychometrika*, 13(3):185–187, DOI: 10.1007/BF02289261.

Eigner, A. E. and Nuppenau, E.-A. (2019). Applied spatial approach of modelling field size changes based on a consideration of farm and landscape interrelations. *Agricultural Systems*, 176:102648, DOI: 10.1016/j.agsy.2019.102648.

El Hajj, M., Baghdadi, N., Zribi, M., Rodríguez-Fernández, N., Wigneron, J. P., Al-Yaari, A., Al Bitar, A., Albergel, C., and Calvet, J.-C. (2018). Evaluation of SMOS, SMAP, ASCAT and Sentinel-1 Soil Moisture Products at Sites in Southwestern France. *Remote Sensing*, 10(4):569, DOI: 10.3390/rs10040569. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

El Hajj, M., Bégué, A., Guillaume, S., and Martiné, J.-F. (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, DOI: 10.1016/j.rse.2009.04.009.

El Merabet, Y. (2013). *Segmentation d'images couleur par combinaison LPE-régions/LPE-contours et fusion de régions. Application à la segmentation de toitures à partir d'orthophotoplans.* PhD thesis, Université de Technologie de Belfort-Montbeliard; Université Ibn Tofai.

Elsawah, S., Pierce, S. A., Hamilton, S. H., van Delden, H., Haase, D., Elmahdi, A., and Jakeman, A. J. (2017). An overview of the system dynamics process for integrated modelling of socio-ecological systems: Lessons on good modelling practice from five case studies. *Environmental Modelling et Software*, 93:127–145, DOI: 10.1016/j.envsoft.2017.03.001.

Escadafal, R., Girard, M.-C., and Courault, D. (1989). Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. *Remote Sensing of Environment*, 27(1):37–46, DOI: 10.1016/0034-4257(89)90035-7.

Evans, T. P., Phanvilay, K., Fox, J., and Vogler, J. (2011). An agent-based model of agricultural innovation, land-cover change and household inequality: the transition from swidden cultivation to rubber plantations in Laos PDR. *Journal of Land Use Science*, 6(2-3):151–173, DOI: 10.1080/1747423X.2011.558602. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/1747423X.2011.558602.

Fall, A. and Fall, J. (2001). A domain-specific language for models of landscape dynamics. *Ecological Modelling*, 141(1):1–18, DOI: 10.1016/S0304-3800(01)00334-9.

FAO (2015). *La situation mondiale de l'alimentation et de l'agriculture. Ouvrir l'agriculture familiale à l'innovation.* Rome, fao edition.

FAO and FIDA (2019). *Décennie des nations unies pour l'agriculture familiale 2019-2028 Plan d'action global.* Rome.

FAO, IFAD, UNICEF, WFP, and WHO (2020). *The State of Food Security and Nutrition in the World 2020*. FAO, IFAD, UNICEF, WFP and WHO, Rome, fao edition, DOI: 10.4060/ca9692en.

FAO and PAM (2013). Mission FAO/PAM d'évaluation de la sécurité alimentaire à Madagascar. Technical report, FAO/PAM, Rome.

Fasbender, D., Radoux, J., and Bogaert, P. (2008). Bayesian Data Fusion for Adaptable Image Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 46(6):1847–1857, DOI: 10.1109/TGRS.2008.917131.

Fern, C. J. S. and Warner, T. A. (2002). Scale and Texture in Digital lmage Classification. *Photogrammetric Engineering et Remote Sensing*, 68(1):51–63.

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15:3133–3181.

FewsNet (2013). Carte des zones de moyens d'existence à Madagascar. PAM/USAID Réseau des systèmes d'alerte précoce contre la famine. Technical report, USAID, Washington.

Fisher, P. F. and Comber, A. J. (2005). Land use and Land cover: Contradiction or Complement. In *Re-Presenting GIS*, pages 85–98. Wiely edition.

Flichman, G., editor (2011). *Bio-Economic Models applied to Agricultural Systems*. Springer Netherlands, Dordrecht.

FOFIFA (2019). Calendriers culturaux du FOFIFA.

Fofifa (n.d.). Carte des variétés de riz pluvial.

Fofifa and Cirad (2015). Des variétés améliorées de riz pluvial adaptées à la haute et moyenne altitude à Madagascar. Technical report, FOFIFA, CIRAD.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., and Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, 478(7369):337–342, DOI: 10.1038/nature10452.

Ford, A. and Ford, F. A. (1999). *Modeling the Environment: An Introduction to System Dynamics Models of Environmental Systems*. Island Press. Google-Books-ID: HLYsAQAAMAAJ.

Forkuor, G., Conrad, C., Thiel, M., Ullmann, T., and Zoungrana, E. (2014). Integration of Optical and Synthetic Aperture Radar Imagery for Improving Crop Mapping in Northwestern Benin, West Africa. *Remote Sensing*, 6(7):6472–6499, DOI: 10.3390/rs6076472.

Forrester, J. W. (1969). *Urban dynamics*. Cambridge, Massachusetts, m.i.t. press edition. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ncr.4100580716.

Foundation, P. S. (2018). Python Language Reference.

Friedl, M. A. and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3):399–409, DOI: 10.1016/S0034-4257(97)00049-7.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Aziz, S. A., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D. R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., Velde, M. v. d., Dunwoody, A., Kraxner, F., and Obersteiner, M. (2015). Mapping global cropland and field size. *Global Change Biology*, 21(5):1980–1992, DOI: 10.1111/gcb.12838. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.12838.

Fritz, S., See, L., You, L., Justice, C., Becker-Reshef, I., Bydekerke, L., Cumani, R., Defourny, P., Erb, K., Foley, J., Gilliams, S., Gong, P., Hansen, M., Hertel, T., Herold, M., Herrero, M., Kayitakire, F., Latham, J., Leo, O., McCallum, I., Obersteiner, M., Ramankutty, N., Rocha, J., Tang, H., Thornton, P., Vancutsem, C., Velde, M. v. d., Wood, S., and Woodcock, C. (2013). The Need for Improved Maps of Global Cropland. *Eos, Transactions American Geophysical Union*, 94(3):31–32, DOI: https://doi.org/10.1002/2013EO030006. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013EO030006.

Funk, C., Raghavan Sathyan, A., Winker, P., and Breuer, L. (2020). Changing climate - Changing livelihood: Smallholder's perceptions and adaption strategies. *Journal of Environmental Management*, 259:109702, DOI: 10.1016/j.jenvman.2019.109702.

Gaetano, R., Cozzolino, D., D'Amiano, L., Verdoliva, L., and Poggi, G. (2017). Fusion of sar-optical data for land cover monitoring. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5470–5473. DOI: 10.1109/IGARSS.2017.8128242. ISSN: 2153-7003.

Gaetano, R., Dupuy, S., Lebourgeois, V., Le Maire, G., Tran, A., Jolivot, A., and Bégué, A. (2019). The MORINGA Processing Chain: Automatic Object-based Land Cover Classification of Tropical Agrosystems using Multi-Sensor Satellite Imagery.

Gafsi, M., Dugué, P., Jamin, J.-Y., and Brossier, J. (2007). *Exploitations agricoles familiales en Afrique de l'Ouest et du Centre - Enjeux, caractéristiques et éléments de gestion.* Versaille, quae edition.

Gamba, P. and Chanussot, J. (2008). Foreword to the Special Issue on Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1283–1288, DOI: 10.1109/TGRS.2008.919761.

Gao, B.-c. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257–266, DOI: 10.1016/S0034-4257(96)00067-3.

Gardner, M. (1970). Mathematical Games - The fantastic combinations of John Conway's new solitaire game "life" - M. Gardner - 1970. *Scientific American*, (223):120–123.

Garedew, E., Sandewall, M., and Soderberg, U. (2012). A Dynamic Simulation Model of Land-Use, Population, and Rural Livelihoods in the Central Rift Valley of Ethiopia. *Environmental Management*, 49(1):151–162, DOI: 10.1007/s00267-011-9783-4.

Gasser, L., Braganza, C., and Herman, N. (1987). IMPLEMENTING DISTRIBUTED AI SYSTEMS USING MACE. *Unknown Host Publication Title*, pages 315–320. Publisher: IEEE.

Gaucherel, C., Boudon, F., Houet, T., Castets, M., and Godin, C. (2012). Understanding Patchy Landscape Dynamics: Towards a Landscape Language. *PLOS ONE*, 7(9):e46064, DOI: 10.1371/journal.pone.0046064. Publisher: Public Library of Science.

GBAD and INSTAT (2013). Dynamique de la pauvreté à Madagascar de 2005 à 2013. Technical report, Groupe de la Banque Agricaine de Développement.

Ghosh, A., Fassnacht, F. E., Joshi, P. K., and Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, 26:49–63, DOI: 10.1016/j.jag.2013.05.017.

Gibson, C. C., Ostrom, E., and Ahn, T. (2000). The concept of scale and the human dimensions of global change: a survey. *Ecological Economics*, 32(2):217–239, DOI: 10.1016/S0921-8009(99)00092-0.

Giordano, S. (2015). *Démélange d'images radar polarimétrique par séparation thématique de sources*. PhD thesis, Université Paris-Est, Paris.

Goodin, D. G., Anibas, K. L., and Bezymennyi, M. (2015). Mapping land cover and land use from object-based classification: an example from a complex agricultural landscape. *International Journal of Remote Sensing*, 36(18):4702–4723, DOI: 10.1080/01431161.2015.1088674.

Goudet, M. (2003). Antsampanimahazo: Caractéristiques agraires d'un territoire villageois des Hautes Terres malgaches et conditions d'adoption des systèmes de culture à base de couverture végétale. Technical report.

Grau, H. R., Gasparri, N. I., and Aide, T. M. (2005). Agriculture expansion and deforestation in seasonally dry forests of north-west Argentina. *Environmental Conservation*, 32(2):140–148, DOI: 10.1017/S0376892905002092.

Griffiths, P., Nendel, C., and Hostert, P. (2019). Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, 220:135–151, DOI: 10.1016/j.rse.2018.10.031.

Grillo, J. and Holt, J. (2009). L'Application de Cartes de Zone et de Profils de Moyens d'Existence dans les Analyses de la Sécurité Alimentaire et d'Alerte Précoce.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., and Cresson, R. (2017). Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, 2(1), DOI: 10.1186/s40965-017-0031-6.

GSDM and Cirad (n.d.). Les principales productions agricoles à Madagascar. État des lieux. Technical Report 7.

GSDM and CIRAD (n.d.). Pratiques de l'agricuture de conservation, Effets agronomiques et socio-économiques à long terme. Technical Report 5.

GTDR and PADR (2007). Programme Régional de Développement Rural. GTDR Vakinankaratra. Technical report.

Guetzkow, H. (1962). *Simulation in social science: readings*. Prentice-Hall, Englewood Cliffs, N.J.

Gómez, C., White, J. C., and Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116:55–72, DOI: 10.1016/j.isprsjprs.2016.03.008.

Górecki, T. and Łuczak, M. (2013). Using derivatives in time series classification. *Data Mining and Knowledge Discovery*, 26(2):310–331, DOI: 10.1007/s10618-012-0251-4.

Górecki, T. and Łuczak, M. (2015). Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications*, 42(5):2305–2312, DOI: 10.1016/j.eswa.2014.11.007.

Hadavand, A., Saadatseresht, M., and Homayouni, S. (2017). Segmentation parameter selection for object-based land-cover mapping from ultra high resolution spectral and elevation data. *International Journal of Remote Sensing*, 38(12):3586–3607, DOI: 10.1080/01431161.2017.1302107. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/01431161.2017.1302107.

Hagolle, O., Huc, M., Desjardins, C., Auer, S., and Richter, R. (2017). MAJA Algorithm Theoretical Basis Document. Technical report, CNES.

Hajj, M. E., Bégué, A., Guillaume, S., and Martiné, J.-F. (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, DOI: 10.1016/j.rse.2009.04.009.

Hall-Beyer, M. (2017). Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, 38(5):1312–1338, DOI: 10.1080/01431161.2016.1278314.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160):850–853, DOI: 10.1126/science.1244693. Publisher: American Association for the Advancement of Science Section: Report.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceeding of the IEEE*, 67(5):786–804, DOI: 10.1109/PROC.1979.11328.

Haralick, R. M. and Shapiro, L. G. (1985). Image Segmentation Techniques. *COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING*, 29:100–132.

Harvey, C. A., Rakotobe, Z. L., Rao, N. S., Dave, R., Razafimahatratra, H., Rabarijohn, R. H., Rajaofara, H., and MacKinnon, J. L. (2014). Extreme vulnerability of smallholder farmers to agricultural risks and climate change in Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1639):20130089, DOI: 10.1098/rstb.2013.0089.

Hauswirth, D., Pham, T. S., Wery, J., Tittonell, P., Jourdain, D., and Affholder, F. (2015). Exploiting farm typologies for designing conservation agriculture systems: a case study in northern Vietnam. *Cahiers Agricultures*, 24(2):102–112, DOI: 10.1684/agr.2015.0744.

Hewitt, C. (1977). Viewing control structures as patterns of passing messages. *Artificial Intelligence*, 8(3):323–364, DOI: 10.1016/0004-3702(77)90033-9.

HLPE (2013). Paysans et entrepreneurs: investir dans l'agriculture des petits exploitants pour la sécurité alimentaire. Technical Report 6, FAO, Rome.

HLPE (2017). Nutrition and food systems. Technical report, Rome.

Hoeschele, W. (2000). Geographic Information Engineering and Social Ground Truth in Attappadi, Kerala State, India. *Annals of the Association of American Geographers*, 90(2):293–321, DOI: 10.1111/0004-5608.00196.

Holland, M. B., Shamer, S. Z., Imbach, P., Zamora, J. C., Medellin Moreno, C., Hidalgo, E. J. L., Donatti, C. I., Martínez-Rodríguez, M. R., and Harvey, C. A. (2017). Mapping adaptive capacity and smallholder agriculture: applying expert knowledge at the landscape scale. *Climatic Change*, 141(1):139–153, DOI: 10.1007/s10584-016-1810-2.

Holzmann, P., Boudreau, T., Holt, J., Lawrence, M., and O'Donnell, M. (2008). *The Household Economy Approach: A guide for programme planners and policy-makers.* FEG Consulting, Save the Childrn.

Holzmann, R. and Jørgensen, S. (2001). Social Risk Management: A New Conceptual Framework for Social Protection, and Beyond. *International Tax and Public Finance*, 8:529–556, DOI: https://doi.org/10.1023/A:1011247814590.

Hong, T. and Rosenfeld, A. (1984). Compact Region Extraction Using Weighted Pixel Linking in a Pyramid. *IEEE Transactions on Pattern Analysis et Machine Intelligence*, 6(02):222–229, DOI: 10.1109/TPAMI.1984.4767505.

Hossain, M. D. and Chen, D. (2019). Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:115–134, DOI: 10.1016/j.isprsjprs.2019.02.009.

Houet, T., Schaller, N., Castets, M., and Gaucherel, C. (2014). Improving the simulation of fine-resolution landscape changes by coupling top-down and bottom-up land use and cover changes rules. *International Journal of Geographical Information Science*, 28(9):1848–1876, DOI: 10.1080/13658816.2014.900775. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/13658816.2014.900775.

Husson, O., Charpentier, H., Naudin, K., Moussa, N., Michellon, R., Andrianasolo, H., Razanamparany, C., Rakotoarinivo, C., Enjalric, F., and Seguy, L. (2009). *Le choix des itinéraires techniques*, volume 2.

Ienco, D., Interdonato, R., Gaetano, R., and Ho Tong Minh, D. (2019). Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:11–22, DOI: 10.1016/j.isprsjprs.2019.09.016.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., and Koetz, B. (2015). Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sensing*, 7(9):12356–12379, DOI: 10.3390/rs70912356.

Inglada, J., Vincent, A., Arias, M., and Marais-Sicre, C. (2016). Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series. *Remote Sensing*, 8(5):362, DOI: 10.3390/rs8050362.

INSTAT (2011). Enquête Périodique auprès des Ménages 2010. Rapport Principal EPM-2010, INSTAT, Antananarivo.

INSTAT (2019). Troisème recensement général de la population et de l'habitation. Résultats provisoires. Technical Report RGPH-3, INSTAT, Antananarivo.

Interdonato, R., Ienco, D., Gaetano, R., and Ose, K. (2019). DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:91–104, DOI: 10.1016/j.isprsjprs.2019.01.011.

IPCC (2019). *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. Intergovernmental Panel on Climate Change.

Irmischer, I. J. and Clarke, K. C. (2018). Measuring and modeling the speed of human navigation. *Cartography and Geographic Information Science*, 45(2):177–186, DOI: 10.1080/15230406.2017.1292150.

Islam, K., Rahman, M. F., and Jashimuddin, M. (2018). Modeling land use change using Cellular Automata and Artificial Neural Network: The case of Chunati Wildlife Sanctuary, Bangladesh. *Ecological Indicators*, 88:439–453, DOI: 10.1016/j.ecolind.2018.01.047.

Jahel, C. (2016). *Analyse des dynamiques des agroécosystèmes par modélisation spatialisée et utilisation d'images satellitaires, Cas d'étude de l'ouest du Burkina Faso*. PhD thesis, AgroParisTech, Montpellier, France.

Jahel, C., Augusseau, X., and Lo Seen, D. (2018). Modelling cropping plan strategies: What decision margin for farmers in Burkina Faso? *Agricultural Systems*, 167:17–33, DOI: 10.1016/j.agsy.2018.08.004.

Janicot, S., Caniaux, G., Chauvin, F., de Coëtlogon, G., Fontaine, B., Hall, N., Kiladis, G., Lafore, J.-P., Lavaysse, C., Lavender, S. L., Leroux, S., Marteau, R., Mounier, F., Philippon, N., Roehrig, R., Sultan, B., and Taylor, C. M. (2011). Intraseasonal variability of the West African monsoon. *Atmospheric Science Letters*, 12(1):58–66, DOI: 10.1002/asl.280.

Janssen, L. L. F. and Middelkoop, H. (1992). Knowledge-based crop classification of a Landsat Thematic Mapper image. *International Journal of Remote Sensing*, 13(15):2827–2837, DOI: 10.1080/01431169208904084. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/01431169208904084.

Jayne, T. S., Chamberlin, J., Traub, L., Sitko, N., Muyanga, M., Yeboah, F. K., Anseeuw, W., Chapoto, A., Wineman, A., Nkonde, C., and Kachule, R. (2016). Africa's changing farm size distribution patterns: the rise of medium-scale farms. *Agricultural Economics*, page 18.

Jiao, L., Sun, W., Yang, G., Ren, G., and Liu, Y. (2019). A Hierarchical Classification Framework of Satellite Multispectral/Hyperspectral Images for Mapping Coastal Wetlands. *Remote Sensing*, 11(19):2238, DOI: 10.3390/rs11192238. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Jouve, P. (2006). La dimension spatiale des systèmes de culture : comparaison entre agriculture tempérée et agriculture tropicale. *Cahiers Agricultures*, 15(3):255–260 (1). Number: 3.

JPL, N. (2014). U.S. Releases Enhanced Shuttle Land Elevation Data. NASA Jet Propulsion Laboratory (JPL).

Jurgens, C. (1997). The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. *International Journal of Remote Sensing*, 18(17):3583–3594, DOI: 10.1080/014311697216810.

Kelly (Letcher), R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., Hamilton, S. H., Henriksen, H. J., Kuikka, S., Maier, H. R., Rizzoli, A. E., van Delden, H., and Voinov, A. A. (2013). Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling et Software*, 47:159–181, DOI: 10.1016/j.envsoft.2013.05.005.

Kim, M., Madden, M., and Warner, T. A. (2009). Forest Type Mapping using Object-specific Texture Measures from Multispectral Ikonos Imagery: Segmentation Quality and Image Classification Issues. *Photogrammetric Engineering et Remote Sensing*, 75(7):819–829.

Kozak, J., Gimmi, U., Houet, T., and Bolliger, J. (2017). Current practices and challenges for modelling past and future land use and land cover changes in mountainous regions. *Regional Environmental Change*, 17(8):2187–2191, DOI: 10.1007/s10113-017-1217-2.

Kuivanen, K., Alvarez, S., Michalscheck, M., Adjei-Nsiah, S., Descheemaeker, K., Mellon-Bedi, S., and Groot, J. (2016). Characterising the diversity of smallholder farming systems and their constraints and opportunities for innovation: A case study from the Northern Region, Ghana. *NJAS - Wageningen Journal of Life Sciences*, 78:153–166, DOI: 10.1016/j.njas.2016.04.003.

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 14(5):778–782, DOI: 10.1109/LGRS.2017.2681128. Conference Name: IEEE Geoscience and Remote Sensing Letters.

Labatut, V. and Cherifi, H. (2011). Accuracy Measures for the Comparison of Classifiers. *The 5th Inter-national Conference on Information Technology*, page 5.

Lam, L. and Suen, S. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 27(5):553–568, DOI: 10.1109/3468.618255.

Lambert, M.-J., Traoré, P. C. S., Blaes, X., Baret, P., and Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, 216:647–657, DOI: 10.1016/j.rse.2018.06.036.

Landmann, T., Eidmann, D., Cornish, N., Franke, J., and Siebert, S. (2019). Optimizing harmonics from Landsat time series data: the case of mapping rainfed and irrigated agriculture in Zimbabwe. *Remote Sensing Letters*, 10(11):1038–1046, DOI: 10.1080/2150704X.2019.1648901. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/2150704X.2019.1648901.

Langlois, A. and Phipps, M. (1997). *Automates cellulaires: Application à la simulation urbaine*. Lavloisier.

Laniak, G. F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P. W., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., and Hughes, A. (2013). Integrated environmental modeling : a vision and roadmap for the future. *Environmental modelling et software*, 39:3–23, DOI: 10.1016/j.envsoft.2012.09.006,. Publisher: Elsevier.

Larson, D. F., Matsumoto, T., Otsuka, K., and Kilic, T. (2012). Should African rural development strategies depend on smallholder farms ? an exploration of the inverse productivity hypothesis. Technical Report 6190, The World Bank. Publication Title: Policy Research Working Paper Series.

Lassalle, P. (2015). *Etude du passage à l'échelle des algorithmes de segmentation et de classification en télédétection pour le traitement de volumes massifs de données*. PhD thesis, Université de Toulouse, UMR CESBIO.

Le Moigne, J.-L. (1977). *La Théorie du Système Général, Théorie de la Modélisation*. PUF.

Lebourgeois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S., and Bégué, A. (2017). A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sensing*, 9(3):259, DOI: 10.3390/rs9030259.

Leenhardt, D., Angevin, F., Biarnès, A., Colbach, N., and Mignolet, C. (2010). Describing and locating cropping systems on a regional scale. A review. *Agronomy for Sustainable Development*, 30(1):131–138, DOI: 10.1051/agro/2009002.

Lefebvre, A. (2011). *Contribution de la texture pour l'analyse d'images à très haute résolution spatiale: application à la détection de changement en milieu périurbain*. PhD thesis, Rennes 2 Haute-Bretagne.

Leite, J. G. D. B., Silva, J. V., and van Ittersum, M. K. (2014). Integrated assessment of biodiesel policies aimed at family farms in Brazil. *Agricultural Systems*, 131:64–76, DOI: 10.1016/j.agsy.2014.08.004.

Lelong, C. C., Tshingomba, U. K., and Soti, V. (2020). Assessing Worldview-3 multispectral imaging abilities to map the tree diversity in semi-arid parklands. *International Journal of Applied Earth Observation and Geoinformation*, 93:102211, DOI: 10.1016/j.jag.2020.102211.

Leroux, L., Jolivot, A., Bégué, A., Seen, D. L., and Zoungrana, B. (2014). How Reliable is the MODIS Land Cover Product for Crop Mapping Sub-Saharan Agricultural Landscapes? *Remote Sensing*, 6(9):8541–8564, DOI: 10.3390/rs6098541. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Levinshtein, A., Stere, A., Kutulakos, K., Fleet, D., Dickinson, S., and Siddiqi, K. (2009). TurboPixels: Fast Superpixels Using Geometric Flows. *IEEE transactions on pattern analysis and machine intelligence*, 31:2290–7, DOI: 10.1109/TPAMI.2009.96.

Li, M., Ma, L., Blaschke, T., Cheng, L., and Tiede, D. (2016). A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *International Journal of Applied Earth Observation and Geoinformation*, 49:87–98, DOI: 10.1016/j.jag.2016.01.011.

Lipton, M. (2005). *The family farm in a globalizing world: the role of crop science in alleviating poverty*. Number 40 in 2020 discussion paper. Internat. Food Policy Research Inst, Washington, DC. OCLC: 255197512.

Lobell, D. B. and Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2(1):014002, DOI: 10.1088/1748-9326/2/1/014002.

Losch, B., Fréguin-Gresh, S., and White, E. T. (2013). Transformations rurales et developpement: Les defis du changement structurel dans un monde globalise. OCLC: 1091716501.

Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice*. PhD thesis, Université de Liège, Liège. arXiv: 1407.7502.

Lowder, S. K., Skoet, J., and Raney, T. (2016). The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development*, 87:16–29, DOI: 10.1016/j.worlddev.2015.10.041.

Lu, D. and Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, DOI: 10.1080/01431160600746456. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/01431160600746456.

Lévy, J. and Lussault, M. (2013). *Dictionnaire de géographie et de l'espace des sociétés*. Pouvoirs Locaux. Belin.

Löw, F., Schorcht, G., Michel, U., Dech, S., and Conrad, C. (2012). Per-field crop classification in irrigated agricultural regions in middle Asia using random forest and support vector machine ensemble. *Earth Resources and Environmental Remote Sensing/GIS Applications III*, Proc. SPIE 8538, DOI: 10.1117/12.974588. Conference Name: Earth Resources and Environmental Remote Sensing/GIS Applications III ISBN: 9780819492784.

Macron, C., Richard, Y., Garot, T., Bessafi, M., Pohl, B., Ratiarison, A., and Razafindrabe, A. (2016). Intraseasonal Rainfall Variability over Madagascar. *Monthly Weather Review*, 144(5):1877–1885, DOI: 10.1175/MWR-D-15-0077.1.

MAEP (2004). Filière maïs. Technical Report 108, Filières de l'Agriculture, de l'Elevage et de la Pêche, et Actions du Ministère de l'Agriculture, de l'Elevage et de la Pêche.

MAEP (2007). Monographie de la région de Vakinankaratra. Technical report.

Mangiarotti, S., Sharma, A., Corgne, S., Hubert-Moy, L., Ruiz, L., Sekhar, M., and Kerr, Y. (2018). Can the global modeling technique be used for crop classification? *Chaos, Solitons et Fractals*, 106:363–378, DOI: 10.1016/j.chaos.2017.12.003.

Manson, S. M. (2005). Agent-based modeling and genetic programming for modeling land change in the Southern Yucatán Peninsular Region of Mexico. *Agriculture, Ecosystems et Environment*, 111(1):47–62, DOI: 10.1016/j.agee.2005.04.024.

Matton, N., Canto, G. S., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., and Defourny, P. (2015). An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series. *Remote Sensing*, 7(10):13208–13232, DOI: 10.3390/rs71013208. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Matus, S. L. S., Cimpoies, D., and Ronzon, T. (2013). Typology and Indicators to Characterize Agricultural Holdings for Improved Policy Formulation. A World Agricultures Watch report, FAO, IFAD, CIRAD, République Française.

Maxwell, T. and Costanza, R. (1997). A language for modular spatio-temporal simulation. *Ecological Modelling*, 103(2):105–113, DOI: 10.1016/S0304-3800(97)00103-8.

McCarty, J. L., Neigh, C. S. R., Carroll, M. L., and Wooten, M. R. (2017). Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery. *Remote Sensing of Environment*, 202:142–151, DOI: 10.1016/j.rse.2017.06.040.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, DOI: 10.1007/BF02295996.

Meinel, G. and Neubert, M. (2004). A comparison of segmentation programs for high resolution remote sensing data. In *International Archives of Photogrammetry and Remote Sensing*, volume 35, pages 1097–1105, Istanbul, Turkey.

Meiyappan, P., Dalton, M., O'Neill, B. C., and Jain, A. K. (2014). Spatial modeling of agricultural land use change at global scale. *Ecological Modelling*, 291:152–174, DOI: 10.1016/j.ecolmodel.2014.07.027.

Mertens, B. and Lambin, E. F. (2000). Land-Cover-Change Trajectories in Southern Cameroon. *Annals of the Association of American Geographers*, 90(3):467–494, DOI: 10.1111/0004-5608.00205.

Mladenoff, D. J. (2004). LANDIS and forest landscape models. *Ecological Modelling*, 180(1):7–19, DOI: 10.1016/j.ecolmodel.2004.03.016.

Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the National Academy of Sciences*, 104(50):19680–19685, DOI: 10.1073/pnas.0701855104.

Muddarisna, N., Yuniwati, E. D., Masruroh, H., and Oktaviansyah, A. R. (2020). An Automated Approach Using Topographic Position Index (TPI) for Landform Mapping (Case Study: Gede Watershed, Malang Regency, East Java, Indonesia). *IOP Conference Series: Earth and Environmental Science*, 412:012027, DOI: 10.1088/1755-1315/412/1/012027.

Muetzelfeldt, R. and Massheder, J. (2003). The Simile visual modelling environment. *European Journal of Agronomy*, 18(3-4):345–358, DOI: 10.1016/S1161-0301(02)00112-0.

Muller, D. and Zeller, M. (2005). Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. *Agricultural Economics*, 27(3):333–354, DOI: 10.1111/j.1574-0862.2002.tb00124.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1574-0862.2002.tb00124.x.

Müller-Hansen, F., Schlüter, M., Mäs, M., Donges, J. F., Kolb, J. J., Thonicke, K., and Heitzig, J. (2017). Towards representing human behavior and decision making in Earth system models – an overview of techniques and approaches. *Earth System Dynamics*, 8(4):977–1007, DOI: https://doi.org/10.5194/esd-8-977-2017. Publisher: Copernicus GmbH.

Nachtergaele, F. and Biancalani, R. (2013). *Mapping Land Use Systems at global and regional scales for Land Degradation Assessment Analysis*. FAO, Rome, land degradation assessment in drylands edition.

Nassor, A. and Jury, M. R. (1997). Intra-seasonal climate variability of Madagascar. Part 2: Evolution of flood events. *Meteorology and Atmospheric Physics*, 64(3-4):243–254, DOI: 10.1007/BF01029696.

Nassor, A. and Jury, M. R. (1998). Intra-seasonal climate variability of Madagascar. Part 1: Mean summer conditions. *Meteorology and Atmospheric Physics*, 65(1-2):31–41, DOI: 10.1007/BF01030267.

Nations, U. (2020). *World Population Prospects 2019 - Volume II: Demographic Profiles*. UN, DOI: 10.18356/7707d011-en.

Negoita, C. (1985). Expert systems and fuzzy systems. Publisher: Benjamin-Cummings Pub. Co.,Menlo Park, CA.

Neigh, C. S., Carroll, M. L., Wooten, M. R., McCarty, J. L., Powell, B. F., Husak, G. J., Enenkel, M., and Hain, C. R. (2018). Smallholder crop area mapped with wall-to-wall WorldView sub-meter panchromatic image texture: A test case for Tigray, Ethiopia. *Remote Sensing of Environment*, 212:8–20, DOI: 10.1016/j.rse.2018.04.025.

Neuwirth, C., Hofer, B., and Schaumberger, A. (2016). Object view in spatial system dynamics: a grassland farming example. *Journal of Spatial Science*, 61(2):367–388, DOI: 10.1080/14498596.2015.1132641. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/14498596.2015.1132641.

Neuwirth, C., Peck, A., and Simonović, S. P. (2015). Modeling structural change in spatial system dynamics: A Daisyworld example. *Environmental Modelling et Software*, 65:30–40, DOI: 10.1016/j.envsoft.2014.11.026.

Novelli, A., Aguilar, M. A., Nemmaoui, A., Aguilar, F. J., and Tarantino, E. (2016). Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almería (Spain). *International Journal of Applied Earth Observation and Geoinformation*, 52:403–411, DOI: 10.1016/j.jag.2016.07.011.

ONE (2013). Evolution de la Couverture de forêts naturelles à Madagascar 2005-2010. Technical report, Office National pour l'Environnement, Antananarivo.

Oppelt, N., Scheiber, R., Gege, P., Wegmann, M., Taubenboeck, H., Berger, M., Scheiber, R., Gege, P., Wegmann, M., Taubenboeck, H., and Berger, M. (2015). Fundamentals of Remote Sensing for Terrestrial Applications: Evolution, Current State of the Art, and Future Possibilities. In Thenkabail, P. S., editor, *Remote Sensing Handbook - Three Volume Set*, volume 1. CRC Press, DOI: 10.1201/b19355-10. Pages: 95-118 Publisher: CRC Press.

Ortigosa-Hernández, J., Inza, I., and Lozano, J. A. (2017). Measuring the class-imbalance extent of multi-class problems. *Pattern Recognition Letters*, 98:32–38, DOI: 10.1016/j.patrec.2017.08.002.

Osman, J., Inglada, J., and Dejoux, J.-F. (2015). Assessment of a Markov logic model of crop rotations for early crop mapping. *Computers and Electronics in Agriculture*, 113:234–243, DOI: 10.1016/j.compag.2015.02.015.

Pacifici, F., Chini, M., and Emery, W. J. (2009). A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, 113(6):1276–1292, DOI: 10.1016/j.rse.2009.02.014.

Pal, N. R. and Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, DOI: 10.1016/0031-3203(93)90135-J.

Panda, S. S., Rao, M. N., Thenkabail, P. S., and Fitzerald, J. E. (2015). Remote Sensing Systems—Platforms and Sensors: Aerial, Satellite, UAV, Optical, Radar, and LiDAR. In Thenkabail, P. S., editor, *Remote Sensing Handbook - Three Volume Set*, volume 1. CRC Press, DOI: 10.1201/b19355-8. Pages: 37-92 Publisher: CRC Press.

Parihar, J. S., Justice, C., Soares, J., Leo, O., Kosuth, P., Jarvis, I., Williams, D., Bingfang, W., Latham, J., and Becker-Reshef, I. (2012). GEO-GLAM: A GEOSS-G20 initiative on Global Agricultural Monitoring. In *39th COSPAR Scientific Assembly*, volume 39, page 1451, Mysore, India.

Persello, C., Tolpekin, V., Bergado, J., and de By, R. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 231:111253, DOI: 10.1016/j.rse.2019.111253.

Petitjean, F., Inglada, J., and Gancarski, P. (2012). Satellite Image Time Series Analysis Under Time Warping. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3081–3095, DOI: 10.1109/TGRS.2011.2179050.

Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., and Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6):1301–1316, DOI: 10.1016/j.rse.2011.01.009.

PNUD (2019). Rapport sur le développement humain 2019. Au-delà des revenus, des moyennes et du temps présent : les inégalités de développement humain au XXIe siècle. Technical report, PNUD.

Pontius, R. G. and Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15):4407–4429, DOI: 10.1080/01431161.2011.552923.

QGIS.org (2020). QGIS Geographic Information System. Open Source Geospatial Foundation Project.

Qiu, J., Tang, H., Frolking, S., Boles, S., Li, C., Xiao, X., Liu, J., Zhuang, Y., and Qin, X. (2003). Mapping Single-, Double-, and Triple-crop Agriculture in China at 0.5° × 0.5° by Combining County-scale Census Data with a Remote Sensing-derived Land Cover Map. *Geocarto International*, 18(2):3–13, DOI: 10.1080/10106040308542268. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/10106040308542268.

Rabiatsarafara, P. (2006). Système de culture sous couverture végétale et Agriculture traditionnelle : Cas de la région de Vakinankaratra.

Raharison, T., Razafimahatratra, H. M., Bélières, J.-F., and Autfray, P. (2018). Mieux connaitre la diversite des exploitations agricoles et leurs modes de fonctionnement...un élément indispensable pour orienter les actions de développement. page 12.

Rakotofiringa, A. and Tokarski, Y. (2007). Caractérisation des exploitations agricoles dans la commune rurale d'Andranomanelatra Région Vakinankaratra, hauts plateaux de Madagascar.

Randrianarison, N. (2006). Diagnostic agraire et msie au point d'une méthodologoie de suivi et d'analyse des succès et abandons des sysètmes à base de smis direct sous couerture végétale (SCV) : cas du Fokontany d'Antsapanimahazo - Madagascar.

Randrianarison, N. (2013). *Processus d'innovationet conditions de maintien desagriculteurs : la diffusion des innovations dans la paysannerie malgache, cas de la région du lac Alaotra.* PhD thesis, Université de Montpellier 1.

Rasmussen, L. V., Rasmussen, K., Reenberg, A., and Proud, S. (2012). A system dynamics approach to land use changes in agro-pastoral systems on the desert margins of Sahel. *Agricultural Systems*, 107:56–64, DOI: 10.1016/j.agsy.2011.12.002.

Raunet (1997). *Les ensembles morphopedologiques de Madagascar*. CIRAD, fofifa,cirad,anae edition.

Reardon, T. and Vosti, S. A. (1995). Links between rural poverty and the environment in developing countries: Asset categories and investment poverty. *World Development*, 23(9):1495–1506, DOI: 10.1016/0305-750X(95)00061-G.

Regniers, O. (2014). *Méthodes d'analyse de texture pour la cartographie d'occupations du sol par télédetection très haute résolution : application à la fôret, la vigne et les parcs ostréicoles*. PhD thesis, Université de Boredeaux.

Rembold, F., Meroni, M., Urbano, F., Csak, G., Kerdiles, H., Perez-Hoyos, A., Lemoine, G., Leo, O., and Negre, T. (2019). ASAP: A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis. *Agricultural Systems*, 168:247–257, DOI: 10.1016/j.agsy.2018.07.002.

Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., and Chookolingo, B. (2018). How much of the world's food do smallholders produce? *Global Food Security*, 17:64–72, DOI: 10.1016/j.gfs.2018.05.002.

Richards-Rissetto, H. and Landau, K. (2014). Movement as a means of social (re)production: using GIS to measure social integration across urban landscapes. *Journal of Archaeological Science*, 41:365–375, DOI: 10.1016/j.jas.2013.08.006.

Riquier, J. (1968). Carte pédologique de Madagascar. One mosaic of 3 map sheets. Scale of 1:1 000 000.

ROR, FIFAMANOR, and FAMPITA (2003). Résultats d'enquêtes des ménages ruraux de l'obervatoire rural d'Antsirabe. page 86.

Rouse, W., Haas, R. H., Schell, J., and Deering, D. (1974). Monitionring Vegetation Systems in the Great Plains with ERTS. *Third Earth Resources Technology Satellite- 1 Symposium*, pages 309–317.

Roy Chowdhury, R. (2006). Landscape change in the Calakmul Biosphere Reserve, Mexico: Modeling the driving forces of smallholder deforestation in land parcels. *Applied Geography*, 26(2):129–152, DOI: 10.1016/j.apgeog.2005.11.004.

Roy Chowdhury, R. (2010). Differentiation and concordance in smallholder land use strategies in southern Mexico's conservation frontier. *Proceedings of the National Academy of Sciences*, 107(13):5780–5785, DOI: 10.1073/pnas.0905892107.

Ruas, A. and Sanders, L. (2015). Regards croisés sur la modélisation des dynamiques spatiales. *Revue Internationale de Géomatique*, 25(3):275–300, DOI: 10.3166/RIG.25.275-300. Number: 3 Publisher: Lavoisier SAS.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, DOI: 10.1109/TASSP.1978.1163055. Conference Name: IEEE Transactions on Acoustics, Speech, and Signal Processing.

Sall, N., Diop, P. A., and Barbedette, L. (2010). Comment les exploitations familiales peuvent-elles nourrir le Sénégal ? Évaluation de la portée stratégique de la problé- matique de la productivité des exploitations familiales. Synthèse d'étape, Fédération des organisations non gouvernementales du Sénégal.

Saltelli, A., editor (2008). *Global sensitivity analysis: the primer*. John Wiley, Chichester, England ; Hoboken, NJ. OCLC: ocn180852094.

Salton, G. and Buckley, C. (1991). Global text matching for information retrieval. *Science (New York, N.Y.)*, 253(5023):1012–1015, DOI: 10.1126/science.253.5023.1012.

Sanders, P. and Sanders, F. (2004). Spatial urban dynamics. A vision on the future of urban dynamics:Forrester revisited. *International System Dynamics Conference*, page 32.

Sandker, M., Puntodewo, A., Sitorus, F., Purnomo, H., Yumte, Y., Ruiz-Pérez, M., and Campbell, B. M. (2010). Spatial Projections of Participatory System Dynamics Modeling Outcomes: Exploring Oil Palm and REDD consequences for Local Livelihoods in Papua, Indonesia. In *LAND-MOD2010*, page 15, Montpellier, France. Symposcience.

Sathyan, A., Funk, C., Aenis, T., and Breuer, L. (2018). Climate Vulnerability in Rainfed Farming: Analysis from Indian Watersheds. *Sustainability*, 10(9):3357, DOI: 10.3390/su10093357.

Schiewe, J. (2002). Segmentation of high resolution remotely sensed data-concepts, applications and problems. In *International Archives of Photogrammetry Remote SensingAnd Spatial Information Sciences*, volume 34, pages 380–385, Ottawa.

Schneider, L. C. and Gil Pontius, R. (2001). Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems et Environment*, 85(1):83–94, DOI: 10.1016/S0167-8809(01)00189-X.

Schoneveld, G. C., van der Haar, S., Ekowati, D., Andrianto, A., Komarudin, H., Okarda, B., Jelsma, I., and Pacheco, P. (2019). Certification, good agricultural practice and smallholder heterogeneity: Differentiated pathways for resolving compliance gaps in the Indonesian oil palm sector. *Global Environmental Change*, 57:101933, DOI: 10.1016/j.gloenvcha.2019.101933.

SCRID (2003). Variétés de riz pluvial pour les Hautes Terres de 1300 à 1800 m. Technical report.

Sebillotte, M. (1990). Systeme de culture, un concept operatoire pour les agronomes. In *Un Point sur... - INRA*, pages 165–196. INRA Editions, Paris, inra edition.

Sharma, A. K., Hubert-Moy, L., Sriramulu, B., Sekhar, M., Ruiz, L., Bandyopadhyay, S., Mohan, S., and Corgne, S. (2019). Evaluation of Radarsat-2 quad-pol SAR time-series images for monitoring groundwater irrigation. *International Journal of Digital Earth*, 12(10):1177–1197, DOI: 10.1080/17538947.2019.1604834.

Sherrill, K. R., Frakes, B., and Schupbach, S. (2010). Travel time cost surface model: standard operating procedure. Natural Resource Report NPS/NRPC/IMD/NRR 2010/238, ational Park Service, Fort Collins, Colorado.

Shiferaw, H., Bewket, W., Alamirew, T., Zeleke, G., Teketay, D., Bekele, K., Schaffner, U., and Eckert, S. (2019). Implications of land use/land cover dynamics and Prosopis invasion on ecosystem service values in Afar Region, Ethiopia. *Science of The Total Environment*, 675:354–366, DOI: 10.1016/j.scitotenv.2019.04.220.

Shubik, M. (1960). Bibliography on Simulation, Gaming, Artificial Intelligence and Allied Topics. *Journal of the American Statistical Association*, 55(292):736–751, DOI: 10.1080/01621459.1960.10483374. Publisher: Taylor et Francis.

Siebert, S., Portmann, F. T., and Döll, P. (2010). Global Patterns of Cropland Use Intensity. *Remote Sensing*, 2(7):1625–1643, DOI: 10.3390/rs2071625. Number: 7 Publisher: Molecular Diversity Preservation International.

Sivasankar, T., Kumar, D., Srivastava, H. S., and Patel, P. (2018). Advances in Radar Remote Sensing of Agricultural Crops: A Review. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4):1126, DOI: 10.18517/ijaseit.8.4.5797.

Skakun, S., Kussul, N., Shelestov, A. Y., Lavreniuk, M., and Kussul, O. (2016). Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8):3712–3719, DOI: 10.1109/JSTARS.2015.2454297. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Song, Q., Hu, Q., Zhou, Q., Hovis, C., Xiang, M., Tang, H., and Wu, W. (2017). In-Season Crop Mapping with GF-1/WFV Data by Combining Object-Based Image Analysis and Random Forest. *Remote Sensing*, 9(11):1184, DOI: 10.3390/rs9111184.

Sorel, L., Viaud, V., Durand, P., and Walter, C. (2010). Modeling spatio-temporal crop allocation patterns by a stochastic decision tree method, considering agronomic driving factors. *Agricultural Systems*, 103(9):647–655, DOI: 10.1016/j.agsy.2010.08.003.

Sourisseau, J.-M., Bélières, J.-F., Bonnal, P., Bosc, P.-M., Losch, B., and Marzin, J. (2014). Politiques publiques pour l'agriculture familiale. Caractériser pour mieux soutenir. *Perspective*, (4):4.

Sourisseau, J.-M., Rasolofo, P., Bélières, J.-F., Guengant, J.-P., Ramanitriniony, H. K., Bourgeois, R., Razafimiarantsoa, T. T., Andrianantoandro, V. T., Ramarijaona, M., Burnod, P., Rabeandriamaro, H., and Bougnoux, N. (2016). Diagnostic Territorial de la Région du Vakinankaratra à

Madagascar. Prospective territoriale sur les dynamiques démographiques et le développement rural en Afrique subsaharienne et à Madagascar, AFD.

Story, M. (1986). Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, 5(3):397–399.

Suzuki, K. and Zupanski, M. (2018). Coupled data assimilation in climate research: A brief review of applications in ocean and land. *Satellite Oceanography and Meteorology*, 3:doi:10.18063/som.v3i2.599, DOI: 10.18063/som.v3i2.599.

Sweeney, S., Ruseva, T., Estes, L., and Evans, T. (2015). Mapping Cropland in Smallholder-Dominated Savannas: Integrating Remote Sensing Techniques and Probabilistic Modeling. *Remote Sensing*, 7(11):15295–15317, DOI: 10.3390/rs71115295. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Tadross, M., Randriamarolaza, L., Rabefitia, Z., and Yip, Z. K. (2008). Climate change in Madagascar; recent past and future. Technical report, World Bank, Washington, DC.

Team, G. D. (2018). Geographic Resources Analysis Support System (GRASS) Software. @Manual{GRASS_GIS_software, title = {Geographic Resources Analysis Support System (GRASS GIS) Software}, author = {{GRASS Development Team}}, organization = {Open Source Geospatial Foundation}, address = {USA}, year = {2018}, url = {https://grass.osgeo.org}, }.

Teluguntla, P., Thenkabail, P. S., Xiong, J., Gumma, M. K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R. G., and Tilton, J. (2015). Global Food Security Support Analysis Data (GFSAD) at Nominal 1 km (GCAD) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities. In Thenkabail, P. S., editor, *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*, pages 131–160. CRC Press. Num Pages: 849.

Temme, A. and Verburg, P. (2011). Mapping and modelling of changes in agricultural intensity in Europe. *Agriculture, Ecosystems et Environment*, 140(1-2):46–56, DOI: 10.1016/j.agee.2010.11.010.

TFNAC (n.d.). Association Maïs et légumineuses en rotation avec le riz pluvial.

Thenkabail, Ph.D., P. S. (2015). *Remote Sensing of Water Resources, Disasters, and Urban Studies*. CRC Press, 1 edition, DOI: 10.1201/b19321.

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108(50):20260–20264, DOI: 10.1073/pnas.1116437108.

Tittonell, P. (2014). Livelihood strategies, resilience and transformability in African agroecosystems. *Agricultural Systems*, 126:3–14, DOI: 10.1016/j.agsy.2013.10.010.

Tobler, W. (1993). Three presentations on geographical analysis and modeling. Rapport technique 93-1, National Center for Geographic Information and Analysis.

Tonneau, J.-P., Begue, A., Leroux, L., Faure, J.-F., Mertens, B., Pinet, C., and Tomasini, L. (2020). Geospatial information for African agriculture: a key investment for agricultural policies. In *2020 World Bank Conference on Land and Poverty*, volume 21, page 19, Washington, DC. DOI: https://doi.org/10.19182/agritrop/00068.

Tonneau, J.-P., Bégué, A., Leroux, L., Augusseau, X., Faure, J.-F., Mertens, B., Pinet, C., and Tomasini, L. (2019). L'information géospatiale au service de l'agriculture africaine, un investissement essentiel pour les politiques agricoles. *Perspective*, (51):1–4, DOI: 10.19182/agritrop/00067. Number: 51.

Tropek, R., Sedláček, O., Beck, J., Keil, P., Musilová, Z., Šímová, I., and Storch, D. (2014). Comment on "High-resolution global maps of 21st-century forest cover change". *Science*, 344(6187):981–981, DOI: 10.1126/science.1248753. Publisher: American Association for the Advancement of Science Section: Technical Comments.

Tsangaratos, P., Loupasakis, C., Nikolakopoulos, K., Angelitsa, V., and Ilia, I. (2018). Developing a landslide susceptibility map based on remote sensing, fuzzy logic and expert knowledge of the Island of Lefkada, Greece. *Environmental Earth Sciences*, 77(10):363, DOI: 10.1007/s12665-018-7548-6.

USAID and NET, F. (2013). Madagascar, zones et descriptions de moyens d'existence. Technical report, USAID, FEWS NET.

Vakinankaratra, R. (n.d.). Rapport sur les données de la région Vakinankaratra.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., and Koetz, B. (2016). Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions. *Remote Sensing*, 8(1):55, DOI: 10.3390/rs8010055. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

van der Laan, C., Budiman, A., Verstegen, J., Dekker, S., Effendy, W., Faaij, A., Kusuma, A., and Verweij, P. (2018). Analyses of Land Cover Change Trajectories Leading to Tropical Forest Loss: Illustrated for the West Kutai and Mahakam Ulu Districts, East Kalimantan, Indonesia. *Land*, 7(3):108, DOI: 10.3390/land7030108.

van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., and Hochman, Z. (2013). Yield gap analysis with local to global relevance—A review. *Field Crops Research*, 143:4–17, DOI: 10.1016/j.fcr.2012.09.009.

Vannier, C. (2011). *Observation et modélisation spatiale de pratiques agricoles territorialisées à partir de données de télédétection: application au paysage bocager*. PhD thesis, Université Européenne de Bretagne.

Veldkamp, A. and Fresco, L. O. (1996). CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological Modelling*, 91(1):231–248, DOI: 10.1016/0304-3800(95)00158-1.

Verburg, P. H. (2006). Simulating feedbacks in land use and land cover change models. *Landscape Ecology*, 21(8):1171–1183, DOI: 10.1007/s10980-006-0029-4.

Verburg, P. H. and Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecology*, 24(9):1167, DOI: 10.1007/s10980-009-9355-7.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., and Mastura, S. S. A. (2002). Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management*, 30(3):391–405, DOI: 10.1007/s00267-002-2630-x.

Verburg, P. H. and Veldkamp, A. (2005). Introduction to the Special Issue on Spatial modeling to explore land use dynamics. *International Journal of Geographical Information Science*, 19(2):99–102, DOI: 10.1080/13658810410001713362. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/13658810410001713362.

Vieira, M. A., Formaggio, A. R., Rennó, C. D., Atzberger, C., Aguiar, D. A., and Mello, M. P. (2012). Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sensing of Environment*, 123:553–562, DOI: 10.1016/j.rse.2012.04.011.

Vintrou, E. (2012). *Cartographie et caractérisation des systèmes agricoles au Mali par télédétection à moyenne résolution spatiale*. PhD thesis, AgroParisTech, Montpellier, France.

Vintrou, E., Desbrosse, A., Bégué, A., Traoré, S., Baron, C., and Lo Seen, D. (2012a). Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):83–93, DOI: 10.1016/j.jag.2011.06.010.

Vintrou, E., Soumaré, M., Bernard, S., Bégué, A., Baron, C., and Lo Seen, D. (2012b). Mapping Fragmented Agricultural Systems in the Sudano-Sahelian Environments of Africa Using Random Forest and Ensemble Metrics of Coarse Resolution MODIS Imagery. *Photogrammetric Engineering et Remote Sensing*, 78(8):839–848, DOI: 10.14358/PERS.78.8.839.

Voiron, C. and Chéry, J.-P. (2005). Espace géographique, spatialisation et modélisation en dynamique des systèmes. *Congré Européen de Science es Systèmes*, (6):10.

von Bertalanffy, L. (1968). *General System theory : Foundations, Development, Applications*. New York, george braziller edition.

Von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. Champaign, IL, USA, university of illinois press edition.

Waldner, F., Bellemans, N., Hochman, Z., Newby, T., de Abelleyra, D., Verón, S. R., Bartalev, S., Lavreniuk, M., Kussul, N., Maire, G. L., Simoes, M., Skakun, S., and Defourny, P. (2019a). Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *International Journal of Applied Earth Observation and Geoinformation*, 80:82–93, DOI: 10.1016/j.jag.2019.01.002.

Waldner, F., Chen, Y., Lawes, R., and Hochman, Z. (2019b). Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sensing of Environment*, 233:111375, DOI: 10.1016/j.rse.2019.111375.

Waldner, F., Hansen, M. C., Potapov, P. V., Löw, F., Newby, T., Ferreira, S., and Defourny, P. (2017). National-scale cropland mapping based on spectral-temporal features and outdated land cover information. *PLOS ONE*, 12(8):e0181911, DOI: 10.1371/journal.pone.0181911. Publisher: Public Library of Science.

Wani, S. P., Rockström, J., Oweis, T. Y., and Ebrary, I. (2009). *Rainfed agriculture unlocking the potential*. CABI, Wallingford, Oxfordshire, UK, cabi edition. OCLC: 703213904.

Warrens, M. J. (2015). Properties of the quantity disagreement and the allocation disagreement. *International Journal of Remote Sensing*, 36(5):1439–1446, DOI: 10.1080/01431161.2015.1011794.

WB (2015). Diagnostic Systématique de Pays (SCD): Madagascar. Technical report, Banque Mondiale.

WB (2016). Transforming Vietnamese Agriculture: Gaining More from Less. Vietnam Development Report, Banque Mondiale, hanoi.

WB (2020). Madagascar : La Banque mondiale fournit 75 millions de dollars pour atténuer les impacts du COVID-19 et soutenir le redressement.

Weiss, A. D. (2001). Topographic Position and Landforms Analysis.

WFP (2020). COVID-19: Potential impact on the world's poorest people. A WFP analysis of the economic and food security implications of the pandemic. Technical report, World Food Programme, Rome.

Whitcraft, A. K., Becker-Reshef, I., and Justice, C. O. (2015a). Agricultural growing season calendars derived from MODIS surface reflectance. *International Journal of Digital Earth*, 8(3):173–197, DOI: 10.1080/17538947.2014.894147. Publisher: Taylor et Francis _eprint: https://doi.org/10.1080/17538947.2014.894147.

Whitcraft, A. K., Becker-Reshef, I., and Justice, C. O. (2015b). A Framework for Defining Spatially Explicit Earth Observation Requirements for a Global Agricultural Monitoring Initiative (GEOGLAM). *Remote Sensing*, 7(2):1461–1481, DOI: 10.3390/rs70201461. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Whitcraft, A. K., Vermote, E. F., Becker-Reshef, I., and Justice, C. O. (2015c). Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sensing of Environment*, 156:438–447, DOI: 10.1016/j.rse.2014.10.009.

White, D. A. and Barber, S. B. (2012). Geospatial modeling of pedestrian transportation networks: a case study from precolumbian Oaxaca, Mexico. *Journal of Archaeological Science*, 39(8):2684–2696, DOI: 10.1016/j.jas.2012.04.017.

White, S. H., Rey, A. M. d., and Sánchez, G. R. (2007). Modeling epidemics using cellular automata. *Applied Mathematics and Computation*, 186(1):193, DOI: 10.1016/j.amc.2006.06.126. Publisher: Elsevier.

Whiteside, T. and Ahmad, W. (2005). A COMPARISON OF OBJECT-ORIENTED AND PIXEL-BASED CLASSIFICATION METHODS FOR MAPPING LAND COVER IN NORTHERN AUSTRALIA.

Wickizer, V. D. (1960). The Smallholder in Tropical Export Crop Production. *Food Research Institute Studies*, 1(1):49–99.

Wiggins, S. (2009). Can the smallholder model deliver poverty reduction and food security for a rapidly growing population in Africa? *FAC Working Paper*, 8:25.

Woodcock, C. E. and Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3):311–332, DOI: 10.1016/0034-4257(87)90015-0.

Wu, B. and Li, Q. (2012). Crop planting and type proportion method for crop acreage estimation of complex agricultural landscapes. *International Journal of Applied Earth Observation and Geoinformation*, 16:101–112, DOI: 10.1016/j.jag.2011.12.006.

Wu, M., Peng, D., Qin, Y., Niu, Z., Yang, C., Li, W., Hao, P., and Zhang, C. (2018). An index of non-sampling error in area frame sampling based on remote sensing data. *PeerJ*, 6:e5824, DOI: 10.7717/peerj.5824. Publisher: PeerJ Inc.

Wyman, M. S. and Stein, T. V. (2010). Modeling social and land-use/land-cover change data to assess drivers of smallholder deforestation in Belize. *Applied Geography*, 30(3):329–342, DOI: 10.1016/j.apgeog.2009.10.001.

Xie, H., You, L., Wielgosz, B., and Ringler, C. (2014). Estimating the potential for expanding smallholder irrigation in Sub-Saharan Africa. *Agricultural Water Management*, 131:183–193, DOI: 10.1016/j.agwat.2013.08.011.

Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., Yadav, K., and Thau, D. (2017). Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126:225–244, DOI: 10.1016/j.isprsjprs.2017.01.019.

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14):3025–3033, DOI: 10.1080/01431160600589179.

Xu, X., Du, Z., and Zhang, H. (2016). Integrating the system dynamic and cellular automata models to predict land use and land cover change. *International Journal of Applied Earth Observation and Geoinformation*, 52:568–579, DOI: 10.1016/j.jag.2016.07.022.

Yao, H., Jia, X., Kumar, V., and Li, Z. (2020). Learning with Small Data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery et Data Mining*, KDD '20, pages 3539–3540, New York, NY, USA. Association for Computing Machinery, DOI: 10.1145/3394486.3406466.

You, L., Wood, S., Wood-Sichra, U., and Wu, W. (2014). Generating global crop distribution maps: From census to grid. *Agricultural Systems*, 127:53–60, DOI: 10.1016/j.agsy.2014.01.002.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28, DOI: 10.1016/0165-0114(78)90029-5.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., and Simonson, D. (2001). Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. *Soil Science Society of America Journal*, 65(5):1463–1472, DOI: https://doi.org/10.2136/sssaj2001.6551463x. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2001.6551463x.

Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., and Scambos, T. A. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, 224:382–385, DOI: 10.1016/j.rse.2019.02.016.

ANNEXES

Annexe A

Bandes spectrales

Satellite	Résolution spatiale	Résolution temporelle	N°	Bandes spectrales	Date d'acquisition
Pléiades	2 m	mono-date annuel	1	430–550 nm (B)	01/02/2019
			2	490–610 nm (V)	
			3	600–720 nm (R)	
			4	750–950 nm (PIR)	
	0,5 m		PAN	480–830 nm (panchromatique)	
Sentinel-2	60 m	5 jours	**1**	421-463 nm (B)	2018-2019
	10 m		2	458-523 nm (B)	
	10 m		3	543-578 nm (V)	
	10 m		4	650-680 nm (R)	
	20 m		5	698-713 nm (PIR 1)	
	20 m		6	733-748 nm (PIR 2)	
	20 m		7	773-793 nm (PIR 3)	
	10 m		8	785-900 nm (PIR large)	
	20 m		8a	855-875 nm (vegetation red edge)	
	60 m		**9**	935-955 nm (vapeur d'eau)	
	60 m		**10**	1360-1390 nm (cirrus)	
	20 m		11	1565-1655 nm (MIR 1)	
	20 m		12	2100-2280 nm (MIR 2)	
Venμs	5 m	2 jours	**1**	400-440 nm (Correction atmosphérique)	2018-2019
			2	423-463 nm (aérosols)	
			3	470-510 nm (B)	
			4	535-575 nm (V)	
			5	600-640 nm (qualité image)	
			6	600-640 nm (R 1)	
			7	652-682 nm (R 2)	
			8	690-714 nm (PIR 1)	
			9	734-750 nm (PIR 2)	
			10	774-790 nm (PIR 3)	
			11	845-885 nm (PIR 4)	
			12	900-920 nm (PIR 5)	

Table 21 Synthèses des bandes spectrales des satellites utilisés. En gras les bandes spectrales qui n'ont pas été utilisées.

Annexe B

Indices spectraux

Indice	Sentinel-2	Venμs
NDVI	$\frac{B8 - B4}{B8 + B4}$	$\frac{B11 - B7}{B11 + B7}$
NDWI	$\frac{B3 - B8}{B3 + B8}$	$\frac{B4 - B11}{B4 + B11}$
NDRE	$\frac{B8 - B5}{B8 + B5}$	$\frac{B11 - B8}{B11 + B8}$
mNDVI	$\frac{B8 - B11}{B8 + B11}$	/
MNDWI	$\frac{B3 - B11}{B3 + B11}$	/

(17)

Sentinel-2 BRI :

$$BIR = \Big\{ B1 * B1 + B2 * B2 + B3 * B3 + B4 * B4 + B5 * B5 + B6 * B6 + B7 * B7 + B8 * B8 + B8a * B8a + B11 * B11 + B12 * B12 \Big\}^{1/2} \tag{18}$$

Venμs BRI :

$$BRI = \Big\{ B1 * B1 + B2 * B2 + B3 * B3 + B4 * B4 + B5 * B5 + B6 * B6 + B7 * B7 + B8 * B8 + B9 * B9 + B10 * B10 + B11 * B11 + B12 * B12 \Big\}^{1/2} \tag{19}$$

Annexe C

Sélection d'ouvrages techniques et scientifiques utilisés pour l'identification et la définition des connaissances

Table 22 Sélection d'ouvrages techniques et scientifiques utilisés pour l'identification et la définition des règles

Sujet	Ouvrages
Général	ROR et al. (2003); Goudet (2003); Randrianarison (2006); Rakotofiringa and Tokarski (2007); Husson et al. (2009); Randrianarison (2013); Sourisseau et al. (2016)
Maïs	ROR et al. (2003); Goudet (2003); MAEP (2004)
Riz pluvial	Andriandrahona et al. (2007); Fofifa (nd); ROR et al. (2003); SCRID (2003); Chauvigne (2005); Rabiatsarafara (2006); Rakotofiringa and Tokarski (2007); Husson et al. (2009); Fofifa and Cirad (2015)
Riz irrigué	Rakotofiringa and Tokarski (2007)
Légumes	Goudet (2003); Rakotofiringa and Tokarski (2007)
Pomme de terre	Bélières (2020a)
Fruits	ROR et al. (2003); Rakotofiringa and Tokarski (2007)
Cultures potagères	Rakotofiringa and Tokarski (2007)
Cultures associées	GSDM and Cirad (nd); MAEP (2007); GSDM and CIRAD (nd); GTDR and PADR (2007); Vakinankaratra (nd); FOFIFA (2019); TFNAC (nd); BVLac (nd); Husson et al. (2009)

Annexe D

Questionnaire socio-économique utilisé lors de la saison culturale 2017-2018 auprès des agriculteurs

Lieu : **Date :** **n°**

Exploitant

Nom/Prénom – Sexe – *H / F* Âge –
Nombre de personnes dans le ménage – Nombre en âge de travailler –
Appartenance à une organisation de producteur – *Oui / Non*
Nom de l'association –

Exploitation

Surface moyenne (estimation) – Nombre de champs –
Nb de tanety – Nb de basfonds –
Nb de RMME – Nb de jachères –
Travail de la terre – *manuel / animaux / machines*
Nb de zébus – Nb de vaches laitières –
Nb charrettes – Nb charrues –
Commentaires :

Cultures

Quelles sont vos principales cultures durant la **saison des pluies** (par ordre d'importance)?
Connaître les cultures principales du lieu (-> zonage de l'organisation spatiale des cultures de saison principale et probabilité de les rencontrer)

Vos principales cultures durant la **contre-saison** (par ordre d'importance) ?
Connaître les cultures de contre-saison du lieu (-> zonage de l'organisation spatiale des cultures de contre saison et probabilité de les rencontrer)

Quelles sont vos **principales séquences de culture** (à l'intérieur d'une année ou de la saison principale si pas de contre saison) ?

Connaître l'organisation temporelle des cultures principales du lieu à l'intérieur d'une année (-> noter les périodes de plantation de chaque culture pour en tirer des règles temporelles, voir si celles-ci correspondent à celles de la FAO)

D'une manière générale, quelles **cultures cherchez vous à planter en priorité** et pourquoi ?

Connaître les contraintes et stratégies liées à l'environnement, au marché, au contexte social

A quel moment décidez-vous quelles seront les cultures mises en terre pour la saison ?

Savoir quand l'exploitant met sa stratégie agricole en place

Le tableau suivant sert à avoir une compréhension de l'organisation spatiale et temporelle des cultures, de ce qui peut influencer les choix de positionnement par rapport aux contraintes (physiques, économiques, sociétales…)

Parcelle (identifiant du point GPS ?)	Taille (environ, en m*m)	Saison culturale (*principale / contre saison)*	Culture, ou séquence de cultures (noter aussi les associations de culture) ex. Tom // Riz-Mais (Tomate puis Association Riz et Maïs avec le Riz qui domine et le Maïs en associé)	Culture l'année précédente	Début-fin de période de semis (ou repiquage)	Début et fin de période de récolte	Type de sol (1-8)	Irrigation (*oui/ non/ mauvaise maitrise*)	Topographie (*Tanety/ bas-fond / plateaux/ bas de tanety/ plaine*)	Autre contrainte (*exposition, érosion, …)*	Traitement de la terre(*engrais, fertilisation, fumier, …)*

Années marquantes

Année / Saison	Cause (climatique, économique, phytosanitaire, sociale, …)	Effet sur les cultures	Adaptations

Mettre en évidence ce qui peut perturber la production, le choix de cultures, les stratégies et quelles sont les adaptations mises en œuvre

Marché et vente

Obtenir des informations sur le contexte économique, sur le positionnement des lieux de vente mais aussi connaître lesquels sont privilégiés (distance, prix, variation selon le type de produit vendu ou acheté).

La production permet-elle de nourrir la famille – *autosuffisant / non*

L'année passée vente de produits – *oui / non*

Quelle culture vendez-vous ? –

Culture	Autoconsommation (%)	Vente (%)

Quel(s) est(sont) votre(vos) lieu(x) de vente –

Si multiples, pourquoi –

Quelle part de vos récoltes conservez-vous pour ressemer ?

Achetez-vous des semences, des denrées alimentaires aux marchés ? – *oui / non*

Si oui, à quel(s) marché(s) –

Comment vous déplacez vous aux marchés ?

Echangez-vous des denrées avec d'autres personnes ? – *oui / non*

Si oui, quels types de denrées ?

Avec quels personnes (localisation) ? –

Annexe E

Résumé des connaissances issues des enquêtes auprès des agriculteurs en 2017-2018 pour les « principales » cultures

Maïs

- souvent associé aux haricots ou soja
- peut se trouver de partout
- cultivé principalement sur les tanety
- proche des routes passantes
- en général distance maximum de 3 à 5 km de l'habitation
- éloigné des habitations d'environ 100, 200 mètres (à cause des zébus)
- sur tout type de tanety
- sols ferrallitiques
- associé aux haricots ou soja et parfois au taro
- commune d'Andranomanelatra occupe 60 % des tanety (55 % d'associé, 5 % de pur)
- presque toujours associé à l'haricot, au soja ou au manioc
- le maïs pur est très rare, il est presque toujours associé au riz pluvial. Il suit les mêmes stratégies que le riz pluvial
- peut être associé aux pommes de terre ou au riz pluvial
- cultivé en zone non irriguée
- loin sur les plateaux
- sur tout type de zone
- maximum à 1km de l'habitation, très grand maximum à 2km sur les collines (Alakamisy)
- plantation en août, récolte en janvier / février. Est souvent laissé dans le champ pour sécher sur pied
- le maïs est plus important que le riz pluvial, car c'est une culture historique

Riz irrigué

- grande partie de la production des bas-fonds
- dans les bas-fonds
- production de bas-fond

- dans les bas-fond ne pouvant être cultivés en maraîchage (Alakamisy)
- sols hydromorphes noirs
- les agriculteurs attendent les premières pluies pour semer

Soja

- souvent associé au maïs
- cultivé dans les mêmes zones que le maïs
- récolté environ 2 semaines après le maïs

Haricots

- cultivé dans les mêmes zones que le maïs
- si pur en bas des pentes
- si associé sur les plateaux
- récolté souvent en décembre.
- presque toujours cultivé avec le maïs
- dans la commune d'Andranomanelatra il correspond à 35 % des cultures associées

Petites cultures

- proche des routes passantes

- **patate douce**
  - cultivé sur les tanety
  - servent à la rotation des cultures
  - servent à ameublir la terre
  - temps de croissance de 1 an et demi
  - petites parcelles pour consommation personnelle

- **taro sauvage**
  - sur tanety en terrasse ou en bord de cuvette inondée (lorsque principale production)
  - planté après les pommes de terre qui bénéficient de fertilisants
  - la croissance dure 1an

- **Manioc**
- cultivé de manière familiale (donc petites parcelles) sauf si besoin en manioc important
- très petites parcelles, pour la famille uniquement
- sur sol blanc
- temps de croissance de 1an et demi

Pomme de terre

- culture possible toute l'année
- pour la saison pluviale principale:

    - plantation en janvier, récolte en avril (3-4 mois de cycle)
    - on peut les cultiver de partout sur les tanety
    - il n'y a pas d'emplacements préférentiels
    - en saison des pluies, cultivée pour avoir des semences pour la contre-saison
    - c'est une culture prioritaire
    - cultivé sur les pentes et plateaux
    - cultivés dans les zones plus froides
    - même dynamique que pour les légumes (Alakamisy)
- Surtout en contre-saison dans les bas-fonds

Légumes

- principalement culture de rente, dans les hauts-fonds des bas-fonds, les terres jamais inondées
- haut fond dans les bas-fonds
- sol humide
- sol humide et volcanique
- petites surfaces
- coûteuses à cause des traitements
- sur les surfaces planes et irriguées (principal déterminant de cette culture)
- carottes en altitude
- à plus de 2-3 km de la RN7
- dans les bas-fonds et les bahibo
- proche des cours d'eau.
- à maximum 500m de l'habitation
- cultivés toute l'année
    - temps de culture 3-4 mois
    - au bout de 3-4 ans rotation pendant 1 an avec du soja ou du maïs
- en plaine sur les zones de maraîchage tout au long de l'année
- culture principale (Alakamisy)
- dans les bas-fonds
- faibles pentes
- plateaux irrigués
- en bas des pentes
- minimum à 30m, maximum à 1km, très grand maximum à 2km en plaine
- plus proche des habitations que le riz et le maïs, car demandent un travail régulier
- surtout en contre-saison dans les bas-fonds
- généralement dans les bas-fonds à 2km de la RN7
- carottes cultivées sur les hauts de pente, plateaux en contre-saison

Vergers

- sol de bonne qualité

- petite culture familiale (peu de vente en gros sauf zone spécialisée)
- peu nombreux, collés aux habitations
- petite taille, consommation personnelle
- en altitude, dans les zones froides

Pins

- à 500m – 1km des habitations
- sur les pentes raides

Riz pluvial

- là où le maïs est bien implanté il est peu présent
- surtout des petites surfaces
- en hauteur (plateaux, haut de pente), là où l'eau ne reste pas.
- introduit récemment (il y a 10 ans), donc moins présent que les autres cultures sauf zones où il a été implanté tôt
- les cultures proches des routes comme le riz pluvial demandent moins d'entretiens (maïs manioc ...)
- sur les plateaux et pentes pas trop importantes (sinon érosion, le riz retient mal la terre)
- planté pur ou bien bordé de maïs
- sur les tanety
- sur sol ferrallitique
- principalement associé au maïs
- sur les hauts et pentes des tanety
- correspond à environ 40 % de la culture sur tanety dans la commune d'Antranomanelatra (10 % maximum des champs en mono-culture)
- planté début novembre, récolté 3-4 mois après
- principalement associé au maïs (au moins les bords)
- quelques parcelles de pur
- surtout les pentes aménagées, les plateaux manquent d'eau pour cette culture.
- pivot de rotation (Alakamisy)
- cultivé en hauteur, comme le maïs
- sur tout type de zone (plaine non irriguée, pentes, plateaux)
- maximum à 1km, très grand maximum à 2km sur les collines (Alakamisy)
- planté selon la pluie, entre fin août et début octobre, récolté de mars à avril.
- le semis est d'octobre à novembre (si pluie tardive) et récolte de mars à avril.

Général pour les cultures

- grand maximum 8 km de là où la personne habite (mais rare, car métayage), moins de 10 % des agriculteurs
- généralement cultivés par soi-même jusqu'à 1 ou 2 km

- au bord de la RN7, les cultures sont moins importantes, car moins d'agriculteurs et risque de vol. À partir de 2/3 km de la RN7 il y a plus de cultures
- peut de cultures au bord de la RN7, car beaucoup d'habitations
- cultures à environ 1km de l'habitation, mais habitat très dispersé
- cours d'eau uniquement important pour la contre-saison
- grand maximum à 18 km de là où la personne habite (mais distance à pieds, car beaucoup de détours), sinon à vol d'oiseau la distance est courte
- les personnes cultivent en général à 1-2 km de chez eux, mais surtout à 500m voir 1km
- pas d'emplacement préférentiel pour les cultures sur tanety, car il y a des rotations tous les ans
- les champs non cultivés sont pour les zébus
- pente maximum à 50 %
- si les pentes sont trop raides pas de culture, mais parfois des plantations de pins.
- pente maximum de 25 % (Alakamisy)

Zones non cultivées

- là où il y a de l'érosion

Annexe F

Résumé des connaissances identifiées par communes à partir des questionnaires réalisés en 2019

Cette annexe est une retranscription des entretiens avec les personnes clés dans les Fokontany et les communes. Ici sont résumées les connaissances que nous avons pu obtenir lors de la campagne de terrain du 9 avril 2019 au 15 mai 2019.

En complément de la retranscription des entretiens :
Les **tanenty** peuvent être rattachés à des collines cultivées. **CS** signifie « contre saison », soit la saison sèche où sont cultivés les bas-fonds et autres parcelles irriguées, souvent avec des légumes ou des pommes de terre.
Lorsque les **successions culturales** sont abordées, « / » signifie « contre saison » et « // » signifie « saison culturale principale suivante ».
La **période de soudure** est définie par Sourisseau et al. (2016) (p.112) comme : « période durant lesquelles ils [*ndlr. les agriculteurs*] n'ont pas suffisamment de nourriture pour l'alimentation de la famille. Pour les ménages en zone rurale, cela correspond à la période avant les récoltes, quand les stocks de l'année précédente se terminent. Pour les ménages urbains sans activité agricole, ces périodes ne sont pas liées aux récoltes, mais sont fonction des activités et des revenus. »

1 Commune Ambohibary

Analyse de la commune issue de la discussion avec Rija Ramananraibe, responsable de l'agriculture et des statistiques de production dans la commune. [enquête 1]

1.1 Principales cultures de la commune

- riz irrigué
- maïs
- haricots
- soja
- pomme de terre (CS)
- carottes (CS)
- choux (CS)

- Vergers

Le **riz irrigué** occupe une grande partie de la production en bas fond, car la commune occupe une partie de la cuvette rizicole. Le riz irrigué est surtout dans les Fokontany de **Tsarazazamandimby**, **Ankeniheny III** et **Mahatsinjo** (commune voisine).

Le **maïs** ne dépend pas de l'élevage, il est consommé par les agriculteurs. On le retrouve dans la plaine avant la cuvette et un peu sur les tanety (colines). Il est associé au **soja et haricots**. On trouve du maïs de partout, la distance à l'habitation est de **maximum 3-5 km**.

Le **maraîchage** (équivalent de la classe *légumes*) est surtout composé de choux et carottes qui sont seulement destinés à la vente, car le prix de vente est beaucoup plus élevé que celui du riz. Le maraîchage est pratiqué sur sol humide, dans les hauts fonds du bas-fond là où cette zone n'est jamais inondée et principalement dans le Fokontany de **Ambatomainty Fihaonana**.

Il y a un peu de **vergers** dans tous les Fokontany, mais il y en a beaucoup plus dans le Fokontany de Madera Ambohimadinika, car le sol est favorable à ces cultures, il est de bonne qualité.

Le **riz pluvial** est peu présent, car le maïs est bien implanté. Cette culture a débuté en 2015 dans le Fokontany de **Mahatsinjo Miaradia** par une seul personne. Cette culture s'est ensuite diffusée à des agriculteurs proches. Surtout cultivé sur de petites surfaces, en hauteur, là où l'eau ne reste pas (altitude, haut de tanety) (Fokontany : **Mahatsinjo Miaradia et Antsofobato Telomiray**).

1.2 Marchés

Le principal marché est **Ambohibary** pour l'achat vente aux locaux de **riz** et **maïs**. Les cultures **maraîchères** et les **pommes** sont achetés par des **collecteurs**.

1.3 Particularités

Lorsqu'il y a des inondations, les Fokontany les plus touchés sont **Ankeniheny III, Tsarazazamandimby** et la commune de **Mandrosohasina**. Les inondations ont lieu en février-mars. Après ces inondations, il y a peu de cultures qui sont replantées directement, les agriculteurs attendent que l'eau parte, puis il attendent la nouvelle saison culturale (soit pendant la contre-saison, soit l'année suivante) pour replanter.

1.4 Spécificités de Fokontany

Vergers : **Madera Ambohimadinika** **Maraîchage** : **Ambatomainty Fihaonana** surtout des carottes et des choux dans les hauts fonds de la cuvette.

1.5 Précision de distance de culture

Grand maximum 8 km de la où la personne habite, mais souvent métayage ou les parcelles sont cultivées par des personnes habitant proche de celles-ci. L'étendue maximum peut être considérée comme au Fokontany voisin. **Pour ce que la personne cultive elle-même ses parcelles, la distance maximum est plutôt de 1 ou 2 km**. Ici, le transmission des parcelles se fait par héritage, donc les descendants cultivent comme les ancêtres le faisaient et là où ils avaient des terres.

1.6 Zones non cultivées

Les zones non cultivables le sont soit à cause de l'érosion, sois à cause de conflits foncier (comme les hauts de tanety autour d'Ambohibary).

1.7 Différence avec les autres communes

Les communes d'**Antsoatany** et de **Mandrosohasina** ont les mêmes pratiques que la commune d'Ambohibary. Mais, plus loin, le sol et la disponibilité en eau est différente, les pluies arrivent plus tôt, il y a des différences (ici, le responsable de l'agriculture parle de notions plus abstraites tel que la tradition).

2 Fokontany de Sahabe Tetezana

Analyse du Fokontany issue de la discussion avec Rasoarimandimly Marie Henriette, agricultrice. [enquête 2]
Cette enquête, courte et avec peu d'informations reflète les différences entre les Fokontany que nous avons pu enquêter. Certains disposent de personnels dédiés à l'agriculture et sont en contact avec des organismes de recherche tel que le Centre National de Recherche Appliquée au Développement Rural (FOFIFA) (équivalent de l'INRA malgache), d'autres sont dépourvus de ressources et connaissent peu les pratiques qui peuvent être mobilisés.

2.1 Principales cultures de la commune

- riz irrigué
- maïs (toujours associé avec des haricots ou du soja)
- maraîchage pour la vente
- patate douce (1/10ème des terres, uniquement des petites cultures), jamais en grande parcelle

2.2 Succession de cultures

Bas-fond choix 1 : Riz irrigué / pomme de terre // Riz irrigué / pomme de terre // ...
Bas-fond choix 2 : Riz irrigué / maraîchage // Riz irrigué/ pomme de terre // ...
Tanety : Riz pluvial // patate douce // Maïs-haricot // Maïs-haricot // Riz Pluvial // ...
Pour les tanety (collines) Il n'y a pas de succession de riz pluvial de suite, car le rendement diminue.

3 Commune d'Antsoatany

Discussion en 2018 avec M. Paulain, responsable rural de la commune.

La commune est composée de 5 Fokontany avec environ 16300 habitants. La commune a des partenariats avec le Fiompiana Fambolena Malagasy Norvéziana (FIFAMANOR), le Agence japonaise de coopération internationale (JICA) et différents projets sur le riz ou l'amélioration des conditions de vie.

3.1 Principales cultures de la commune

- riz irrigué
- maïs
- pomme de terre (CS)
- riz pluvial
- vergers

Répartition des cultures :
Environ 70 % de la surface de la commune est cultivée et 30% n'est pas cultivée.

- 60 % de riz, maïs, haricots
- 15 – 20 % arbres
- 10 % arbres fruitiers
- 10 % autres cultures.

Le **maïs** est plus important que le **riz pluvial**, car c'est une culture historique. Le riz pluvial n'a été introduit que récemment (depuis 10 ans) dans la zone grâce à une vulgarisation de l'apprentissage. Depuis que le riz pluvial est cultivé, la quantité de champs de maïs a un peu diminué au profit du riz pluvial. Le maïs est peu vendu, il est surtout destiné à l'alimentation. Le maïs est le 2ème aliment en période de soudure. Le riz pluvial et le maïs sont **cultivés sur les tanety**. **Maintenant il y a environ autant de riz pluvial que de maïs dans la commune**.

Les vergers sont des petites productions familiales, mais il y a aussi de la vente.

Il y a peu de **maraîchage**, sur de petites surfaces et principalement en contre-saison. Ces cultures coûtent cher, car il faut faire de nombreux traitements et encore plus lors de la saison des pluies.

Si la surface est plane et irrigué on peut y cultiver des cultures maraîchères (soit classe *légumes* ou classe *pomme de terre*). C'est surtout cela qui détermine le type de culture.

Il est plus facile de cultiver des **carottes** en **altitude** (au Nord d'Ambohibary par exemple), il y a plus d'**irrigation**, ici, il y a un manque de volonté de la part des agriculteurs.

3.2 Marchés

Il y a un petit marché tous les lundi. La vente principale commercent les cultures issues de l'agriculture maraîchère grâce à des groupements locaux de paysans. Il y a aussi des collecteurs qui viennent acheter les légumes.

3.2.1 Précision de distance de culture

En étant au **bord de la route nationale n°7**, les habitants ont plutôt tendance à avoir des emplois complémentaires pour gagner plus d'argent. **Les cultures sont donc en partie délaissées**. **A partir de 2-3 km, il y a plus de cultures et le maraîchage est plus fréquemment pratiqué**. Les cultures proches des **routes** sont des cultures qui demandent moins d'entretiens (maïs, manioc). Les cultures au bord de la route nationale n°7 et au bord des routes ou chemins où les charrettes passent facilement sont moins cultivées à cause de l'insécurité (vol des récoltes).

Le maraîchage est pratiqué dans les bas-fonds et bahibo (espace avant les tanenty). La pratique du maraîchage en contre-saison et intersaison est majoritairement située à environ 2km de la route nationale n°7.

Les parcelles **cultivées** sont au maximum à 8 km de l'habitation (mais cela est très rare, moins de 10 % des agriculteurs). Les **cultures sont en majorité à 2-3 km des habitations**. La pente maxime est autour de **50 %**. Les cours d'eau sont importants uniquement pour la contre-saison et les légumes.

Les **pins** sont au moins à 500 m-1k m des habitations. Le **maïs** et le **riz pluvial** sont éloignés des habitations à cause des zébus (soit 100–200 m). Les **légumes** sont au maximum des habitations à 500 m. Les **vergers** sont peu nombreux (quelques pieds) et sont très proche des habitations.

Les cultures sont à certains endroits dans le paysage :

- le **riz irrigué** dans les bas-fonds ;
- le **riz pluvial** sur les plateaux et pentes, mais les pentes importantes ne sont pas idéales, car il y a une importante érosion et le riz retient mal la terre ;
- le **maïs** est sur les tanety ;
- le **soja** est cultivé sur les mêmes zones que là où le maïs est cultivé ;
- les **haricots** sont situés aux mêmes zones que là où le maïs est cultivé. S'ils sont cultivés sans association, les champs sont en bas de pente, s'il est associé, les champs sont majoritairement sur les plateaux.
- Les **patates douces** sont cultivés sur les tanety et servent surtout dans les rotations avec les autres cultures, pour ameublir la terre.

3.3 Homogénéité des Fokontany

Tous les Fokontany cultivent du riz irrigué, du maïs, des haricots et du riz pluvial.

3.4 Rotation des cultures

Bas-fond : riz irrigué / légumes // riz irrigué / légumes // ...

Riz pluvial :

- riz pluvial // riz pluvial // ... (70 % des agriculteurs)
- riz pluvial // patate douce ou jachère ou maïs-haricot (30 % des agriculteurs)
- , mais il est difficile de faire 2 années de suite des céréales, les rendements diminuent.

Maïs (90 % des champs sont associés, parfois avec du riz pluvial, mais surtout avec du soja) :

- Maïs // autre culture ou maïs.
- Lorsque la culture est moins productive (au bout de 4-5 ans, une année de jachère est mise en place).

4 Commune de Mandrosohasina

Entretien avec l'adjoint au maire, Rabeoaimalele Edouard, le 19/04/2019.

4.1 Principales cultures de la commune

- riz irrigué
- carottes
- choux
- maïs

Cultures secondaires :

- riz pluvial
- taro sauvage
- manioc

Le **riz irrigué** occupe une grande partie de la production en bas fond, car la commune occupe une partie de la cuvette rizicole. Basé sur les sols hydromorphes noirs, les champs sont toujours cultivés avec du riz irrigué et durant la contre-saison, le type de culture est toujours la pomme de terre, car il y a un fort marché [rotation : **riz/pdt//riz/pdt**...]. En général le semi est entre août et octobre et le repiquage en novembre (si pluie précoce), mais toujours lors des 1ères pluies. Récolte en mars – avril. Cette année, repiquage en janvier.

Le **maraîchage** est surtout composé de choux et carottes, qui sont uniquement destinés à la vente, car leur valeur commerciale est beaucoup plus élevée que celle du riz. Sol humide ou volcanique, haut fond dans le bas-fond. Le maraîchage est pratiqué toute l'année. En général les rotations sont [**carottes / choux / carottes /** …] et au bout de 3-4 ans, il a une plantation de soja ou de maïs. Le temps de culture est de 3-4 mois.

Les **pommes de terre** sont surtout cultivées en contre-saison dans les bas-fonds. Lors de la saison des pluies, ces cultures sont principalement destinées à obtenir des semences pour la contre-saison. Le reste de la production est acheté par des collecteurs. Ces cultures sont possibles toute l'année, mais pour la saison des pluies, plantation en janvier et récolte en avril (3-4 mois de maturation). On les cultive de partout sur les tanety. Les rotations sont généralement **PdT // riz pluvial // maïs-soja // PdT**. Il n'y a donc pas d'emplacement préférentiel. **C'est une culture prioritaire**.

Le **maïs** est cultivé sur les tanety, sol ferrallitique. Il est associé aux **haricots ou au soja**, parfois au **taro**. On le retrouve sur tous les types de tanety. Plantation en août et récolte en janvier/février (parfois plus tard, comme maintenant (mi-avril) pour qu'il sèche sur pied). Le **soja** est récolté environ 2 semaines après le maïs et les **haricots** environ en décembre.

Le **riz pluvial** est planté pur ou bien bordé de maïs. Sur les tanety ferrallitiques, il est planté selon la pluie entre fin août et début octobre. Il est récolté de mars à avril.

Le **taro sauvage** est planté sur les tanety en terrasse ou bien en bord de cuvette (principale production). Souvent planté ici après les **pommes de terre**, qui bénéficient de fertilisants. [rotation : **Pdt // taro // PdT**]. La culture dure 1 an.

Le **manioc** est surtout cultivé de manière familiale, mais est surtout important s'il y a des « besoins ».

4.2 Marchés

Il n'y a pas de marché dans la commune. Il y a surtout des collecteurs qui viennent acheter la production. La vente au marché est rare. Les achats sont généralement effectués aux autres marchés proches (Ambohibary ou autre)

4.3 Particularités

Les productions sont plutôt bonnes.
Sol : le sol **volcanique** est surtout destiné aux **légumes**, le **ferrallitique** au **maïs** ou au **riz pluvial**, et l'**argileux** au **riz irrigué**.

4.4 Homogénéité des Fokontany

La commune est assez homogène. Partout où il y a la route, il y a les mêmes pratiques et les mêmes techniques grâce à la diffusion d'informations. Dans les Fokontany les plus enclavés, il y a des connaissances techniques, mais moins de moyens financiers et moins de fertilisation (**donc moins de PdT**) (Fokonatny de **Betampere** et de **Rianjanoloanga**).

4.5 Spécificités de Fokontany

Maraîchage : **Tsaramody**, car le sol est volcanique.
Il y a plus de tanety dans les Fokontanys de **Betampere**, **Rianjanoloanga**, dans une partie de **Soahazo** et d'**Amparihy**.

4.6 Précision de distance de culture

Grand maximum 18 km de là où la personne habite (en passant par les chemins, court à vol d'oiseau). Mais souvent les parcelles sont cultivées par des personnes habitantes proches de celles-ci, par métayage. **Pour ce que la personne cultive elle-même, la distance est plus proche de 1-2 km maximum, surtout entre 500m et 1km**. Si les pentes sont trop raides, il y a de l'érosion, donc pas de culture, mais des pins peuvent parfois être plantés. Il n'y a **pas d'emplacement préférentiel pour les cultures en tanety**, car il y a une rotation tous les ans.

5 Commune d'Ambatolahy

Peu de réponse dans cette commune urbaine.

5.1 Principales cultures de la commune

- riz
- maïs
- patates douces
- pomme de terre (CS)

5.2 Homogénéité des Fokontany

Les Fokontany sont assez homogènes, sauf un, qui a beaucoup de grenadelles.

5.3 Particularités

En février-mars, les routes sont cassées.

6 Commune d'Andranomanelatra

Discussion et enquête avec M. Randrianarivony (dit Rasoul) , responsable agricole de la commune. [enquête 3 A et B]

6.1 Principales cultures de la commune

- riz irrigué
- riz pluvial
- maïs
- haricots

Cultures secondaires :

- manioc (petites parcelles, très peu)
- soja (associés au maïs, mais faible ici)
- tomates (CS)
- pomme de terre (CS)
- patates douces (presque pas)

Environ **60 % des terres de la commune sont cultivées**. Il y a un manque d'argent pour exploiter plus, mais il n'y a pas de limitation technique.

Dans cette commune le **sol est plutôt ferrallitique**.

Pour le **riz irrigué**, les agriculteurs attendent la pluie pour semer.

Le **riz pluvial** est cultivé depuis environ 1995-1997, mais son essor se fait en 2012. Il est principalement **associé au maïs**. On le retrouve sur les **hauts et pentes des tanety** et il correspond à environ **40 % de la culture sur les tanety**. Parmi les champs cultivés avec du **riz pluvial**, il y en a un maximum de 10 % qui sont en mono-culture. La culture est plantée environ du 01/11 au 10/11 puis récolte 3-4 mois après.

Le maïs est presque toujours associé aux haricots, au riz pluvial, au soja ou au manioc. Il n'est cultivé que pour la consommation personnelle des agriculteurs. Cette culture est présente sur environ **60 % des tanety (55 % d'associées et 5 % de pures)**.

Les **haricots** sont presque toujours cultivés avec le maïs. Ils correspondent à **35 % des cultures associées**.

Il y a un peu de vergers dans la commune, mais les parcelles sont de petite taille et pour la consommation familiale.

6.2 Marchés

La production est surtout pour la consommation personnelle. Il y a un petit marché ici, mais surtout pour les besoins des locaux. Il y a des collecteurs clandestins qui viennent acheter dans les villages. La vente se fait surtout durant la contre-saison.

6.3 Homogénéité des Fokontany

Les Fokontany sont assez homogènes

6.4 Précision de distance de culture

Peu de cultures au bord de la route nationale n°7, car il y a beaucoup d'habitations. Les **cultures sont à environ 1km des habitations, mais l'habitat est très dispersé**. Il n'y a qu'un seul cours d'eau qui passe par la commune donc il n'y a **presque aucune irrigation**, les **cultures sont principalement situées sur les tanety**.

6.5 Particularités

La commune, qui a une vocation agropastorale, ne peut aider les agriculteurs, car ils manquent de fonds.
Il y a des champs pour les zébus.

6.6 Différence avec les autres communes

D'un point de vue personnel de la personne enquêtée, dans **tout Antsirabe II il y a les mêmes dynamiques et les cultures sont similaires**. Les **dates de semi** sont aussi similaires.
Ambano est différent, avec beaucoup de maraîchage, car les sols sont très fertiles.
Antsoatany est plus ou moins similaire qu'Andranomanelatra.
Ambohibary ressemble aussi à la commune enquêtée (après vérification à la mairie, nous avons constaté qu'il y a beaucoup moins de riz pluvial à Ambohibary).

7 Commune d'Alakamisy

Discussion et enquête avec M. Razafimohatratra Fananezantre, conseiller au développement rural de la commune et Rakotardiane Armand, membre d'une association sur l'agriculture locale et réceptionniste de la commune.

La commune peut être divisée en deux. Le Nord, plus enclavé, avec plus de pentes et plus frais, et le Sud, avec un développement plus important et plus chaud.
Au Nord, Fokontany de : Ambohitromby, Marofangady, Mahandraza (le plus enclavé) et Soamatamana (le plus développé des 4, seul à faire de la contre-saison).
Au Sud, Fokontany de : Tsaratanana (majorité de vergers), Alakamisy et Andohafarihy (avec le plus de rizières).

Environ **70 % des terres des Fokontany Nord sont cultivées et 90 % des terres dans les Fokontany du Sud**.
Au Nord, le sol est plutôt noir et au Sud rouge.

7.1 Cultures principales

Au Nord :

- riz irrigué
- riz pluvial
- maïs

De manière secondaire :

- petit pois
- pomme de terre

De manière tertiaire :

- manioc (petites parcelles, très peu).
- carottes
- patates douces (presque pas)

* légumes = pomme de terre, petits poids, haricots verts.

Pour le **riz irrigué**, en temps normal les semis sont en octobre puis le repiquage autour du **20 novembre**, dans ce cas, les **récoltes sont d'avril à mai**. Cette année, la pluie a été tardive donc semi en octobre, repiquage en janvier et récolte en mai - début juin. Il y a toujours de l'eau dans les bas-fonds donc la contre-saison est impossible. Pour les rares occasions où cela est possible la rotation riz irrigué // pomme de terre // riz irrigué // ... est pratiquée.

Le **riz pluvial** est principalement **associé au maïs (au moins sur les bords)** bien qu'il y ait quelques parcelles de riz pluvial pur. Il y en a **surtout à Soamatamana, à Ambohitromby et à Mahandraza** où le riz pluvial commence à être cultivé. Les rotations sont **riz // légumes // riz //**

légumes. Les zones cultivées en **riz pluvial** sont surtout les **pentes aménagées** et **les plateaux qui manquant d'eau**. Le semi est d'octobre à novembre (si pluie tardive) et la récolte de mars à avril.

Le **maïs pur est très rare**, il est **presque toujours associé au riz pluvial** et est donc cultivés dans les mêmes conditions que cette culture. Le semi est d'octobre à novembre (si pluie tardive) et récolte en mai.

Les **petits poids** et les **pommes de terre** sont cultivés dans les mêmes proportions dans le Nord. La rotation est similaire à celle du riz pluvial : riz pluvial // légumes // riz pluvial // ... . Ils sont cultivés sur les **pentes et plateaux**. Les petits pois sont semés en octobre et récoltés en décembre (saison 1), puis de janvier à avril (saison 2). Les pommes de terre sont semées vers le 15 octobre et récoltées en janvier (environ 3 mois et 8 jours).

Les **carottes** sont surtout cultivées sur les plateaux (hauts de pente) avec une rotation avec le riz pluvial. En plaine, elles sont cultivées sur les zones dédiées au maraîchage (champs irrigués et plutôt plat) tout au long de l'année (avec choux et haricots verts). Semées d'octobre à novembre puis récolté en mars – avril.

Le **manioc** est cultivé sur de très petites parcelles, pour la famille uniquement. Plutôt sur sol blanc. Le temps de croissance est d'environ 1 an et demi. Il en va de même pour les **patates douces**, mais le temps de croissance est d'environ 1 an.

Il n'y a pas de vergers au Nord. **Au Nord, la contre-saison sur tanety** n'est pratiqué qu'à **Marofangady** et sur les **bas-fond** pour le Fokontany de **Soamatamana**.

Au Sud :

- légumes
- pommes de terre
- maïs
- riz pluvial
- riz irrigué
- vergers

Les **légumes**, (surtout les carottes) sont la culture principale. Ils sont **cultivés toute l'année**, surtout dans les **bas-fonds**, sinon en **pente faible** et sur les **plateaux pouvant être irrigués**. Il y a beaucoup de petits canaux de partout pour l'irrigation.

Les **pommes de terre** suivent la même dynamique que les légumes.

Comme dans le Nord, le **maïs** est **souvent associé au riz pluvial** ou bien aux pommes de terre. Il est cultivé dans les zones non irriguées, souvent **loin sur les plateaux**. Le riz est toujours le pivot de rotation.

Le **riz pluvial** est surtout cultivé en **hauteur**, comme le maïs, sur tout type de zone (plaine non irriguée, pente, plateaux). Les **légumes** sont en bas des tanety et le riz pluvial en haut. Le **riz**

irrigué est cultivé dans les mêmes proportions que le riz pluvial. Il est dans les bas-fonds éloignés ou bien ne pouvant être cultivés en maraîchage.

7.2 Répartition des cultures

Pour le Nord :

Selon le responsable de l'association :

- 15 % de bas-fond
- 25 % de riz pluvial
- 50 % de légumes
- 10 % de maïs

Selon le responsable du développement rural :

- 15 % de bas-fond
- 25 % de riz pluvial + maïs
- 50 % de légumes
- 10 % de quelques vergers et pins

Pour le Sud: Selon le responsable de l'association :

- 35 % de légumes
- 20 % de riz
- 15 % de maïs
- 30 % pins et vergers

Selon le responsable du développement rural :

- 40 % de légumes
- 25 % de riz pluvial + maïs
- 10 % de vergers
- 25 % pins et sans cultures (dont 5 % sans cultures)

7.3 Marchés

Il y a beaucoup de collecteurs pour les légumes, car la commune est peu accessible (peu de routes praticables). Il y a un petit marché local tous les matins. Les petits poids sont vendus à Antsirabe. Samedi, il y a une vente de zébus et de porcs à Antsirabe donc les achats se font en même temps.

7.4 Homogénéité des Fokontany

Les Fokontany sont divisés entre le Nord (plus enclavé) et le Sud (plus développé).

7.5 Précision de distance de culture

Si les agriculteurs ont peu de parcelles, ils vont tout de même toutes les cultiver même si celles-ci sont loin.
La **pente maximale est d'environ 25 %**.
L'élevage est proche des maisons, donc un peu plus loin on retrouve le maraîchage (minimum environ 30 m), maximum 1 km, grand maximum 2 km en plaine (2 fokontany – certains cultivent à Ambano) Pour le **riz et maïs environ la même distance, mais sur les collines**. Les **légumes** sont plus proches que le maïs et le riz, car demande un travail régulier.

8 Commune d'Antanifotsy

Discussion et enquête avec M. Tiane, responsable de l'agriculture pour le district d'Antanifotsy Personne qui qui a travaillé sur ce domaine et qui à différents endroits. Il fait aussi des relevés GPS pour relever l'occupation des sols.

Il y a 3 saisons culturales dans la commune, de septembre à décembre, de janvier à avril et de mai à août pour la contre-saison. Les pratiques culturales sont bonnes. Il y a beaucoup de culture associée pour plus de diversité dans la consommation des ménages.

8.1 Cultures principales

- maïs
- riz irrigué
- pomme de terre (peu en saison des pluies)
- poulet gasy

Cultures secondaires :

- soja
- haricots
- vergers
- lentilles
- voanemba (type de poids)
- poids du cap (petites parcelles, faible production)
- taro

Le **maïs** est dominant dans le paysage et dans les choix des agriculteurs. Il est quelquefois pur, mais la plupart du temps **associé au soja, au haricot** et quelques fois à la patate douce. Sa séquence de rotation est **maïs // tubercules ou patate douce**. On peut le retrouver sur **tous les types de reliefs**. Il est semé en septembre puis récolté en décembre (4 mois), puis semé en janvier et récolté en avril. On retrouve surtout des variétés ayant un cycle court.

Le **riz irrigué** se trouve dans les **bas-fonds** et est en rotation avec les légumes (surtout **pomme de terre**). Il est **semé en septembre** puis **repiqué entre 30 et 60 jours après** et **récolté 150 jours après**.

La culture de **pomme de terre** est souvent précurseuse du maïs dans le cycle de rotation. **Il y en a peu durant la saison des pluies**, car le sol est trop humide. On les retrouve sur les **tanety**, en terrasse, ou dans les **plaines**.

Le **soja** et les **haricots** sont cultivées en même temps et en association avec le **maïs**. Le **soja** est **récolté au bout de 5 mois** et les **haricots au bout de 3-4 mois** et sont semés un peu avant. On les retrouve sur **tous les types de reliefs lorsque associé**, mais sur les **plaines et plateaux pour la monoculture**.

Les lentilles et le voanemba occupent environ 20 ares par Fokontany. Le Voanemba est assez récent avec GSDM.

Les **vergers** sont surtout présents dans les **6 Fokontany au Nord de la commune**. Ils sont cultivés sur les **tanety, mais surtout sur les plateaux**.

Le **taro** est cultivé en **contre-saison dans les bas-fonds** et sur les **plateaux** avec une durée **maturation de 8 mois**.

Les rotations sont généralement **graines // tubercules**.

8.2 Répartition des cultures

15 % du domaine cultivé est composé de monoculture et 75 % de cultures associées.

8.3 Marchés

Les prix sont moins chers dans les zones les plus enclavées, car seuls les camions peuvent y accéder et il y a moins d'acheteurs. Il y a environ 40 collecteurs différents pour Antananarivo et Antrisrabe. Les petits collecteurs revendent au district. Les producteurs font aussi de la vente directe dans la commune (car c'est le plus gros marché du district). Achat local lorsque les prix sont peu élevés et vente parfois (à Antanifotsy ou ailleurs).

8.4 Homogénéité des Fokontany

Il y a 11 Fokontany accessibles (autour du Fokontany d'Antanifotsy), pour le reste la route est en très mauvais état. Toutes les zones sont cultivées lorsque cela est possible. La commune est homogène, mais le district connaît des variations. Dans les autres districts, il y a moins de poulet et plus de haricots.

8.5 Précision de distance de culture

Les **grandes exploitations** sont plutôt **loin des villages**, à environ **10 km**. Les **petites** sont proches, **entre 500 m et 1 km du bâti**. Elles sont de plus en plus nombreuses à cause des héritages et divisions. La **pente maximum** cultivée est de **25 %**. Le **long de la route nationale n°7** et de la **rivière** il y a surtout du **maïs** et des **haricots**, car peur du gel pour les autres cultures. Le maïs et les haricots résistent bien au gel, on les retrouve donc le long de la rivière.

8.6 Particularités

Le maïs est cultivé toute l'année, aussi en contre-saison. Les nouvelles variétés n'ont pas trop fonctionnées.

9 Commune d'Ambano

Discussion et enquête avec M. Rakotamalala Joseph, chef du Fokontany d'Ambano avec une certaine expérience.

Il y a 2 types de sols dans la commune, un sombre très fertile qui reçoit de plus des fertilisants (légumes) et un plus pauvre qui est rouge. Le sombre n'est jamais en jachère.

9.1 Cultures principales

- légumes (carottes, choux, pomme de terre, haricot vert)
- vergers
- riz irrigué
- maïs
- blé
- haricots

Les **légumes** (**1/3 des cultures** selon lui) sont la culture dominante dans la commune avec les **vergers**. Ils sont cultivés tout le long de l'année au même endroit. Les zones préférentielles sont sur les **plateaux et faibles pentes (légumes)**. Si l'**exploitation est petite**, les **parcelles sont proches des maisons**. Si elle est **grande**, elles sont à environ **500 m**. Pour les **vergers**, on peut les retrouver de **partout**, proche ou loin des maisons.

Le **blé** est une culture relativement récente, cultivée de **février à mai et de mai à septembre** pour la contre-saison. Le blé peut être cultivé sur des sols riches ou peu riches. Pour lui, le blé représente environ 1/3 des cultures. Il est en **rotation avec les légumes**.

Le **riz irrigué** est surtout situé dans les **bas-fonds** lorsque qu'il n'y a pas de **maraîchage**. Mais cette culture est **secondaire par rapport aux légumes**.

Le **maïs** est **souvent pur**. Il y a environ 10 ans, il était cultivé en **association avec le soja**, mais le rendement était moins bon pour les deux. On peut le cultiver sur tous les sols et sa rotation est **taro ou patate douce // maïs // soja ou légumes**.

Le **soja** pur est majoritaire, il peut être associé avec le maïs. On peut aussi le **cultiver de partout**.

9.2 Répartition des cultures

Il y a environ 90 % de la surface agraire qui est utilisée pour les cultures, les 10 % restants sont repartis entre le reboisement ou les trop mauvais sols (mais il ne sait pas vraiment dire donc peu fiable). Pour lui, s'il a un are il y a environ 50 % de vergers et 50 % de riz et légumes (exemple).

9.3 Marchés

Il y a tous les jours un marché à Ambano et il y a des petits marchés dans les communes du Nord. Avant il y avait plus de collecteurs, maintenant les producteurs vendent eux même, les petites productions à Antsirabe et les plus importantes dans tout le pays.

9.4 Homogénéité des Fokontany

Il y a **3 Fokontany** avec **peu de riz et une majorité de vergers** : Ankerambe, Andrakodavaka et Amparihindramananiolona (les **3 plus au Nord et enclavés**), car plus **froids**. Sinon le reste des Fokontany est homogène avec du riz irrigué et pluvial.

9.5 Précision de distance de culture

Il n'y a pas de différence de distance entre les cultures, mais si certains champs doivent être plus loin ce sera plutôt les champs de maïs. Les cultures sont à un **maximum de 1 km** de distance, 7 km si cultivées par quelqu'un d'autre, mais cela sera proche de sa maison. La **pente maximum** cultivée est de **25 %**.

9.6 Particularités

Il y a beaucoup de vols de cultures donc des gardiens sont payés pour surveiller les cultures lorsqu'elles sont à maturité.

10 Commune de Soanindrariny

Discussion et enquête avec M. Rakotamandimly Celestin responsable du développement rural. En présence du maire et d'autres adjoints. Ils nous ont fourni les données SIG de la commune.

Il y a 3 types de sols, rouge, sableux et de podzol.

10.1 Cultures principales

- pommes
- maïs
- riz pluvial
- soja
- riz irrigué

Secondaires :

- haricots
- tomates

Minoritaires :

- manioc
- ail

Les **vergers** sont souvent **plantés avec du maïs ou du riz** en dessous. Parfois du soja ou des haricots. Cela est du à la faible surface qu'on les agriculteurs d'où l'optimisation maximale. Ils sont situés sur les **plus hautes altitudes, surtout versants Nord et Est**.

Le **maïs** est cultivé sur tous les types de sol et est principalement associé avec le **soja** ou les **haricots**. Il n'est pas cultivé en contre-saison et sa période de **semi est en octobre et sa récolte fin avril – mai**. On le retrouve sur **tout type de reliefs**. Sa rotation est souvent **maïs // maïs // riz pluvial // maïs // maïs // riz pluvial**.

Le **riz pluvial** est cultivé sur les mêmes **zones que le maïs** et selon les **mêmes caractéristiques**. Il est **semé lorsque la pluie arrive**, en général **fin octobre – début novembre** et **récolté en avril**. Même rotation que pour le maïs.

Le **soja** est souvent **pur** (car meilleur rendement) et **parfois associé au maïs**. Il n'y a **pas schéma de rotation** pour cette culture. Le **semi** est en **même temps que le maïs (octobre)** et la **récolte en avril**. On le retrouve sur tous les types de reliefs.

Le **riz irrigué** est dans les **bas-fonds**. En **contre-saison**, les **pommes de terre** sont cultivées (**ou les choux, oignons ou l'ail**). Le **semi dépend de la quantité d'eau dans les rizières**, en général en **septembre**. Le **repiquage** se faire entre **octobre et novembre** dans la majorités des cas. La **récolte est en mars-avril**.

Le **haricot** est planté en **association avec le maïs** et **planté en même temps que celui-ci**, **fin octobre**. Il est **récolté 3 mois après**. S'il est planté en février (peu de cultures), il est récolté en mai. On le retrouve de partout et en contre-saison sur les bas de pente.

Les **tomates** ne sont que au tout début de leur implantation durant la saison des pluies. Elles ne peuvent pas être cultivées en contre-saison, car il fait trop froid. Cette culture a commencé à être cultivée il y a 2 ans. Plusieurs récoltes sont possibles et cette culture est considérée comme une aide pendant la période de soudure en apportant un revenu important et en n'oblige pas à vendre directement le riz. Si les tomates sont **semées en septembre, le récolte est en décembre (3 mois 1/2)**. Possibilité d'une seconde plantation juste après. On les retrouve dans les plaines et les vallées. Il faut du fumier en plus que pour les autres cultures, mais les rendements sont 3 fois supérieurs au riz ou maïs. Sa rotation est **maïs // tomates // riz pluvial (car fertilisants l'année d'avant)**.

Le **manioc** n'est présent que sur de petites parcelles de manière familiale, tout comme l'**ail**.

10.2 Répartition des cultures

La commune est assez intensément cultivée, mais il y a aussi beaucoup de bois et de pommiers. La surface non cultivée est assez faible, il ne reste que les sols trop pentus ou trop pauvres.

10.3 Marchés

Il y a une très forte autoconsommation des produits et ventes pour les devoir sociaux (morts, études, santé …). Il y a deux petits marchés, le mercredi à Soanindrariny et le mardi à Bemasoandro. La vente est surtout locale, mais si la production est importante la vente se fait plus loin à l'aide de camions. Les pommes sont exportées dans tout Madagascar, mais localement ce sont les collecteurs locaux qui sont présents.

10.4 Homogénéité des Fokontany

Les Fokontany sont homogènes, mais il y a des spécificités dans chaque Fokontany. Il y a quelques zones qui sont plus propices pour les pommiers (en altitude) ou bien pour le riz pluvial (plus bas).

10.5 Précision de distance de culture

En général, les cultures sont à **1 km des habitations**, mais presque tous les espaces sont cultivés donc les champs peuvent être plus loin. Les vergers sont à côté des maisons, d'une distance allant du jardin et à **500 m maximum**. Les **pins et eucalyptus** sont plus loin (**5 km environ**). Il n'y a pas de hiérarchie spatiale entre les cultures. Cela dépend où l'agriculteur possède ses parcelles. Les **routes n'ont pas vraiment d'impact**, car elles relient toujours les habitations. **Proches des cours d'eau il y a plus de cultures**. Les maisons sont souvent construites au milieu des parcelles. La **pente maximum est entre 30 et 35 %**.

10.6 Particularités

Les routes sont très abîmées, les camions ne passent plus à certains endroits et les charrettes sont les seules à faire la liaison. Il y a une volonté de lancer la culture de tomates durant la saison des pluies ainsi que l'ail pour augmenter les revenus. Les pommiers déclinent, car ils sont vieux (moins de production) et il y a des ravageurs (champignons – bêtes) qui affaiblissent les vergers. Les responsables ont peur pour l'avenir commercial de la commune.

11 Commune d'Ambatomena

Discussion et enquête avec Mme. Ravaoharisoa Marie Jacqueline, 1ère adjointe au maire.

Il y a 3 types de sols, rouge, noir et sableux (moins fertile).

11.1 Cultures principales

Principales :

- riz pluvial
- riz irrigué
- maïs
- haricots
- pommes
- soja

Secondaires :

- patates douces
- pomme de terre

le **riz pluvial** est surtout cultivé sur les sols rouges avec du maïs le long des bordures. Le **semi** se fait en **septembre** (automne) et la **récolte** en **mars**. La **pluie n'est pas forcément attendue** pour le semi. On retrouve surtout le riz sur les **plateaux**. La rotation est **riz // jachère (si sols pauvres) ou maïs-haricots**.

Le **riz irrigué** est **semé en septembre** puis **repiqué en octobre**. La récolte est en **mars (environ 6 mois)**. En contre-saison on retrouve des pommes de terre, mais cela n'est valable que pour certains agriculteurs.

Le **maïs** peut être cultivé sur tous les sols, **souvent associé avec des haricots**, il est rare d'observer des parcelles pures. Il est **semé** en **juillet** et **récolté** en **décembre dans les plaines**. Pour les **plateaux** (et un peu partout), il est **planté selon la pluie entre octobre et novembre**, puis **récolté en avril – mai**. Sa rotation est **maïs // maïs // riz pluvial // maïs // maïs // riz pluvial**.

Les **haricots** sont presque toujours **associés au maïs**. Il y a **3 saisons possibles** : semi en juillet, récolte en décembre, semi en décembre, récolte en mars, semi en mars, récolte en mai.

Les **patates douces** sont cultivées en **petites parcelles**, mais sont **très nombreuses**. Elles n'ont **pas de place précise dans la rotation des cultures**. Si le **sol est pauvre** en général des patates douces peuvent être plantées **ou bien du soja**. Elles sont **plantées selon la pluie**, entre **décembre et mars**. Et sont **récoltées entre 6 mois et un 1 an**. On les retrouve surtout en **plaine**.

Les **pommes de terre** ne sont **pas très nombreuses pendant la saison des pluies** et servent à **ouvrir les nouveaux champs**. Elles sont surtout cultivées sur les **plateaux, car il y a moins de cultures**. On ne les cultive qu'**une fois sur une parcelle**. Elles sont **semées** en **octobre** et **récoltées** en **avril** (4-5 mois). **Après, des patates douces ou du maïs peut être plantés**.

Les **vergers** sont très souvent **associées avec d'autres cultures** (maïs, haricots, soja, patate douce ...) par peur que les zébus mangent les pommes. Les vergers sont surtout présents dans les **plaines**.

Le **soja** est souvent **associé au maïs**, mais est **dominant dans le champ**. Il est **semé** en **décembre** et **récolté** en **mai -début juin** (6 mois). On trouve cette culture de **partout**.

11.2 Répartition des cultures

La savane est assez rare, il y a environ 10 % de la surface de la commune qui est occupée plantations de bois et environ 90 % de zones de cultures. Presque toutes les terres de la commune sont cultivées.

11.3 Marchés

Il y a un marché local tous les vendredi. Il y a surtout des collecteurs qui achètent les produits et la vente directe est assez rare. Les pommes sont achetées à l'avance.

11.4 Homogénéité des Fokontany

Les Fokontany sont homogènes.

11.5 Précision de distance de culture

Presque toutes les pentes sont cultivées (le maximum n'a pas été communiqué). En **plaine**, on retrouve le **riz irrigué**, sur les **tanety** un **mélange de cultures et de vergers**, quelques **pins** aussi. **Les bois sont assez isolés (5km)**. Le **grand maximum de distance est de 5-6 km de l'habitat** (car il y a de l'héritage et une très forte pression foncière). Les **vergers** sont au **maximum à 2 km**, mais majoritairement **proche des habitations**. Il n'y a pas de liens avec les routes ni avec les cours d'eau.

11.6 Particularités

Pas de cartes ni de monographie.

12 Commune d'Ambohimiarivo

Discussion et enquête avec M. Randriamihaje Riconiaina, maire.

Il y a 3 types de sols dans la commune : sableux (à l'Est de la rivière), rouge (le plus fertile) et noir qui est dans les bas-fonds.

12.1 Cultures principales

Principales :

- riz pluvial
- soja - maïs
- riz irrigué

Secondaires :

- haricots verts
- patates douces
- arachides
- poids de terre
- artémisia
- haricots
- pomme de terre
- tomates

Le **riz pluvial** est principalement pur et planté sur les sols fertiles. Il est **semé** entre **mi-septembre** et **novembre** (avant pas d'attente de la pluie, mais maintenant si) et **récolté en mars, début avril**. On les retrouve surtout sur les **tanety**, mais de **partout**. Sa rotation est **tomates (ou pdt ou patates) // riz // riz // tomates (ou autres)**.

Le **soja-maïs** est cultivé sur **tous les sols**. Le **soja est dominant**. Les 2 sont semés en **octobre**, le **maïs récolté en mars/fin avril** (car bien sec) et le **soja en mars-avril**. On les retrouve sur les **tanety** de **partout**. Si le sol est pauvre les rotations sont **soja-maïs // jachère // soja-maïs**.

Le **riz irrigué** a 2 saisons possibles.
La première : **semé en septembre puis repiqué en novembre**. La **récolte est mi-avril**.
La seconde : semé en juin puis repiqué en août et la récolte en janvier (sur les zones proches de la rivière, car il y a tout le temps de l'eau. En contre-saison on retrouve des pommes de terre ou haricots ou brèdes, mais cela n'est valable que pour certains.

Les **haricots verts** ont été lancés par Lecofruit. Le maire les voit peu, mais il voit beaucoup les camions des collecteurs.

La **patate douce** est cultivée sur des **petites parcelles**, mais il y en a beaucoup. Souvent en alternance avec le riz ou le maïs (**riz ou maïs // patates douces // riz ou maïs**). On peut les

retrouver de **partout**. Le **semi** est effectué soit en **novembre** soit en **avril** puis **récolte 3-4 mois après**.

Les **arachides** sont cultivées en **petites parcelles** et en alternance avec le **maïs-soja ou maïs-haricots // arachides**. On les retrouve de partout. Semi en **octobre, récolte en avril / mai**.

Les **pois de terre** sont sur les **petites parcelles** pour commencer les rotations avant le maïs. Semi en octobre – novembre et récolte en mars-avril. On peut les cultiver de partout (c'est-à-dire plus sur les plateaux, car nouvelles parcelles).

Les **haricots** sont surtout **associés au maïs**. Ils sont purs en contre-saison dans les bas-fonds. Le **semi** est en **novembre** et la **récolte** en **mars**. Sinon, semi en février et récolte en mai. On les retrouve de **partout**.

Les **tomates** peuvent être plantées à **tout moment**, sur les **tanety** la maturité arrive **3 mois après le semi**. Souvent semi entre septembre et février. Rotation : **tomates // riz pluvial (car fertilisants)**.

12.2 Répartition des cultures

Il y a environ **80 %** des terres de la commune qui sont en **culture (pins compris)**. Il y a quelques pâturages dans la commune. Il y a environ **50 h de pêchers**, mais peu de producteur de grandes surfaces, surtout agriculture familiale.

12.3 Marchés

Il y a un marché local, mais les collecteurs viennent acheter les cultures.

12.4 Homogénéité des Fokontany

Le Fokontany d'**Ambatolahimarina** cultive des tomates durant la saison des pluies et en contre-saison. Cela n'est lié qu'à des différences culturelles. Les autres Fokontany font plus de riz.

12.5 Précision de distance de culture

Les **cultures** sont entre **300 m et 5 km maximum des habitations** ; le maximum « moyen » est d'environ 3 km. Il n'y a pas de hiérarchie spatiale entre les cultures. Pas de privilèges non plus pour les zones proches de routes ou des cours d'eau. Il n'y a pas de vraie limite de pente dans la commune.

12.6 Particularités

Pas de monographie.

Annexe G

Position topographique ou relief

Cette méthode consiste à générer un positionnement du pixel par rapport à la topographie proche définie par le MNT (500 mètres) et plus éloignés (2000 mètres). Ces deux grilles sont combinées avec les pentes afin de créer une typologie du relief en 10 classes. Dans cette thèse, nous avons assimilée cette position topographique à la géomorphologie.

Table 23 Classes de la position topographique de de chaque pixel selon la méthode Weiss (2001)

Code	Description
1	Vallée de montagne ou vallée un peu profonde, parfois utilisée pour l'agriculture
2	Bas-fond ou limite de bas fond
3	Haut de pente de talweg de montagne
4	Fond de vallée proche de montagne, englobant la partie irriguée et le bas de pente des collines ou de la montagne. Généralement proche de reliefs élevés. Les plaines ne sont pas forcément irriguée
5	Plaines ou bas-fonds, cuvettes planes
6	Pentes de colline ou de cours d'eau incisé
7	Plateaux
8	Petite colline
9	Haut de colline entre des petits cours d'eau
10	Sommets de montagne et hauts plateaux

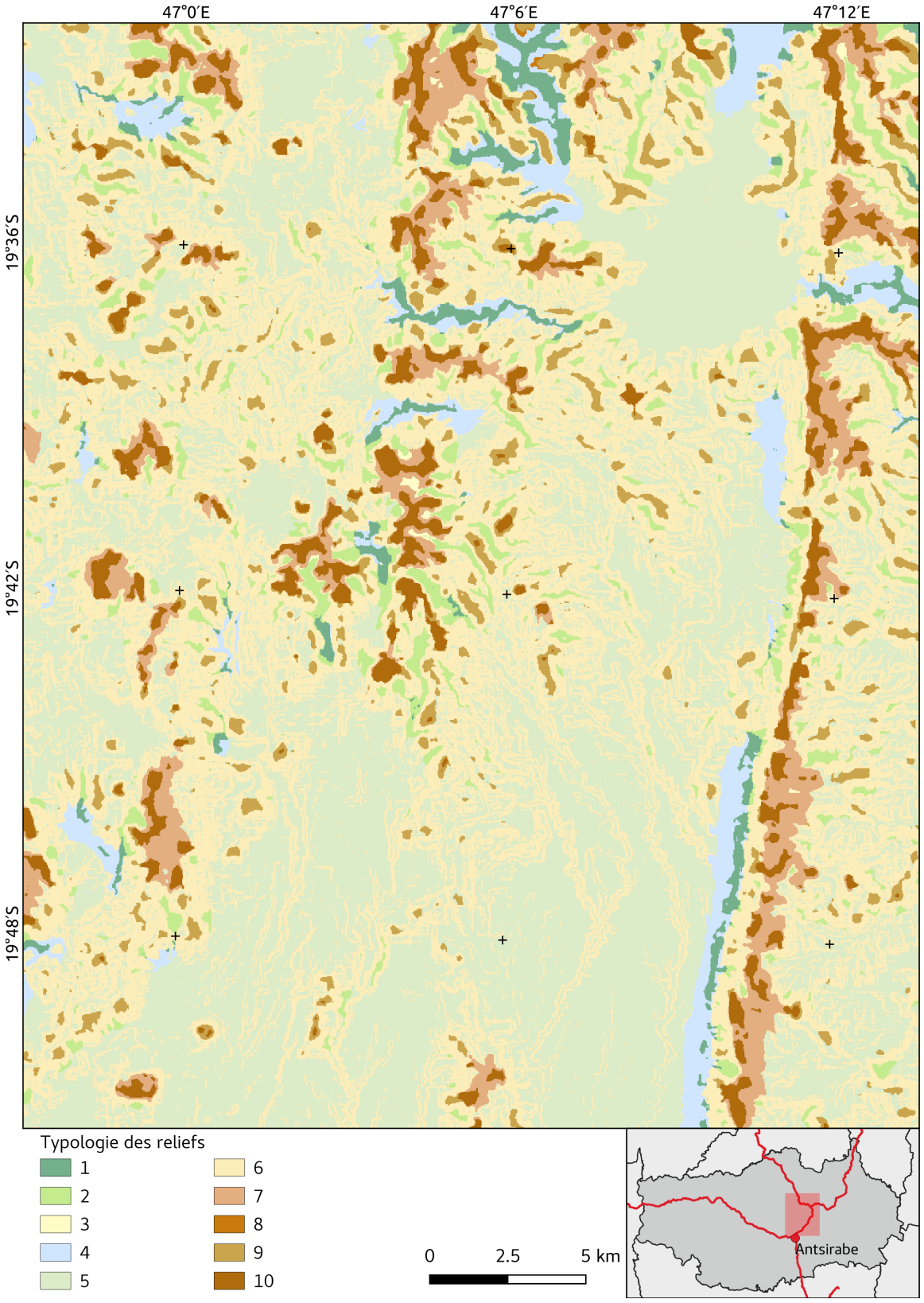


Figure 33 Carte des types de reliefs de notre zone d'étude selon la méthode (Weiss, 2001). Les classes sont décrivent dans le tableau 23

Annexe H

Production de la carte du temps de trajet à pied entre un segment et un bâtiment

Dans cette annexe, nous présentons donc la méthode de calcul de la variable « temps de trajet entre un segment et un bâtiment » nécessaire à l'application de règles spatiales dans le module 2. Le calcul de cette donnée se fait en trois étapes : 1) l'identification des segments qui peuvent être labellisés comme des bâtiments à partir de la classification des données de télédétection, 2) le choix d'une fonction de coût de déplacement pour estimer le temps de trajet à pied entre un bâtiment et un segment et, 3) l'implémentation de l'automate cellulaire dans la plateforme Ocelet. La dernière sous-section est consacrée à l'intégration des données support dans les segments pour pouvoir ensuite appliquer les règles.

1 Identification des segments de *bâti*

La solution pour créer une donnée représentant le temps de trajet entre les habitations et les éléments non considérés comme du bâti a été de simplifier le type « habitation » par le type « bâti » qui correspond à notre nomenclature de données (tableau 3). Notre choix est de se baser sur notre propre classification de l'occupation du sol, avec tous les biais qui lui sont liés. La classe *bâti* obtient parmi les meilleurs résultats de classification avec une précision et un rappel satisfaisants (précision : 0,92 ; rappel : 0,99 et F-mesure : 0,96, tableau 1.2)[1]. Ainsi, pour éviter de trop nombreux faux positifs, seuls les segments ayant une probabilité d'appartenance à la classe *bâti* supérieure à 0,9 dans le $VPAC_{RF}$ sont conservés.

2 Temps de déplacement

Pour estimer le temps de trajet entre deux segments, la principale méthode existante est de calculer la vitesse de marche en fonction de l'occupation des sols (Balstrøm, 2002), soit une fonction de coût de déplacement, mais cette méthode n'est pas utilisable dans notre cas, car nous cherchons à cartographier cette même occupation des sols. Notre fonction de coût de déplacement s'est donc basée sur l'équation de déplacement de Tobler (1993) (équation 20) représentant la variation de la vitesse de marche en fonction de la pente. Cette équation est utilisée dans plusieurs études

[1] Ces résultats sont présentés dans la section 1.2 Analyse des résultats de classification par classe du chapitre 5

de temps de trajet ou de trajets optimaux (Sherrill et al., 2010; White and Barber, 2012; Richards-Rissetto and Landau, 2014). L'utilisation de l'équation 20 ne s'appuie pas sur les mêmes temps de trajet en fonction d'études, nous avons décidé de choisi une vitesse moyenne mesurée comme vitesse maximale (ws). C'est le résultat de l'étude de Buchmüller and Weidmann (2006), soit une vitesse moyenne de 4,824 km/h sur du plat qui est utilisé comme vitesse maximale (ws). Cette vitesse est choisie, car elle représente une vitesse de marche acceptable sur une surface plane pour de courtes distances avec des outils et des enfants pour aller travailler au champ.

$$W = ws\, e^{-3.5\left|\frac{dh}{dx} + 0.05\right|} \tag{20}$$

W est la vitesse de marche en m/s,
ws est la vitesse de marche maximale en m/s,
dh est la différence d'altitude en mètre,
et dx est la distance en mètre, soit $\frac{dh}{dx} = pente$ en radians.

3 Implémentation

Le MNT présenté dans le chapitre 3, section 2.3 Données support pour l'application des connaissances est utilisé pour spatialiser l'équation 20. Pour déterminer le chemin le plus court, l'algorithme de Dijkstra (1959) a été appliqué à un voisinage de Moore d'ordre 1 (8 cellules) correspondant aux pixels du MNT. La fonction de coût de l'algorithme de Dijkstra (1959) correspond au calcul de l'équation 20 entre les centroïdes de chaque cellule, où la valeur d'une cellule va changer en fonction de ses voisines. Le processus de modélisation débute par l'initialisation des cellules avec une valeur extrême théorique, correspondant au temps de trajet, ainsi qu'une initialisation de la valeur des cellules de bâti à 0, donnant les points de départ du processus (figure 34, (1)). À partir des cellules correspondant aux bâtiments, la fonction de coût est appliquée aux cellules voisines (figure 34, (2)). Une fonction d'agrégation est utilisée pour déterminer quelle est la valeur minimale entre les valeurs candidates (pour le voisinage de Moore d'ordre 1, la valeur de coût est choisie entre la valeur initiale et la valeur nouvellement calculée). La valeur retenue devient la nouvelle valeur coût de la cellule si est inférieure à la valeur précédente (figure 34, (3)). Tant que la valeur d'une cellule peut être diminuée, ce processus est itéré jusqu'à ce que la valeur de chaque cellule ne puisse plus être diminuée avec un élargissement du voisinage (figure 34, (4)). Ce processus de modélisation permet d'obtenir une grille qui correspond au temps de trajet le plus court entre chaque cellule voisine et le bâtiment le plus proche en temps de déplacement (figure 34, (5), figure 35 est la carte obtenue).

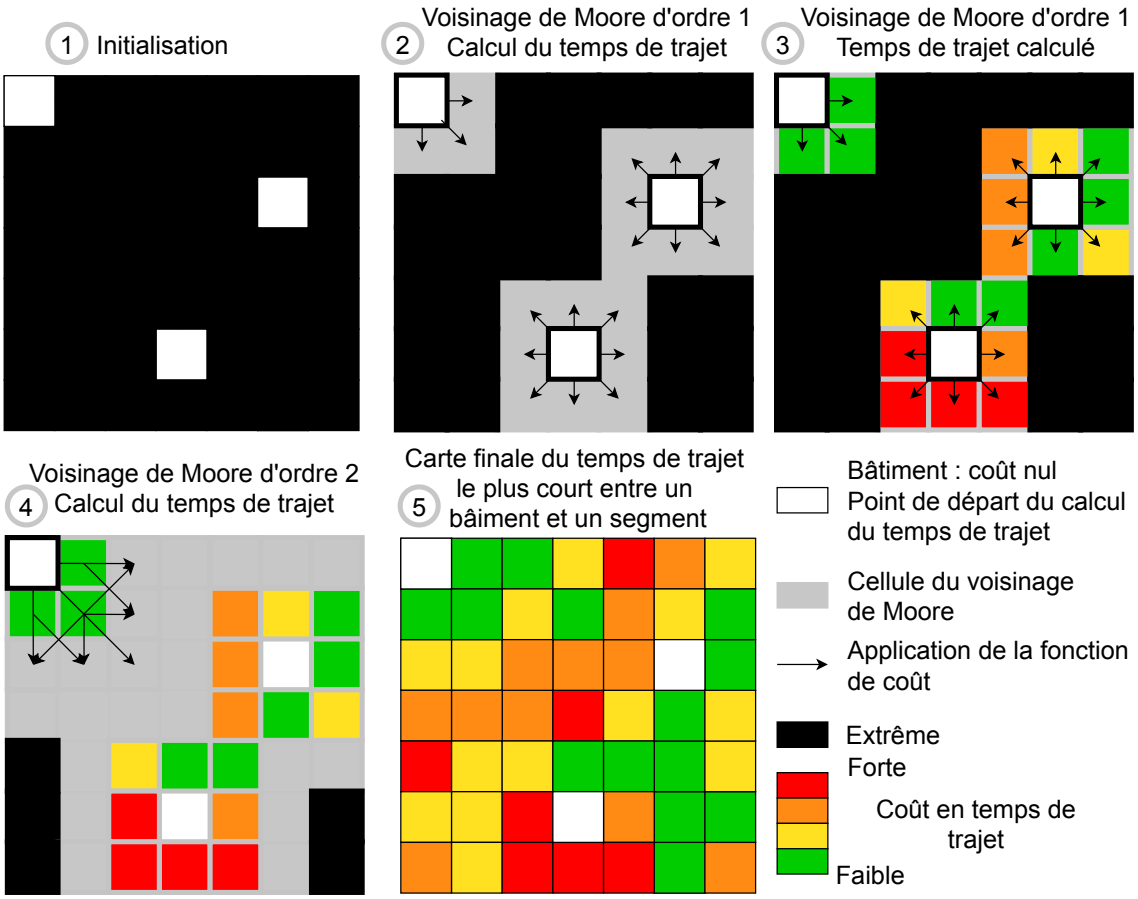


Figure 34 Schéma du calcul du temps de trajet selon l'équation de Tobler (20) appliqué aux pixels d'un MNT selon un voisinage de Moore. Pour une facilité de lecture, les flèches d'application de la fonction de coût du voisinage de Moore d'ordre 2 ne sont représentées que un point de départ.

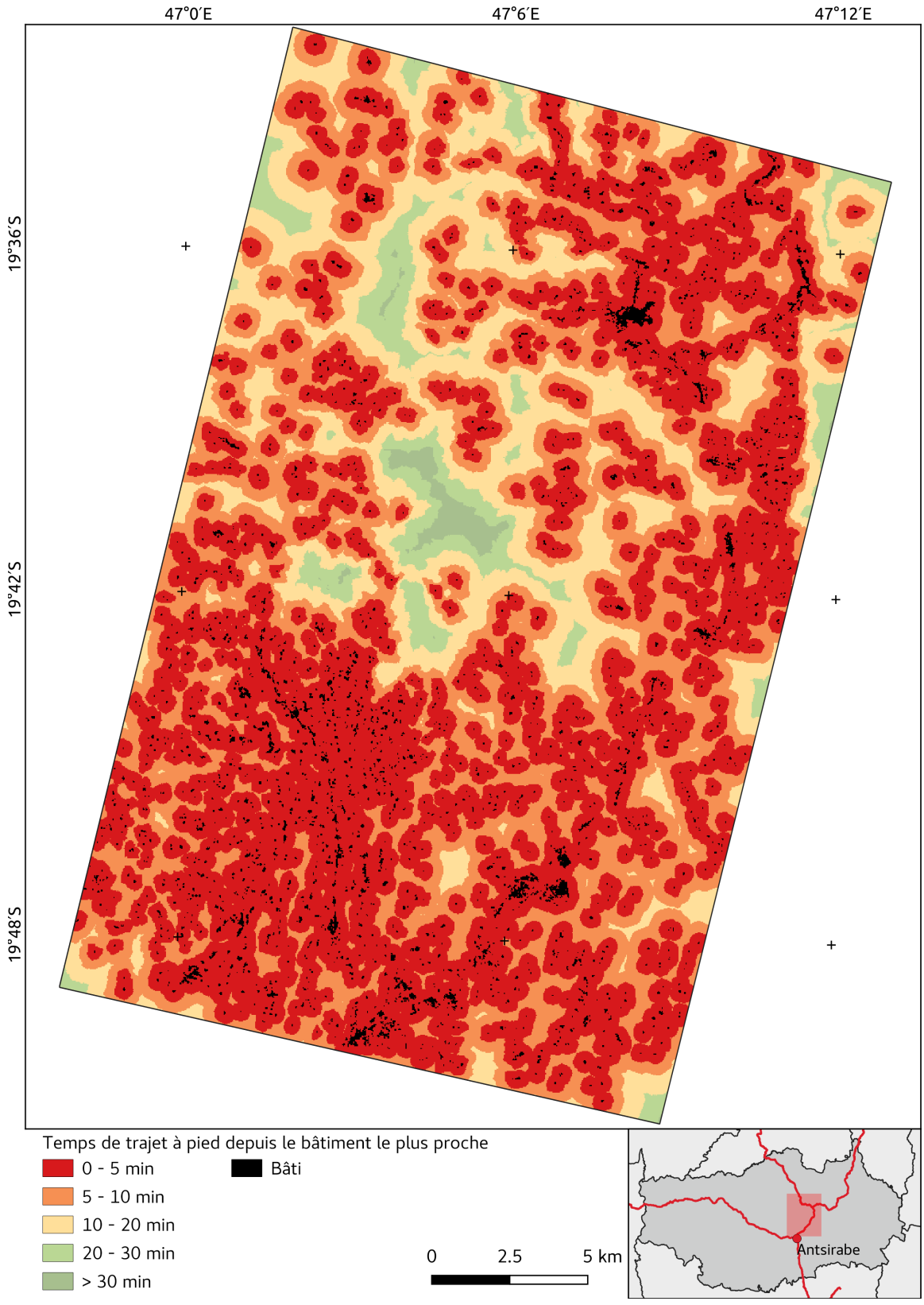


Figure 35 Carte du temps de trajet entre un segment et le bâtiment le plus proche.

Annexe I

Matrices de confusion

Chaque ligne de la matrice de confusion correspond à une classe réelle, chaque colonne correspond à une classe estimée. La cellule correspondant au croisement de la ligne et de la colonne de la même classe correspond à part correctement attribuée.

Table 24 Correspondance entre les classes et les codes pour les matrices de confusion.

Occupation du sol (français)	Code
Maïs	1
Riz irrigué	2
Riz pluvial	3
Légumes	6
Soja	7
Pomme de terre	10
Herbes et autre cultures fourragères	21
Verger	23
Culture boisée de rente	24
Sol nu agricole	25
Cultures potagères	30
Savane	104
Sols rocheux	106
Bâti	108
Eau	109
Maïs associé	1001
Riz pluvial associé	1002

Table 25 Matrice de confusion de la classification des données de télédétection. Les résultats sont exprimé en pourcentage d'aire. La correspondance des classes est présentée dans le tableau 24.

	1	**2**	**3**	**6**	**7**	**10**	**21**	**23**	**24**	**25**	**30**	**104**	**106**	**108**	**109**	**1001**	**1002**	**Somme**
1	7,55	0,09	0,40	0,05	0,00	0,00	0,03	0,00	0,00	0,00	0,01	0,17	0,00	0,00	0,00	0,11	0,02	8,44
2	0,00	5,92	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	5,96
3	0,00	0,05	0,17	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,06	0,07	0,00	0,00	0,00	0,22	0,30	0,94
6	0,02	0,01	0,00	3,07	0,08	0,44	0,00	0,01	0,04	0,00	0,19	0,00	0,00	0,00	0,00	0,00	0,00	3,87
7	0,22	0,06	0,00	0,08	0,93	0,02	0,01	0,09	0,06	0,00	0,12	0,75	0,00	0,01	0,00	0,17	0,00	2,51
10	0,04	0,02	0,00	0,48	0,00	1,91	0,01	0,03	0,04	0,00	0,13	0,05	0,00	0,07	0,00	0,01	0,02	2,81
21	0,00	0,00	0,00	0,00	0,00	0,00	6,84	0,00	0,01	0,00	0,00	0,29	0,00	0,00	0,00	0,02	0,00	7,16
23	0,02	0,00	0,00	0,14	0,00	0,02	0,00	2,57	0,27	0,00	0,07	0,06	0,00	0,01	0,00	0,05	0,01	3,22
24	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,04	11,79	0,00	0,01	0,50	0,00	0,31	0,00	0,00	0,00	12,66
25	0,00	0,00	0,00	0,09	0,00	0,17	0,00	0,00	0,00	0,00	0,02	0,02	0,00	0,00	0,00	0,00	0,01	0,31
30	0,00	0,00	0,00	0,48	0,01	0,15	0,07	0,03	0,04	0,00	0,70	0,11	0,00	0,01	0,00	0,01	0,01	1,62
104	4,45	0,03	0,04	0,12	0,10	0,01	0,02	0,02	3,12	0,00	0,23	30,58	0,00	0,02	0,00	0,07	0,01	38,81
106	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,19	0,20	0,06	0,00	0,00	0,00	0,44
108	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	6,32	0,00	0,00	0,00	6,36
109	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	2,28	0,00	0,00	2,34
1001	0,05	0,00	0,13	0,04	0,04	0,00	0,00	0,14	0,03	0,00	0,12	0,24	0,00	0,00	0,00	0,82	0,13	1,73
1002	0,08	0,00	0,09	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,04	0,13	0,00	0,00	0,00	0,15	0,30	0,81
Somme	12,43	6,23	0,82	4,60	1,23	2,71	6,98	2,92	15,42	0,01	1,71	33,17	0,20	6,84	2,28	1,64	0,81	100,00

Table 26 Matrice de confusion de la fusion des $VPAC_{RF}$ (classification des données de télédétection) et $VPAC_{RS}$ (module 2 : règles spatiales en un $VPAC_{Fusion}$ selon la méthode de la moyenne pondérée avec les paramètres $\Delta RF = 0,5$ et $coef\ wRS = 1$. La correspondance des classes est présentée dans le tableau 24.

	1	**2**	**3**	**6**	**7**	**10**	**21**	**23**	**24**	**25**	**30**	**104**	**106**	**108**	**109**	**1001**	**1002**	**Somme**
1	7,55	0,09	0,37	0,05	0,00	0,03	0,03	0,00	0,00	0,00	0,01	0,17	0,00	0,00	0,00	0,11	0,02	8,44
2	0,00	5,92	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	5,96
3	0,00	0,05	0,17	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,06	0,07	0,00	0,00	0,00	0,24	0,28	0,94
6	0,02	0,01	0,00	3,07	0,06	0,46	0,00	0,01	0,04	0,00	0,19	0,00	0,00	0,00	0,00	0,00	0,00	3,87
7	0,19	0,05	0,00	0,06	0,95	0,09	0,00	0,09	0,05	0,00	0,12	0,73	0,00	0,01	0,00	0,19	0,00	2,51
10	0,02	0,02	0,00	0,48	0,00	2,01	0,00	0,01	0,04	0,00	0,11	0,04	0,00	0,07	0,00	0,01	0,00	2,81
21	0,00	0,00	0,00	0,00	0,00	0,00	6,84	0,00	0,00	0,00	0,00	0,29	0,00	0,00	0,00	0,02	0,00	7,16
23	0,00	0,00	0,00	0,14	0,00	0,00	0,00	2,62	0,27	0,00	0,07	0,06	0,00	0,00	0,00	0,05	0,01	3,22
24	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,04	11,79	0,00	0,01	0,50	0,00	0,31	0,00	0,00	0,00	12,66
25	0,00	0,00	0,00	0,09	0,00	0,17	0,00	0,00	0,00	0,00	0,01	0,03	0,00	0,00	0,00	0,00	0,01	0,31
30	0,00	0,00	0,00	0,46	0,01	0,16	0,06	0,04	0,04	0,00	0,70	0,12	0,00	0,00	0,00	0,02	0,00	1,62
104	4,45	0,03	0,04	0,12	0,10	0,01	0,02	0,02	3,12	0,00	0,22	30,58	0,00	0,02	0,00	0,07	0,01	38,81
106	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,19	0,20	0,06	0,00	0,00	0,00	0,44
108	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	6,32	0,00	0,00	0,00	6,36
109	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	2,28	0,00	0,00	2,34
1001	0,09	0,00	0,13	0,04	0,09	0,03	0,00	0,14	0,03	0,00	0,04	0,20	0,00	0,00	0,00	0,81	0,14	1,73
1002	0,10	0,00	0,09	0,00	0,00	0,03	0,00	0,00	0,01	0,00	0,03	0,13	0,00	0,00	0,00	0,16	0,27	0,81
Somme	12,42	6,22	0,80	4,56	1,28	2,98	6,96	2,97	15,38	0,01	1,58	33,12	0,20	6,83	2,28	1,69	0,73	100,00

Annexe J

Résumé des règles utilisées dans le modèle par type de connaissance

1 Pédologie

Table 27 Correspondances entre les classes et les règles appliquées d'après une simplification de la carte de (Riquier, 1968). Les classes simplifiées correspondent à : **1** : nature Sols peu évolués / sols hydromorphes ou code II6,X3 ; **2** : type 2 sols lessivés ou code VIII11 ; **3** : type 2 sols minéraux bruts ou code I1 ou code I7 ; **4** : nature sols a mull ou type 4 « volcan », ou code VI1, ou code VIII12, ou code VIII13 ; **5** : type 1 sols ferralitiques et type 2 bruns - rouge, ou type 3 rouge, ou code VIII9 ; **10** : nature plan d'eau.

	Pédo-géomorphologie					
	1	**2**	**3**	**4**	**7**	**10**
Maïs	faible	faible	très faible	moyenne	moyenne	/
Riz pluvial	faible	faible	très faible	moyenne	moyenne	/
Soja	faible	faible	très faible	moyenne	moyenne	/
Légumes	moyenne	très faible	très faible	moyenne	faible	/
Pomme de terre	moyenne	très faible	très faible	moyenne	faible	/
Fruit	faible	faible	très faible	élevée	faible	/
Maïs associé	faible	faible	très faible	moyenne	moyenne	/
Riz pluvial associé	faible	faible	très faible	moyenne	moyenne	/

2 Distance à un cours d'eau

Table 28 Correspondances entre les classes et les règles appliquées d'après les cours d'eau dérivées du MNT de l'USGS à l'aide du logiciel Team (2018).

	Distance à un cours d'eau (mètre)				
	0 – 15	**15 – 25**	**25 – 50**	**50 – 100**	**100 – 150**
Riz irrigué	extrême	très élevée	élevée	moyenne	moyenne
Légumes	moyenne	moyenne	faible	faible	/

3 Altitude

Table 29 Correspondances entre les classes et les règles appliquées d'après l'altitude dérivée du MNT de l'USGS.

	Altitude					
	800 -1300	**1300-1500**	**1500 – 1700**	**1700 – 1850**	**1850 – 1900**	**> 1900**
Riz irrigué	élevée	élevée	moyenne	moyenne	/	/
Riz pluvial	élevée	moyenne	faible	faible	/	/
Pomme de terre	moyenne	élevée	élevée	très élevée	très élevée	très élevée
Fruit	/	/	faible	moyenne	moyenne	élevée
Riz pluvial associé	élevée	moyenne	faible	faible	/	/
Autre	faible	moyenne	élevée	élevée	élevée	élevée

4 Pente

Table 30 Correspondances entre les classes et les règles appliquées basées sur les pentes dérivées du MNT de l'USGS.

	Pente (°)								
	0 – 2	**2 – 5**	**5 – 7**	**7 – 10**	**10 – 15**	**15 – 20**	**20 – 30**	**30 – 35**	**> 35**
Maïs	moyenne	moyenne	élevée	élevée	élevée	élevée	faible	/	/
Riz irrigué	faible	faible	très faible	très faible	très faible	très faible	très faible	/	/
Riz pluvial	moyenne	moyenne	faible	faible	faible	/	/	/	/
Soja	moyenne	moyenne	élevée	élevée	élevée	élevée	faible	/	/
Légumes	élevée	élevée	moyenne	/	/	/	/	/	/
Pomme de terre	moyenne	moyenne	moyenne	faible	faible	très faible	très faible	/	/
Fruit	moyenne	moyenne	moyenne	moyenne	faible	très faible	très faible	très faible	très faible
Maïs associé	élevée	élevée	moyenne	moyenne	faible	faible	très faible	/	/
Riz pluvial associé	moyenne	moyenne	faible	faible	faible	/	/	/	/
Cultures potagères	élevée	faible	très faible	/	/	/	/	/	/
Autre	faible	faible	faible	faible	faible	élevée	élevée	élevée	très élevée

5 Relief ou géomorphologie

Table 31 Correspondances entre les classes et les règles appliquées basées sur les types de reliefs à partir de la méthode de Weiss (2001), la correspondance des types de reliefs est présenté en Annexe G.

	Relief									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Maïs	/	/	faible	moyenne	moyenne	élevée	élevée	élevée	élevée	élevée
Riz irrigué	très élevée	très élevée	très faible	moyenne	faible	faible	/	/	/	/
Riz pluvial	/	/	faible	faible	moyenne	moyenne	moyenne	moyenne	faible	faible
Soja	/	/	faible	moyenne	moyenne	élevée	élevée	élevée	élevée	élevée
Légumes	moyenne	moyenne	très faible	faible	élevée	moyenne	/	/	/	/
Pomme de terre	très faible	très faible	moyenne	faible	moyenne	moyenne	moyenne	moyenne	moyenne	moyenne
Fruit	très faible	très faible	faible	moyenne	moyenne	faible	faible	faible	faible	faible
Maïs associé	/	/	faible	moyenne	moyenne	élevée	élevée	élevée	élevée	élevée
Riz pluvial associé	/	/	faible	faible	moyenne	moyenne	moyenne	moyenne	faible	faible
Autre	/	/	moyenne	très faible	très faible	faible	faible	faible	faible	élevée

6 Exposition

Table 32 Correspondances entre les classes et les règles appliquées basées sur l'exposition dérivée du MNT de l'USGS.

	Exposition (°)			
	0 – 45	**45 -135**	**135 – 315**	**315 – 360**
Fruit	faible	très faible	/	faible

7 Temps de déplacement depuis un bâtiment

Table 33 Correspondances entre les classes et les règles appliquées basées sur le temps de déplacement à pieds depuis un bâtiment. La méthode de calcul est présenté section 2 Temps de déplacement.

	Temps de trajet (minute)				
	0 – 5	**5 – 10**	**10 – 20**	**20 – 30**	**30 – 90**
Maïs	faible	moyenne	moyenne	moyenne	/
Riz irrigué	élevée	élevée	moyenne	faible	/
Riz pluvial	faible	moyenne	moyenne	moyenne	/
Soja	faible	moyenne	moyenne	moyenne	/
Légumes	très élevée	élevée	faible	/	/
Pomme de terre	élevée	moyenne	faible	/	/
Fruit	faible	très faible	très faible	/	/
Maïs associé	faible	moyenne	moyenne	moyenne	/
Riz pluvial associé	faible	moyenne	moyenne	moyenne	/
Autre	faible	faible	faible	faible	moyenne

8 Distance à une route autre que la route nationale 7

Table 34 Correspondances entre les classes et les règles appliquées basées sur la distance entre un segment et une route (source OpenStreetMap, 2020).

	Distance à une route (mètre)		
	150 – 1 km	**1 km – 1,5 km**	**1,5 – 3 km**
Légumes	moyenne	faible	très faible

9 Distance à la route nationale 7

Table 35 Correspondances entre les classes et les règles appliquées basées sur la distance entre un segment et la route nationale 7 (source OpenStreetMap, 2020).

	Distance à la route nationale 7 (mètre)				
	0 - 100	**100 – 200**	**200 – 500**	**500 – 1 km**	**1 km – 3 km**
Mais	très élevée	élevée	moyenne	/	/
Riz pluvial	élevée	moyenne	faible	/	/
Légumes	/	/	très faible	faible	moyenne
Maïs associé	élevée	moyenne	moyenne	/	/
Riz pluvial associé	élevée	moyenne	faible	/	/

10 Distance visuelle

Table 36 Correspondances entre les classes et les règles appliquées basées sur la distance visuelle entre un segment et un polygone considéré comme du « bâti » selon type de relief. Les polygones de bâti sont extraits de la classification des données de télédétection (section 1 Télédétection). Le relief est dérivé du MNT de l'USGS selon la méthode Weiss (2001). La classe **7** correspond à des plateaux, la classe **8** a des petites collines, la classe **9** aux hauts des collines entre de petits cours d'eau et la classe **10** aux sommets de montagne et hauts plateaux.

	Distance visuelle **Si la classe de relief est 7, 8, 9 ou 10**	
	1 km < distance au bâti >= 2 km	**2 km < distance au bâti >= 3 km**
Maïs	moyenne	très faible
Riz pluvial	moyenne	très faible
Soja	moyenne	très faible
Maïs associé	faible	très faible
Riz pluvial associé	faible	très faible

Annexe K

Implémentation avec Ocelet du module 4 : modélisation de la répartition spatiale des classes

De manière détaillée voici le fonctionnement du module 4 : modélisation de la répartition spatiale des classes pour un Fokontany :

Soit :
$X = \{seg1, ..., seg\ n\}$ est l'ensemble des segments d'un modèle, où n est le nombre de segments dans l'unité spatiale étudiée,
$k \in [1 : c]$, où c est égal au nombre de classes,

1. Initialisation de la répartition théorique des classes, du seuil de fiabilité et du seuil de variabilité des aires théoriques et observées.
2. L'ensemble des segments X du groupe administratif (groupement de Fokontany) sont séparés en deux listes selon le seuil de fiabilité :
   2.1. si la probabilité la plus forte du $VPAC_{Fusion}$ du segment est inférieure au seuil, ce segment est ajouté à la **liste « non fiable »** $Lnonfiable$;
   2.2. si la probabilité la plus forte du $VPAC_{Fusion}$ du segment est égal ou supérieur au seuil, ce segment est ajouté à la **liste « fiable »** $Lfiable$.
3. Pour $Lfiable$, l'aire totale de chaque classe est calculée puis est comparée à l'aire théorique de chaque classe. Un vecteur est obtenu avec les aires excédentaires ou déficitaires.
4. Pour les classes avec une aire excédentaire, les segments avec la fiabilité la plus basse (*i.e.* probabilités les plus faibles pour la classe étudiée dans le $VPAC_{Fusion}$) sont supprimés de $Lfiable$ puis ajoutés à $Lnonfiable$. Ce processus est itéré jusqu'à ce que le rapport entre les deux aires soit équivalent, plus ou moins le seuil de variabilité appliqué rapport entre les aires.
5. Lorsque l'itération est terminée, la liste de segments $Lfiable$ est répartie en c classes. Ces nouvelles listes temporaires correspondent à l'ensemble des segments « fiables » pour chaque classe. Le déficit de chacune de ces listes sert à attribuer les segments de la liste $Lnonfiable$ à une classe appropriée.
6. À partir de la liste de segments $Lnonfiable$, une liste pour chaque classe est créée et chaque liste est triée selon un ordre croissant des probabilités de la classe concernée.

7. Tant que l'intégralité des segments n'a pas été attribuée à une classe, les segments sont assignés à une classe déficitaire de manière itérative, selon ces étapes :
   7.1. sélection de la classe à traiter k, sauf si $k > n$ (dans ce dernier cas, la classe à traiter sera la classe $k = 1$) ;
   7.2. si la liste « non fiable » de la classe k n'est pas vide, et que la classe k est déficitaire, alors ;
   7.3. le segment avec la plus forte probabilité de la classe k, seg_i, est ajouté à la liste « temporaire » de la classe k ;
   7.4. Le déficit de la classe k est recalculé en lui ajoutant l'aire de seg_i :
   7.5. seg_i est retiré de toutes les autres listes « non fiables » ;
   7.6. La classe $k + 1$ est ensuite traitée.
8. Si des segments n'ont pas été reclassés, alors leur classe est déterminée en fonction de leur probabilité la plus forte.

Annexe L

Article : Agriculturally consistent mapping of smallholder farming systems using remote sensing and spatial modelling

AGRICULTURALLY CONSISTENT MAPPING OF SMALLHOLDER FARMING SYSTEMS USING REMOTE SENSING AND SPATIAL MODELLING

A. Crespin-Boucaud[1,2,*], V. Lebourgeois[1,2], D. Lo Seen[1,2], M. Castets[1,2], A. Bégué[1,2]

[1] CIRAD, UMR TETIS, F-34398 Montpellier, France – (arthur.crespin-boucaud, valentine.lebourgeois, danny.lo_seen, agnes.begue)@cirad.fr

[2] TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France.

KEY WORDS: Landscape, Crop type mapping, Cropping practices, Sentinel-2, Spatio-temporal modelling, Smallholder agriculture

ABSTRACT:

Smallholder agriculture provides 90 % of primary food production in developing countries. Its mapping is thus a key element for national food security. Remote sensing is widely used for crop mapping, but it is underperforming for smallholder agriculture due to several constraints like small field size, fragmented landscape, highly variable cropping practices or cloudy conditions. In this study, we developed an original approach combining remote sensing and spatial modelling to improve crop type mapping in complex agricultural landscapes. The spatial dynamics are modelled using Ocelet, a domain-specific language based on interaction graphs. The method combines high spatial resolution satellite imagery (Spot 6/7, to characterize the landscape structure through image segmentation), high revisit frequency time series (Sentinel 2, Landsat 8, to monitor the land dynamic processes), and spatiotemporal rules (STrules, to express the strategies and practices of local farmers). The method includes three steps. First, each crop type is defined by a set of general STrules from which a model-based map of crop distribution probability is obtained. Second, a preliminary crop type map is produced using satellite image processing based on a combined Random Forest (RF) and Object Based Image Analysis (OBIA) classification scheme, after which each geographical object is labelled with the class membership probabilities. Finally, the STrules are applied in the model to identify objects with classes locally incompatible with known farming constraints and strategies. The result is a map of the spatial distribution of crop type mapping errors (omission or commission) that are subsequently corrected through the joint use of spatiotemporal rules and RF class membership probabilities. Combining remote sensing and spatial modelling thus provides a viable way to better characterize and monitor complex agricultural systems.

1. INTRODUCTION

The ESA (European Spatial Agency) Sentinel 2 (S2) mission provides free imagery with an unprecedented combination of high spatial (10-60m), temporal (5 days) and spectral (13 bands) resolutions, having promising applications in land use/cover characterization and crop monitoring (Drusch et al., 2012). A study carried out on 12 contrasted agricultural sites around the world showed that image time series acquired at a decametric resolution are adapted to field crop characterization for the majority of intensive agricultural systems (Inglada et al., 2015). However, methods usually used in northern countries are not efficient in the context of smallholder tropical agriculture. In such complex systems, the small size of fields and the landscape fragmentation are significant constraints to crop characterization using remote sensing. To overcome these limitations, studies proposed OBIA (Objects Based Image Analysis) approaches to improve crop identification (Peña-Barragán et al. 2011), or combined with high-resolution time series, very high resolution imagery and auxiliary spatialized data (i.e. Digital Elevation Model; Lebourgeois et al., 2017). However, limitations remain for a crop characterization in smallholder agriculture areas.

To improve current approaches, several studies combine remote sensing with other disciplines. For example, the number of papers where remote sensing is coupled with neural networks or deep learning is increasing. They integrate satellite data, and environmental or socio-economic data (called hereafter context data) as inputs into the network, in order to extract essential variables for each class. However, these researches have mainly focused on cultivated areas with relatively large fields, a large number of ground truth points, and have not, or only slightly, been applied to smallholder agricultural areas.

Only a few studies focused on the combination of remote sensing and expert rules or spatial modelling. Bailly et al. (2018) highlight that taking into account the spatio-temporal structures of fields improves small objects classification. Two main approaches can be used for modelling agro-systems. The first one consists in using expert knowledge to create crop rotation models (e.g. Dury et al. 2012). To do so, it is necessary to understand the economic, social and environmental stakes of these agroecosystems and therefore to rely on experts in these disciplines. The second consists in using machine learning methods (such as deep learning and neural network) to free from the limitations of expert knowledge. The direct integration of context data into the classifier allows the latter to learn automatically from the ground truth data. Nevertheless, this approach is particularly suitable to address one issue at a time, such as for crop rotation in Bailly et al. (2018) and with a large and diversified dataset.

To go beyond traditional approaches in remote sensing, this research aims at investigating the complementarity of remote sensing and spatio-temporal modelling to study:

- How can spatial modelling help converting remote sensing data into more reliable thematic products?
- How can remote sensing provide evidence of the landscape processes expressed as formalized knowledge in models?

Such an approach is expected to be useful for a better understanding of complex fragmented and heterogeneous

* Corresponding author

landscapes, such as those shaped by smallholder tropical agriculture in developing countries. To illustrate our method, we describe its application for the characterization of smallholder agriculture in Madagascar Highlands. This agricultural landscape is an interesting case study because of its high complexity that led to mapping difficulties when remote sensing approaches were used alone, as shown in Inglada et al., (2015) and Lebourgeois et al. (2017).

The approach is divided into three main parts: the first one concerns remote sensing data processing (producing a first land use map), the second one focuses on the spatio-temporal modelling (based on the implementation of spatio-temporal rules expressing farmers' strategies and practices), and the last one on the combination of remote sensing and spatio-temporal modelling (for obtaining a more agriculturally consistent map).

2. STUDY AREA AND DATA

2.1 Study Area

Our study area is 20 km x 30 km in size and is part of the JECAM network (Joint Experiment for Crop Assessment and Monitoring, www.jecam.org). It is located, close to Antsirabe, the administrative center of Vakinankaratra (coordinates: 19°31'56" to 19°51'32"S, 46°57'24" to 47°13'14"E) in Madagascar Highlands. This area is characterized by complex heterogeneous and predominantly agricultural landscapes. Within the cropland, the complexity is mainly due to the very low average area cultivated per household (ranging from 0.1 to 0.5 ha, according to Sourisseau et al. (2016)). There is also a large variety of crop types with different crop calendars, and the practice of intercropping is widespread. Moreover, the landscape is fragmented by the presence of natural vegetation between the fields that is phenologically synchronized with the crops. The primary growing season occurs during the rainy season, from November to April, constraining the acquisition of optical cloud-free satellite images. The crops are diversified, but rice and maize dominate. They are cultivated under rainfed (in slopes and plateaus) or irrigated conditions (mainly lowland rice paddies).

2.2 Data

For this study, we used several types of data: (i) both high spatial resolution and high revisit frequency time series satellite data, (ii) ground truth data, (iii) agro-environmental and socio-economic knowledge data on the farmers' strategies (for the spatio-temporal rules definition), and (iv) auxiliary data (for the rules implementation).

2.2.1 Remote sensing data and pre-processing: The remote sensing dataset corresponds to (i) one very high spatial resolution SPOT 6 image, and (ii) a high spatial resolution (10m-30m) time series of 37 Sentinel 2 and 9 Landsat 8 images, for the period between October 1st, 2017 and April 29th, 2018, corresponding to the main growing season.

The SPOT 6 image was acquired on February 14th 2017 around the peak of the growing season. The image was ortho-rectified, converted to Top Of Atmosphere reflectance and pansharpened using the open-source Orfeo ToolBox (OTB; Grizonnet et al., 2019) to obtain multispectral images at 1.5 m spatial resolution. Sentinel-2 image time series (Donadieu, L'Helguen, 2016) were acquired from THEIA website (`https://theia.cnes.fr`). These images are provided in surface reflectance, and with a cloud mask, thanks to the MAJA (MACCS ATCOR Joint Algorithm) processing chain (Hagolle et al., 2017). Then, additional preprocessing steps were performed using the MORINGA chain (Gaetano et al., 2019). The functions used in MORINGA include: a) the automatic resampling (with bicubic interpolation) of S2 20 m bands to 10 m resolution, b) a multi-temporal gap-filling (Inglada et al., 2017), and c) a co-registration of S2 and L8 time series on the SPOT 6 image used as a reference.

Data	Spatial resolution / scale	Temporal resolution	Data source
Satellite data			
SPOT 6	1.5	Once	CNES / Airbus Defense & Space
Sentinel 2	10 m	Every 5 days	ESA
Landsat 8	30 m	Every 16 days	NASA / USGS
Auxiliary data			
Altitude	1 arc second	Once	SRTM v3.0 Global 1 arc second
Slope	1 arc second	Once	Computed from SRTM v3.0 Global
Landform (TPI)	1 arc second	Unique	Produced from SRTM v3.0 Global
Pedology	Area at Madagascar scale	Unique	Riquier (1968)
Hydrology	1 arc second	Unique	Hydrological modelling (GRASS GIS)
Fokontany	Municipality scale	Unique	Personal communication

Table 1. Satellite and auxiliary data used in the study.

2.2.2 *In situ* data on land use: A ground truth database on actual land use was built during field surveys around the peak growing season (beginning of February to end of March 2018). GPS waypoints were recorded following an opportunistic sampling approach. Waypoints were chosen according to their accessibility (but some parts of the area could not be reached due to heavy rains affecting the road conditions), and the size of the object (at least 20 * 20 m, so that they can be identified on satellite images). We recorded land uses according to a modified version of the JECAM nomenclature (see http://jecam.org/wp-content/uploads/2018/10/JECAM_Guidelines_for_Field_Data_Collection_v1_0.pdf) to better take into account the specificities of tropical agriculture. In the case of intercropping, we recorded the main and the second crop in the field. Once the waypoints were acquired, the boundaries of each field were digitized on the SPOT 6 image, and the class labels (and other attributes) were joined to the polygon database. Additional non-cropland polygons were also added from photo-interpretation of the SPOT 6 image for easily recognizable classes such as built-up surfaces, water bodies, wetlands, mineral soils, or natural forests. At the end, 2170 polygons were obtained (1759 crop and 511 non-crop).

2.2.3 Knowledge on farmer strategies and practices: Our approach aimed at integrating agronomic, environmental or socio-economic expert knowledge that express farmer strategies and practices. To do so, three sources of information were considered: (i) documents (research papers, books, reports, technical documents, territorial analysis, etc.) describing the agricultural functioning of the region, (ii) interviews of a set of experts with different backgrounds (agronomy, geography or socio-economy), (iii) semi-directive interviews of farmers, local officials or persons involved in rural development. These last interviews were conducted at the Fokontany (municipality/village) scale using a questionnaire. The questions of the surveys concerned the local agricultural practices (crop calendars, cropping constraints and spatial distribution), and the socio-economic context that has an impact on farmer strategies.

2.2.4 Auxiliary Data: These include existing spatial information such as altitude, slope and crop calendars. The full list of auxiliary data used is given in Table. The topographic position index (TPI) was used to generate the local and global topography using a multi-scale grid from the digital elevation model (Weiss, 2001) . The hydrological network was modelled with GRASS GIS (GRASS Development Team, 2017) to obtain an active network for the rainy season.

3. METHODOLOGY

The overall approach involves remote sensing data processing, spatio-temporal modelling and their combination (Figure 1). It is divided into five steps:

(i) A preliminary RS-based crop type map is produced using satellite image processing based on a combined OBIA and Random Forest (RF) classification scheme. Each object is labelled with its corresponding class membership probabilities, i.e. the probability that the object pertains to each of the classes (hereafter referred to as CMVPsat, for class membership vector of probabilities based on satellite image processing).

(ii) Each crop type is defined by a set of general spatio-temporal rules (STrules) that are formalized from the analysis of knowledge gathered on farmer strategies and practices.

(iii) The STrules are implemented in a modelling platform to produce a model-based map of crop distribution probability (CMVPrule for STrules-based class membership probabilities) of each object.

(iv) Possible misclassified objects are detected following specific rules.

(v) The membership probability of the selected objects is recalculated (to obtain a corrected class membership vector probabilities, CMVPcorr) to produce a corrected and more agriculturally consistent land use map (Figure 1).

3.1 Remote sensing

The preliminary land use map was produced from remote sensing. This involved the following steps: (i) segmentation of the very high spatial resolution (VHSR) image into homogeneous objects, (ii) extraction of spectral and textural features from satellite images and other products (DEM), and (iii) training and validation of a Random Forest classifier which was then applied to the whole study area. For this study, all steps were performed using the MORINGA processing chain (Gaetano et al. 2019), which uses OTB and Gdal functions driven by Python scripts.

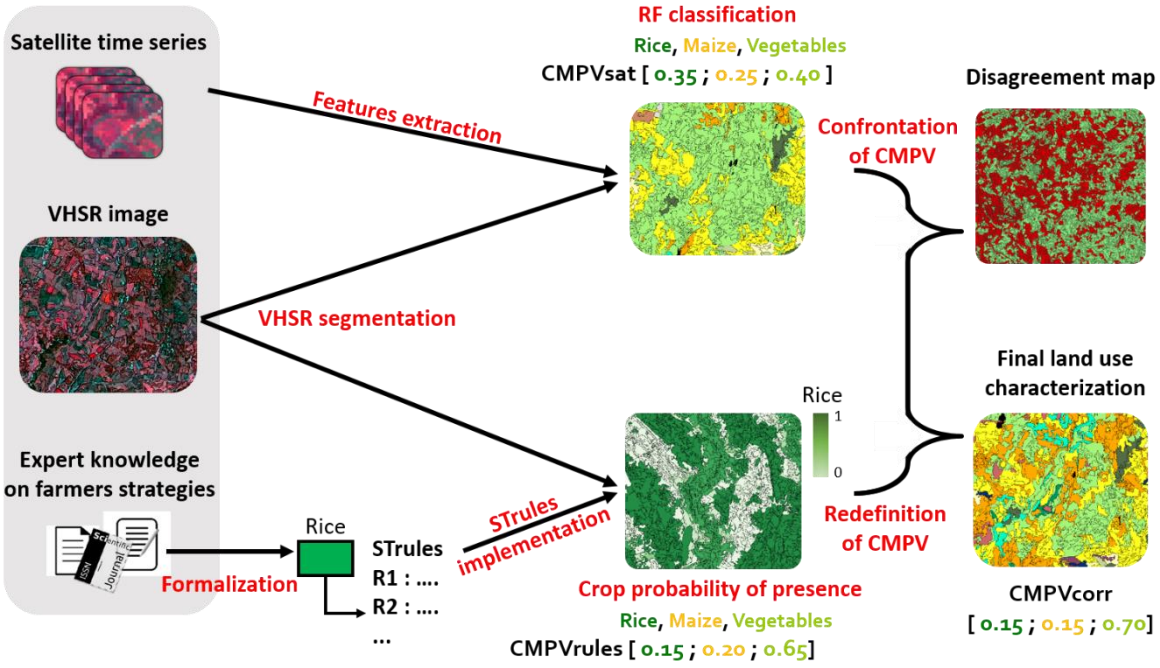


Figure 1: Framework of the method. The processes are presented in red color. CMVPsat corresponds to class membership vector probabilities obtained after random forest (RF) classification of satellite images. CMVPrules refers to the same vector obtained after spatio-temporal rules (STrules) implementation in the model. Finally, CMVPcorr presents the redefinition of the class membership vector probabilities from the two previous ones using fusion methods.

3.1.1 OBIA: According to the OBIA approach (Blaschke et al., 2014), the segmentation of the VHSR image aimed at producing objects delineating the boundaries of agricultural fields, or at least groups of homogeneous fields to characterize the landscape structure. This segmentation determines the landscape structure of the study area, i.e. the boundaries of the "entities" that will be used during the complete workflow. The Baatz and Schäpe segmentation algorithm (Baatz and Schäpe, 2000) implemented in OTB was used with the following parameters: 0.7 for the homogeneity criteria, 0.5 for the spectral homogeneity and the threshold was fixed at 200.

3.1.2 Feature extraction: Satellite features were extracted for each entity of the ground truth database to build a learning database. This step involved the extraction of the following features: (i) reflectance from the Sentinel-2 and Landsat-8 time series (471 features), (ii) spectral indices from the same time series (267 features), (iii) textural indices from the Spot 6 panchromatic image (6 features), and (iv) altitude and slope from the digital elevation model (2 features). Those features were chosen according to a previous study by Lebourgeois et al. (2017). Concerning the spectral indices, we computed the most common ones: NDVI (Rouse et al., 1974), NDWI (Gao, 1996), MNDWI (Xu, 2006), MNDVI (Jurgens, 1997), Brightness Index, and RNDVI (Jorge et al., 2019) for each image of Sentinel-2 and Landsat 8. Haralick textures (Haralick et al. 1973) were computed using mean and correlation for internal textures and contrast as edge texture. Two windows were chosen (7 pixels and 14 pixels) to traduce the diversity of fields organisation???. SRTM-30m ("SRTM V3.0, 1 arcsec") (NASA JPL, 2014) was used to extract altitude and to compute slope.

3.1.3 Classification: In the final step of the remote sensing approach, the objective was to obtain a land use map and to extract and a class membership vector of probabilities for each object (CMVPsat). This vector is useful for understanding the probabilities distribution between classes, and in the following steps, for redefining class membership probabilities for selected objects. The Random forest (RF) algorithm (Breiman, 2001) is widely used in remote sensing (e.g. Lebourgeois et al., 2017; Debats et al., 2016) and has been shown to outperform other classifiers in benchmark studies (e.g. Fernández-Delgado et al., 2014; Inglada et al., 2015). The RF classifier was used to produce a land-use classification based on the learning database at the end of the growing season. The main output, the class membership vector of probabilities (CMVPsat), was extracted for each object.

3.2 Spatio-temporal modelling

The model was built with the Ocelet Modelling Platform (OMP, www.ocelet.org). OMP allows the spatial modelling of a system and their dynamics (Degenne and Lo Seen, 2016) using both raster and vector data. Ocelet is based on the concept of interaction graphs to describe how entities composing a system (e.g. fields, rivers, weather, municipality) are inter-connected. Each entity is characterized by a number of properties that can evolve through time with the application of interaction functions.

3.2.1 Formalization and selection of STrules: The knowledge of farmer strategies and practices gathered during literature review and interviews was formalized so that crop type can be described by a set of general spatio-temporal rules (STrules). The spatial rules can concern distances or intersection with specific entities (such as roads, villages, markets and hydrographic network) or zones (such as soil map units, altitude and slope classes). The temporal rules can be linked to the cropping calendars or rainfall regime during the growing season, or the previous crop rotation on the same field. The spatio-temporal rules for each crop type were selected according to the following criteria: (i) the rule should be representative of the whole study area, (ii) the rule should be frequently mentioned by the persons interviewed or in the documents, (iii) the rule must be consistent with agronomic knowledge and observable in the field, and (iv) the data must be available for the rule to be applied. A crop type can be defined by exclusive rules (e.g. maximum distance from a point) or optional rules (e.g. generally in fertile soil). The assignment of an entity to a given class must comply with all exclusive rules and part (or none) of the optional rules.

3.2.2 Implementation: In the model, the landscape structure of the system is based on the segmentation produced during step 3.1.1, which provide the boundaries of the entities. For each crop, rules are then implemented in the OMP at the entity level to produce a map giving the probability of presence of the corresponding crop in the study area. From this step onwards, the application of all the rules will define a class membership vector of probabilities for each entity, based on formalized knowledge on farmer strategies (referred to as CMVPrule for STrules-based membership probabilities).

3.3 Redefinition of class membership probabilities

Traditional evaluation methods used in remote sensing to validate a classification are based on the computation of indices such as overall accuracy, Kappa, F-score (Labatut, Cherifi, 2012) or the quantity and allocation disagreement (Pontius, Millones, 2011). These methods can be derived from confusion matrices, giving the number of well-classified and misclassified ground truth occurrences per class. These methods allow evaluating the results of a given classifier, but only ground truth information. Other entities are not evaluated. This step proposes a spatial evaluation of the remote sensing classification with respect to farmer strategies (expressed through STrules in the model), and to redefine class membership probabilities to obtain more consistent agricultural maps.

When comparing the CMVPrule to CMVPsat for the whole study area, a map can be generated highlighting areas or entities where the rules disagree, and for which corrections are required. For a given entity, in order to decide which of the two probability vectors is more reliable, other information such as the number of satellite acquisition for an object during the growing season, the set of expert rules concerned, or the confidence value obtained during the random forest vote are used. Then, to redefine class membership, various fusion methods can be used such as fuzzy logic, average between CMVPsat and CMVPrule, or rules weighting.

4. RESULTS

In the following sections, we give examples of STrules used and present three main outputs of the approach. Based on the application of STrules for each entity in the model, probabilities of presence of each crop type were mapped. This map was combined with the map produced during the remote sensing step

to create a disagreement map. The final output of the method was a new land use map with an associated class membership probability for each object.

4.1 Spatio-temporal rules

Literature review and interviews allowed determining a set of STrules for each crop. The type of dominant rules can differ between crops. For example, some crops are strongly dependent on economic rules (e.g. good connection to markets) while other crops are more constrained to a given biophysical environment or linked to a specific crop calendar.

Examples of different exclusive and optional rules are given below for three crop classes of the study area:

Market gardening:
Exclusive rules:
R1. Less than 30 minutes by truck or less than 15 km from markets (the presence of market gardening is strongly linked to the presence of markets).
R2. Distance from villages less than 150 m (to avoid steals because market gardening is a high-value crop).
Optional rules:
R3. The field is usually in fertile soil (sandy soil, recent alluvial soil).
R4. The field could be irrigated or close to a watercourse.
R5. The slope can be between 0% and 7%.

Irrigated rice:
Exclusive rules:
R1. The field must be in an irrigated lowland or in the riverbed.
R2. The parcel must be close to a watercourse.
R3. The crop calendar spreads from 11/21 to 12/20 (for transplantation phase) and harvest occurs from 04/20 to 05/10.
Optional rules:
R4. The field is frequently in a hydromorphic soil.

Maize:
Exclusive rules:
R1. The field is located in poor soil.
R2. The crop calendar is 10/20 - 12/30 (seeding) and 02/20 - 04/30 (harvest) for an intermediary cycle and 10/20 - 12/10 (seeding) and 03/20 - 05/10 (harvest) for a long cycle (most common).
Optional rules:
R3. The crop is frequently in intercropping with rainfed rice or soya (to be implemented, this rule thus relies on the probabilities of presence of rainfed rice and soya)
R4. Maize is frequently present in the livestock sector, which corresponds to slopes or plateaus.

4.2 Map of probabilities

The maps of probabilities are the result of STrules Implementation in the model.
Figure 2 shows the irrigated rice probability map in the study area. For each crop, the same map of probability of presence can be computed based on STrules as defined during step 3.3.1. For irrigated rice, this map highlights high probabilities of presence around watersheds and lowlands. In the center of the map, the Ambohibary lowland mainly consists in rice and market gardening crops. On plateaus and slopes, the probability of the presence of irrigated rice is low as expressed by STrules.

Maps of probabilities allow to analyze and understand the possible spatial distribution of the crop in the landscape, and to identify the main underlying explanatory factors of this repartition.

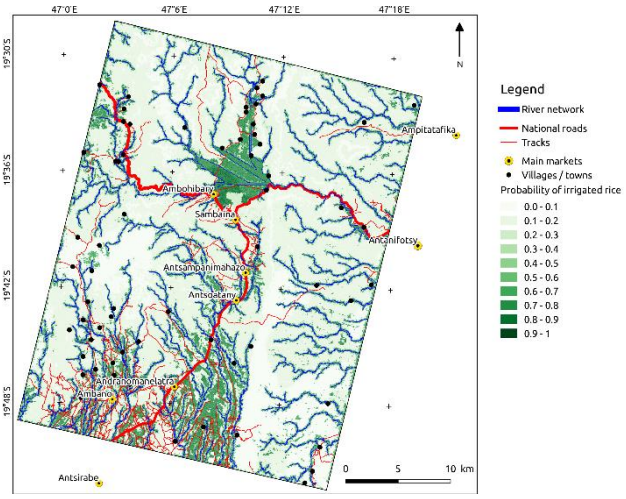


Figure 2. Map of probability for irrigated rice from spatio-temporal modelling (STrules implementation).

4.3 Disagreement map

A comparison between CMVPsat and CMVPrule (see 3.4) allowed producing a disagreement map. This map highlights the entities where remote sensing and STrules seem to be in disagreement. These disagreements mean that either the expert rules or the classifier are not accurate for the concerned entities . Assuming that the STrules have been verified by experts and were found to be consistent, this map thus allows the spatialization of remote sensing image classification errors.

The spatialization of the classification errors can also help to identify production basins or specific zones where remote sensing has more difficulties to discriminate between crop types. These maps support the analysis of the possible factors limiting the success of remote sensing approach in such context. In our study case, the disagreement map presented in Figure 3 shows remote sensing classification errors for maize, irrigated rice, rainfed rice and market gardening.. Majority of irrigated rice is well classified with a predominance in lowlands, but a part of market gardening and other “minor” crops (e.g. cassava, taro, etc.) in lowlands is misclassified. For these minor crops, the weakness of the learning set, leading to an overlearning of the RF classification algorithm can explain the disagreement . In the North-East part of the map, an important area is misclassified. This inaccessible zone had no ground truth data (used in the learning of the RF classification algorithm) to represent its land use variability, and few auxiliary data to support STrules. Consequently, few entities reached an agreement in this zone. A photo-interpretation of this place showed that, except near watercourses, a majority of natural vegetation and rainfed crops were present. In the South-East of the map, another large disagreement area is identified. In this place, ground truth and auxiliary data were present in sufficient quantity. Disagreements are related to the important misclassification of market gardening and orchards with other rainfed crops. An analysis of CMVPsat showed that all these crop types had a very close membership probability. .

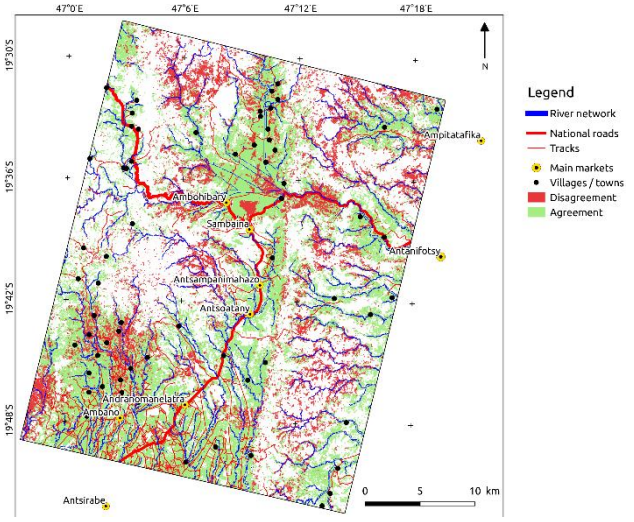


Figure 3. Map showing the disagreements between rule-based (CMVPrule) and remote sensing-based (CMVPsat) class assignment for rice, maize and market gardening.

From this first result, an analysis of the type and characteristics of the entities where a disagreement is found (the type of crop assigned according to CMVPsat and CMVPrule, and values of probability of presence of each crop for the same entity) can be performed to improve the understanding of processes and the characterization of such complex system.

4.4 Corrected map

The main output of the method is a new modified land use map, more consistent with agricultural strategies and practices of the region. An example of a corrected map, produced after the redefinition of probability values by averaging CMVPsat and CMVPrule for each entity in disagreement, is presented in Figure 4. In Madagascar highlands, a first comparison with ground truth data showed that this method allowed better discrimination of rainfed crop types. This crop type is challenging to characterize due to the very small field sizes and the high level of fragmentation of the rainfed cropland area, as shown previously in Lebourgeois et al. (2017).

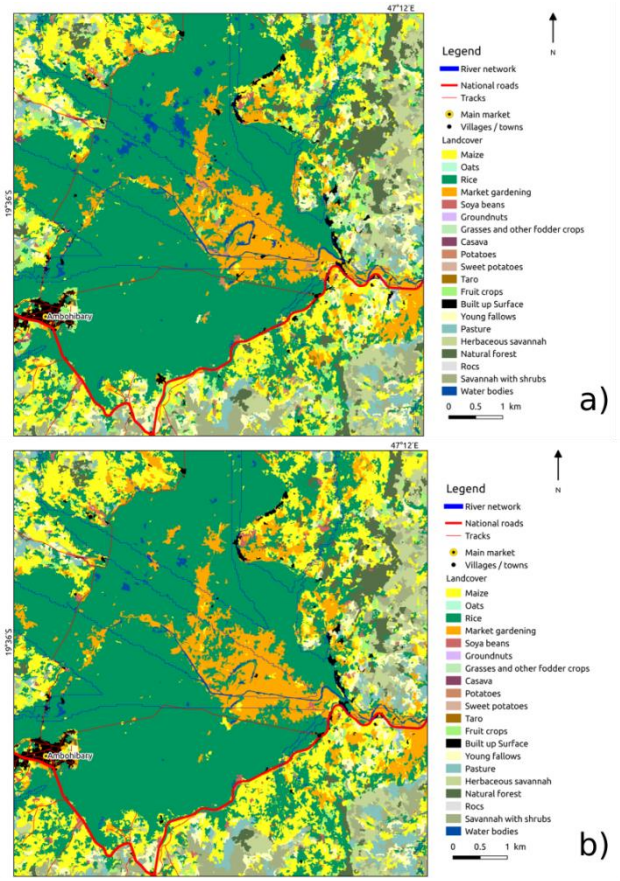


Figure 4. Zoom on land use maps obtained using the remote sensing approach, before (a) and after (b) the redefinition of CMVP values of each entity for which rule-based and remote sensing-based class assignment differs for main crops.

5. CONCLUSION AND OUTLOOKS

This work proposes a new framework in which remote sensing and expert knowledge on farmer practices and strategies are used in a spatio-temporal modelling approach to produce more agriculturally consistent crop type maps in complex agricultural systems. The method has the originality of combining remote sensing and modelling with knowledge from human and social sciences that reflect farmers' behaviour. Such approach can help to improve (i) the understanding of processes driving land use distribution in complex agricultural areas such as in smallholder farming systems and (ii) the accuracy of crop type maps in regions where traditional remote sensing approaches present limitations.

This approach is also being currently tested in the same study area with a new set of satellite images with higher spatial and temporal resolutions (Pléiades instead of SPOT6/7 for the very high spatial coverage, and addition of Venμs data in complement of the Landsat-8 / Sentinel-2 time series), building on the same set of STrules. An analysis of the model sensitivity to the integration of each spatio-temporal rule will also be performed. Beyond its application to the complex smallholder agricultural landscape of Madagascar Highlands, the approach proposes a generic framework that explicitly brings rich thematic expert knowledge into the analysis of satellite time series data. Such a framework can certainly be adapted to other regions of interest or thematic applications (e.g. biodiversity). However, in a broader perspective, this method can be considered as a first attempt to

combine knowledge-driven and data-driven approaches for extracting useful information from one of the most abundant sources of geographical data that are earth observation satellites.

ACKNOWLEDGEMENTS

This study was conducted as part of the SAMSAM project funded by CNES, the French Space Agency (TOSCA APR 2017). This work was also supported by the French National Research Agency under the Investments for the Future Program, referred to as ANR-16-CONV-0004. The authors would like to thank Eloise Rasoamalala and Valérie Andoniaina for their contributions to the field surveys, and the researchers of SPAD, the CIRAD platform in partnership on "Highland production systems and sustainability in Madagascar".

REFERENCES

Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informations-Verarbeitung XII*, 12–23.

Bailly, S., Giordano, S., Landrieu, L., Chehata, N., 2018. Crop-Rotation Structured Classification using Multi-Source Sentinel Images and LPIS for Crop Type Mapping. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Valencia, pp. 1950–1953. https://doi.org/10.1109/IGARSS.2018.8518427

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Matton, N., Morin, D., Popescu, R., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Ahmad, I., Bégué, A., Wu, B., De Abelleyra, D., Diarra, A., Dupuy, S., French, A., Ul Hassan Akhtar, I., Kussul, N., Lebourgeois, V., Le Page, M., Newby, T., Savin, I., Verón, S.R., Koetz, B., Defourny, P., 2015. Building a Data Set over 12 Globally Distributed Sites to Support the Development of Agriculture Monitoring Applications with Sentinel-2. *Remote Sensing* 7, 16062–16090. https://doi.org/10.3390/rs71215815

Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32. Debats, S.R., Luo, D., Estes, L.D., Fuchs, T.J., Caylor, K.K., 2016. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sensing of Environment* 179, 210–221. https://doi.org/10.1016/j.rse.2016.03.010

Degenne, P., Lo Seen, D., 2016. Ocelet: Simulating processes of landscape changes using interaction graphs. *SoftwareX* 5, 89–95. https://doi.org/10.1016/j.softx.2016.05.002

Donadieu, J., L'Helguen, C., 2016. SENTINEL-2A L2A Products description (Technical note). CNES, Direction du Centre Spatial de Toulouse.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Dury, J., Schaller, N., Garcia, F., Reynaud, A., Bergez, J.E., 2012. Models to support cropping plan and crop rotation decisions. A review. *Agronomy for Sustainable Development* 32, 567–580. https://doi.org/10.1007/s13593-011-0037-x

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15, 3133–3181.

Gaetano, R., Dupuy, S., Lebourgeois, V., Le Maire, G., Tran, A., Jolivot, A., Bégué, A., 2019. The MORINGA Processing Chain: Automatic Object-based Land Cover Classification of Tropical Agrosystems using Multi-Sensor Satellite Imagery. ESA Living Planet Symposium (LPS 2019), Milan, Italy.

Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

GRASS Development Team, 2017. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data, Software and Standards* 2. https://doi.org/10.1186/s40965-017-0031-6

Hagolle, O., Huc, M., Desjardins, C., Auer, S., Richter, R., 2017. MAJA Algorithm Theoretical Basis Document. https://doi.org/10.5281/zenodo.1209633

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics SMC*-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sensing* 7, 12356–12379. https://doi.org/10.3390/rs70912356

Jurgens, C., 1997. The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. *International Journal of Remote Sensing* 18, 3583–3594. https://doi.org/10.1080/014311697216810

Labatut, V., Cherifi, H., 2011. Accuracy Measures for the Comparison of Classifiers. *The 5th Inter-national Conference on Information Technology*. Amman, Jordanie.

Lebourgeois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S., Bégué, A., 2017. A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sensing* 9(3), 259. https://doi.org/10.3390/rs9030259

NASA JPL, 2014. U.S. Releases Enhanced Shuttle Land Elevation Data. NASA Jet Propulsion Laboratory (JPL). http://www.jpl.nasa.gov/news/news.php?release=2014-321 (28 October 2019).

Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E., Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment* 115, 1301–1316. https://doi.org/10.1016/j.rse.2011.01.009

Pontius, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Riquier, J., 1968. Carte pédologique de Madagascar. One mosaic of 3 map sheet. Scale of 1 :1 000 000. ORSTORM.

Rouse, W., Haas, R.H., Schell, J., Deering, D., 1974. Monitionring Vegetation Systems in the Great Plains with ERTS. *Third Earth Resources Technology Satellite-1 Symposium, Technical presentations*, section A, vol. I, National Aeronautics and Space Administration, Washington, DC, 309–317.

Sourisseau, J.-M., Rasolofo, P., Bélières, J.-F., Guengant, J.-P., Ramanitriniony, H.K., Bourgeois, R., Razafimiarantsoa, T.T., Andrianantoandro, V.T., Ramarijaona, M., Burnod, P., Rabeandriamaro, H., Bougnoux, N., 2016. Diagnostic Territorial de la Région du Vakinankaratra à Madagascar (Prospective territoriale sur les dynamiques démographiques et le développement rural en Afrique subsaharienne et à Madagascar). AFD.

Weiss, A.D., 2001, Topographic position and landform analysis. Poster presentation, ESRI User Conference, San Diego, California, USA.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27, 3025–3033. https://doi.org/10.1080/01431160600589179

RÉSUMÉ

Télédétection et intégration de connaissances via la modélisation spatiale pour une cartographie plus cohérente des systèmes agricoles complexes : Application aux Hautes Terres, à Madagascar

Les méthodes de classification des données de télédétection pour la cartographie de l'occupation des sols reposent sur des mesures de proximité entre pixels ou objets dans un hyperespace spectral dans lequel ils sont projetés en fonction de leur réflectance dans plusieurs bandes spectrales de l'image. La projection en retour des pixels ou des objets vers l'espace géographique permet d'obtenir la carte d'occupation du sol recherchée, avec autant de classes que de régions identifiées comme significatives dans l'hyperespace. Le raisonnement qui sous-tend l'analyse dans l'hyperespace des données satellitaires est que dans ce dernier, les relations de proximité de "signatures spectrales" des objets et des pixels sont plus faciles à mettre en évidence que dans l'espace géographique. Ces méthodes sont particulièrement efficaces pour la cartographie des espaces agricoles caractérisés par de grandes parcelles et par des pratiques culturales homogènes.

Cependant, pour une grande partie des Suds, les espaces agricoles sont plus complexes, avec des parcelles généralement plus petites, et une diversité d'usages et d'occupations des sols qui reflète les coutumes et leur adaptation au climat et à la géographie locale. L'efficacité des méthodes de télédétection diminue fortement pour ces systèmes agricoles plus complexes. Afin de repousser ces limites, les recherches actuelles portent sur des méthodes combinant à la fois des séries temporelles d'images satellites et des données contextuelles, tel que des indices de texture, l'altitude ou la pente, dont certaines sont basées sur l'apprentissage profond.

Ces développements méthodologiques et techniques poussés utilisent principalement l'information spectrale et n'intègrent que peu les autres types de connaissances agricoles disponibles. Par exemple, on sait que certaines cultures ne poussent qu'à partir d'une altitude ou à proximité des habitations. Aisément identifiables dans l'espace géographique, ces connaissances, à la fois spatiales et temporelles, sont plus difficilement identifiables dans l'hyperespace des données. Cela suggère intuitivement qu'une meilleure prise en compte des connaissances agricoles pourrait améliorer les méthodes de classification pour obtenir des cartes de l'occupation de l'usage des sols plus cohérentes sur le plan agricole.

Dans cette thèse, nous explorons la possibilité d'utiliser les connaissances en agriculture, formalisées sous forme de règles, pour améliorer une méthode de classification supervisée pour la cartographie de l'occupation et de l'usage des sols des systèmes agricoles complexes. Dans un premier temps, cette thèse propose un modèle conceptuel permettant de combiner à la fois des méthodes de classification de données de télédétection et de la modélisation spatio-temporelle. Ce modèle est décomposé en quatre modules spatiaux et temporels, chacun correspondant à une méthode visant à améliorer la caractérisation de l'occupation et de l'usage des sols, et pouvant être utilisé indépendamment. Les deux modules spatiaux du modèle sont ensuite appliqués à une zone d'étude agricole située dans la région du Vakinankaratra, sur les Hautes Terres de Madagascar afin d'évaluer l'approche développée.

D'un point de vue quantitatif, l'application des deux modules spatiaux n'améliore que très peu la caractérisation des classes d'occupation et d'usage des sols de la zone d'étude, notamment du fait du manque de données de qualité supportant l'application des règles spatiales identifiées. Néanmoins, l'application des modules spatiaux permet une meilleure discrimination entre les cultures pluviales et les espaces de savanes, source de nombreuses confusions avec les méthodes utilisées en traitement de données de télédétection. L'analyse de ces résultats permet de proposer des améliorations pour le modèle conceptuel ainsi que pour son application plus générale aux systèmes agricoles complexes.

Mots clés : *Télédétection, Modélisation spatio-temporelle, Agriculture familiale, Cartographie de l'occupation et de l'usage des sols*

ABSTRACT

Remote sensing and knowledge integration through spatial modelling for a more consistent mapping of complex agricultural systems: Application to the Highlands region in Madagascar

Classification methods for remote sensing data for land cover mapping are based on proximity measurements between pixels or objects in a spectral hyperspace in which they are projected according to their reflectance in several spectral bands of the image. Then projecting the pixels or objects from the hyperspace back into geographical space yields the land cover map sought, with as many land cover classes as regions identified as meaningful in the hyperspace. The rationale behind the analysis within the satellite data hyperspace is that in the latter, relationships involving the "spectral signatures" of objects and pixels are easier to retrieve. Such methods have proved very efficient in mapping agricultural landscapes in many parts of the world characterised by large, well-managed intensive agricultural plots.

However, for large parts of the developing world, agricultural landscapes are more complex, with typically smaller plots, and a diversity of land use and land cover that reflect local customs and their adaptation to local climate and geography. The efficiency of remote sensing methods sharply decrease for the more complex landscapes. In order to push back these limits, current researches are focusing on methods combining both time series of satellite images and contextual data, such as texture indices, altitude or slope, some of which are based on deep learning.

These methodological developments and in-depth techniques use mainly spectral information and integrate little of the other types of agricultural knowledge available. For example, it is known that some crops only grow above a given altitude or close to households. These types of (spatial and temporal) relationships are easy to formalise in geographical space, but much less so in the data hyperspace. This intuitively suggests that better taking into account agricultural knowledge could improve classification methods to obtain agriculturally consistent land use maps.

In this thesis, we explore the possibility of using agricultural knowledge, formalised as rules to improve an existing machine learning classification method for land cover and land use mapping in complex agricultural systems. First, this thesis proposes a conceptual model allowing to combine both remote sensing methods and spatio-temporal modelling. This model is decomposed into four spatial and temporal modules, each corresponding to a method aimed at improving the characterization of land cover and land use, and which can be used independently. The two spatial modules of the model are then applied to an agricultural study area located in the Vakinankaratra region, in the highlands of Madagascar, in order to evaluate the approach developed.

In quantitative terms, the application of the two spatial modules only slightly improves the characterisation of the land use classes of the study area, mainly due to the lack of quality data supporting the application of the spatial rules. Nevertheless, the application of spatial modules allows a better discrimination between rainfed crops and savannah areas, which is a source of much confusion with the methods used in remote sensing. The analysis of these results leads to improvements for the conceptual model as well as for its more general application to complex agricultural systems.

Key words : *Remote sensing, Spatio-temporal modeling, Family farming, Land cover and land use mapping*