Université de Toulouse

MASTER 2 GEOMATIQUE

« **Sc**I**ences** **G**éomatiques en environne**M**ent et **A**ménagement »
(**SIGMA**)

http://sigma.univ-toulouse.fr

MÉMOIRE DE RECHERCHE

ÉVALUATION DU POTENTIEL D'UNE SÉRIE TEMPORELLE D'IMAGES MULTI-SOURCE À HAUTE RÉSOLUTION SPATIALE POUR LE SUIVI DES CULTURES ANNUELLES EN PETITE AGRICULTURE FAMILIALE : CAS DU BASSIN ARACHIDIER SÉNÉGALAIS

GBODJO Yawogan Jean Eudes

Maîtres de stage : LEROUX Louise (Cirad) et DIOUF Abdoul Aziz (CSE)
Tuteur-enseignant : SHEEREN David

Septembre 2018

Résumé

L'intensification écologique des pratiques agricoles est assurément aujourd'hui l'une des voies les plus efficaces pour garantir la sécurité alimentaire de tous, tout en favorisant une agriculture durable préservant les écosystèmes et la biodiversité notamment dans les pays en voie de développement. Le présent travail s'inscrit dans le cadre de recherches portant sur l'évaluation spatialisée des pratiques d'intensification écologique des systèmes de cultures à base de mil et d'arachide, à Diohine (région de Fatick) dans la zone écogéographique du bassin arachider sénégalais. Il a pour objectif d'évaluer les dates de semis et de calibrer des modèles statistiques pour l'estimation de leur biomasses et rendements sur la saison agricole 2017, en se basant sur des métriques phénologiques dérivées de télédétection. Pour ce faire, nous avons utilisé une série temporelle multisource d'images satellitaires incluant des images PlanetScope, RapidEye et Sentinel-2 et un réseau de 47 parcelles. L'indice de végétation NDVI a été calculé pour les différentes images et la série temporelle de NDVI a été lissée et reconstruite avec un pas de temps journalier en employant deux méthodes de lissage : l'algorithme HANTS et la méthode de Whittaker. Les profils temporels de NDVI obtenus par les 2 méthodes étant difficilement dissociables sur base des critères qualitatifs de sélection définis, ils ont été maintenu pour la phase d'évaluation des dates de semis. Cette dernière a montré que la méthode de Whittaker était plus performante que HANTS compte tenu des variabilités plus faibles otbtenues sur les écarts entre dates de semis observées et dates de début de croissance de la végétation extraites (SOS). Pour le reste, les dates de semis ont été estimées avec plus ou moins 5, 10 et 20 jours de décalage respectivement pour les parcelles d'arachide, de mil en culture pure et de mil en culture associée. En ce qui concerne l'estimation des biomasses et rendements, deux indices de végétation ont été comparés : le NDVI et le GDVI. Les modèles de régression linéaire établis pour l'arachide ont modérément expliqué les biomasses ($R^2 = 0,56$; $RMSE = 1292$ *kg/ha*) et rendements ($R^2 = 0,49$; $RMSE = 404$ *kg/ha*) tandis que pour le mil, les rendements ($R^2 = 0,45$; $RMSE = 540$ *kg/ha*) ont été mieux estimés que les biomasses ($R^2 = 0,27$; $RMSE = 4254$ *kg/ha*). Les variables explicatives dérivées du GDVI ont été plus performantes que celles du NDVI traditionnellement utilisé dans ce type d'approche. Nous avons expliqué ceci notamment par un changement dans les technologies d'observation de la Terre mobilisées dans le cadre de ce travail par rapport à la littérature existante.

Mots-clefs : Intensification écologique, Evaluation spatialisée, Mil, Arachide, Série temporelle multisource (Sentinel-2, RapidEye, PlanetScope), Métriques phénologiques, Estimation des biomasses et rendements.

Abstract

Nowadays, the ecological intensification of agricultural practices had became undoubtedly one of the most effective way of ensuring food security for all, while promoting sustainable agriculture that preserves ecosystems and biodiversity particularly in developing countries. This work is part of researches on spatial evaluation of the ecological intensification practices of millet and groundnut-based crop systems at Diohine (Fatick region) in the Senegalese groundnut Basin. Its objective is to evaluate sowing dates and to calibrate statistical models for estimating biomass and yields on the 2017 agricultural season, based on phenological metrics. To do this, we used a multisource time series of satellite images including PlanetScope, RapidEye and Sentinel-2 images and a network of 47 field plots. The popular vegetation index NDVI were extracted from the images and the NDVI time series was smoothed and reconstructed with a daily time step using two smoothing methods : HANTS algorithm and Whittaker smoother. The temporal profiles of NDVI obtained by the 2 methods being difficult to dissociate on the basis of our qualitative criteria of selection, we maintained them for the sowing dates evaluation. It was later showed that Whittaker's method was more efficient than HANTS, given the lower variability obtained on the gap between observed sowing dates and starting dates of growth of vegetation (SOS). For the remainder, we managed to estimate sowing dates with more or less 5, 10 and 20 days of interval, for the groundnut, pure millet and mixed millet plots respectively. With regard to the estimation of biomass and yields, we compared NDVI and GDVI indices. The established linear regression models moderately explained groundnut biomass ($R^2 = 0,56$; $RMSE = 1292$ *kg/ha*) and yields ($R^2 = 0,49$; $RMSE = 404$ *kg/ha*) while millet yields ($R^2 = 0,45$; $RMSE = 540$ *kg/ha*) were better estimated than biomass ($R^2 = 0,27$; $RMSE = 4254$ *kg/ha*). The explanatory variables derived from the GDVI were more efficient than those of NDVI usually used in this type of approach. We explained this by a change in earth observation technologies mobilized as part of this work compared to the existing literature.

Keywords : Ecological intensification, Spatial evaluation, Millet, Groundnut, Multisource time series (Sentinel-2, RapidEye, PlanetScope), Phenological metrics, Biomass and Yields estimation.

Remerciements

Par ces simples mots, je viens tout d'abord témoigner ma gratitude à vous tous qui aviez de près ou de loin par vos soutiensmultiformes à l'aboutissement de ce travail qui vient couronner une année académique (2017-2018) de dur labeur et riche d'expériences.

Merci à toi Papa, qui n'a eu point de cesse d' œuvrer pour que je produise un document de qualité.

Ma gratitude va également à tout le corps professoral du Master SIGMA qui par leurs expertises et la qualité de leurs enseignements m'ont formé pour être parmi les meilleurs dans le domaine de la Géomatique.

Un infini merci à mes maîtres de stage, Dr. Louise Leroux et Dr. Abdoul Aziz Diouf avec qui j'ai pris plaisir à travailler tout au long de ce stage. Merci encore pour vos orientations et conseils quand par moment j'étais perdu. J'ai apprécié à leur juste valeur en ces moments précis, vos savoir-faire qui ont été pour moi d'une utilité inexprimable.

Merci également aux agents du CSE qui ont grandement contribué à mon intégration en m'accueillant comme leur frère et ami : Mme. Sene, Mme Soti, M. N'dao, M. Thiaw *Jërëngen jëf.*

Enfin, merci spécial à mon tuteur enseignant M. David Sheeren qui a encadré ce travail et à vous messieurs les membres du jury qui allez le juger, recevez ici la marque de ma sincère reconnaissance.

Sommaire

1. Introduction 1

2. Contexte et Objectifs 3
   2.1. Contexte du stage . . . 3
      2.1.1. Présentation des organismes d'accueil . . . 3
      2.1.2. Projets GloFoodS SERENA, SIIL SIMCo et TOSCA LYSA . . . 4
   2.2. Objectifs et Hypothèses . . . 6
   2.3. Planning de travail . . . 6

3. Synthèse bibliographique 8
   3.1. Extraction des dates de semis par télédétection . . . 8
      3.1.1. Considérations générales . . . 8
      3.1.2. Phénologie de la végétation . . . 9
      3.1.3. Métriques phénologiques . . . 9
   3.2. Estimation des rendements par télédétection . . . 11

4. Matériels et Méthodes 15
   4.1. Présentation générale de la zone d'étude . . . 15
   4.2. Données satellitaires utilisées . . . 15
      4.2.1. Images PlanetScope et RapidEye . . . 15
      4.2.2. Images Sentinel-2 . . . 18
      4.2.3. Chronologie des images acquises . . . 19
   4.3. Données de terrain . . . 19
   4.4. Calcul du NDVI . . . 22
   4.5. Lissage des séries temporelles . . . 22
      4.5.1. Filtre de Savitzky-Golay . . . 23
      4.5.2. HANTS . . . 24
      4.5.3. Méthode de lissage de Whittaker . . . 25
      4.5.4. Evaluation du lissage . . . 26
   4.6. Extraction des métriques phénologiques . . . 30
   4.7. Estimation des biomasses et rendements . . . 30

5. Résultats et Discussions 32
   5.1. Résultats . . . 32
      5.1.1. Extraction du SOS et du EOS . . . 32
      5.1.2. Estimation des biomasses et rendements . . . 36
   5.2. Discussions . . . 42
      5.2.1. Sur l'extraction du SOS et du EOS . . . 42
      5.2.2. Sur l'estimation des biomasses et rendements . . . 43

6. Conclusion et Perspectives 45

A. Diagramme de Gantt 47

B. Essais de lissage avec le filtre de Savitzky-Golay 48

C. Estimation des SOS et EOS sans le jeu ANR CERAO 50

D. Spatialisation des SOS et EOS 53

E. Relations linéaires entre biomasses et rendements 55

Bibliographie 56

Liste des figures

3.1. Illustration de quelques métriques phénologiques 10
3.2. Illustration de quelques métriques phénologiques 12

4.1. Situation géographique de la zone d'étude 16
4.2. Chronologie des images acquises 20
4.3. Parcelles terrain 21
4.4. Moyenne et Ecart type du NDVI pour 6 parcelles 28
4.5. Résultats du lissage de la série complète pour 6 parcelles 29
4.6. Résultats du lissage de la série rectifiée pour 6 parcelles 29

5.1. Distribution des écarts entre SOS et dates de semis 33
5.2. SOS – RMSE vs CV 34
5.3. Distribution des écarts entre EOS et dates de récoltes 35
5.4. EOS – RMSE vs CV 36
5.5. Régressions entre Cumuls de NDVI et Biomasses 37
5.6. Régressions linéaires entre Cumuls de GDVI et Biomasses 38
5.7. Régressions linéaires entre interactions et biomasses 38
5.8. Modèles finaux pour l'estimation des biomasses 39
5.9. Régressions linéaires entre cumuls de NDVI et rendements 40
5.10. Régressions linéaires entre cumuls de GDVI et rendements 40
5.11. Régressions linéaires entre interactions et rendements 41
5.12. Modèles finaux pour l'estimation des rendements 41

Liste des tableaux

4.1. Caractéristiques spectrales des images PlanetScope 17
4.2. Caractéristiques spectrales des images RapidEye 17
4.3. Caractéristiques spectrales de l'instrument MSI des Sentinel-2 19
4.4. Statistiques descriptives des données de terrain 21
4.5. Paramétrage des méthodes de lissage 28

Glossaire

BISE Best Index Slope Extraction.

Cirad Centre de Coopération Internationale pour la Recherche Agronomique et le Développement.

CSE Centre de Suivi Ecologique.

CV Coefficient de variation qui est le rapport de l'écart type à la moyenne exprimé souvent en pourcentage.

ENVISAT ENVIronment SATellite — Satellite d'observation de la Terre de l'ESA lancé en 2002 dont l'objectif est de mesurer de manière continue à différentes échelles les principaux paramètres environnementaux de la Terre relatifs à l'atmosphère, l'océan, les terres émergées et les glaces.

EOS End of Season — Fin de croissance de la végétation.

ESA European Spatial Agency — Agence Spatiale Européenne.

EVI Enhanced Vegetation Index.

FAPAR Fraction of Absorbed Photosynthetically Active Radiation qui désigne la fraction de rayonnement solaire absorbée par les plantes dans le domaine spectral permettant la photosynthèse. Le FAPAR est une variable biophysique directement reliée à la productivité primaire de la végétation.

GMES Global Monitoring for Environment and Security.

GSL Growing Season Length — Durée de croissance de la végétation = EOS – SOS.

HANTS Harmonic Analysis of Time Series.

IDR Iterative interpolation for Data Reconstruction.

INRA Institut National de la Recherche Agronomique.

IRD Institut de Recherche pour le Développement.

ISRA Institut Sénégalais de Recherche Agronomique.

LAI Leaf Area Index ou indice de surface foliaire est une grandeur sans dimension, qui exprime la surface foliaire d'un arbre, d'un peuplement, d'un écosystème ou d'un biome par unité de surface de sol. Il est déterminé par le calcul de l'intégralité des surfaces des feuilles de la plante sur la surface de sol que couvre cette plante.

LSP Land Surface Phenology.

MACCS Multi-sensor Atmospheric Correction and Cloud Screening.

MAJA MACCS-ATCOR Joint Algorithm.

MSAVI Modified Soil Adjusted Vegetation Index.

MSI MultiSpectral Instrument.

MUSCATE Atelier de production MUlti Satellite, multi-CApteurs, pour des données multi-TEmporelles mise en place par le CNES et le CESBIO au sein de Theia.

MVC Maximum Value Composite.

NDVI Normalized Difference Vegetation Index — Indice de végétation par différence normalisé.

Persyst Performances des systèmes de production et de transformation tropicaux est un département scientifique du Cirad qui conduit des études sur les productions tropicales à l'échelle de la parcelle, de l'exploitation et de la petite entreprise de transformation.

POS Peak of Season — Maximum de croissance de la végétation.

RMSE Root Mean Square Error, $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$ avec y la variable prédite, x la variable observée et n le nombre total d'échantillons.

SAVI Soil Adjusted Vegetation Index.

SIIL Sustainable Intensification Innovation Lab.

SOS Start of Season — Démarrage de croissance de la végétation.

Theia Pôle de données et de services surfaces continentales — a pour objectif d'accroître l'utilisation par la communauté scientifique et les acteurs publics de la donnée spatiale en complémentarité d'autres types de données, notamment les données in situ et aéroportées.

TOA Top of Atmosphere — Radiance ou Réflectance au sommet de l'atmosphère.

UCAD Université Cheikh-Anta-Diop, Dakar-Sénégal.

UR Aïda Unité de Recherche Agroécologie et intensification durable des cultures annuelles.

UTM Projection Transverse Universelle de Mercator; Zone 28N sur Niakhar.

VCI Vegetation Condition Index.

1. Introduction

La *sécurité alimentaire* désigne « une situation garantissant à tout moment, à tous les êtres humains, la possibilité physique, sociale et économique de se procurer une nourriture suffisante, saine et nutritive leur permettant de satisfaire leurs besoins et préférences alimentaires pour mener une vie saine et active » tel que défini par le Comité de la Sécurité alimentaire mondiale (Comité de la Sécurité Alimentaire Mondiale, 2012). On considère généralement *quatre piliers* ou dimensions de la sécurité alimentaire : l'accès, la disponibilité (la plus mise en avant), la qualité et la stabilité. Aujourd'hui la notion d'Excès est également évoquée quand on sait qu'un régime alimentaire malsain peut être la cause de l'obésité, du surpoids ou de maladies comme l'hypertension artérielle. Selon (FAO et al., 2017), 815 millions de personnes en 2016, soit plus d'une personne sur dix dans le monde, étaient en situation d'insécurité alimentaire. Les facteurs de l'insécurité alimentaire sont divers. La pénurie d'eau, la dégradation des sols, le changement climatique, les épidémies, l'explosion démographique ou encore les situations de conflits sont autant de sources antérieures ou concomitantes à une telle situation.
L'un des défis majeurs de notre temps est de garantir à une population toujours plus grandissante (près de 10 milliards en 2050 selon les projections), une alimentation suffisante pour couvrir ses besoins nutritionnels. Ainsi, nourrir 2 milliards de personnes de plus en 2050 nécessitera de revoir à la hausse la production alimentaire mondiale qui devra être globalement augmentée de 50% . Face à ce défi notamment dans les pays en développement, le concept « d'*Intensification écologique* ou d'*Agriculture écologiquement intensive* » trouve parfaitement son application. L'intensification écologique est un concept agronomique initié et promu par le Cirad dans le contexte de l'agriculture des pays du Sud, caractérisés par des rendements agricoles faibles. Ce concept se veut de mettre au point des systèmes de production agricole utilisant de façon intensive les processus biologiques et écologiques ainsi que leurs fonctionnalités naturelles, plutôt que d'utiliser de façon intensive les intrants (énergies fossiles, engrais chimiques, pesticides). L'utilisation intensive de ces facteurs de production naturel et écosytémique permettrait ainsi de maintenir des niveaux de rendements élevés, préservant les ressources naturelles et assurant de ce fait une durabilité des écosystèmes cultivés (Goulet, 2012).
À Niakhar au Sénégal, dans la région du bassin arachidier, l'une des voies de l'intensification écologique repose sur les associations culturales entre céréales et légumineuses. Ainsi, l'on retrouve souvent des cultures de *mil* entresemées de *niébé* (haricot) ou d'*arachide*. Plus encore, l'on note la présence d'espèces arborées sur les parcelles cultivées telle que le *Faidherbia albida* qui est également une légumineuse. Ces différentes pratiques permettraient non seulement à moyen terme d'accroître la fertilité des sols et par conséquent la productivité des systèmes de cultures mais aussi de diversifier les sources de revenus et d'alimentation dans des régions rurales où les moyens d'existence des populations dépendent étroitement des productions annuelles réalisées. Par ailleurs, elles permettraient également de limiter l'impact des fluctuations

climatiques sur la production agricole. Une évaluation spatialisée des rendements des principales cultures alimentaires s'avère donc indispensable pour évaluer les performances agronomiques et environnementales de ces systèmes de cultures. Cela étant, l'intérêt de recourir aux techniques de télédétection s'impose comme une nécessité. La télédétection s'est révelée comme un outil d'une extrême utilité ces dernières décennies pour le suivi des cultures, la prévision des récoltes ou encore l'estimation de la biomasse et des rendements (Kogan et al., 2013; Johnson, 2014; Leroux et al., 2016; Battude et al., 2016; Sibley et al., 2014), sur des échelles régionales et globales. De nos jours, la mise au point d'instruments combinant à la fois haute ou très haute résolution spatiale et haute fréquence temporelle à l'instar de *Sentinel-2*, ouvre la voie à des applications à des échelles parcellaires voire infra-parcellaires. Pour le suivi de l'agriculture dans le contexte tropical africain, ceci se traduit par l'affranchissement des contraintes liées à l'hétérogénéité des pratiques, aux parcelles de petites tailles ou encore à la présence d'arbres dans les parcelles.

2. Contexte et Objectifs

2.1. Contexte du stage

Le présent stage s'inscrit dans le cadre de recherches portant sur l'évaluation spatialisée par télédétection des pratiques d'intensification écologique des systèmes de cultures à base de mil et d'arachide au Sénégal, dans la région du bassin arachidier. Ces travaux sont relatifs aux projets GloFoodS SERENA, SIIL SIMCo et TOSCA LYSA et sont conduits par l'UR Aïda (Cirad — Montpellier, France) conjointement avec ses partenaires sénégalais le CSE et l'ISRA (Dakar, Sénégal).

2.1.1. Présentation des organismes d'accueil

Ce stage proposé par l'UR Aïda du Cirad s'est déroulé dans les locaux du CSE à Dakar au Sénégal.

L'UR Aïda [1] est une unité de recherche du Cirad créée en 2014 et faisant partie du département Persyst. Elle se positionne sur l'intensification et la durabilité de la production des cultures annuelles en milieu tropical contraint et ses recherches visent la pleine valorisation des ressources disponibles, en mobilisant les processus écologiques qui régissent leur dynamique au sein des agrosystèmes. Elle a pour objectif l'étude, la conception et la proposition de systèmes de culture annuels (canne à sucre, cotonnier, riz . . .) répondant aux exigences de performances agronomiques, technologiques et environnementales. L'unité est structurée en 5 équipes de recherche dont *Artists* (Télédétection, systèmes d'information, techniques de simulations et analyses spatiales) pour couvrir les différents aspects de l'intensification écologique et répondre à la demande sociétale et aux besoins du développement. Ces équipes interagissent entre elles pour relever ensemble trois grands défis de recherche liés aux systèmes de culture innovants : la compréhension du fonctionnement de l'agro-écosystème complexe, l'émergence de systèmes de cultures innovants, efficaces et pertinents pour les agricultures familiales des pays du Sud et l'évaluation de ces systèmes de culture comme élément essentiel pour la décision et l'action. L'unité dispose d'une large capacité d'expertise dans les domaines des filières agricoles, de l'environnement, de la gestion de données ou encore des systèmes d'information et propose des outils, logiciels et analyses à destination des chercheurs et professionnels des filières agricoles.

Le CSE [2] est un organisme étatique créé en 1986. Il est placé sous la tutelle technique du ministère en charge de l'environnement et est doté d'une personnalité morale lui permettant de jouir d'une autonomie administrative et financière. Le CSE a pour mission de contribuer à la connaissance et à la gestion durable des ressources naturelles

1. https://ur-aida.cirad.fr/
2. https://www.cse.sn/index.php/fr/

et de l'environnement, par la production et la diffusion de produits et services d'aide à la décision notamment pour l'Etat, les collectivités locales, le secteur privé, la société civile, les institutions de recherche et de développement, les organisations de producteurs et les partenaires au développement. Les domaines d'activités du CSE couvrent la gestion du littoral, le suivi des zones de parcours, des feux de brousse et de la production agricole, les études de vulnérabilité et d'adaptation aux changements climatiques, la séquestration de carbone, le suivi à long terme des écosystèmes ou encore les problématiques d'environnement et santé. Le CSE dispose d'un important réseau de partenariat au niveau national (ministères, universités, institutions de recherches basées à Dakar à l'instar de l'IRD ou du Cirad) comme international (en Afrique notamment le Centre Agro-hydro-météorologique (AGRHYMET) ou ailleurs dans le monde comme la FAO ou le PNUE).

L'ISRA [3] est un établissement public à caractère scientifique et technologique fondé en 1974, actuellement sous la tutelle du ministère de l'Agriculture mais disposant de son propre conseil d'administration. L'ISRA est considéré comme la principale organisation de recherche au Sénégal employant plus de 70% de chercheurs. L'ISRA a pour principales missions : la conception et l'exécution de programmes de recherche sur les productions végétales, forestières, animales et halieutiques et en économie rurale ; la création de connaissances scientifiques, l'innovation technologique et la mise au point d'outils d'aide à la décision pour l'amélioration du secteur agricole ; la valorisation et le transfert des résultats de la recherche ; la promotion et la formation à la recherche par la recherche et le développement de la coopération scientifique aussi bien interafricaine et internationale qu'avec les institutions de recherche et universités sénégalaises.

2.1.2. Projets GloFoodS SERENA, SIIL SIMCo et TOSCA LYSA

SERENA [4] (de la biodiverSité des paysagEs agRicoles à la sEcurité alimeNtAire des ménages ruraux) est un projet financé dans le cadre du métaprogramme GloFoodS (Transitions pour la sécurité alimentaire mondiale), un métaprogramme conjoint INRA et Cirad qui a pour objectif de mobiliser les forces scientifiques pluridisciplinaires des deux établissements pour contribuer à éclairer les mécanismes, relevant des écosystèmes et des systèmes socio-économiques, qui sous-tendent les différentes dimensions de la sécurité alimentaire. Le projet SERENA a été proposé par l'UR Aïda et ses partenaires sénégalais (le CSE, l'ISRA, l'UCAD avec l'objectif d'éclairer la contribution d'un paysage agricole sénégalais diversifié à la sécurité alimentaire et nutritionnelle (SAN) des ménages ruraux. Ce projet est construit autour de l'hypothèse que la production d'aliments au sein de paysages agricoles polyvalents et diversifiés peut permettre d'accroître significativement la productivité des systèmes agricoles et favoriser les sources de revenus et/ou l'accès à des produits diversifiés. Il s'agit donc de mieux prendre en compte la dimension paysagère dans les études portant sur la SAN. L'intérêt et l'originalité du projet SERENA est de proposer une démarche de recherche mobilisant de façon innovante l'écologie du paysage, la télédétection, la modélisation spatialisée et des enquêtes socio-économiques de ménage pour traiter plusieurs aspects de la SAN à l'échelle d'un paysage sénégalais. Pour ce faire, le

3. https://www.isra.sn
4. https://www.projects.igeo.fr/projet-1/

projet s'articule autour de 4 axes : l'axe *diversité paysagère* (proposer un plan d'échantillonnage spatialisé à partir d'indicateurs de diversité paysagère et cartographier plus finement des indicateurs de la biodiversité des paysages agricoles à partir d'images satellite), l'axe *impact sur les rendements* (analyse du lien entre diversité paysagère et rendements à l'échelle des parcelles et du paysage), l'axe *SAN et moyens d'existence* (combiner enquêtes socio-économiques sur les ménages agricoles et indicateurs de biodiversité paysagère pour renseigner le lien entre diversité paysagère observée et les stratégies des ménages associées en matière de sécurité alimentaire) et l'axe *Valorisation* (diffusion des résultats du projet auprès de la communauté scientifique et des différents acteurs, à travers un site web intégrant une composante de webmapping et formations prévues)

SIMCo (Sustainable Intensification of Millet based agrosystems using Cowpea in the Groundnut Basin - Senegal) est un projet financé par le SIIL[5], laboratoire de l'Université du Kansas. Les recherches de ce projet visent une meilleure compréhension des mécanismes impliqués dans les associations culturales entre mil et niébé et l'estimation de leurs rendements sur divers sites du bassin arachidier sénégalais. De manière spécifique, les objectifs du projet SIMCo sont : (1) concevoir des agrosystèmes durables à base de céréales et légumineuses en calibrant et en validant des modèles de cultures associées ; et (2) estimer les rendements des cultures associées au niveau local et régional. Pour ce faire, le projet s'articule autour de 3 grandes étapes : l'acquisition des données de terrain pour la calibration des modèles de cultures, une étude comparative entre les systèmes de cultures à base de mil et les systèmes de cultures associées afin de mettre en lumière l'impact de l'association culturale sur les rendements et l'évaluation spatialisée des rendements estimés à l'échelle régionale. Le projet SIMCo est coordonné par l'ISRA et ses partenaires l'IRD et le Cirad.

LYSA[6] (from Landscape diversity to crop Yield monitoring in complex Smallholder Agricultural systems : a remote sensing approach based on multi-source dense time series of high spatial resolution imagery) est un projet financé dans le cadre de l'appel à projet de recherches du TOSCA (CNES). Le projet LYSA se propose d'étudier le lien entre la diversification des paysages agricoles et la sécurité alimentaire des ménages ruraux en se focalisant notamment sur l'impact de la diversité des cultures et/ou diversité/hétérogénéité du parc arboré sur la disponibilité alimentaire au travers du prisme des rendements des principales céréales alimentaires d'Afrique sahélienne. En particulier, l'objectif général du projet LYSA est de tester les potentialités offertes par des séries temporelles multi-sources denses et à haute résolution spatiale pour : (1) améliorer la caractérisation des paysages agricoles, en considérant notamment la structure du parc arboré et la diversité des espèces ; (2) améliorer l'estimation et la prévision des rendements des principales cultures céréalières alimentaires et ; (3) évaluer l'effet de la diversité paysagère (composition et structuration) sur les rendements. La démarche du projet LYSA s'articule autour de 4 grandes étapes : le prétraitement des données satellitaires, la caractérisation de la diversité paysagère (diversité des cultures et pratiques agricoles, caractérisation du parc arboré, création d'indicateurs synthétiques de diversité paysagère), l'estimation des rendements céréaliers à partir

5. http://www.k-state.edu/siil/index.html
6. https://www.projects.igeo.fr/projet-2/

de séries temporelles optique et radar en se basant sur un réseau de parcelles sur le terrain et la caractérisation de l'impact de la diversité des paysages sur les rendements à l'échelle de la parcelle et du paysage. Ce projet est coordonné par l'UR Aïda du Cirad et le CSE.

2.2. Objectifs et Hypothèses

L'objectif de ce stage est de mener une première analyse des potentialités offertes par une série temporelle multisource à haute résolution spatiale, pour un suivi spatialisé des systèmes de cultures à base de mil et d'arachide, à Diohine (près de la ville de Niakhar, région de Fatick) dans le bassin arachidier sénégalais. De manière plus concrète, il s'agira :

— d'évaluer les dates de semis sur les différentes parcelles
— et d'estimer les biomasses végétatives et rendements grains du mil et gousses de l'arachide.

Pour ce faire, une série temporelle multisource composée d'images *PlanetScope*, *RapidEye* et *Sentinel-2* couvrant presque entièrement la saison agricole 2017 sera utilisée. Nous disposons également de données de terrain décrivant, sur une quarantaine de parcelles de la zone, les pratiques agricoles adoptées ou encore les caractéristiques agronomiques dont les rendements observés. La principale contrainte pour ce travail que l'on peut considérer toutefois comme anodine, se trouve dans l'utilisation de solutions libres et open-source. Ce travail vient en amont d'une étude plus complète visant à évaluer à l'échelle d'un paysage, l'impact de la biodiversité paysagère sur la productivité des systèmes de cultures et les potentiels d'intensification des pratiques.

Nos hypothèses pour ce travail sont les suivantes :

— La pratique des agriculteurs est de semer le mil à sec avant les premières pluies et l'arachide dès la première pluie significative. Le début de la saison pluvieuse survient généralement vers la mi ou fin Juin, par conséquent les dates de semis estimées devraient donc se situer en Juin et les dates de fin de saison au plus tard en fin Septembre ou début Octobre selon les dates de récolte observées.
— L'évaluation spatialisée des rendements des systèmes de culture mixte (arachide/mil — niébé) devrait être moins évidente que celle des systèmes de culture pure en raison de la mixité du signal capté par les satellites.

2.3. Planning de travail

Afin d'assurer sa bonne marche, le présent travail a été subdivisé en une série de 5 tâches principales à réaliser sur la période des 6 mois de stage :

— État de l'art sur les méthodes d'extraction des dates de semis et d'estimation des rendements par télédétection
— Mise en place d'une chaîne de prétraitements des données et d'extraction de variables permettant de décrire l'évolution de la végétation et les caractéristiques biophysiques des couverts
— Évaluation des dates de semis
— Estimation des rendements

— Rédaction du mémoire de stage.

La planification de ces tâches et leur détail sont illustrés à travers le diagramme de Gantt présenté dans l'annexe A.

Dans les lignes qui suivent, nous débuterons par un exposé synthétique sur les méthodes d'extraction des dates de semis et d'estimation des rendements par télédétection. Ceci aboutira à la présentation des méthodes adoptées dans ce travail et nous conclurons notre étude après la présentation des résultats obtenus et leur analyse critique.

3. Synthèse bibliographique

3.1. Extraction des dates de semis par télédétection

3.1.1. Considérations générales

Les périodes ou dates les plus appropriés pour semer dépendent de nombreux facteurs. Nous pouvons considérer les espèces et leurs variétés, les températures selon la zone de production, l'humidité du sol, les objectifs de la production (date de récolte souhaitée) ou encore les pratiques culturales (semi direct sans labour, semi en sec avant les premières pluies . . .). Cependant, l'on peut s'accorder sur le rôle prépondérant que joue le facteur climatique et plus précisément le début de la saison des pluies (Ingram et al., 2002; Barbier et al., 2009). En effet, au Sénégal comme dans d'autres pays d'Afrique où l'agriculture est pluviale, le comportement des agriculteurs est généralement celui de semer après les premiers épisodes pluvieux importants (Bacci et al., 1999). Néanmoins, il existe un risque d'échec des premières semences en cas d'avènement d'une sécheresse (Marteau et al., 2011) qui peut emmener les agriculteurs à semer de nouveau.
Étant donné l'étroite relation entre les rendements agricoles et la durée du cycle de développement des céréales comme le mil, des semis tardifs sont susceptibles de conduire à des rendements faibles en fin de saison (Sivakumar, 1990). Le suivi du démarrage de la saison agricole permet donc de fournir aux décideurs, une évaluation précoce des menaces potentielles à la production agricole et à la sécurité alimentaire. L'une des méthodes les plus courantes pour l'estimation des dates de semis dans les pays d'Afrique de l'Ouest repose sur une approche agrométéorologique ou agroclimatiques qui consiste à appliquer des seuils sur les quantités de précipitations survenues pendant une période définie (Marinho et al., 2014). Cependant, cette méthode est limitée par les inconvénients qu'elle présente : la résolution spatiale des données qui est souvent grossière (de l'ordre des kilomètres) et l'estimation elle-même des quantités de précipitations qui peut être imprécise. Compte tenu de notre échelle de travail qui est celle de la parcelle, cette méthode n'est pas adaptée à notre étude et ne peut être considérée.
Une autre approche par télédection cette fois-ci consiste à dériver les dates du début de croissance de la végétation ou en général les *métriques phénologiques* à partir de séries temporelles d'indices de végétation comme le NDVI ou le EVI et à en déduire les dates de semis. Ces indices de végétation rendent compte entre autres de l'activité photosynthétique de la végétation, de l'intensité de son métabolisme ainsi que de sa verdure (Duarte et al., 2018). Cette approche par télédétection présente non seulement l'avantage d'avoir une résolution spatiale beaucoup plus élevée en fonction de l'image satellitaire utilisée mais d'intégrer également la réponse spectrale de la végétation aux divers facteurs externes comme les pratiques agricoles. Bien évidemment, les contraintes intrinsèques à l'acquisition des images satellitaires ne sont pas à négliger. C'est notamment le cas des bruits induits par les perturbations atmosphériques

(nébulosité). Il faudra donc avant l'exploitation des images, considérer par exemple l'application d'une méthode de lissage. Une autre précaution est à prendre vis-à-vis de l'influence des sols nus dans la réponse spectrale du couvert végétal notamment quand celui ci est clairsemé tel pour des cultures peu couvrantes comme le mil. Il faudra considérer par exemple l'utilisation d'un indice de végétation tenant compte de l'effet des sols comme le SAVI ou le MSAVI. Il existe une foultitude de méthodes dans la littérature pour estimer le démarrage de la végétation en étudiant sa phénologie par observation satellitaire.

3.1.2. Phénologie de la végétation

La *phénologie* est l'étude de l'apparition d'événements périodiques (annuels le plus souvent) dans le monde vivant, déterminée par les variations saisonnières du climat. Elle se traduit au niveau de la végétation, par l'ensemble des stades de développement intervenant dans le cycle de vie des plantes en l'occurrence le bourgeonnement, la croissance, la floraison et la sénescence (Kimball, 2014). La phénologie du mil peut être par exemple décomposée en une phase *végétative*, une phase *reproductrice* et une phase de *maturation*. Selon les variétés de mil cultivées, la longueur du cycle de développement peut être de 90 à 100 jours pour les *Souna* (variétés à cycle court — petits grains, épis non aristés et peu photosensibles) ou de 130 à 150 jours pour les *Sanio* (variétés à cycle long — grains plus gros, épis aristés et variété photopériodique) (Diouf, 2001).
L'étude de la phénologie des plantes sur une échelle régionale à globale à partir d'observations satellitaires est désignée par LSP (Helman, 2018). Le terme LSP tient compte du fait que le signal intercepté par les satellites provient d'une surface hétérogène et n'est pas représentatif de la réponse spectrale d'une seule espèce (Kimball, 2014). L'analyse de la LSP contribue entre autres à diverses applications comme l'étude des changements climatiques (Bégué et al., 2014) ou encore au suivi des cultures. Les séries temporelles d'indices de végétation provenant de capteurs à basse résolution comme AVHRR, MODIS, SPOT Vegetation ou encore PROBA-V permettent d'évaluer la phénologie de la végétation sur une échelle régionale à globale. Plus récemment, quelques auteurs ont étudié la phénologie de la végétation à une échelle beaucoup plus fine en utilisant des données à haute et très haute résolution comme ceux du satellite chinois HJ-1 A/B (Pan et al., 2015) ou du satellite RapidEye (Vrieling et al., 2017).

3.1.3. Métriques phénologiques

Le démarrage de la croissance de la végétation et plus généralement les métriques phénologiques peuvent être utilisés comme proxy à l'évaluation des dates de semis. Les métriques phénologiques ou variables phénologiques désignent l'ensemble des stades phénologiques du cycle saisonnier de la végétation dérivés par observation satellitaire (Helman, 2018). Elles fournissent donc des indications sur la dynamique des écosystèmes. Il s'agit usuellement (Figure 3.1) :

— du démarrage de croissance de la végétation ou SOS,
— du pic ou du maximum de croissance de la végétation POS,
— de la fin de croissance de la végétation EOS,

— et de la durée de croissance de la végétation GSL : différence de temps entre le SOS et le EOS.

Pour ces premières métriques citées, il importe de noter que l'on considère à la fois la date (jour exact dans l'année ou nombre de jours dans le cas du GSL) et la valeur de l'indice de végétation utilisé par exemple le NDVI. D'autres métriques en plus peuvent être également dérivées :

— le niveau de base (le plus bas) au cours du cycle de développement de la végétation
— le milieu de croissance de la végétation
— le taux de croissance de la végétation au début du cycle
— le taux de décroissance de la végétation à la fin du cycle
— la petite intégrale saisonnière (cumul du SOS au EOS pour les valeurs au dessus de la base)
— la grande intégrale saisonnière (cumul du SOS au EOS en totalité).

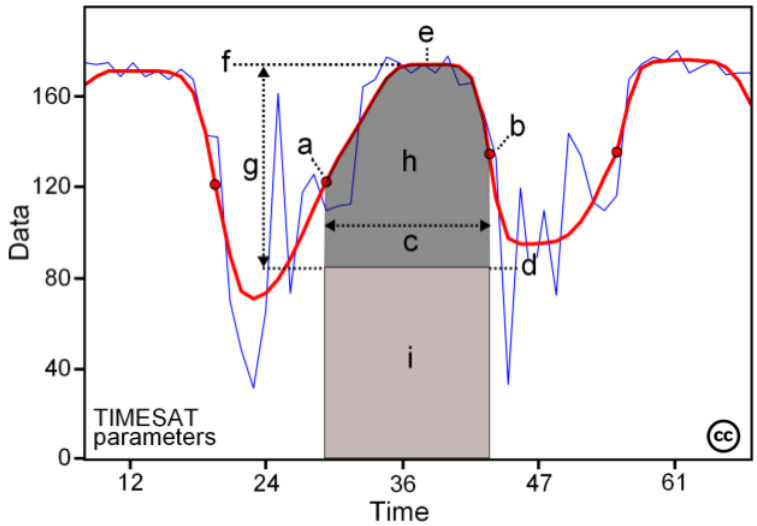


Figure 3.1. – Illustration de quelques métriques phénologiques : *(a)- SOS (b)- EOS (c)- Durée de la saison (d)- Valeur de Base (en unité d'indice de végétation) (e)- Milieu de la saison (f)- Maximum de l'indice de végétation (g)- Amplitude de l'indice de végétation (h)- Petite intégrele (h et i)- Grande intégrale*, Source : Eklundh and Jönsson (2017)

Outre leur utilisation comme proxy à l'évaluation des dates de semis, les métriques phénologiques peuvent être employées à d'autres fins comme l'estimation de rendements, exploitant la forte corrélation avec la biomasse en fin de saison ou encore la détection d'anomalies dans le cycle de la végétation, en considérant une référence ou une normale parmi un historique de saisons culturales.

Comme évoqué dans la section 3.1.1, plusieurs méthodes ont été mises au point pour dériver les métriques phénologiques à partir de séries temporelles d'indices de végétation. Beck et al. (2006) et Atzberger (2013) les ont classé en 2 catégories :

— un premier groupe de méthodes qui estime le timing des transitions phénologiques de manière isolée et indépendante les unes des autres

— et un second groupe qui modélise la série temporelle entière par une fonction mathématique pour estimer les métriques phénologiques.

Dans le premier groupe, sont classées surtout les méthodes par seuillage sur les valeurs d'indices de végétation ou sur leurs amplitudes. Dans leur revue de Beurs and Henebry (2010) en ont décrit quelques unes comme la méthode par seuillage basée sur les ratios de NDVI (White et al., 1997). Les méthodes par seuillage sont les plus simples et les plus communes (Pan et al., 2015). En effet, elles supposent qu'un stade phénologique a commencé quand la valeur de l'indice de végétation atteint un certain seuil fixé (Jönsson and Eklundh, 2002). Cependant, cette approche peut s'avérer inconsistante quand le couvert végétal n'est pas homogène ou quand il s'agit d'écosystèmes cultivés avec plusieurs cycles culturaux au cours de l'année. De plus, un seuil n'est pas toujours transposable du fait qu'il soit bien souvent lié à sa zone d'application. La seconde catégorie regroupe les fonctions gaussiennes, les modèles logistiques et quadratiques (Zhang et al., 2003; Jönsson and Eklundh, 2004), l'analyse en composantes principales, l'analyse harmonique ou encore la transformation en ondelettes. Ces dernières méthodes sont plus poussées et plus complexes à mettre en œuvre. Il existe également une autre catégorie de méthodes dite de dérivées de courbes qui définissent généralement le SOS et le EOS respectivement comme les moments où l'on observe la plus grande hausse et la plus grande chute dans les valeurs de l'indice de végétation (Moulin et al., 1997; Tateishi and Ebata, 2004).

Bon nombre de logiciels et applications ont été développés pour traiter les séries temporelles d'indices de végétation et en extraire les métriques phénologiques souhaitées. Nous avons identifié les suivants :

— TIMESAT [1] (Eklundh and Jönsson, 2017)
— SPIRITS [2]
— PhenoSat [3] (Rodrigues et al., 2013)
— Plugin QGIS QPhenoMetrics (Duarte et al., 2018)
— Plugin QGIS VERSAO VegaMonitor [4].

3.2. Estimation des rendements par télédétection

L'évaluation des rendements agricoles, qu'elle intervienne très tôt dans la saison (prévision) ou en fin de saison (estimation), est une information d'une extrême utilité pour de nombreux acteurs du secteur agricole, particulièrement dans les pays dont l'économie est fortement influencée par la production agricole annuelle. L'estimation des rendements couplée à l'évaluation des superficies cultivées fournit notamment les statistiques agricoles nécessaires aux gouvernements dans leur planification des budgets d'importation et d'exportation de denrées alimentaires. L'estimation des rendements agricoles intéresse également les Systèmes d'Alerte Précoce (SAP) qui doivent effectuer un suivi exhaustif des conditions de développement des cultures au cours de la saison agricole afin notamment d'estimer la production des principales cultures et déterminer l'organisation éventuelle d'aides alimentaires.

En agriculture, le rendement est la quantité de produit récolté sur une surface cultivée

1. http://web.nateko.lu.se/timesat/timesat.asp?cat=0
2. http://spirits.jrc.ec.europa.eu/
3. http://www.fc.up.pt/PhenoSat/software.html
4. https://github.com/Xdarii/VERSAO_VegaMonitor/wiki

souvent exprimée en kilogrammes, quintaux ou tonnes par hectares. Dans le cas d'une culture céréalière, le rendement est le poids des grains récoltés par unité de surface. Le rendement d'une culture dépend des interactions entre 3 paramètres : le potentiel génétique ou caractéristiques variétales de la plante cultivée, l'environnement naturel réel de la plante et les pratiques agricoles adoptées (Figure 3.2). En fonction des interactions entre ces 3 paramètres, on distingue 3 types de rendements (Neumann et al., 2010; Tittonell and Giller, 2013; Van Ittersum et al., 2013) :

— le *rendement potentiel* qui est le rendement obtenu pour une culture si tous les facteurs déterminants, notamment les conditions environnementales, sont reunies ;
— le *rendement atteignable* qui est le rendement maximum atteignable en fonction des facteurs limitants (contraintes hydriques dans le cas de cultures pluviales) ;
— et le *rendement observé* est le rendement obtenu dans les conditions réelles, déterminé par différents facteurs de réduction comme la présence de mauvaises herbes, de ravageurs ou de maladies ...

Au final, le rendement observé est déterminé par l'action conjointe des facteurs déterminants, limitants et des facteurs de réduction.

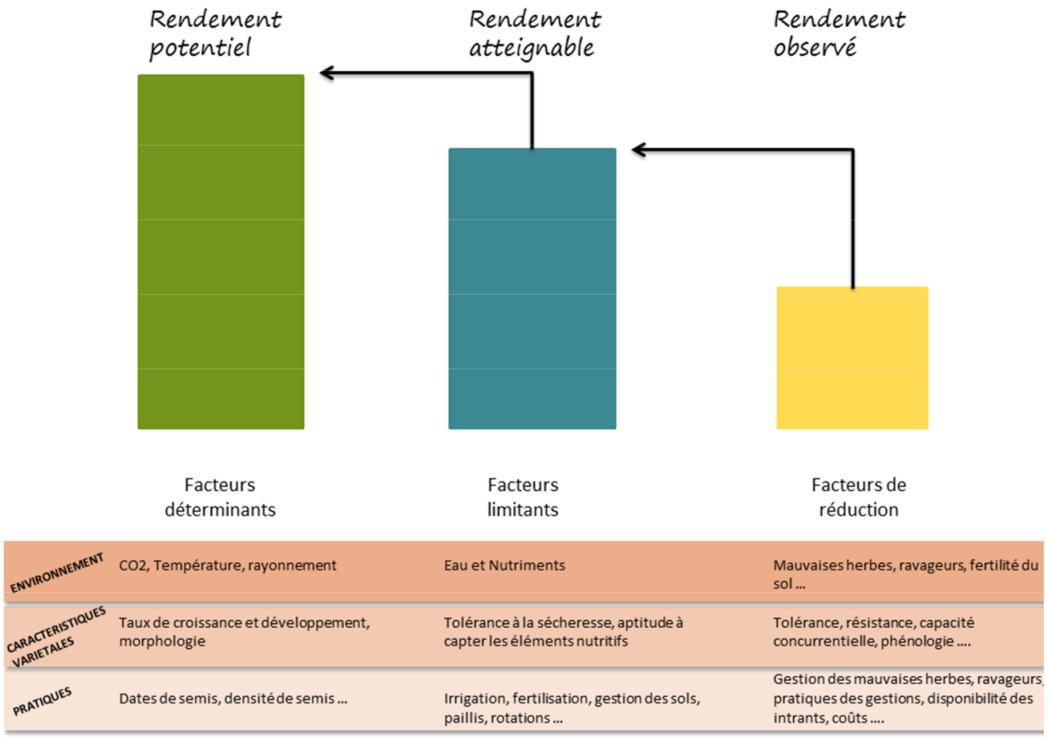


Figure 3.2. – Définitions du rendement potentiel, atteignable et observé et des principaux facteurs déterminants, Source : Leroux (2015)

Du fait qu'elles intègrent l'ensemble des facteurs affectant les rendements observés, les données de télédétection apparaissent comme un outil pertinent pour l'estimation des rendements agricoles. De plus, grâce à la vue synoptique qu'offrent les images satellitaires, la télédétection est complémentaire des méthodes traditionnelles que sont la modélisation agro-météorologique et les enquêtes terrains. La relation entre les

propriétés spectrales des cultures et leur biomasse et rendement a été mise en évidence très tôt, peu après le lancement du premier satellite civil Landsat-1 en 1972. Tucker fut l'un des premiers à montrer que la combinaison des réflectances spectrales dans les bandes du rouge et du proche infrarouge, communément appelé indices de végétation, était corrélée à la biomasse totale aérienne sous la forme d'une relation linéaire (Tucker, 1979) et à estimer les rendements de cultures à partir d'observations satellitaires (Tucker et al., 1981; Tucker and Sellers, 1986). D'autres études ayant suivi ont montré que le NDVI était étroitement lié à l'activité photosynthétique de la végétation et aux variables biophysiques comme le LAI et le FAPAR. En raison de sa relation quasi linéaire avec le FAPAR, le NDVI est une mesure indirecte de la productivité primaire de la végétation (Prince, 1990). Dans leur article de synthèse, Rembold et al. (2013) ont distingué 2 types d'approches pour le suivi des rendements : une approche qualitative et des approches quantitatives. Le suivi qualitatif des rendements repose sur l'interprétation des indices de végétation tout au long de la saison agricole notamment par rapport aux données historiques afin de détecter des anomalies de production de biomasse. Le VCI (Kogan, 1995) est un indice de végétation bien connu pour ce cas de figure. Les approches quantitatives pour estimer le rendement d'une culture sur base de données de télédétection se présentent soit sous la forme de modèles de régression entre observations satellitaires et rendements observés ou sous la forme de modèles de simulation de la croissance des cultures.

Modèles de régression Il s'agit d'établir une relation empirique entre les rendements observés sur le terrain et les indices de végétation dérivés d'observations satellitaires, généralement le NDVI à une date donnée ou cumulé sur un intervalle de temps caractéristique des stades phénologiques des cultures. Dès lors, la fiabilité de cette relation est grandement déterminée par la qualité des données récoltées sur le terrain. D'autres types d'informations peuvent être rajoutés aux observations satellitaires pour mieux affiner les modèles établis, notamment des variables biophysiques comme le LAI ou des variables bio-climatiques et météorologiques à l'instar de l'humidité du sol, des quantités de précipitations et des indices thermiques (Leroux et al., 2016). Par ailleurs, l'utilisation de données historiques sur les rendements des saisons agricoles précédentes peut aussi aider à mieux déterminer cette relation. L'utilisation d'un masque de cultures provenant notamment d'une carte d'occupation du sol est également une pratique courante dans l'estimation des rendements. Ceci peut aider à améliorer les estimations en réduisant la part de variabilité du NDVI due à la présence de la végétation naturelle. Les modèles de régression les plus simples pour l'estimation des rendements sont linéaires avec une ou plusieurs variables (Johnson, 2014; Lambert et al., 2018). D'autres méthodes plus complexes par fouille de données peuvent être également utilisées comme les forêts d'arbres décisionnels ou les réseaux de neurones artificiels (Fieuzal et al., 2017).

Modèles de simulation de la croissance des cultures Dans ce cas de figure, les observations satellitaires sont couplées avec des modèles de simulation de la croissance des cultures. Ces modèles mathématiques basés sur des équations déterministes permettent de formaliser les connaissances que l'on a sur les mécanismes physiologiques de croissance des végétaux et leurs interactions avec les principaux facteurs limitants comme le climat ou le niveau de fertilité du sol, tout en prenant en compte les différentes pratiques culturales (Delécolle et al., 1992). Dans ces modèles, les variables

biophysiques dérivées des données de télédétection (LAI, FAPAR) sont utilisées pour (re)-paramétrer, (re)-calibrer ou forcer les modèles (Jin et al., 2018). Ceci permet à la fois de bénéficier d'une estimation continue dans le temps de la croissance des cultures et d'une spatialisation des variables d'état du couvert végétal et des variables de sorties comme la biomasse ou le rendement (Rembold et al., 2013).

4. Matériels et Méthodes

4.1. Présentation générale de la zone d'étude

Notre zone d'étude (Figure 4.1) est située dans le bassin arachidier au centre ouest du Sénégal à cheval entre les arrondissements de N'goye (département de Bambey, région de Diourbel), de Niakhar et Tattaguine (département et région de Fatick) et de Mbour (département de Fissel et région de Thiès). Le principal centre urbain de notre zone d'intérêt est la ville de Niakhar située au sud-est. Notre zone d'étude est essentiellement agricole avec principalement des cultures de mil, d'arachide, de niébé et de sorgho. On y trouve également de nombreuses espèces arborées dont les principales sont le *Faidherbia albida*, le Baobab (*Adansonia digitata*) et le Neem (*Azadirachta indica*).
La zone est située dans une vaste plaine dont les altitudes maximales ne dépassent pas les 100 mètres. Il est courant de retrouver quelques dépressions communément appelées bas-fonds. Deux types de sols y sont rencontrés : les sols *Dior* et les sols *Deck*. *Dior* est le terme vernaculaire désignant des sols ferrugineux tropicaux lessivés, meubles et très perméables. Ces sols sont propices aux cultures d'arachide et de mil. Les *Deck* sont des sols ferrugineux tropicaux non lessivés à forte teneur en limons et argiles. De ce fait elles ont une plus grande capacité de rétention en eau. La nappe phréatique existante est salée et empêche toute activité d'irrigation en plus du problème sans cesse accru de la salinisation des sols.
Notre zone d'étude étant située plus vers le Sud du bassin arachidier, son climat est plutôt du type soudano-sahélien caractérisé par une saison sèche et une saison pluvieuse. La saison sèche relativement longue s'étend d'Octobre à Juin avec des températures mensuelles moyennes comprises entre 25 et 40°C influencées par l'harmattan (vent chaud et sec). La saison pluvieuse (hivernage) s'étend de Juillet à Octobre avec des pluies fortes et intenses pendant les mois d'Août et Septembre associées au vent de la Mousson. La pluviométrie interannuelle est très variable avec une moyenne autour de 700 mm.

4.2. Données satellitaires utilisées

Notre série temporelle est composée de 3 types d'imageries satellitaires. Il s'agit d'images PlanetScope, RapidEye et Sentinel-2.

4.2.1. Images PlanetScope et RapidEye

Les images PlanetScope sont produites par la société privée américaine Planet Labs, Inc. [1]. Fondée en 2010 et basée à San Francisco en Californie, cette société est spécialisée dans l'observation de la terre par imagerie satellitaire. Planet Labs conçoit et

1. https://www.planet.com/

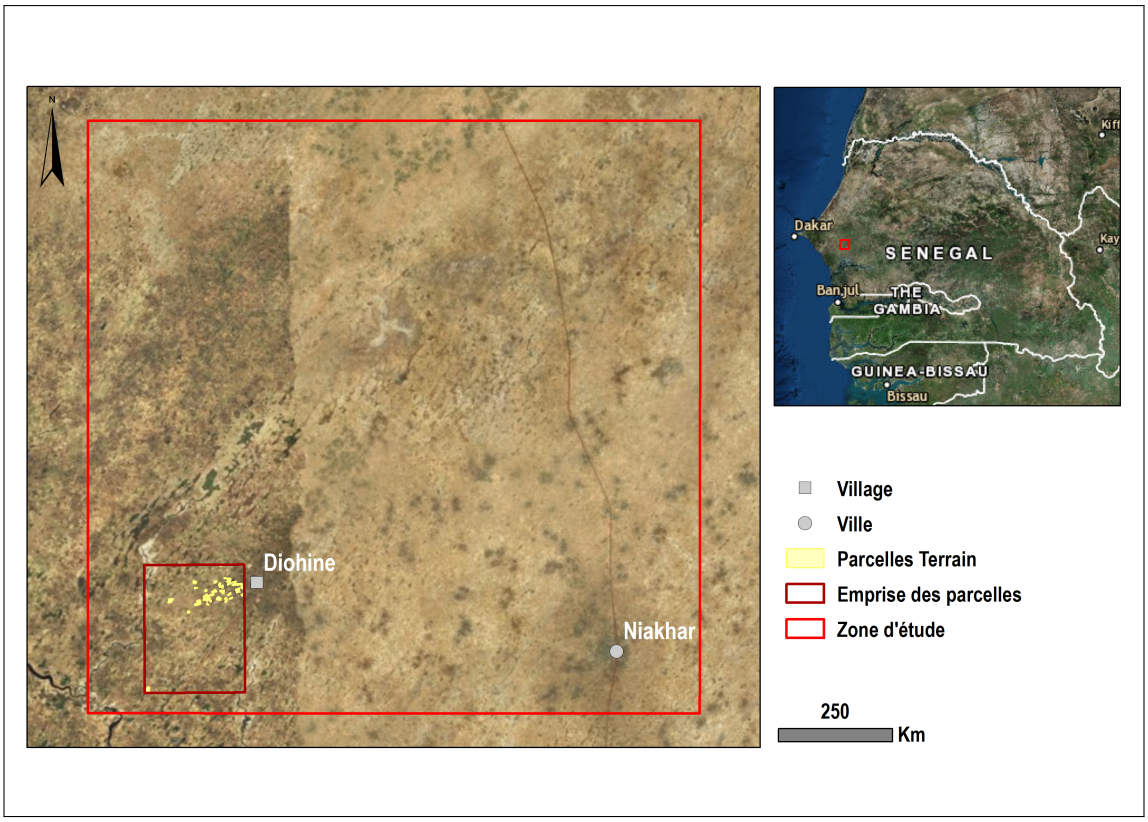


Figure 4.1. – Situation géographique de la zone d'étude

fabrique des *nanosatellites* ou *CubSats* ($10 \times 10 \times 30\,cm$) appelés *Doves* qui sont placés sur orbite en tant que charge utile secondaire sur d'autres missions de lancement de fusée. La société dispose ainsi d'une constellation de nanosatellites surnommée *Flock*. La constellation est composée actuellement de 175 nanosatellites positionnés sur une orbite héliosynchorne à 475 km d'altitude. Ces nanosatellites produisent des images complètes de la Terre une fois par jour, à une résolution spatiale variant de 3 à 5 mètres. Ces images fournissent des informations permettant de suivre les changements climatiques, de prévoir les récoltes, de gérer les catastrophes ou encore de mettre au point des applications urbaines. Les images recueillies par les Doves sont accessibles en ligne et certaines disponibles dans le cadre de l'Open Data.
En 2015, Planet Labs a acquis la constellation RapidEye auprès de la société allemande BlackBridge. Les images RapidEye qui ont une résolution spatiale de 5 mètres sont fournies par une constellation de 5 satellites. Leur période de revisite est de 5 jours et demi au nadir ou journalière sinon.

Spécifications des produits PlanetScope Trois niveaux de traitements sont disponibles pour les images PlanetScope. Ce sont les niveaux 1B, 3A et 3B. Le niveau 1B correspond aux produits basiques. Les données numériques ont été calibrées en radiance TOA mais les images ne sont pas géoréférencées. Le niveau 3B qui est celui de nos images correspond à des produits orthorectifiés et projetés en UTM. Comme pour le niveau 1B, les valeurs numériques ont été calibrées en radiance TOA. Le niveau 3A est similaire au 3B à la différence que les images sont tuilées pour couvrir un système de grilles de 25×25 kilomètres.

Les images PlanetScope sont distribuées en format *GeoTiff*. Au niveau 3B, elles ont une résolution au sol de *3 mètres* et une résolution radiométrique de *12 bits* s'il s'agit de compte numérique ou *16 bits* dans le cas des radiances TOA. Elles disposent de 4 bandes spectrales (Tableau 4.1).

Tableau 4.1. – Caractéristiques spectrales des images PlanetScope

Bande spectrale	Domaine spectral	Longueurs d'onde (µm)
1	Bleu	0,455 — 0,515
2	Vert	0,500 — 0,590
3	Rouge	0,590 — 0,670
4	Proche Infrarouge	0,780 — 0,860

La conversion des données PlanetScope en réflectance TOA s'effectue par la formule suivante :

$$Reflectance(i) = DN(i) \times reflectanceFactor(i) \tag{4.1}$$

où :
i = Numéro de la bande spectrale de l'image,
DN = Valeurs numériques (brutes) de l'image,
et *reflectanceFactor(i)* = Facteur de conversion en réflectance pour une bande spectrale donnée, renseigné dans les metadonnées de l'image.

Spécifications des produits RapidEye Les images RapidEye sont pour leur part, disponibles en 2 niveaux de traitements : 1B et 3A. Ces niveaux de traitements sont identiques à ceux des produits PlanetScope. Les images RapidEye utilisées sont traitées au niveau 3A. Elles sont distribuées également en format *GeoTiff*. Leur résolution au sol est de *5 mètres* et leur résolution radiométrique de *16 bits*. Elles disposent de 5 bandes spectrales (Tableau 4.2).

Tableau 4.2. – Caractéristiques spectrales des images RapidEye

Bande spectrale	Domaine spectral	Longueurs d'onde (µm)
1	Bleu	0,440 — 0,510
2	Vert	0,520 — 0,590
3	Rouge	0,630 — 0,685
4	Red Edge	0,690 — 0,730
5	Proche Infrarouge	0,760 — 0,850

La conversion des données RapidEye en réflectance TOA s'effectue en 2 étapes à savoir la conversion en radiance puis la conversion en réflectance :

$$Radiance(i) = DN(i) \times radiometricScaleFactor(i) \tag{4.2}$$

$$Reflectance(i) = Radiance(i) \times \frac{\pi \times SunDist^2}{EAI(i) \times \cos(SolarZenith)} \tag{4.3}$$

où :
i = Numéro de la bande spectrale de l'image
DN = Valeurs numériques (brutes) de l'image
radiometricScaleFactor(i) = Facteur de conversion en radiance pour une bande spectrale donnée, renseigné dans les métadonnées de l'image
SunDist = Distance Terre-Soleil en unités astronomiques
EAI(i) = Irradiance exo-atmosphérique pour une bande spectrale donnée, renseignée dans les métadonnées de l'image
SolarZenith = Angle zénithal solaire en degrés (90°- Elévation solaire)

4.2.2. Images Sentinel-2

Sentinel est une famille de 6 satellites d'observation de la Terre, développée par l'ESA et destinée à assurer la continuité des données de la mission ENVISAT arrivée à terme en 2012. Sentinel représente le volet spatial du programme Copernicus (ex GMES) de l'Union Européenne qui vise à doter l'Europe d'une capacité autonome et opérationnelle en matière d'observation de la Terre notamment pour la surveillance de l'environnement et la sécurité (Drusch et al., 2012). La mission Sentinel-2 est la composante spatiale du programme devant fournir une imagerie optique haute résolution permettant l'observation des sols (utilisation, végétation, zones côtières, fleuves ...) ainsi que la mise en place de services de traitement des situations d'urgence notamment les catastrophes naturelles. Cette mission est composée des satellites *Sentinel-2A* et *Sentinel-2B* qui circulent en déphasage de 180° sur la même orbite héliosynchrone de 10h30. Ces 2 satellites sont identiques et embarquent l'instrument MSI qui fournit des images dans 13 bandes spectrales du Visible à l'Infrarouge (Tableau 4.3) avec une résolution spatiale comprise entre 10 et 60 mètres et une résolution radiométrique de 12 bits. Les satellites Sentinel-2 sont configurés pour une période de revisite de 5 jours au nadir et doivent acquérir au moins une donnée claire par mois sur la plupart des terres émergées. Cette grande richesse spectrale couplée à cette capacité d'observation temporelle élevée constituent le véritable apport de la mission Sentinel-2. Les données sont principalement utilisées pour l'agriculture, la sylviculture, l'occupation des sols, la biodiversité, la caractérisation des habitats ou encore l'observation et la prévention des catastrophes naturelles notamment les inondations.

Les images Sentinel-2 utilisées dans ce travail sont produites par Theia/MUSCATE [2]. Il s'agit de données Copernicus Sentinel-2 de niveau 1C (données ortho-rectifiées en réflectance TOA) qui sont traités au niveau 2A (données ortho-rectifiées en réflectance de surface après correction atmosphérique notamment par MACCS/MAJA [3]). Au niveau 2A, des masques de nuages et d'ombres ainsi que des surfaces d'eau et de neige sont fournis. Les données de réflectance fournies sont de 2 types :

— le SRE pour Surface REflectance qui incluent les corrections atmosphériques et effets d'environnement
— le FRE pour Flat REflectance qui incluent en plus des corrections des données SRE, les corrections liées aux effets de pente.

Notons qu'à terme, seules les données FRE seront fournis par Theia, afin de diminuer les volumes à distribuer.

2. http://www.theia-land.fr/
3. http://www.cesbio.ups-tlse.fr/multitemp/?p=6050

Tableau 4.3. – Caractéristiques spectrales de l'instrument MSI des Sentinel-2

Bande spectrale	Domaine spectral	Longueurs d'onde (µm)	Résolution spatiale (m)
1	Aérosols	0,433 — 0,453	60
2	Bleu	0,4575 — 0,5225	10
3	Vert	0,5425 — 0,5775	10
4	Rouge	0,65 — 0,68	10
5	Red Edge	0,6975 — 0,7125	20
6	Red Edge	0,7325 — 0,7475	20
7	Red Edge	0,773 — 0,793	20
8	Proche Infrarouge	0,83625 — 0,84775	10
8A	Red Edge	0,855 — 0,875	20
9	Vapeur d'eau	0,935 — 0,955	60
10	Cirrus	1,365 — 1,395	60
11	Infrarouge moyen	1,565 — 1,655	20
12	Infrarouge moyen	2,181 — 2,199	20

Nos images Sentinel-2 sont des données SRE. Elles sont codées en entiers sur du 16 bits signé. L'obtention des vraies valeurs de réflectance est donnée par :

$$Reflectance(i) = \frac{ReflectanceCodee(i)}{10\,000} \tag{4.4}$$

Nous avons mis en place une chaîne de prétraitements, écrite avec le langage de programmation `Python`, pour effectuer la calibration de nos images en données de réflectance (TOA pour les images PlanetScope et RapidEye et réflectance de surface pour les images Sentinel-2) et leur découpage selon l'emprise de notre zone d'étude. Les bibliothèques `gdal` et `rasterio` ont été utilisées à cet effet. Notons que pour les images PlanetScope et certaines images RapidEye, nous avons d'abord dû effectuer un mosaiquage de toutes les images correspondant à la même date d'acquisition avant de les découper. Les bandes Red Edge des images Sentinel-2 (20 mètres de résolution) et celles de 60 mètres, plus relatives aux applications météorologiques n'ont pas été considérées dans le cadre de ce stage.

4.2.3. Chronologie des images acquises

Les 26 images utilisées ont été acquises entre Juin et Novembre 2017, soit en théorie une image tous les 7 jours sur la période couvrant la saison agricole 2017. Nous ne disposons, cependant, que d'une image (PlanetScope) en Septembre. La chronologie des acquisitions est représentée sur la figure 4.2.

4.3. Données de terrain

Notre jeu de données de terrain est basé sur des enquêtes agronomiques réalisées au cours de la saison agricole 2017 sur 47 parcelles d'environ un demi hectare de

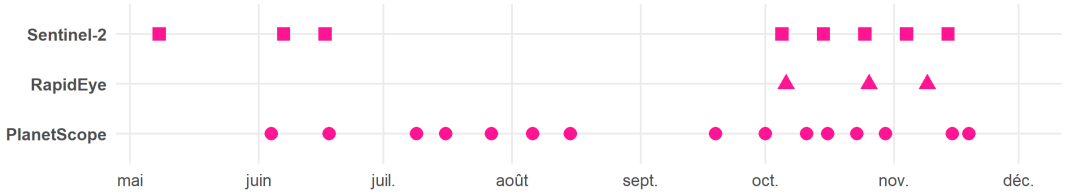


Figure 4.2. – Chronologie des images acquises

surface, situées à l'ouest du village de Diohine. Ces enquêtes ont été menées dans le cadre des projets Oracle[4] et ANR CERAO[5]. Le jeu de données du projet Oracle provient de la thèse de Djiba Sophie et consiste en un réseau de 30 parcelles enquêtées en milieu paysan où les pratiques agricoles sont axées sur des rotations entre céréales et légumineuses (mil/arachide et/ou niébé) et des associations entre légumineuses (arachide/niébé). Par contre, celui du projet ANR CERAO, mis au point par Tounkara Adama lors de sa thèse, est un réseau de 17 parcelles enquêtées également en milieu paysan mais présentant chacune sur une partie, des essais relatifs à l'efficience de l'utilisation de l'azote par le mil en fonction de différents niveaux d'apports ou non en fertilisants, et ce indépendamment des pratiques de l'agriculteur. Nous émettons l'hypothèse que les résultats de ces essais notamment les données de biomasse et rendement sont représentatifs des parcelles concernées, néanmoins, conscients de l'incertitude potentielle dans cette généralisation, nous envisageons d'analyser nos résultats séparement en tenant compte de la source des données de terrain. Pour chaque parcelle enquêtée, différentes informations sont collectées. Celles qui nous interessent principalement sont :

— le système de culture : espèces cultivées et pratiques culturales (fertilisant, dates de semis et de récoltes)
— la biomasse (fraîche et sèche) et le rendement (frais et sec)
— la géolocalisation des parcelles
— le nombre et les espèces d'arbres présentes au sein des parcelles.

Pour les parcelles de mil, les données de biomasse représentent le poids des pailles (fraîches ou sèches) sur une surface de 15 m² (3 placettes) extrapolé à l'hectare, calculé à partir du poids total des pailles pour 10 pieds de mil et du nombre de pieds se trouvant sur une surface de 15 m². Les rendements représentent le poids des grains (frais ou secs) une fois extraits de leur épis et calculé par le même procédé. Pour les parcelles d'arachide, la biomasse totale (fraîche ou sèche) est mesurée sur une surface de 15 m² puis extrapolée à l'hectare et les rendements correspondent au poids (frais ou sec) des gousses matures pour 15 m² extrapolé également à l'hectare.

Le tableau 4.4 présente les principales statistiques descriptives pour les parcelles enquêtées et comme nous pouvons nous en apercevoir, les rendements du mil sont meilleurs quand il y'a association culturale avec une légumineuse (1670,21 kg/ha en système de culture mixte contre 989,51 kg/ha en système de culture pure). La figure 4.3 présente la localisation des parcelles terrain suivant leur système de culture.

4. https://afrique-ouest.cirad.fr/actualites/projet-oracle-2017-2020
5. http://www.agence-nationale-recherche.fr/Projet-ANR-13-AGRO-0002

Tableau 4.4. – Statistiques descriptives des données de terrain

Culture majoritaire	Système de culture	Nombre de parcelles	Biomasse fraiche (kg/ha)		Rendement frais (kg/ha)	
			$\bar{X}$	σ	$\bar{X}$	σ
Arachide	Mixte	12	6288,89	2041,67	1487,76	596,52
Mil	Pure	9	5296,09	2662,75	989,51	398,27
	Mixte	26	8981,77	5381,86	1670,21	757,49

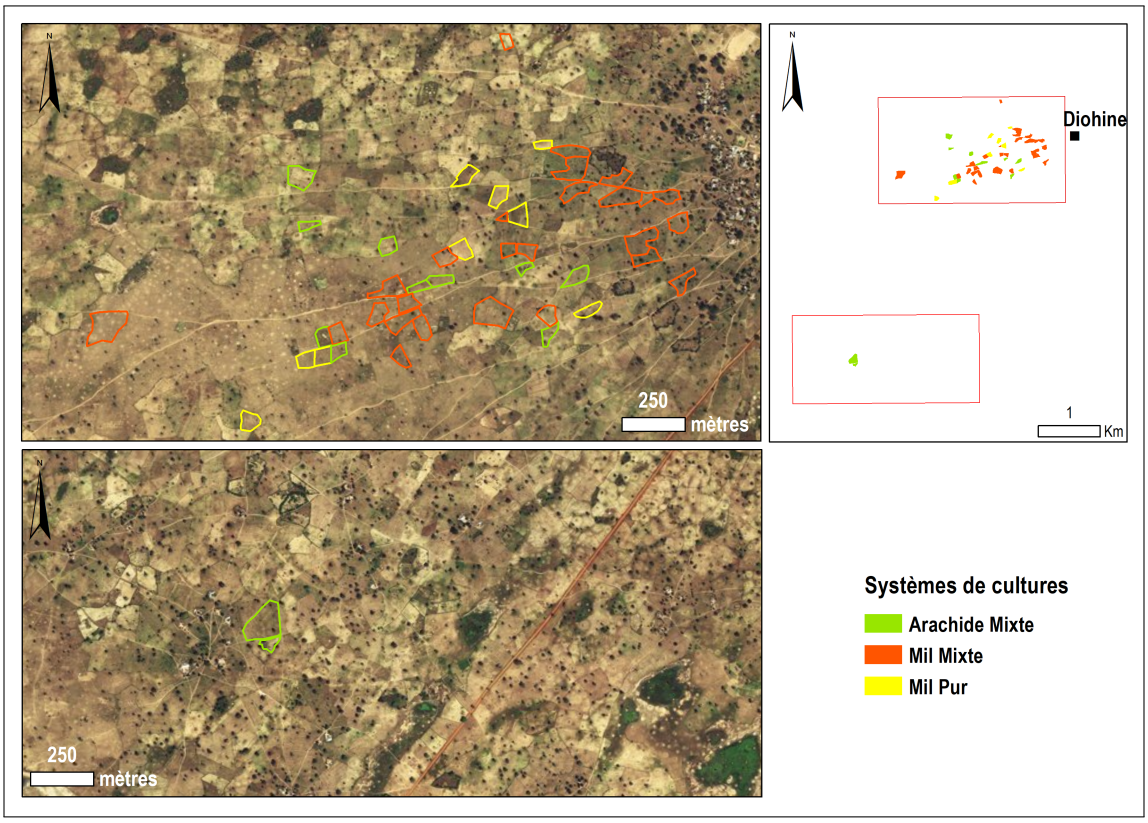


Figure 4.3. – Localisation des parcelles terrain suivant leur culture et système de culture avec leurs biomasses fraîches et rendements frais

4.4. Calcul du NDVI

Sur base des images de réflectance, nous avons extrait l'indice de végétation par différence normalisé NDVI. Il importe de noter que les indices dérivés d'images RapidEye (5m) ou Sentinel-2 (10m) ont été rééchantillonnés par la méthode du plus proche voisin à la résolution des indices PlanetScope (3m), ce afin de pouvoir considérer une série temporelle uniforme.

NDVI (Rouse et al., 1973; Tucker, 1979)

$$NDVI = \frac{PIR - R}{PIR + R} \tag{4.5}$$

4.5. Lissage des séries temporelles

À présent que nous disposons d'une série temporelle uniforme de NDVI, nous pouvons l'exploiter pour la suite de notre travail. Cependant, un autre traitement s'impose : le *lissage* ou *smoothing*. En effet, bien que des prétraitements soient effectués sur les images satellitaires : calibrations radiométriques et corrections atmosphériques entre autres, il subsiste du bruit qui affecte l'utilisation des séries temporelles d'images, impactant ainsi les futures analyses et pouvant donc fausser les interprétations données aux résultats obtenus (Chen et al., 2004). Ce bruit résiduel peut être lié à plusieurs facteurs notamment les conditions atmosphériques variables ou la présence de pixels nuageux indétectés et particulièrement dans notre cas à l'utilisation de données multi-capteurs. Les techniques de lissage font l'hypothèse que le bruit résiduel dans les images entraine des chutes soudaines dans le profil temporel des indices de végétation (Bojanowski et al., 2009). Ces valeurs peuvent être ainsi identifiées puis supprimées et des séries temporelles de meilleure qualité reconstruites.
L'une des techniques usuelles adoptée notamment par les fournisseurs d'indices de végétation périodique comme GIMMS-MODIS, SPOT Vegetation ou PROBA-V est le MVC (Holben, 1986). Cette technique consiste à créer des synthèses d'indices de végétation sur une période donnée, généralement une décade en considérant pour chaque pixel la plus grande valeur enrégistrée sur la période. D'autres techniques de lissage de séries temporelles d'indices de végétation ont été mises au point. Nous pouvons citer des méthodes par seuillage comme les algorithmes BISE (Viovy et al., 1992) ou IDR (Julien and Sobrino, 2010) qui utilisent un seuil pour contrôler le degré de lissage des séries reconstruites, des méthodes qui définissent un filtre pour lisser les données dans une fenêtre mobile comme le filtre de Savitzky-Golay (Savitzky and Golay, 1964; Chen et al., 2004) ou le filtre à poids variable de Zhu et al. (2012) ou encore d'autres méthodes qui ajustent des fonctions mathématiques au profil saisonnier de la végétation telles que la fonction asymétrique gaussienne (Jönsson and Eklundh, 2002), les doubles fonctions logistiques (Beck et al., 2006) et les méthodes basées sur l'analyse de Fourier comme HANTS (Verhoef et al., 1996; Roerink et al., 2000) et la transformée de Fourier rapide (Menenti et al., 1993). La méthode de lissage de Whittaker (Eilers, 2003; Atzberger and Eilers, 2011) et la transformation en ondelettes (Lu et al., 2007) sont également d'autres approches existantes.
Plusieurs études ces dernières années ont comparé différentes techniques de lissage

(Jönsson and Eklundh, 2002; Chen et al., 2004; Hird and McDermid, 2009; Kandasamy et al., 2012; Geng et al., 2014; Shao et al., 2016; Liu et al., 2017) et beaucoup ont conclu qu'il n'y avait pas de méthode de lissage idéale, chacune présentant ses avantages et inconvénients. C'est notamment le cas de Geng et al. (2014) qui ont comparé pas moins de 8 méthodes de lissage pour reconstruire des séries temporelles de NDVI. De plus, la comparaison entre les méthodes de lissage n'est pas toujours significative du fait de l'objet de l'étude (simple comparaison, classification ...), des différentes données utilisées et de l'application à des zones parfois caractérisées par des facteurs environnementaux opposés. Néanmoins, s'il est vrai qu'aucun classement absolu des méthodes de lissage n'ait été établi, il n'en demeure pas moins que certaines méthodes reviennent souvent dans la littérature et ont donné de bons résultats. Le filtre de Savitzky-Golay notamment est connu pour ses résultats consistants (Chen et al., 2004; Bojanowski et al., 2009; Kandasamy et al., 2012; Kim et al., 2014; Geng et al., 2014) et apprécié comme l'une des méthodes de lissage qui conserve le mieux la forme du profil temporel des indices de végétation ainsi que le timing et l'amplitude des minima et maxima locaux (Geng et al., 2014). L'algorithme HANTS a également été utilisé avec succès à de nombreuses reprises pour reconstruire des séries temporelles d'indices de végétation comme le NDVI ou de température de surface (Roerink et al., 2000; Jakubauskas et al., 2001; Lunetta et al., 2006; Julien et al., 2006; Zhou et al., 2012, 2015). HANTS a été particulièrement développé pour traiter les séries temporelles d'observations irrégulièrement espacées dans le temps, identifier et supprimer les observations nuageuses (Verhoef et al., 1996; Roerink et al., 2000) ainsi que pour prédire les observations manquantes. Ces points siéent parfaitement aux caractéristiques de nos séries temporelles. Par ailleurs, Atzberger and Eilers (2011) ont montré que la méthode de lissage de Whittaker pouvait parfaitement faire l'équilibre entre fidélité aux données d'origine et lissage tout en restant facile à mettre en oeuvre et en permettant un traitement rapide des données (Eilers, 2003). Atkinson et al. (2012); Geng et al. (2014); Shao et al. (2016) ont également démontré les bonnes performances de cette méthode. Pour les raisons citées, nous avons donc retenu et comparé ces 3 approches pour lisser nos séries temporelles.

4.5.1. Filtre de Savitzky-Golay

Savitzky and Golay (1964) ont proposé une méthode pour lisser et calculer les dérivées successives d'un ensemble de valeurs consécutives et régulièrement espacées dans le temps. Le filtre de Savitzky-Golay est un type de filtre passe-bas qui peut être interprété comme une moyenne mobile ou glissante pondérée dont les cœfficients de pondération sont définis par une fonction polynomiale d'un certain degré (Chen et al., 2004). À chaque glissement de la fenêtre du filtre, un ajustement basé sur les moindres carrés est réalisé grâce au polynôme de degré d. Ce polynôme a pour but de préserver les fortes variations ou pics dans le signal et d'attenuer le biais induit par le filtre. La valeur ajustée est sauvegardée à la mi-largeur de la fenêtre et celle-ci glisse d'un autre point vers la droite. La série $\tilde{Y}$ ajustée par fonction polynomiale au sens des moindres carrés est donnée par :

$$\tilde{Y}_j = \frac{\sum_{i=-m}^{m} C_i Y_{j+i}}{N} \tag{4.6}$$

où Y est la série d'origine, C_i le cœfficient de pondération de la ième valeur de la fenêtre du filtre, N la taille de la fenêtre du filtre $(2m+1)$ avec m comme demi-largeur

de la fenêtre et j l'indice du point en cours de lissage dans la série d'origine.
Le lissage de Savitzky-Golay est contrôlé par 2 paramètres : le degré de la fonction polynomiale d et la demi-largeur m de la fenêtre du filtre. Plus m est grand, plus le lissage sera important au détriment des pics du signal. Une taille entre 4 et 7 est généralement conseillée (Chen et al., 2004; Geng et al., 2014). Le degré du polynôme est généralement compris entre 2 et 4 (Chen et al., 2004). Une petite valeur de d produit un résultat lissé mais peut introduire un biais tandis qu'une valeur élevée de d réduira le biais du filtre mais peut surajuster la série et donner un résultat bruité.

4.5.2. HANTS

HANTS est un algorithme fondé sur l'analyse harmonique, branche des mathématiques qui étudie la représentation de fonctions ou signaux comme superposition de fonctions périodiques (ondes de base ou harmoniques). L'analyse harmonique approfondit et généralise les notions de séries et transformées de Fourier visant respectivement la décomposition de fonctions périodiques en une somme de fonctions trigonométriques (sinusoïdes) dans le domaine fréquentiel et l'extension de cette dernière aux fonctions non périodiques. Une série temporelle d'observations satellitaires y reconstruite par analyse harmonique s'écrit :

$$\tilde{y}(t_j) = a_0 + \sum_{i=1}^{n_f} [a_i \cos(2\pi f_i t_j) + b_i \sin(2\pi f_i t_j)] \tag{4.7}$$

$$y(t_j) = \tilde{y}(t_j) + \varepsilon(t_j) \tag{4.8}$$

où :
$\tilde{y}$ est la série temporelle reconstruite, ε la série d'erreurs, t_j le temps où y est observé dans la série temporelle $(t_1, t_2, \ldots t_N)$ avec N comme nombre maximal d'échantillons dans la série temporelle ; n_f est le nombre de termes périodiques dans la série temporelles ou le nombre d'harmoniques associées à la fréquence f_i ; a_i et b_i sont les cœfficients des composantes trigonométriques de la fréquence f_i et a_0 est le cœfficient à la fréquence zéro de phase nulle et représente la moyenne de la série reconstruite (Zhou et al., 2015).
HANTS considère uniquement les fréquences les plus importantes devant être présentes dans le profil temporel et applique une procédure d'ajustement par moindres carrés basée sur les harmoniques. Pour chaque fréquence, l'amplitude et la phase de l'harmonique sont déterminées par itération : à chaque ajustement, les points qui dévient grandement (positivement comme négativement) sont supprimés en leur assignant un poids nul et un nouvel ajustement reprend sur les points restants. L'itération s'arrête quand l'erreur maximale devient tolérable ou quand le nombre de points restant est trop petit. Les paramètres de l'algorithme HANTS sont :

- la *période* ou nombre d'échantillons de temps dans une période de base (365 jours pour 1 an par exemple)
- le nombre de fréquences n_f
- le paramètre *Hi/Lo* : il détermine si les valeurs fortes (*Hi*) ou faibles (*Lo*) doivent être rejetées pendant l'ajustement
- les valeurs *Low* et *High* qui sont respectivement la valeur minimale et la valeur maximale qui définissent la plage valide des données de la série d'origine

— le *FET* (Fit Error Tolerance) est l'erreur tolérée sur l'ajustement
— le *DOD* (Degree of Overdeterminess) est le nombre minimum de points supplémentaires qui doit être utilisé dans l'ajustement ultime
— et le *Delta* qui est un facteur de suppression des fausses oscillations (Roerink et al., 2000)

Comme l'ont soutenu Roerink et al. (2000), il n'y a pas de règle objective pour fixer ces paramètres et la solution la plus commune est de procéder par tâtonnement. Néanmoins, en se référant aux applications de HANTS dans la littérature, le nombre de fréquence est souvent fixé à 3 ou 4, le *FET* à $0,05$ (en unité de NDVI), le *DOD* à 5 points et le *Delta* à $0,5$ (Roerink et al., 2000; Zhou et al., 2015).

4.5.3. Méthode de lissage de Whittaker

Le lissage de Whittaker est une méthode basée sur les moindres carrés pénalisés dont l'idée est de reconstruire les données en minimisant une quantité Q exprimant l'équilibre entre la fidélité aux données brutes et la rugosité des données reconstruites (Eilers, 2003; Atzberger and Eilers, 2011; Atkinson et al., 2012). Plus les données sont lissées, plus elles dévient des données brutes.

Considérons un ensemble de n observations d'une série y. Les valeurs de la série reconstruite z doivent minimiser la quantité Q exprimée comme suit :

$$Q = S + \lambda R \tag{4.9}$$

$$S = \underbrace{\sum_{i=1}^{n} (y_i - z_i)^2}_{Déviation} \tag{4.10}$$

$$R = \underbrace{\sum_{i=1}^{n-d} (\Delta^d z_i)^2}_{Rugosité} \tag{4.11}$$

où $\Delta^d z_i^2$ représente la différence d'ordre d des valeurs z_i, S la somme des carrés des écarts, R la somme des carrés des différences d'ordre d des valeurs reconstruites et λ le paramètre de lissage qui donne un poids relatif à la rugosité : plus il est élevé, plus le lissage de la série reconstruite sera important.

En adoptant un formalisme matriciel, la série z qui minimise Q devient :

$$z = (I + \lambda D^T D)^{-1} y \tag{4.12}$$

où I représente la matrice identité de taille $n \times n$, D la matrice telle que $Dz = \Delta^d z_i$ et D^T la transposée de la matrice D.

Cette méthode interpole également les valeurs manquantes dans les données brutes. Dans ce cas, un poids w_i doit être affecté aux données : il vaudra 1 pour les points valides et 0 pour les valeurs manquantes. L'équation précédente devient :

$$z = (W + \lambda D^T D)^{-1} W y \tag{4.13}$$

où W est une matrice diagonale contenant les valeurs w_i.

Comme évoqué par Atzberger and Eilers (2011), l'interpolation des valeurs manquantes peut être utilisée pour réduire le pas d'échantillonnage des données (pas

journalier par exemple) ou à des fins d'extrapolation ou de prédiction de données. L'une des façons de fixer le paramètre de lissage λ est de procéder par tâtonnement en appréciant les résultats obtenus. Une autre façon plus objective de procéder consiste à se référer au processus de validation croisée de l'algorithme décrit par Eilers (2003). Quant à l'ordre d de la différence des valeurs z_i, la valeur 2 est assignée par défaut. Quand le lissage est important, la série reconstruite tend vers une ligne horizontale pour $d = 1$, une droite inclinée pour $d = 2$ et une parabole pour $d = 3$.

4.5.4. Evaluation du lissage

Nous avons testé les 3 méthodes décrites ci-dessus pour lisser la série temporelle de NDVI. Nous avons dans un premier temps mené une réflexion sur l'influence de l'utilisation des données multisources sur le profil temporel des cultures. Nous avons tout d'abord masqué les arbres présents dans les parcelles en appliquant un seuillage absolu (0,16) sur les valeurs de NDVI correspondant à l'image PlanetScope du 18/06/2017, la plus contrastée. Par la suite nous avons calculé et tracé les moyennes parcellaires de NDVI et leur écart type en fonction du temps. Il nous est alors apparu que les valeurs Sentinel-2 des 05, 15 et 25 Octobre étaient aberrantes puisqu'elles étaient largement au dessus des autres valeurs situées dans la même plage et ce pour des dates situées après la date de récolte notamment sur les parcelles de mil (Figure 4.4). Nous avons alors considéré 2 séries temporelles de NDVI pour le lissage : la série d'origine et la série sans les valeurs Sentinel-2 jugées aberrantes.
Afin d'optimiser le temps de traitement[6] de chaque série temporelle, le nombre de pixels en entrée a été réduit de façon à ne considérer que l'emprise des parcelles terrain au sud ouest de la zone d'étude (une taille de 1056×1384 soit 1 461 504 pixels sur 41 990 000 ou une taille de 6500×6460 Figure 4.3). Une chaîne de traitement incluant les 3 méthodes de lissage a été mise en place avec le langage `Python`. Le filtre de Savitzky-Golay a été implémenté dans la chaîne de traitement grâce à la fonction `savgol_filter` du module `scipy.signal`. L'algorithme HANTS implémenté est la transcription en `Python` effectuée par Espinoza-Dávalos et al. (2017) du programme originel écrit en `Fortran` et `Matlab`. Enfin, pour implémenter la méthode de Whittaker, nous avons transcrit en langage `Python`, le programme `Matlab` développé par Eilers (2003)[7]. Les valeurs originales et lissées par série temporelle pour l'ensemble des pixels sont stockées comme variables d'un fichier `netCDF4`.
En vue de définir le paramétrage optimal pour chaque méthode de lissage, nous avons adopté 3 critères d'évaluation qualitative tels que définis par Klisch et al. (2006) :

1. Conserver la temporalité des profils notamment que les dates de début et de fin de saison ne soient pas influencées par le lissage
2. Conserver le plus possible les valeurs hautes en raison de la rareté de valeurs aberrantes hautes
3. Réduire le plus possible les bruits à hautes fréquences.

Notons également que nous avons reconstruit les séries temporelles avec un pas journalier en interpolant les dates manquantes. L'algorithme HANTS et la méthode de Whittaker interpolent automatiquement les valeurs manquantes d'une série, le filtre

6. Deux heures approximativement par méthode de lissage avec une station de travail HP possédant un processeur Intel® Core™ i7-7700 CPU @ 3.60 GHz de 8 cœurs et 16 Go de mémoire RAM
7. https://pubs.acs.org/doi/abs/10.1021/ac034173t

de Savitzky-Golay non. Avec ce dernier, nous avons d'abord effectué une interpolation linéaire des valeurs manquantes avant de procéder au lissage comme ont eu à le faire également Chen et al. (2004) et Liu et al. (2017). En dépit de différents essais où nous avons fait varier la largeur de la fenêtre de lissage et le degré du polynôme d'ajustement, le filtre de Savitzky-Golay n'a pas donné les résultats escomptés. Quelques sorties des essais effectués sur des pixels appartenant à des parcelles agricoles sont présentés dans l'annexe B.
Le paramétrage adopté pour les autres méthodes est reporté dans le tableau 4.5. Notons que pour la méthode de Whittaker, nous avons mis en place une procédure d'ajustement à l'enveloppe supérieure du profil temporel des données d'origine. Cette procédure d'ajustement est la même que celle proposée par Chen et al. (2004) qui consiste :

— à lisser une première fois les valeurs de la série d'origine et à considérer le résultat comme un profil de référence
— à attribuer ensuite des poids aux valeurs de la série d'origine selon qu'elles soient au dessus (poids plus important) ou en dessous (poids moins important) du profil de référence
— à générer alors une nouvelle série temporelle reprenant les valeurs d'origine si elles sont au dessus du profil de référence et les valeurs de référence si les valeurs d'origine sont en dessous du profil
— et à procéder de nouveau au lissage à travers une procédure itérative : le premier lissage est effectué sur la série temporelle nouvellement générée ; un indicateur de l'ajustement détermine ensuite si l'itération doit être poursuivie, si oui un second lissage est effectué sur le résultat de la première itération mais en remplaçant les valeurs qui sont toujours inférieures aux valeurs d'origine par ces dernières et ainsi de suite. L'itération se poursuit tant que la valeur de l'indicateur d'ajustement n'est pas supérieure à celle calculée précédemment. Cet indicateur est donné par :

$$F_k = \sum_{i=1}^{n} (|N_i^{k+1} - N_i^0|) \times W_i \tag{4.14}$$

où N_i^{k+1} est la i^{ème} valeur du k^{ème} ajustement de la série générée, N_i^0 la i^{ème} valeur de la série d'origine, et W_i le poids associé à cette i^{ème} valeur. La définition de l'indicateur F_k est telle que plus sa valeur est faible, plus la courbe ajustée aura tendance à se rapprocher des valeurs à fort poids et donc de l'enveloppe supérieure du profil temporel des valeurs d'origine.

Remarquons que cette procédure d'ajustement n'a pas eu d'effet notable sur la reconstruction du signal par analyse harmonique et n'a donc pas été appliquée avec HANTS. Les résultats du lissage des séries temporelles de NDVI avec et sans les valeurs Sentinel-2 jugées aberrantes sont reportés sur les figures 4.5 et 4.6. De même que les valeurs d'origine, les valeurs lissées ont été agrégées à l'échelle parcellaire.

Sur les quelques profils reconstruits présentés, nous constatons plus de similitudes entre les 2 méthodes de lissage sur la série temporelle rectifiée que la série temporelle comprenant les valeurs Sentinel-2 aberrantes. Sur les profils de cette dernière, nous remarquons également que les valeurs Sentinel-2 ont tendance à rapprocher le maximum de NDVI de la date de récolte ou fin de saison. Globalement, la série rectifiée donne des résultats plus cohérents avec le cycle de développement des cultures (on

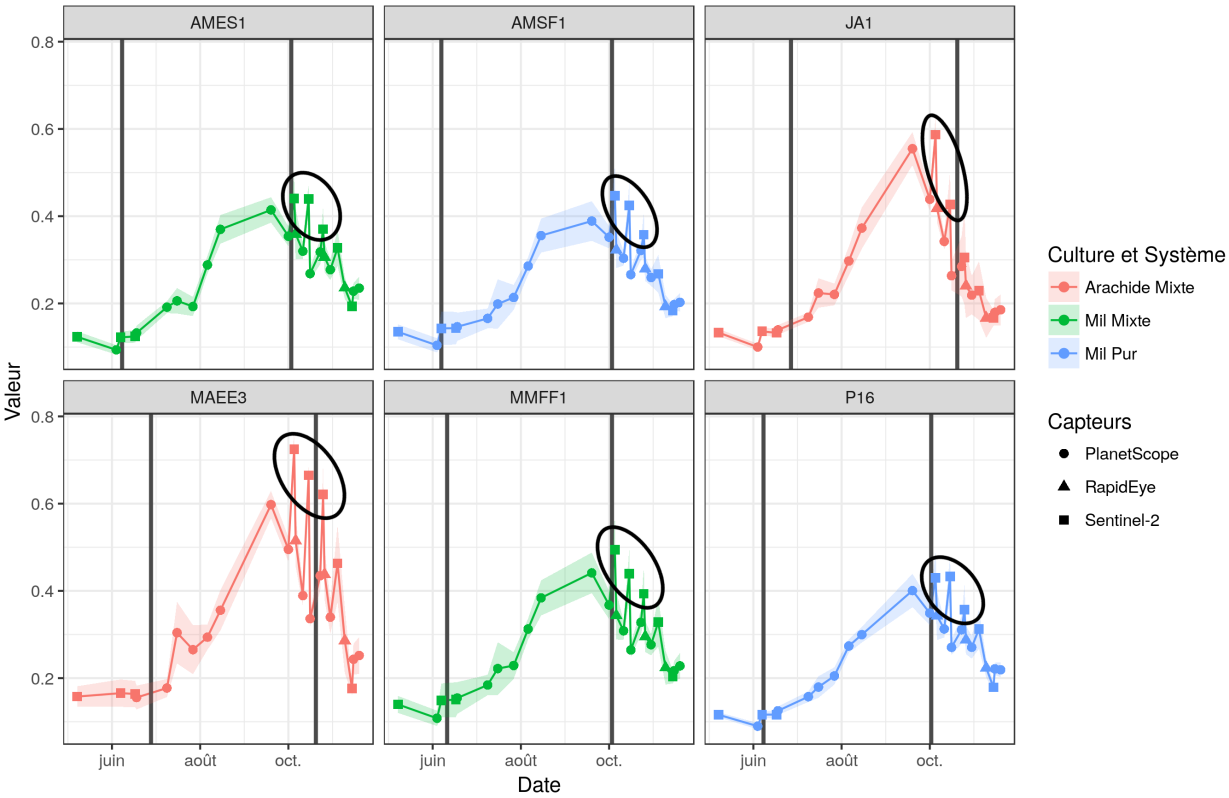


Figure 4.4. – Moyenne et Ecart type du NDVI sur quelques parcelles (*Les traits gris correspondent de la gauche vers la droite respectivement aux dates de semis et de récolte des parcelles dont les identifiants sont marqués au dessus*)

Tableau 4.5. – Paramétrage des méthodes de lissage

Méthode	Paramétrage
HANTS	$Période = 365$; $nf = 3$; $HiLo = Lo$; $Low = -0.3$; $High = 1$; $FET = 0.05$; $DOD = 1$; $Delta = 0.25$
Whittaker	$\lambda = 5000$ (2500 pour les itérations) ; $d = 2$

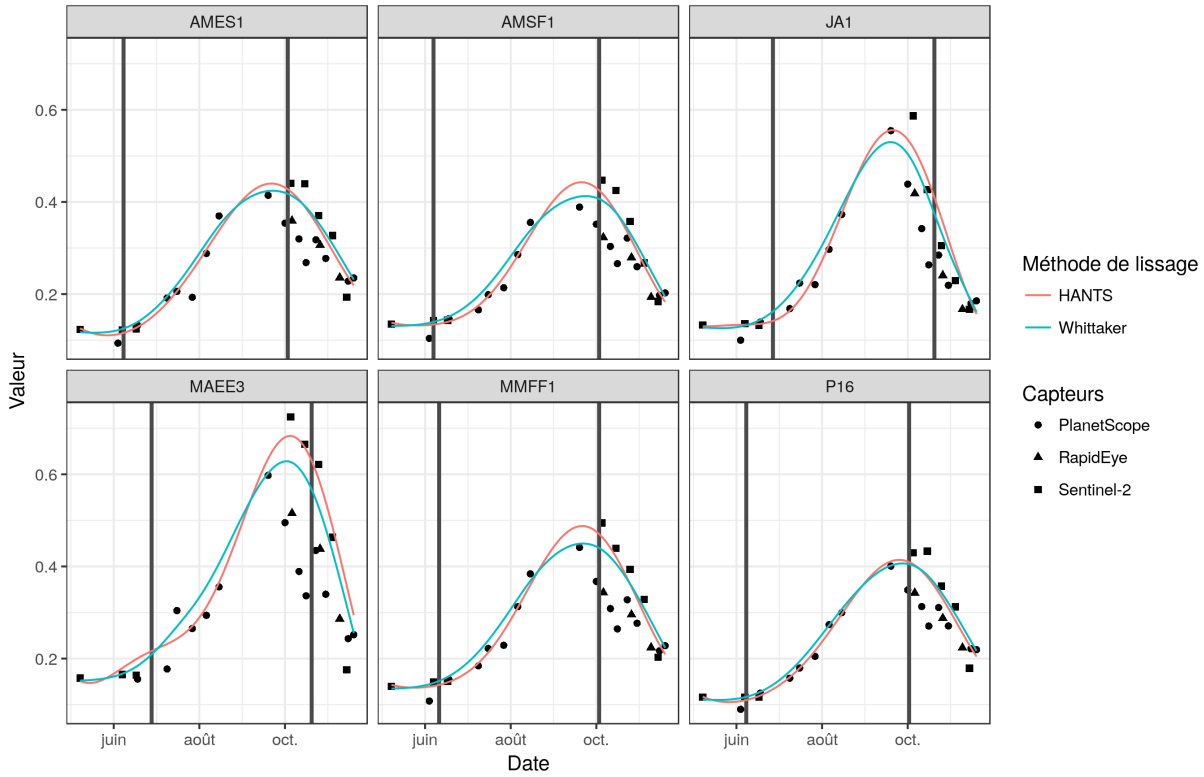


Figure 4.5. – Résultats du lissage de la série complète pour 6 parcelles

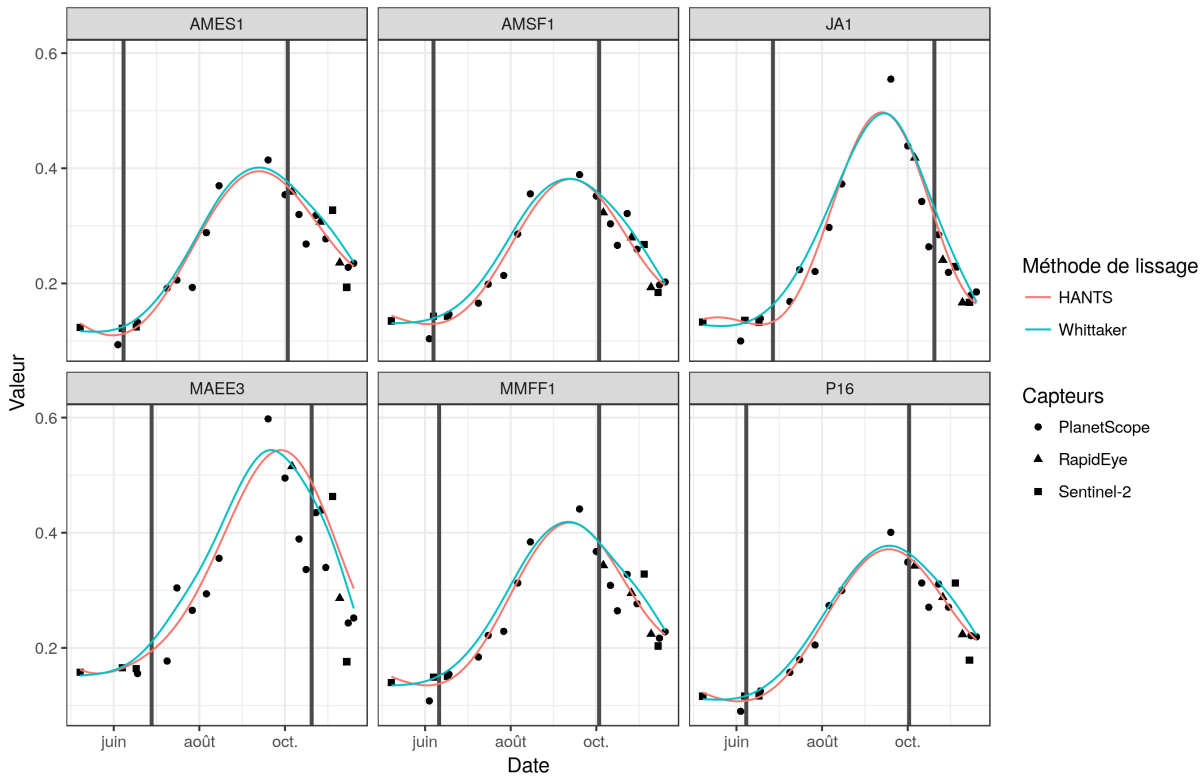


Figure 4.6. – Résultats du lissage de la série rectifiée pour 6 parcelles

ne peut observer un maximum de végétation après la date de récolte) que la série complète et les profils reconstruits par les 2 méthodes sont semblables. Nous avons donc retenu pour l'extraction des métriques phénologiques, les profils temporels de NDVI obtenus en lissant la série temporelle rectifiée par l'algorithme HANTS et la méthode de Whittaker.

4.6. Extraction des métriques phénologiques

L'extraction des métriques phénologiques va nous permettre, entre autres, d'évaluer les dates de semis. Parmi les méthodes d'extraction répertoriées dans notre état de l'art (cf. Section 3.1.3), nous avons opté pour l'approche par seuillage basée sur les ratios de NDVI (White et al., 1997). C'est une méthode simple, facile à mettre en œuvre. Contrairement aux méthodes par seuillage absolu, le seuillage effectué à partir des ratios de NDVI est relatif à l'amplitude des valeurs :

$$NDVI_{ratio}(t) = \frac{NDVI(t) - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \tag{4.15}$$

où $NDVI(t)$ est la valeur de l'indice spectral $NDVI$ à un instant t du profil temporel de la végétation, $NDVI_{min}$ la valeur minimale du profil temporel, $NDVI_{max}$ la valeur maximale du profil et $NDVI_{ratio}(t)$ la valeur de l'indice spectral transformée.
La transformation des valeurs de NDVI en ratio permet dès lors de déterminer les seuils des transitions phénologiques indépendamment du type de couvert ou de la géolocalisation du site d'étude. Le ratio de NDVI varie entre 0 et 1. Avec cette approche, le SOS est identifié comme la date où le ratio de NDVI atteint un seuil fixé (généralement 20% de l'amplitude) avant la valeur maximale du NDVI. La date de fin de saison ou EOS est déterminée similairement généralement avec un seuil de 80% de l'amplitude après la valeur maximale du NDVI.
Pour chaque système de culture, nous avons extrait les métriques suivantes à partir des profils temporels de NDVI lissés : le démarrage de la saison (SOS), la fin de saison (EOS), la durée de la saison (LOS), la date de la valeur maximale du NDVI (MAX), la valeur maximale du NDVI ($NDVI_{MAX}$), l'amplitude du NDVI (AMPL), le cumul des valeurs de NDVI du SOS à la date du MAX ($CUM_{SOS-MAX}$), le cumul des valeurs de NDVI du MAX au EOS ($CUM_{MAX-EOS}$), le taux d'augmentation du NDVI entre le SOS et le MAX ($RATE_{SOS-MAX}$) et le taux de diminution du NDVI entre le MAX et le EOS ($RATE_{MAX-EOS}$).
La détermination du SOS et du EOS est essentielle pour calculer les autres métriques excepté le MAX, le $NDVI_{MAX}$ et le AMPL. Pour ce faire, nous avons testé plusieurs seuils, respectivement 10, 20, 30 puis 50% pour le SOS et 50, 60, 70 puis 80% pour le EOS.

4.7. Estimation des biomasses et rendements

La dernière étape de notre travail consiste à calibrer des modèles statistiques pour estimer les biomasses végétatives et rendements grain ou gousse pour chaque type de culture : arachide et mil. Nous disposons déjà comme variables explicatives des différentes métriques phénologiques calculées en section 4.6. Néanmoins, nous avons

cherché d'autres variables pouvant mieux expliquer les biomasses et rendements observés. Pour ce faire, nous avons tout d'abord introduit un nouvel indice spectral en supplément du NDVI : le GDVI (Difference NIR/Green ou Green Difference Vegetation Index) qui est la différence entre les bandes du proche infrarouge et du vert ($PIR - V$). Tout comme pour le NDVI, les images de GDVI extraits à partir des données RapidEye et Sentinel-2 ont été rééchantillonnées à 3 mètres de résolution spatiale par la méthode du plus proche voisin. La série temporelle de GDVI a ensuite été lissée et reconstruite avec un pas journalier avec la méthode de lissage de Whittaker ($\lambda = 3000\,;\, d = 2$). Ainsi, pour chaque type de culture, les valeurs des indices de végétation provenant des profils temporels reconstruits ont respectivement été cumulées sur un intervalle de 100 jours à partir de la date du SOS, avec des pas et des décalages de 5 jours. Pour le NDVI par exemple, les valeurs sont cumulées à partir du SOS à 5 jours après le SOS, du SOS à 10 jours après le SOS et ainsi de suite jusqu'à 100 jours après le SOS. On reprend ensuite le même procédé en décalant le début de 5 jours c'est à dire en commençant cette fois ci à partir de 5 jours après le SOS et ce jusqu'à 95 jours après le SOS. Pour chaque cumul effectué, un modèle de régression linéaire est établi pour tenter d'expliquer les biomasses et rendements observés. Ceci doit nous permettre d'identifier pour chaque système de culture les périodes d'intégration les plus correlées avec les biomasses et rendements finaux. Une fois ces périodes déterminées et à condition que les régréssions effectuées soient significatives, nous recherchons les interactions possibles entre nos deux jeux de données de variables explicatives c'est-à-dire les métriques phénologiques calculées et les périodes d'intégration, pouvant expliquer les biomasses et rendements observés. Pour cela, nous établissons à nouveau, des modèles de régression linéaire à partir de chaque interaction possible de variables. Ainsi, les interactions les plus correlées avec les biomasses et rendements observés sont gardées à condition que les régressions effectuées soient significatives. Au final, étant donné l'étroitesse de la taille d'échantillons dont nous disposons (12 parcelles d'arachide et 35 de mil) pour établir des modèles multivariés et des échecs répétés de méthodes de sélection de variables comme le VIF (Variable Inflation Factor) ou de transformation de variables comme l'analyse en composantes principales (ACP), nous calibrons nos modèles définitifs pour chaque type de culture, à partir des variables précédentes (indices de végétation intégrés et interaction des variables intégrées et des métriques phénologiques) ayant le mieux expliqué les biomasses et rendements. Également, compte tenu de cette étroitesse d'échantillons, nous nous arrêterons à la calibration des modèles et n'aborderons pas leur validation.

5. Résultats et Discussions

5.1. Résultats

5.1.1. Extraction du SOS et du EOS

À partir des profils temporels de NDVI obtenus par les 2 méthodes de lissage (HANTS et Whittaker), nous avons extrait le SOS et le EOS respectivement avec les seuils de 10, 20, 30 puis 50% avant le MAX et 50, 60, 70 puis 80% après le MAX. Nous avons calculé ensuite pour chacune des parcelles terrain, les écarts en nombre de jours entre les dates de semis et les SOS extraits puis entre les dates de récolte et les EOS estimés. En considérant ces écarts par système de culture, nous avons calculé 2 indicateurs statistiques : la racine carrée de l'erreur quadratique moyenne ou RMSE qui est un indicateur d'écart entre valeurs observées et valeurs prédites et le coefficient de variation ou CV qui mesure la dispersion relative autour de la moyenne.

SOS et évaluation des dates de semis La distribution des écarts entre les dates de semis observées et les SOS extraits par système de culture et méthode de lissage est illustrée sur la figure 5.1. Globalement, les parcelles d'arachide mixte montrent les variabilités les plus faibles entre les écarts calculés (moins de 15 jours d'écart au maximum, tous seuils et méthodes de lissage confondus) suivies des parcelles de mil pur (moins de 20 jours d'écart au maximum, tous seuils et méthodes de lissage confondus) et des parcelles de mil mixte qui présentent les variabilités les plus fortes (près de 30 jours d'écart avec les SOS estimés par HANTS pour un seuil de 10%). En analysant la distribution des écarts par méthode de lissage, nous remarquons que pour l'ensemble des systèmes de culture et presque pour tous les seuils, la plage des écarts obtenus avec les SOS estimés par HANTS est toujours plus importante que celle des écarts obtenus avec les SOS extraits par la méthode de Whittaker avec des écarts plus grands pour HANTS quand on considère le même seuil. Ceci semble indiquer que la méthode de lissage de Whittaker soit plus appropriée pour estimer le timing du démarrage de croissance de la végétation. L'analyse de la distribution des écarts en fonction du seuil d'extraction du SOS montre quant à elle une certaine précocité dans les SOS extraits avec le seuil de 10% (écarts négatifs pour certaines parcelles d'arachide et de mil mixtes). Par contre, le seuil de 50% les extrait trop tardivement (40 à 60 jours après les dates de semis). Le seuil le plus adapté doit extraire les SOS avec des écarts réalistes par rapport aux dates de semis et minimiser au mieux la variabilité entre ces écarts. Afin de déterminer les seuils les plus adaptés et ce par système de culture, référons nous à la figure 5.2 où nous avons représenté le RMSE entre les dates de semis et SOS estimés en fonction du coefficient de variation (CV) de leurs écarts par système de culture et méthode de lissage.

La variation du RMSE en fonction du seuil d'extraction quelque soit le système de culture ou la méthode de lissage pris en compte rejoint l'analyse sur la distribu-

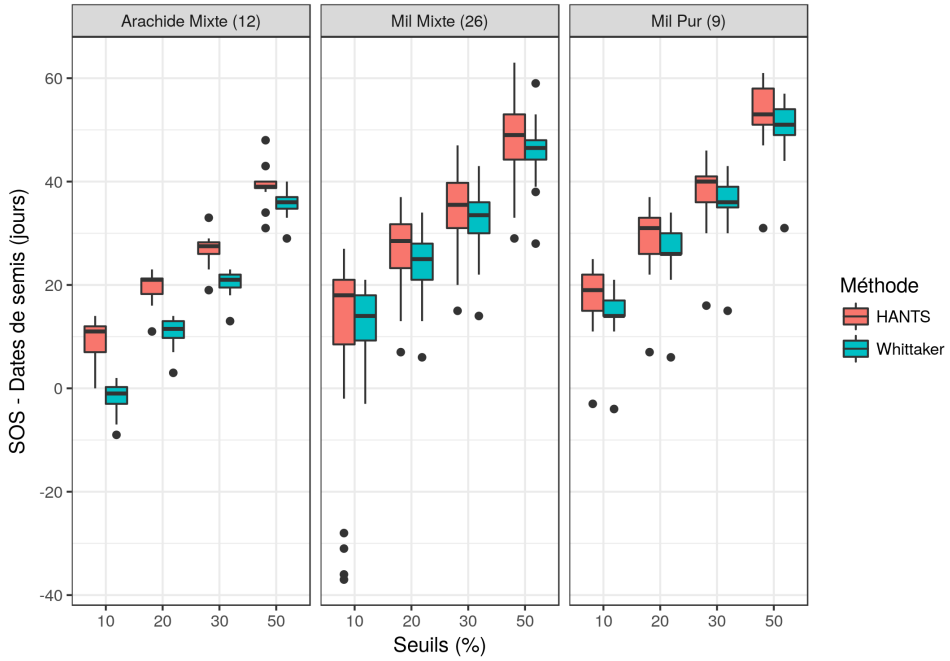


Figure 5.1. – Boîtes à moustaches illustrant la distribution des écarts entre SOS estimés et dates de semis observées en fonction des systèmes de culture et des méthodes de lissage (*le nombre de parcelles est indiqué à la suite du système de culture*)

tion des écarts. En effet, plus le seuil d'extraction croit et plus le RMSE entre SOS et dates de semis est grand, ce dernier étant un indicateur d'écart. La variation du CV en fonction du seuil d'extraction révèle d'abord deux valeurs qui portent à interrogation : -179% (pour un seuil de 10% avec la méthode de Whittaker sur les parcelles d'arachide mixte) et 200% (pour un seuil de 10% avec HANTS sur les parcelles de mil mixte). L'analyse de nos résultats a montré qu'il s'agissait en fait de cas où l'écart-type était supérieure à une moyenne très petite proche de 0 (négative et positive respectivement). Ceci indique que les SOS extraits dans ces cas sont précoces. Pour le reste, la valeur du coefficient de variation diminue quand le seuil d'extraction croît, système de culture et méthode de lissage confondus. Ceci peut s'expliquer par le fait que la moyenne des écarts entre SOS et dates de semis augmente à mesure que le seuil d'extraction s'accroit quand l'écart-type ou la variabilité entre les écarts n'est pas très différente. Finalement, il apparait que les seuils les plus adaptés sont celui de 20% avec la méthode de Whittaker pour les parcelles d'arachide mixte, estimant les SOS avec un RMSE légèrement supérieur à 10 jours et ceux de de 10% avec la méthode de Whittaker pour les parcelles de mil pur et de mil mixte, estimant les SOS avec des RMSE de 15 jours.

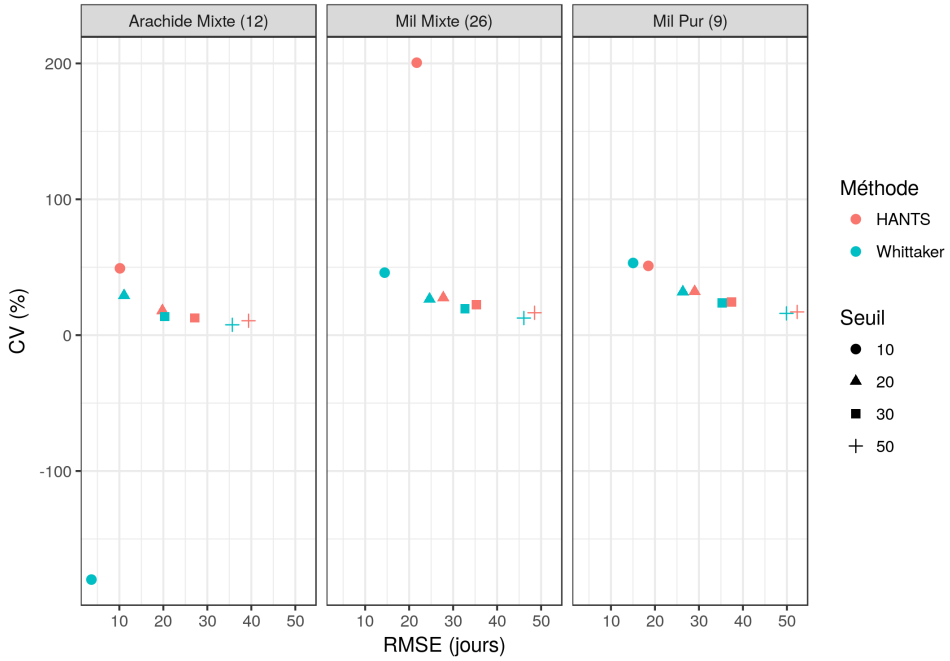


Figure 5.2. – Représentation du RMSE entre dates de semis et SOS estimés en fonction du coefficient de variation de leurs écarts par système de culture et méthode de lissage

EOS Comme pour les SOS, nous avons représenté la distribution des écarts mais cette fois ci entre les dates de récoltes observées et les EOS extraits par système de culture et méthode de lissage (Figure 5.3). Les parcelles d'arachide mixte montrent légèrement moins de variabilité entre les écarts que les parcelles de mil mixte et de mil pur. La plage des écarts obtenus avec les EOS estimés par HANTS est légèrement inférieure à celle des écarts obtenus avec les EOS extraits par la méthode de Whittaker notamment pour les parcelles d'arachide et de mil mixtes tous les seuils compris. En ce qui concerne la variation des seuils, nous avons des tendances différentes pour les parcelles d'arachide et de mil. Certes, plus on augmente les seuils d'extraction du EOS et plus les dates de fin de saison estimées se rapprochent des dates de récoltes mais cette précocité des EOS n'est pas obtenue au même moment pour les deux types de cultures. Ainsi, le seuil de 60% extrait des EOS déjà trop précoces pour les parcelles d'arachide (médiane des écarts autour de 0 pour les 2 méthodes de lissage) tandis que les EOS extraits pour les parcelles de mil avec un seuil de 80% ont au moins 10 jours d'écart avec les dates de récoltes. Le seuil le plus adapté pour chaque système de culture doit déterminer les EOS en minimisant les écarts par rapport aux dates de récoltes et en réduisant au mieux la variabilité de ces écarts puisqu'en fin de compte les périodes de forte corrélation pour l'estimation des rendements qui viendra par la suite, ne doivent pas excéder les dates de récoltes.
Référons nous ensuite à la figure 5.4 où nous avons représenté la variation du RMSE en fonction du CV. Pour les parcelles d'arachide mixte, les RMSE les plus faibles sont obtenus avec les seuils de 60 et 70% ce qui est normal car les écarts entre EOS et dates de récoltes à ces seuils sont faibles (proches de 0). Par contre, leurs coefficients de variations très élevés rappellent le cas de dates précoces comme pour le SOS. Il apparait donc que le seuil le plus adapté pour l'extraction des EOS pour les parcelles

d'arachide mixte est celui de 50% car minimisant à la fois les écarts par rapport aux dates de semis et leurs variabilités. Pour les parcelles de mil (pure et mixte confondus) par contre, le RMSE diminue à mesure que le seuil d'extraction augmente. Ainsi, les seuils de 80% sont les plus adaptés car ils minimisent à la fois les écarts par rapport aux dates de semis et leur variabilité. En ce qui concerne la méthode de lissage associée aux seuils choisis, celle de Whittaker est plus appropriée pour 2 raisons. La première raison se trouve dans l'estimation des SOS qui d'un point de vue agronomique est plus déterminante pour l'estimation des rendements finaux que celle des EOS et la méthode de Whittaker a été plus performante que HANTS sur ce point. L'autre raison se trouve dans le fait de garder un seul type de profil temporel pour l'estimation des rendements en éliminant les profils lissés par HANTS et tout compte fait, les EOS obtenus avec les 2 méthodes de lissage ne sont pas aussi différents que les SOS estimés où il y'a nettement une supériorité des performances de la méthode de Whittaker.

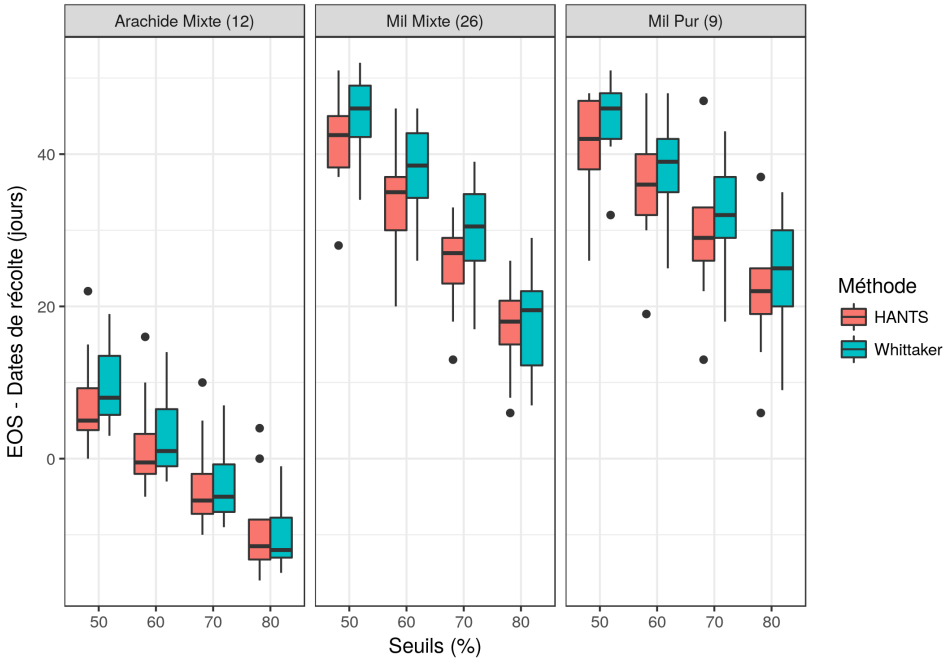


Figure 5.3. – Boîtes à moustaches illustrant la distribution des écarts entre EOS estimés et dates de récoltes observées en fonction des systèmes de culture et des méthodes de lissage

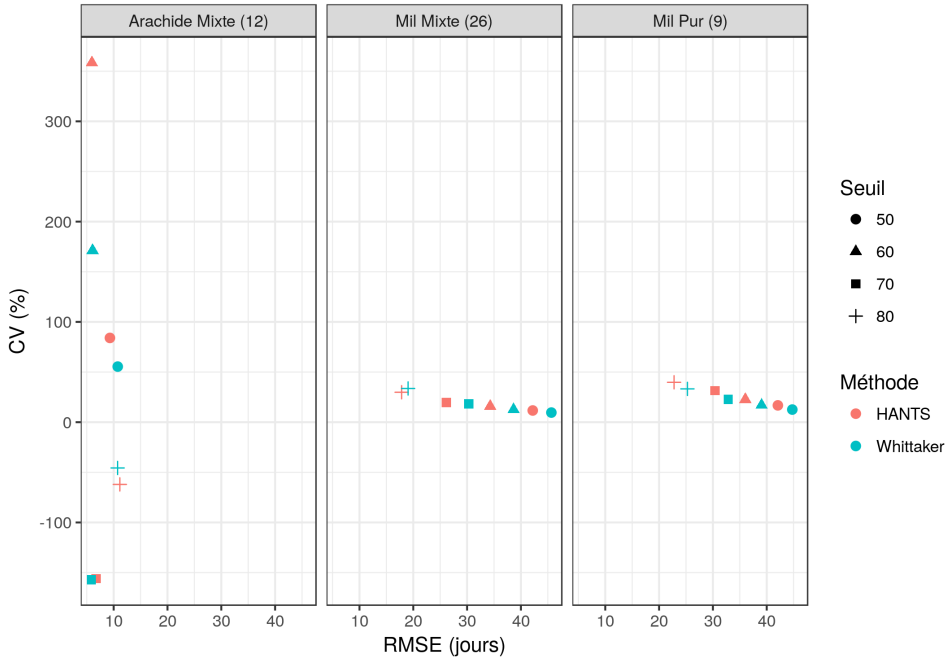


Figure 5.4. – Représentation du RMSE entre EOS estimés et dates de récoltes observées en fonction du coefficient de variation de leurs écarts par système de culture et méthode de lissage

Nous avons repris la même analyse sur l'estimation des SOS et des EOS en excluant les jeux de données de parcelles provenant du projet ANR-CERAO sans que les résultats présentés en annexe C soient significativement différents de ceux ci. La spatialisation des SOS et EOS avec les seuils adoptés est également présentée en annexe D.

5.1.2. Estimation des biomasses et rendements

Nous avons calibré 2 modèles de régression linéaire par type de culture pour estimer respectivement les biomasses et rendements observés. Afin de déterminer les meilleures variables explicatives pour chaque modèle, nous avons procédé en 2 étapes. Premièrement, nous avons cumulé les valeurs de NDVI et GDVI respectivement puis réalisé une régression linéaire à partir de chaque cumul effectué pour expliquer les biomasses et rendements. Nous avons ensuite aussi cherché les interactions possibles entre les métriques phénologiques calculées et les meilleures périodes d'intégration des indices, pouvant expliquer les biomasses et rendements observés. Au final, les variables explicatives retenues pour calibrer les modèles sont celles avec lesquelles nous avons obtenu les meilleures estimations de biomasses végétatives et rendements grain ou gousse, à condition que les régressions effectuées soient significatives.

Biomasses végétatives Les résultats des régressions effectuées entre les biomasses et les cumuls de NDVI et GDVI sont illustrés respectivement par les figures 5.5 et 5.6. Globalement, les cumuls de NDVI et de GDVI sont plus fortement corrélés à la biomasse de l'arachide qu'à celle du mil. Les cumuls de NDVI présentent de plus grandes périodes de fortes corrélations avec la biomasse de l'arachide que les cumuls

de GDVI. Cependant, c'est le cumul des valeurs de GDVI entre 15 et 100 jours après le SOS ($CUM_{GDVI:15-100}$) qui explique mieux la biomasse de l'arachide avec un R^2 de $0,532$ et une p-value de $0,007$ contre un R^2 de $0,437$ et une p-value de $0,019$ pour le cumul des valeurs de NDVI sur un intervalle de 75 jours après le SOS ($CUM_{NDVI:0-75}$). Pour le mil par contre, ce sont les cumuls de GDVI qui sont plus corrélés à la biomasse. C'est le cumul des valeurs de GDVI entre 0 et 60 jours après le SOS ($CUM_{GDVI:0-60}$) qui explique le mieux la biomasse du mil avec un R^2 de $0,273$. La part de biomasse expliquée est certes faible mais reste significative avec une p-value de $0,001$. Par la suite, les variables $CUM_{GDVI:15-100}$ et $CUM_{GDVI:0-60}$ renommées CUM afin d'uniformiser l'axe des x sur la figure 5.7 ont respectivement été mis en interaction avec les métriques phénologiques calculées pour estimer la biomasse de l'arachide et du mil.

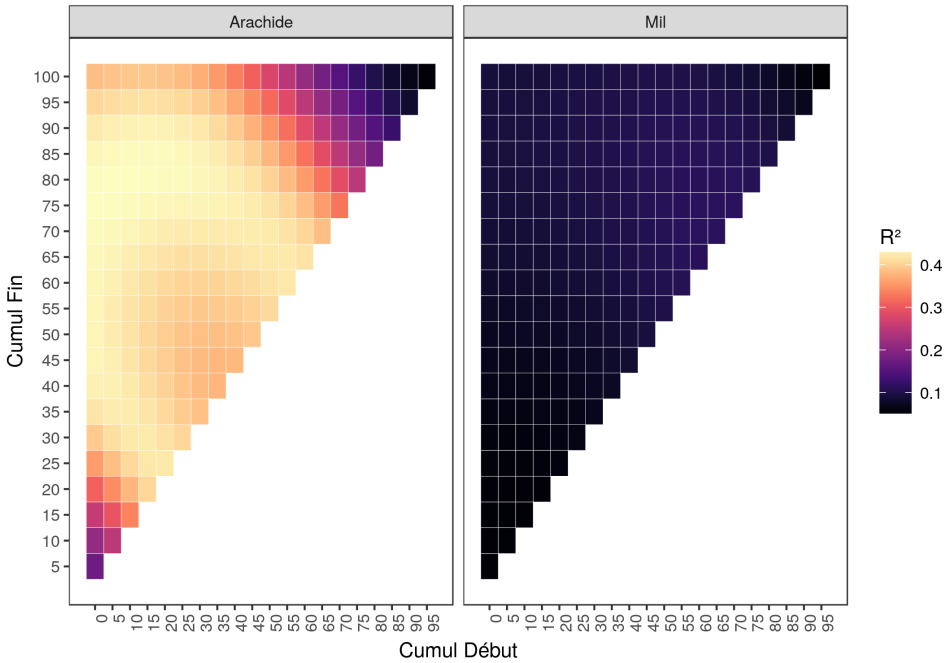


Figure 5.5. – Résultats des régressions entre les cumuls de NDVI et les biomasses (*l'axe x correspond au début du cumul et l'axe y à la fin du cumul ; le résultat de la régression entre les biomasses et les cumuls des valeurs de NDVI sur un intervalle de 5 jours après le SOS* ($CUM_{NDVI:0-5}$) *est donné par le croisement de la lecture de* 0 *sur l'axe des x et* 5 *sur l'axe des y*)

Les résultats des régressions entre les interactions de variables et les biomasses de l'arachide et du mil sont reportés sur la figure 5.7. Les interactions des métriques EOS, MAX et SOS avec la variable $CUM_{GDVI:15-100}$ sont celles qui ressortent pour l'arachide et c'est l'interaction $EOS : CUM_{GDVI:15-100}$ qui explique le mieux la biomasse de l'arachide avec un R^2 de $0,562$ et une p-value de $0,005$. Pour le mil, c'est l'interaction entre le MAX et le EOS qui explique mieux la biomasse avec un R^2 toujours faible de $0,263$ mais significatif avec une p-value de $0,002$. L'interaction $EOS : CUM_{GDVI:15-100}$ explique mieux les biomasses de l'arachide tandis que celles du mil sont plus corrélés avec la période d'intégration du GDVI sur un intervalle de 60 jours après le SOS. Au final, nous avons adopté les modèles de régression linéaire avec les variables $EOS : CUM_{GDVI:15-100}$ et $CUM_{GDVI:0-60}$ respectivement pour l'estimation des biomasses de l'arachide et du mil (Figure 5.8).

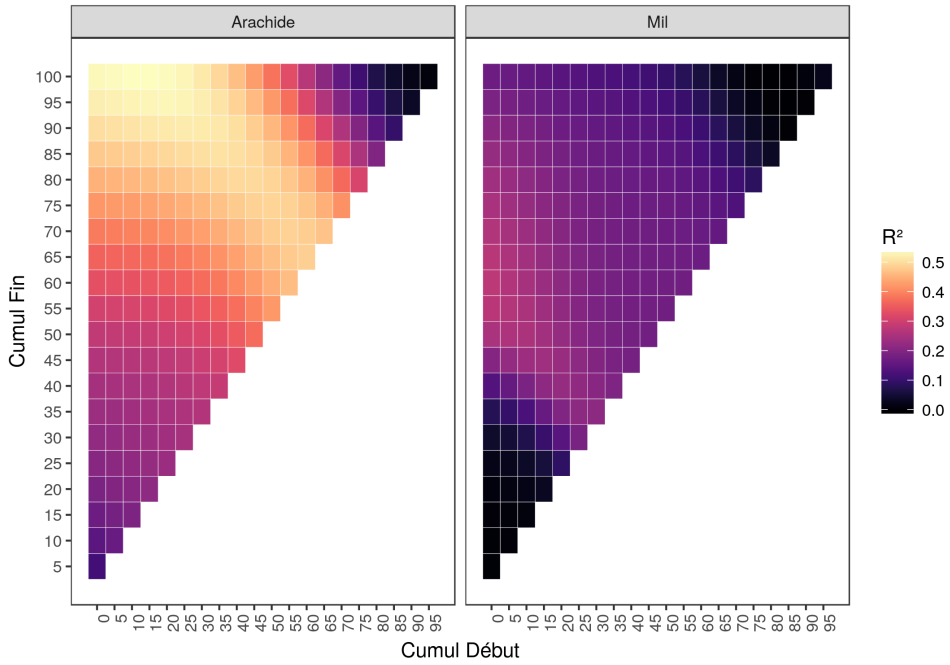


Figure 5.6. – Résultats des régressions linéaires entre les cumuls de GDVI et les biomasses

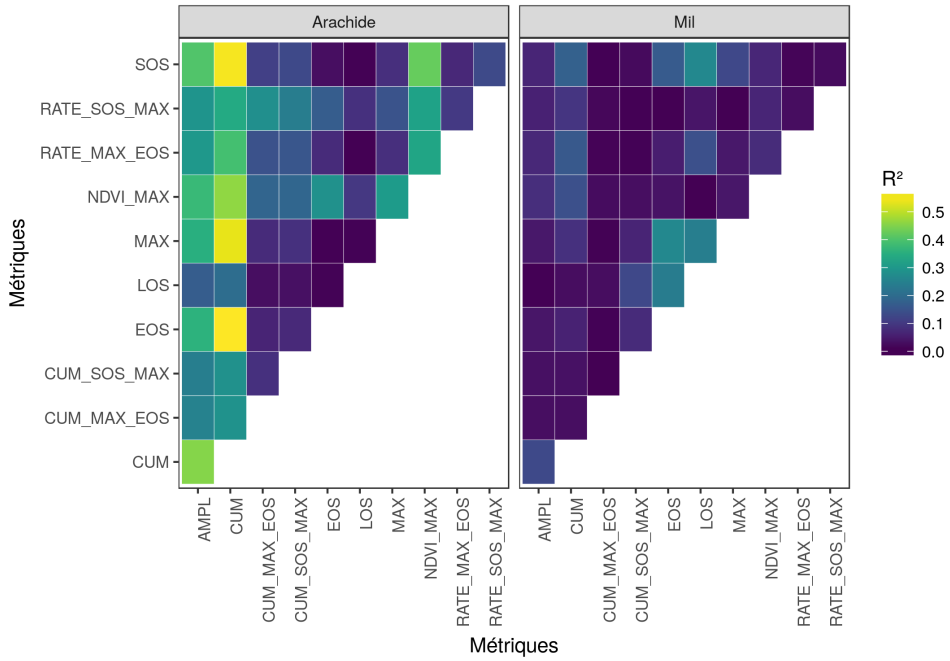


Figure 5.7. – Résultats des régressions linéaires entre les interactions de variables et les biomasses (*l'axe des x donne le premier élément de l'interaction et l'axe des y le second ; le résultat de la régression linéaire entre l'interaction du SOS et du EOS (SOS : EOS) et les biomasses est donné par le croisement de la lecture de SOS sur l'axe des x et EOS sur l'axe des y*)

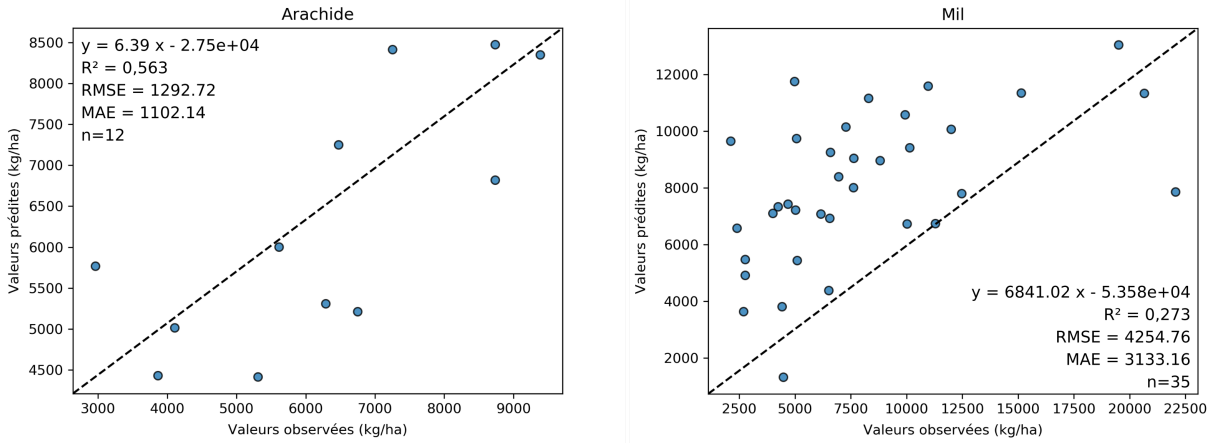


Figure 5.8. – Modèles finaux pour l'estimation des biomasses de l'arachide et du mil

Rendements Les résultats des régressions effectuées entre les rendements et les cumuls de NDVI et GDVI sont illustrés respectivement par les figures 5.9 et 5.10. Globalement, les cumuls de NDVI et de GDVI sont plus corrélés aux rendements du mil qu'à ceux de l'arachide. Toutes les périodes d'intégration du NDVI sont pratiquement corrélées avec les rendements du mil contre une majeure partie au milieu du cycle de développement du mil pour les périodes d'intégration du GDVI. Cependant, c'est le cumul des valeurs de GDVI entre 0 et 65 jours après le SOS ($CUM_{GDVI:0-65}$) qui explique mieux les rendements du mil avec un R² de $0,4187$ et une p-value de 2.7×10^{-5} contre un R² de $0,316$ et une p-value de 4.33×10^{-4} pour le cumul des valeurs de NDVI entre 60 et 65 jours après le SOS ($CUM_{NDVI:60-65}$). Pour l'arachide par contre, ce sont les cumuls de GDVI au tout début du cycle de développement qui sont plus corrélés aux rendements. C'est le cumul des valeurs de GDVI entre 15 et 20 jours après le SOS ($CUM_{GDVI:15-20}$) qui explique le mieux les rendements de l'arachide avec un R² de $0,363$ et une p-value de $0,038$. Par la suite, les variables $CUM_{GDVI:15-20}$ et $CUM_{GDVI:0-65}$ renommées CUM afin d'uniformiser l'axe des x sur la figure 5.11 ont respectivement été mis en interaction avec les métriques phénologiques calculées pour estimer les rendements gousses de l'arachide et grains du mil.

Les résultats des régressions entre les interactions de variables et les rendements de l'arachide et du mil sont reportés sur la figure 5.11. L'interaction entre le EOS et le $CUM_{GDVI:15-20}$ est celle qui ressort pour les rendements de l'arachide avec un R² de $0,497$ et une p-value de $0,01$. Pour le mil, c'est l'interaction entre le MAX et le EOS qui explique le mieux les rendements avec un R² de $0,4532$ et une p-value de 9.36×10^{-6} tout comme c'était le cas pour la biomasse. Les interactions entre variables expliquent mieux les rendements de l'arachide et du mil que les périodes d'intégration des indices seules. Au final, nous avons adopté les modèles de régression linéaire avec les interactions de variables respectivement $EOS : CUM_{GDVI:15-20}$ et $MAX : EOS$ pour l'estimation des rendements de l'arachide et du mil (Figure 5.12).

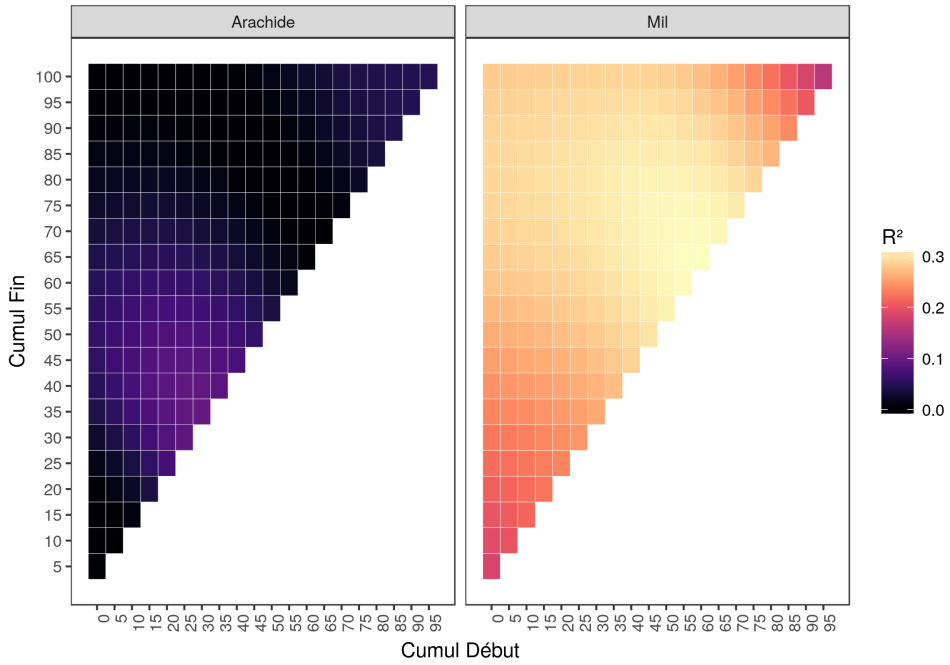


Figure 5.9. – Résultats des régressions linéaires entre les cumuls de NDVI et les rendements (*l'axe x correspond au début du cumul et l'axe y à la fin du cumul ; le résultat de la régression linéaire entre les rendements et les cumuls des valeurs de NDVI entre 5 et 10 jours après le SOS* ($CUM_{NDVI:5-10}$) *est donné par le croisement de la lecture de* 5 *sur l'axe des x et* 10 *sur l'axe des y*)

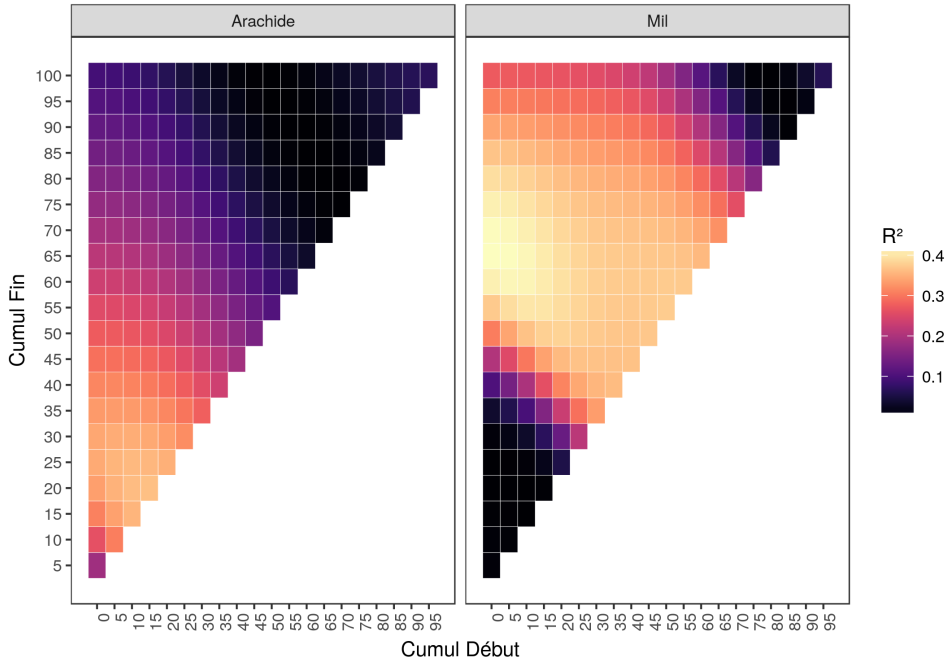


Figure 5.10. – Résultats des régressions linéaires entre les cumuls de GDVI et les rendements

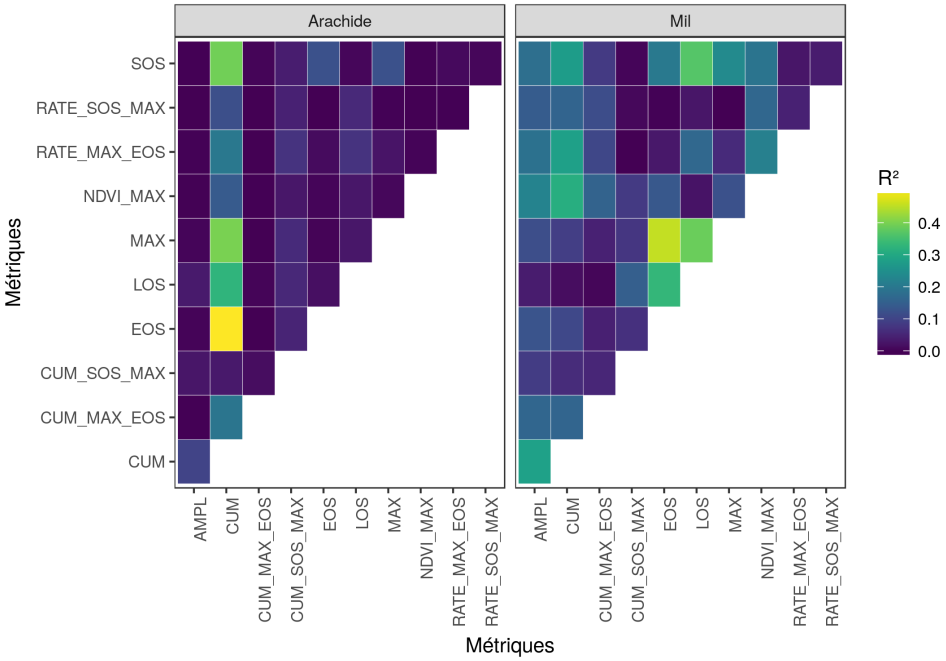


Figure 5.11. – Résultats des régressions linéaires entre les interactions de variables et les rendements (*l'axe des x donne le premier élément de l'interaction et l'axe des y le second ; le résultat de la régression linéaire entre l'interaction du SOS et du LOS (SOS : LOS) et les rendements est donné par le croisement de la lecture de SOS sur l'axe des* x *et LOS sur l'axe des y*)

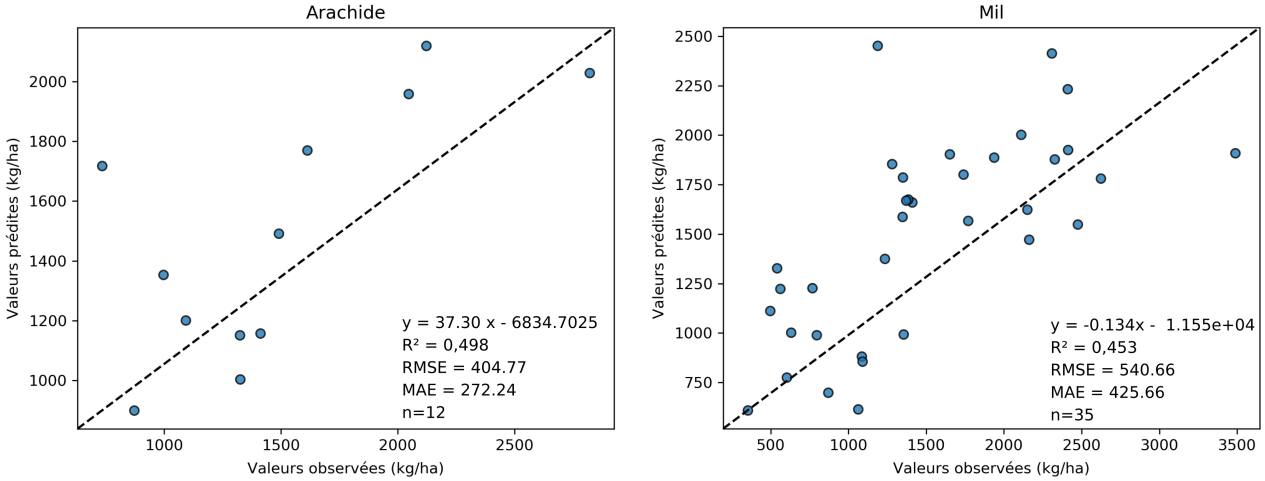


Figure 5.12. – Modèles finaux pour l'estimation des rendements de l'arachide et du mil

5.2. Discussions

5.2.1. Sur l'extraction du SOS et du EOS

L'approche par seuillage relatif sur les amplitudes de NDVI et le test de plusieurs seuils nous ont permis de choisir les meilleures estimations possibles des SOS et EOS. Les parcelles d'arachide ont montré les variabilités les plus faibles des écarts entre SOS et dates de semis (plus ou moins 5 jours) suivies des parcelles de mil en culture pure (plus ou moins 10 jours) et des parcelles de mil en culture associée (plus ou moins 20 jours). Cette différence entre parcelles d'arachide et de mil tout d'abord peut s'expliquer par les pratiques culturales. En effet, les agriculteurs sèment l'arachide dès les premières pluies significatives qui arrivent à la fin Juin ou début Juillet, ce qui a tendance à harmoniser les dates de semis. Les semis du mil par contre sont plus aléatoires. Les agriculteurs sèment le mil à sec avant les premières pluies avec des semis pouvant aller jusqu'à un mois avant le démarrage effectif de la saison des pluies. Par ailleurs, si la première pluie n'est pas suffisamment efficace cela peut entraîner un échec des premiers semis et conduire à la mise en place d'un second semis. Ensuite, l'estimation des SOS est influencée par la mixité du signal capté sur les parcelles de mil en cultures associées, ce qui se traduit par une variabilité plus grande observée sur les parcelles de mil associé. Cette même raison peut en partie expliquer la grande variabilité observée sur les écarts entre dates de EOS estimées et dates de récoltes observées, à l'exception des parcelles de mil en culture pure.
En ce qui concerne les seuils testés, les meilleures estimations des SOS ont été faites avec 20% de l'amplitude du NDVI pour les parcelles d'arachide et 10% pour les parcelles de mil. Les meilleurs EOS ont quant à eux été extraits avec 50% les parcelles d'arachide et 80% pour les parcelles de mil. Le seuil de 20%, correspondant le plus souvent au stade de déploiement des feuilles (Misra et al., 2016), est communément adopté pour les SOS (Brandt et al., 2016) tandis que pour les EOS, il est plus difficile de s'accorder sur un seuil particulier tant la fin de saison peut être influencée par de nombreux facteurs (stress, ravageurs . . .). Cependant, les différences observées entre les seuils d'extraction des EOS adoptés (50% pour l'arachide et 80% pour le mil) sont déterminées par les pratiques de récoltes. En effet, le mil est récolté très tôt, entre la mi-septembre et la mi-octobre selon les années en coupant uniquement les épis (les tiges et feuilles sont généralement laissées sur la parcelle et ramassées au cours de la saison sèche) tandis que les arachides sont arrachées et récoltées plus tard en Octobre, voire début Novembre. Ainsi, en tenant compte du profil temporel de la végétation, c'est à partir de seuils élevés que les EOS extraits pourront se rapprocher des dates de récoltes du mil et à partir de seuils plus faibles que les dates de récoltes des parcelles d'arachide seront approchées.
Outre le test de différents seuils possibles pour la détection des dates de démarrage et de fin de saison, nous avons aussi comparé deux méthodes de lissage : HANTS et Whittaker. Nous avons montré que la méthode de lissage de Whittaker était plus à même d'estimer les dates de début de croissance de la végétation que HANTS. L'algorithme HANTS est basé sur une analyse harmonique (sommes de fonctions sinusoïdales) qui peut conduire à des fausses oscillations des profils temporels reconstruits comme cela a été le cas en début de cycle. Ces fausses oscillations produites par les méthodes basées sur la décomposition de Fourier ont été également soulignées par Hermance (2007); Lu et al. (2007); Roerink et al. (2000); Jönsson and Eklundh (2002);

Chen et al. (2004).

5.2.2. Sur l'estimation des biomasses et rendements

L'estimation des biomasses et rendements de l'arachide et du mil a révélé tout d'abord que l'utilisation du seul indice NDVI et des métriques dérivées était insuffisante pour expliquer une part satisfaisante de leurs variabilités. Si les périodes d'intégration du NDVI sont plus nombreuses à être corrélées à la biomasse et aux rendements de l'arachide et du mil que celles du GDVI, ce sont bien les cumuls de valeurs de GDVI qui expliquent le mieux les biomasses et rendements observés. Les études menées jusqu'à présent sur l'estimation des rendements en Afrique de l'Ouest étaient basées essentiellement sur des séries temporelles NDVI à basse résolution spatiale comme MODIS ou NOAA-AVHRR et ne prenaient pas en considération d'autres indices de végétation (ex. Leroux et al. (2016); Maselli et al. (2000); Rasmussen (1997)). Les récents travaux menés sur le continent africain avec des données à haute résolution spatiale, temporelle et spectrale mieux adaptées à l'hétérogénéité spatiale et temporelle des systèmes de culture en place ont permis de mettre en évidence des indicateurs plus performant que le NDVI pour l'estimation des rendements. Au Mali, (Lambert et al., 2018) ont montré que le red edge NDVI au maximum de la saison dérivés de Sentinel-2 permettait une meilleure estimation des rendements du mil que le NDVI. Au Kenya, (Jin et al., 2017) ont quant à eux montré que les prédictions des rendements du maïs à partir du MERIS Terrestrial Chlorophyll Index dérivés de RapidEye et Sentinel-2 étaient supérieures à celles faites à partir du NDVI. Les résultats de notre étude corroborent donc les tendances observées dans les récents travaux menés dans le domaine.
Nos résultats ont également montré que les rendements en grain du mil étaient mieux estimés par nos modèles que la biomasse végétative. Sachant que la biomasse végétative est linéairement reliée au FAPAR (Monteith, 1972) d'une part qu'il a d'autre part était montré une relation linéaire entre le FAPAR et des indices de végétation comme le NDVI (Myneni and Williams, 1994), les résultats que nous obtenons sont surprenants. Ces résultats peuvent en partie s'expliquer par la mixité du signal reçu sur les parcelles de mil en cultures associées, où l'association avec le niébé qui est une culture couvrante à fort taux d'activité chlorophyllienne peut contribuer de façon significative au signal capté. Concernant le cas spécifique de l'arachide, nous avons constaté globalement une meilleure capacité de nos modèles finaux de télédétection à estimer la biomasse ($R^2 = 0,56$) que les rendements gousses ($R^2 = 0,49$). Ces résultats s'expliquent notamment par le fait que, s'il existe une relation linéaire entre la biomasse végétative et la biomasse grain pour le cas du mil ($R^2 = 0,78$ lorsque l'on regarde les données observées, voir annexe E), cette relation est plus modérée pour le cas de l'arachide entre la biomasse végétative et le rendement gousse ($R^2 = 0,45$, voir annexe E). Ceci traduit donc le fait que la quantité d'activité photosynthétique produite par la plante n'est pas proportionnelle au nombre de gousses qui se développent sous terre.
Enfin concernant les variables explicatives les plus corrélées à la biomasse et au rendement, nous avons vu que dans le cas de l'arachide c'est l'interaction entre le GDVI cumulé sur l'ensemble du cycle de la plante ($CUM_{GDVI:15-100}$) et la date de fin (EOS) qui explique le mieux la biomasse végétative finale. Sachant qu'à la fin de la saison, l'ensemble des biomasses végétatives pour l'arachide sont arrachées et retirées des

parcelles, marquant donc un arrêt quasi-total de l'activité chlorophyllienne, cette interaction n'est pas surprenante. Pour le cas du rendement gousse, il s'agit du GDVI cumulé sur une période allant du 15ème au 20ème jour après l'émergence, correspondant au premier stade de la phase reproductive de l'arachide, soit l'ouverture des premières fleurs (Boote, 1982). Ce résultat est surprenant et demande plus d'investigations car à notre connaissance ce n'est pas le stade de développement le plus déterminant pour le rendement gousse de l'arachide. Pour le mil, la variable qui explique le mieux la biomasse végétative est le GDVI cumulé sur la phase végétative ($CUM_{GDVI:0-60}$), c'est à dire allant de l'émergence jusqu'au LAI maximum, signifiant donc un arrêt de la production de biomasse végétative par la suite. Nos résultats sont en ce sens cohérents avec les résultats déjà obtenus pour du mil sur le Niger par Leroux et al. (2016). Pour l'estimation des rendements grains finaux du mil, le meilleur modèle intègre une interaction entre la date du LAI (ou NDVI) maximum et la date de fin du cycle (EOS). Là encore, les résultats sont assez cohérents avec ce que l'on connaît du cycle de développement du mil, notamment (1) la date du maximum de végétation détermine en partie la phase reproductive du mil qui inclus notamment la floraison et le remplissage des grains, étapes déterminantes pour les rendements finaux, (2) la date de fin de la saison des cultures est déterminée par la fin de la saison des pluies, par conséquent un arrêt précoce des précipitations entraîne une réduction de la période MAX-EOS et une augmentation des risques de stress hydriques en fin de cycle de développement de la plante sur les phases reproductives et de maturation des grains.

6. Conclusion et Perspectives

Ce travail a été réalisé dans le contexte de l'évaluation spatialisée par télédétection des systèmes de cultures à base de mil et d'arachide dans le bassin arachidier du Sénégal. Dans cette perspective, il avait pour objectif principal d'évaluer le potentiel d'une série temporelle multisource (Sentinel-2, RapidEye et PlanetScope) à travers deux sous objectifs : l'évaluation des dates de semis et l'estimation des biomasses végétatives et des rendements grains du mil et gousses de l'arachide.
En lissant les profils temporels de NDVI extraits à partir de la série multisource, notamment avec la méthode de Whittaker, nous avons extrait différentes métriques phénologiques dont les dates de début de croissance de la végétation (SOS). Avec les SOS extraits, nous avons estimé les dates de semis des différents systèmes de culture avec notamment plus ou moins 5 jours de décalage pour les parcelles d'arachide et 10 à 20 jours de décalage pour celles de mil. Pour l'estimation des biomasses et rendements, le NDVI, qui est traditionnellement utilisé, a été comparé à un autre indicateur de productivité de la végétation, le GDVI. Nous avons montré que le NDVI était moins performant que le GDVI pour expliquer la variabilité de la biomasse et des rendements, ceci s'expliquant notamment par un changement dans les technologies d'observation de la Terre par rapport aux études précédentes. Cette étude a montré le réel intérêt de combiner différentes sources d'images satellitaires afin d'améliorer la répétitivité de l'information spatiale. Elle a également montré que la différence de résolutions spatiales entre les types d'images ne représente en aucun cas un frein à leur utilisation combinée au sein d'une même application.
Néanmoins, au vu des résultats obtenus, la méthodologie mise en place pourrait être affinée, notamment pour améliorer la partie estimation des biomasses et rendements. Dans un ordre de priorité, notre première recommandation concerne la phase de prétraitements des images qui pourrait inclure des corrections atmosphériques afin d'uniformiser tout le jeu données. En effet, seules les images Sentinel-2A dans notre série temporelle multisource présentaient des corrections atmosphériques et nombreuses sont les études en télédétection ayant fait état du gain significatif que ces corrections pouvaient apporter. Dans notre série temporelle, nous n'avons utilisé que les images Sentinel-2A. Sentinel-2B étant aujourd'hui opérationnel, un couplage des 2 types d'images viendrait densifier notre série temporelle afin d'éviter des périodes sans informations comme cela a été le cas pendant le mois de Septembre dans notre étude. Également sur ce point, l'utilisation d'images radar pourrait être une solution envisageable quand on sait que le principal avantage de l'imagerie radar se trouve dans l'affranchissement des conditions atmosphériques, nuages notamment. L'adoption d'images radars permettrait également de tester un apport éventuel des indices radar pour l'extraction des SOS et EOS et l'estimation des biomasses et rendements.
En ce qui concerne les méthodes de lissage, s'il est vrai qu'elles ont plutôt donné satisfaction, il n'en est pas moins que cette étape influence fortement les résultats obtenus par la suite. Il serait alors intéressant de tester d'autres types d'approches comme les fonctions asymétriques gaussiennes ou les doubles fonctions logistiques

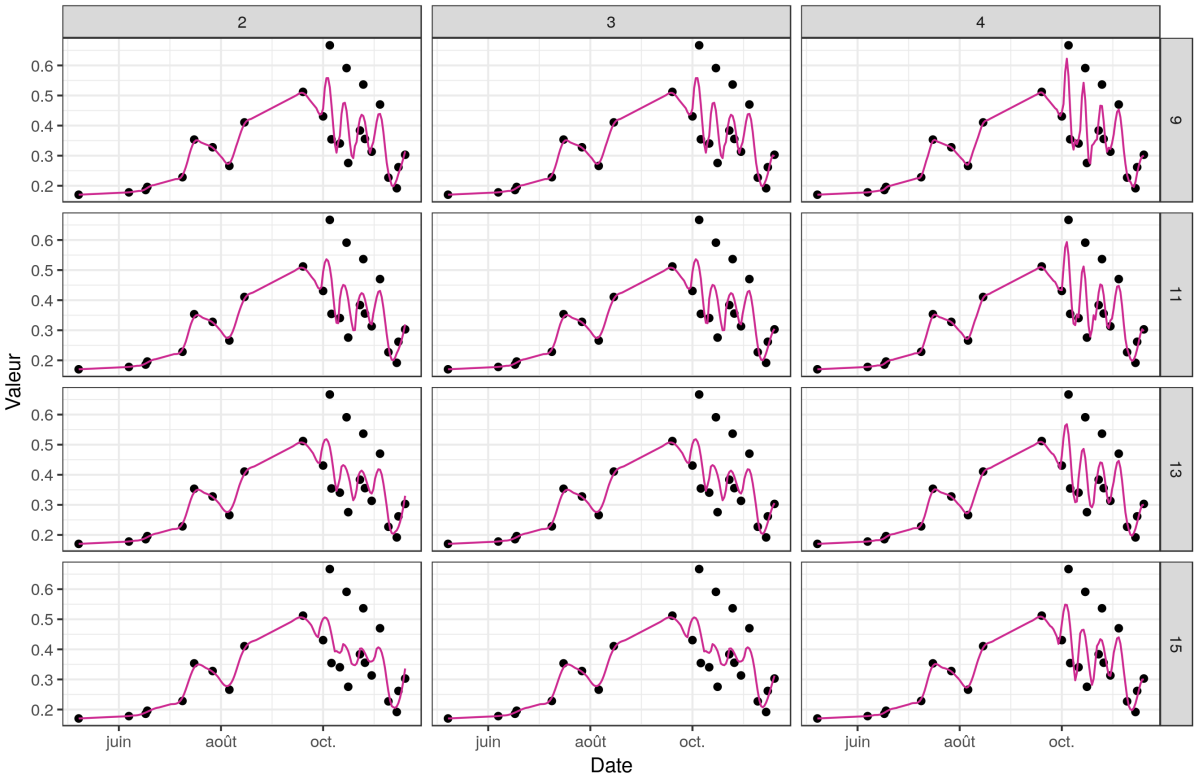
dont les résultats sont jugés tout autant satisfaisants. Aussi, la grande majorité de notre travail a été basé sur l'utilisation du NDVI qui a montré ses limites pour l'estimation des biomasses et rendements. Dès lors, il serait pertinent de se tourner vers d'autres indices de végétation comme cela a déjà pu être le cas avec le GDVI. D'autres indicateurs comme le MSAVI qui tient compte de l'effet des sols (intéressant dans le cas de cultures peu couvrantes comme le mil), le NDWI pouvant détecter la présence de stress hydrique chez les cultures ou encore les indices explorant la bande du red edge comme le Red Edge NDVI pourraient être de réelles alternatives. Par ailleurs, l'extraction des métriques phénologiques a été réalisée uniquement à partir de la méthode de seuillage relatif sur les amplitudes de NDVI et il serait intéressant de tester d'autres méthodes notamment les méthodes dérivatives. Outre cette étape, celle de l'estimation des biomasses et rendements a tenu compte de variables explicatives provenant des seuls indices que sont le NDVI et le GDVI. Nous n'avons ni exploré les variables biophysiques comme le LAI ni les variables climatiques comme l'humidité du sol et les quantités de précipitations bien connues pour améliorer les modèles calibrés. Les indices texturaux pourraient être également interessant à tester. L'autre phase inexplorée également dans notre travail concerne la non prise en compte du pouvoir explicatif des arbres (présents dans les parcelles) dans la productivité des systèmes de cultures. Une des limites de notre travail aura été l'étroitresse de la taille de notre jeu de données terrain rendant impossible la calibration de modèles multivariés plus robustes et leur validation. En effet, une taille d'échantillons plus grande aurait permis l'implication de plusieurs variables dans nos modèles et aurait certainement amélioré les estimations de biomasses et rendements. D'autre part, augmenter la taille de nos échantillons permettrait d'envisager le recours à des méthodes de régression plus poussées par fouilles de données telle que les forêts aléatoires d'arbres décisionnels pour l'estimation des biomasses et rendements. Une autre limite de notre travail repose sur la qualité des données de biomasses et de rendements observés. Effectivement, ces données ayant été collectées dans la cadre d'autres projets et donc pas initialement pour l'estimation des biomasses et rendements, il nous aura été impossible de savoir jusqu'à quel point nous y fier. Sur ce point, un nouveau réseau de parcelles est en cours de suivi pour la campagne agricole 2018. Ce nouveau jeu de données sera ajouté à celui de la campagne 2017 sur laquelle nous avons travaillé et viendra élargir la taille d'échantillons pour les estimations de biomasses et de rendements de 2018. Pour finir, une évaluation spatialisée des rendements dans la région du bassin arachidier sénégalais requiera avant tout un masque des principaux systèmes de cultures présents. Ce type de données est obtenu par classification. Dans le cadre du projet S2-Agri [1], un outil a été mis en place pour la production de masques de cultures ainsi que des cartes d'occupation du sol spécialisées sur l'agriculture.

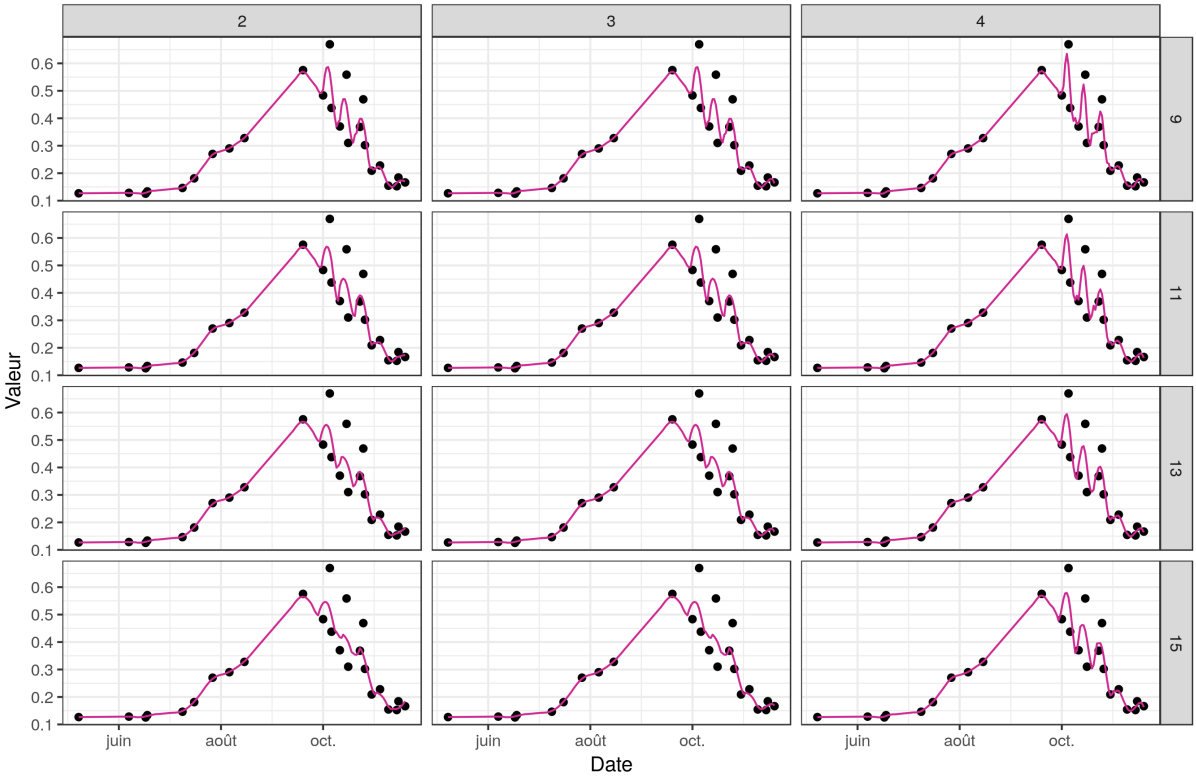
1. http://www.esa-sen2agri.org/

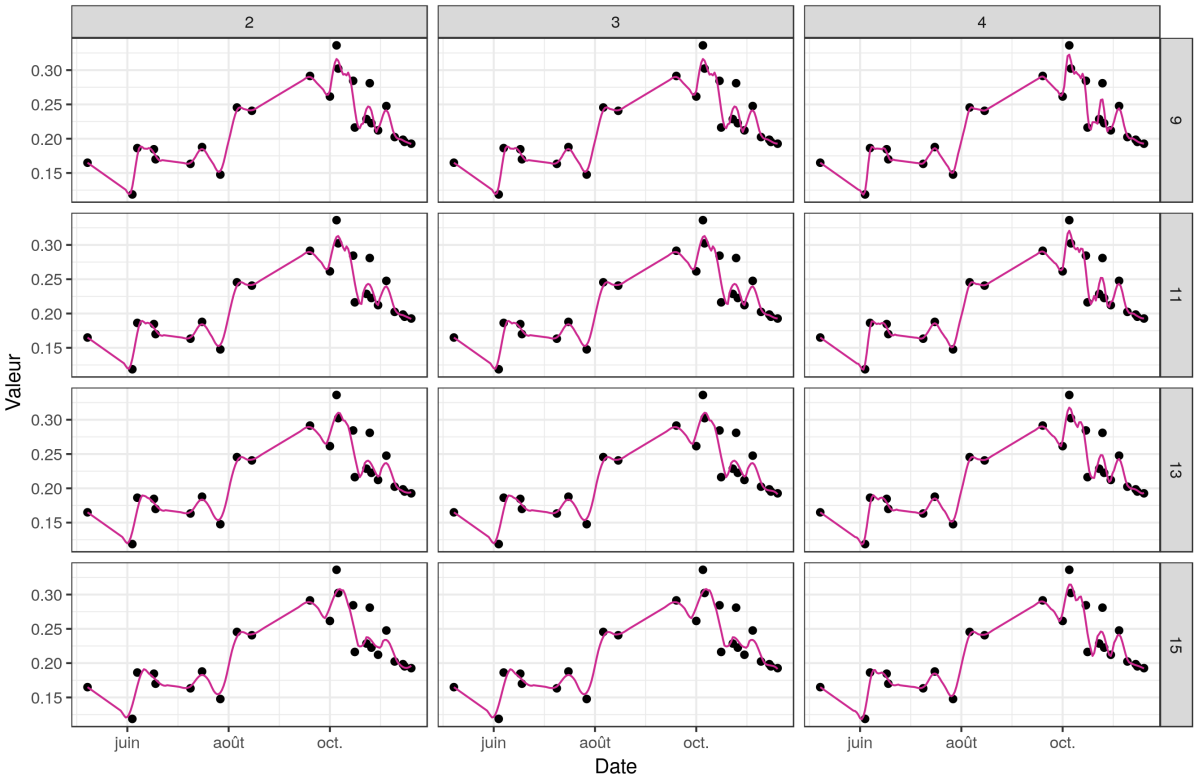
A. Diagramme de Gantt

B. Essais de lissage avec le filtre de Savitzky-Golay

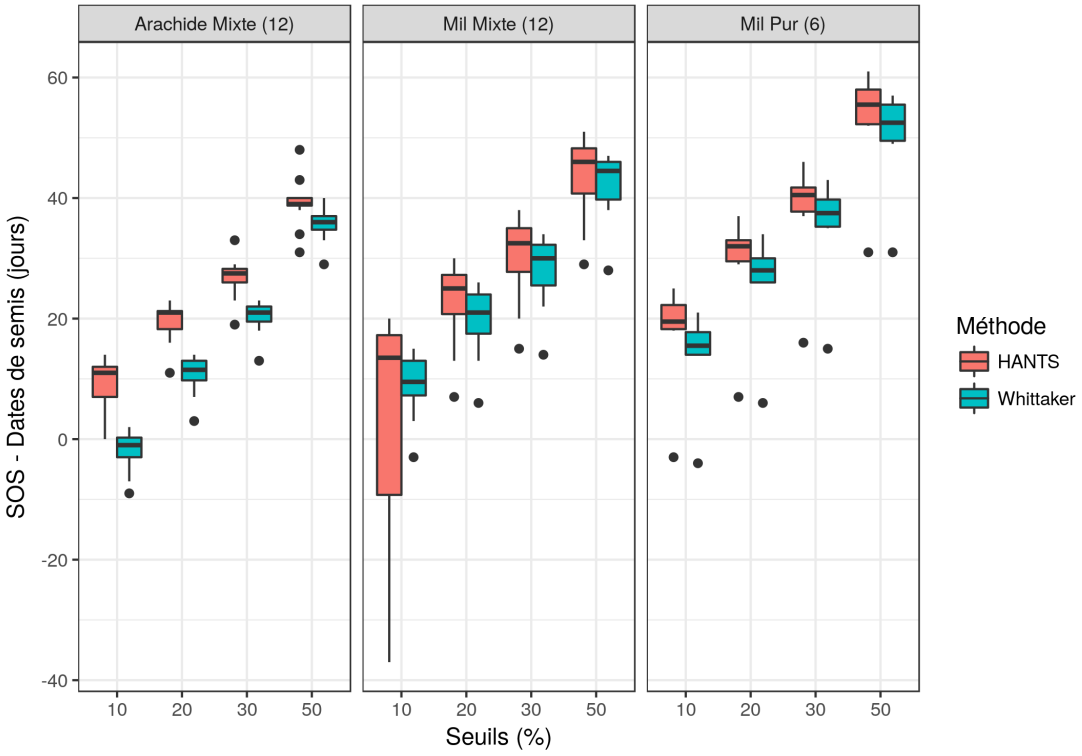
Les figures ci-après représentent des essais de lissage avec le filtre de Savitzky-Golay sur quelques pixels appartenant aux parcelles suivies. Pour chaque pixel, nous faisons varier sur les colonnes le degré du polynôme d'ajustement et sur les lignes la largeur de la fenêtre de lissage, selon les valeurs proposées dans la littérature notamment par Chen et al. (2004).



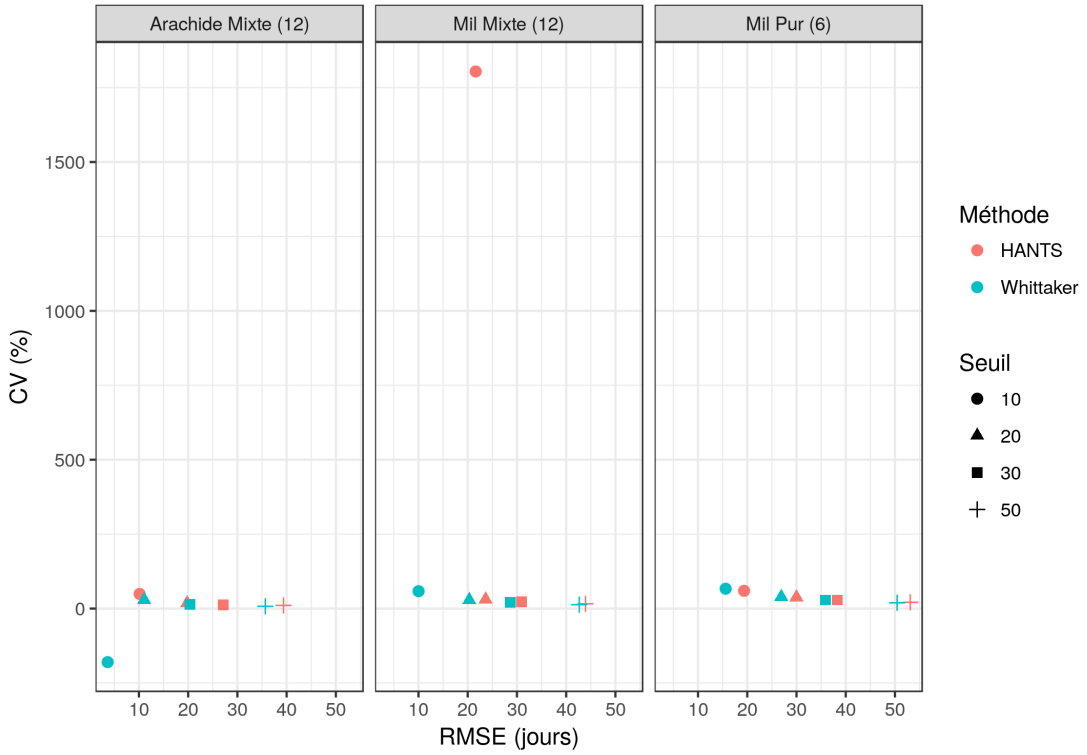

2
3
4
9
11
13
15
Valeur
0.6
0.5
0.4
0.3
0.2
0.1
juin
août
oct.
Date


2
3
4
9
11
13
15
Valeur
0.30
0.25
0.20
0.15
juin
août
oct.
Date

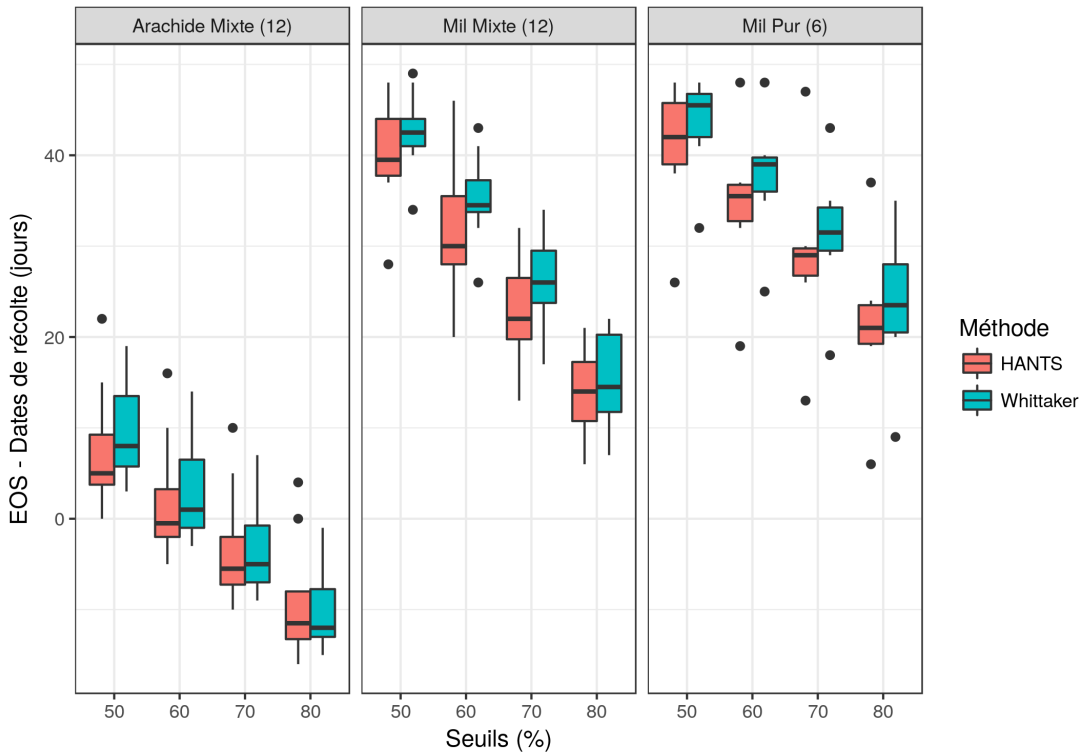
C. Estimation des SOS et EOS sans le jeu ANR CERAO



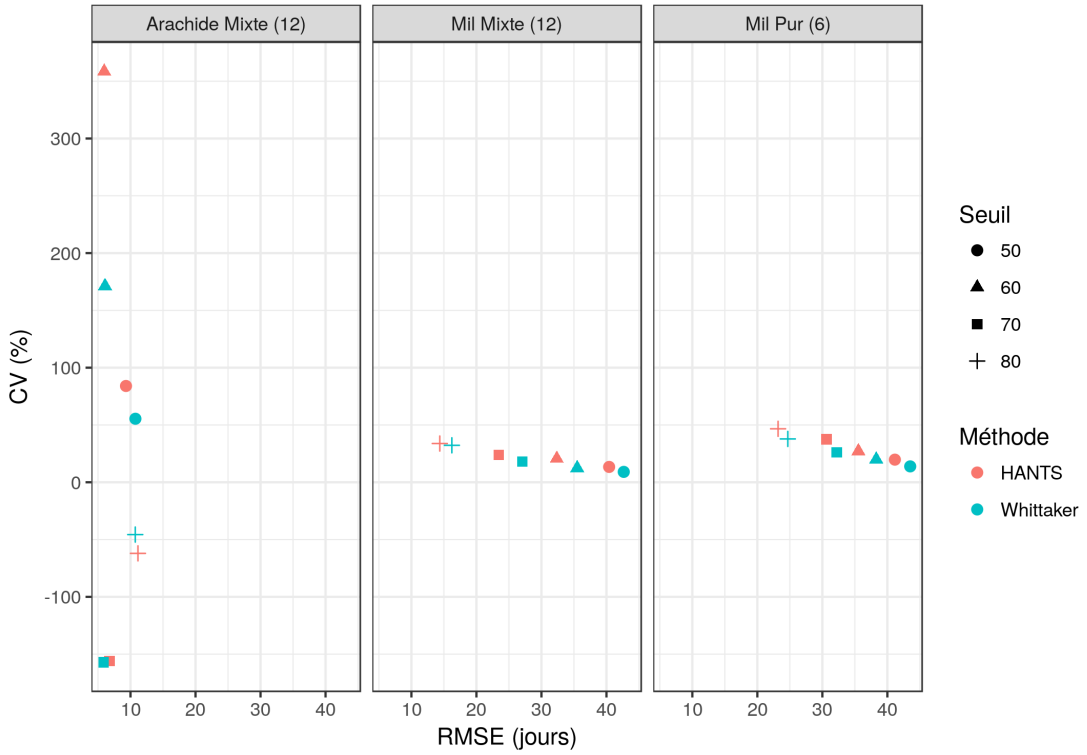
Distribution des écarts entre SOS et dates de semis



SOS : RMSE vs CV



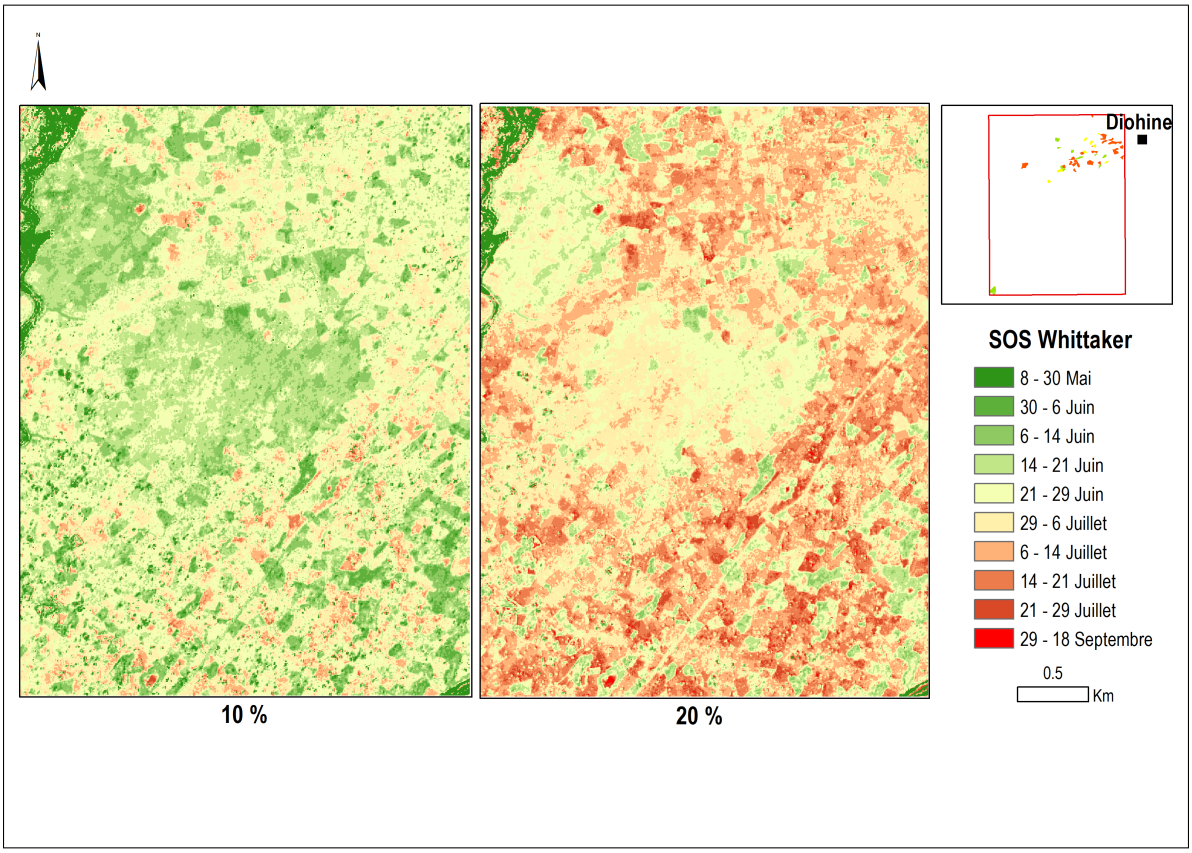
Distribution des écarts entre EOS et dates de récolte



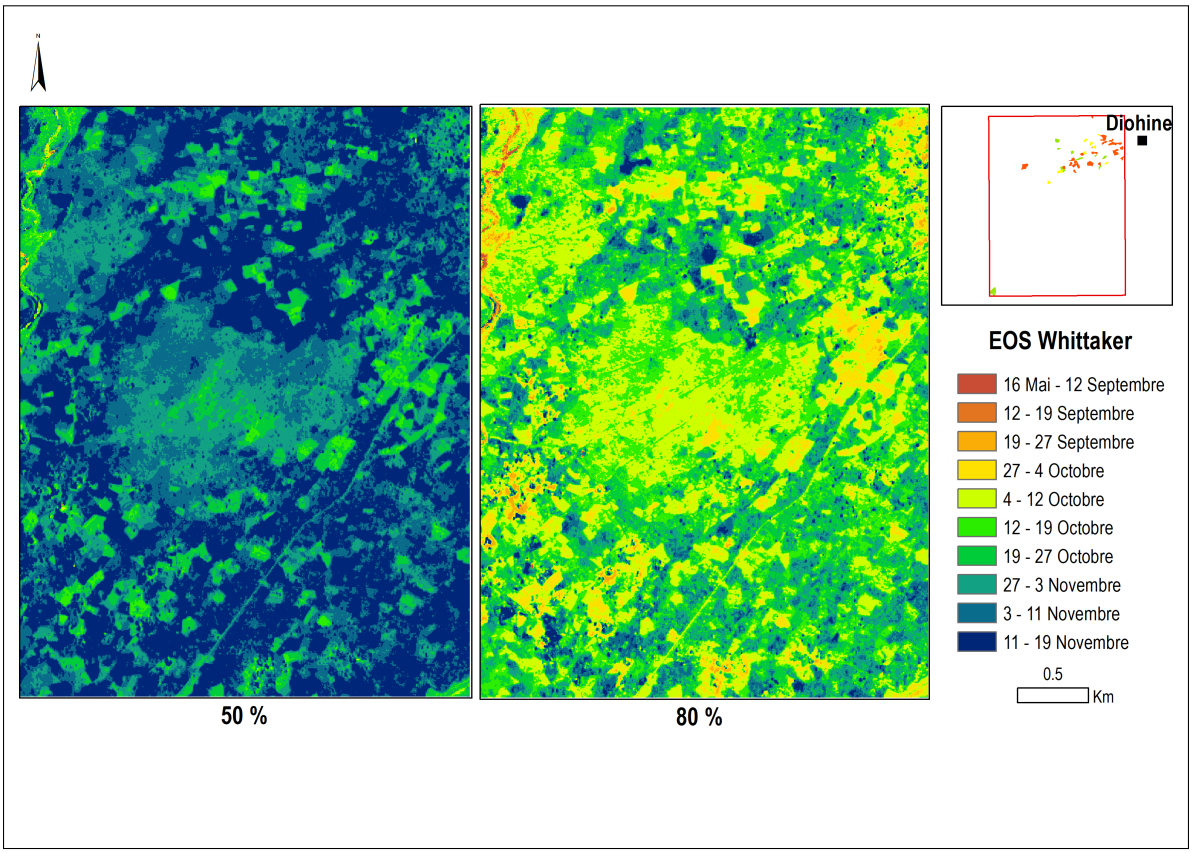
EOS : RMSE vs CV

D. Spatialisation des SOS et EOS

La spatialisation des SOS et EOS à l'échelle pixellaire (3 mètres) avec les seuils adoptés met bien en évidence les zones de jachère sur les cartes produites.



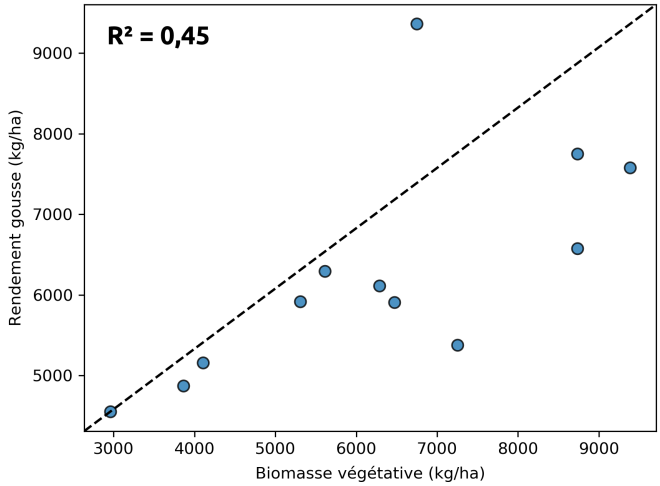
Spatialisation des SOS



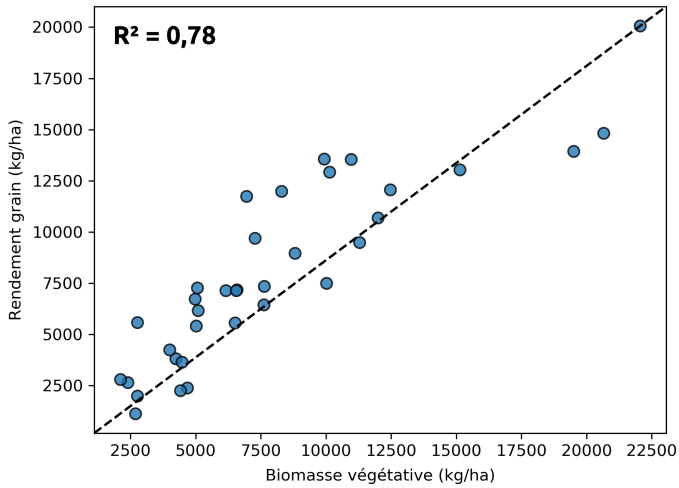
Spatialisation des EOS

E. Relations linéaires entre biomasses et rendements

Régression linéaire entre la biomasse végétative et le rendement gousse de l'arachide



Régression linéaire entre la biomasse végétative et le rendement grain du mil



Bibliographie

Atkinson, P. M., Jeganathan, C., Dash, J., and Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, 123 :400–417.

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.

Atzberger, C. and Eilers, P. H. (2011). Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *International Journal of Remote Sensing*, 32(13) :3689–3709.

Bacci, L., Cantini, C., Pierini, F., Maracchi, G., and Reyniers, F. N. (1999). Effects of sowing date and nitrogen fertilization on growth, development and yield of a short day cultivar of millet (Pennisetum glaucum L.) in Mali. *European Journal of Agronomy*, 10(1) :9–21.

Barbier, B., Yacouba, H., Karambiri, H., Zoromé, M., and Somé, B. (2009). Human vulnerability to climate variability in the sahel : Farmers' adaptation strategies in northern burkina faso. *Environmental Management*, 43(5) :790–803.

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Marais Sicre, C., Le Dantec, V., and Demarez, V. (2016). Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment*, 184 :668–681.

Beck, P. S., Atzberger, C., Høgda, K. A., Johansen, B., and Skidmore, A. K. (2006). Improved monitoring of vegetation dynamics at very high latitudes : A new method using MODIS NDVI. *Remote Sensing of Environment*, 100(3) :321–334.

Bégué, A., Vintrou, E., Saad, A., and Hiernaux, P. (2014). Differences between cropland and rangeland MODIS phenology (start-of-season) in Mali. *International Journal of Applied Earth Observation and Geoinformation*, 31(1) :167–170.

Bojanowski, J., Kowalik, W., and Bochenek, Z. (2009). Noise Reduction of Ndvi Time-Series : a Robust Method Based on Savitzky-Golay Filter Redukcja Szumów W Szeregach Czasowych Ndvi Z Zastosowaniem Filtra Savitzky-Golay. *International Journal of Remote Sensing*, 13 :1585–1590.

Boote, K. J. (1982). Growth Stages of Peanut (Arachis hypogaea L.) 1. *Peanut Science*, 9(1) :35–40.

Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A. A., Mbow, C., Mougin, E., and Fensholt, R. (2016). Woody plant cover estimation in drylands

from Earth Observation based seasonal metrics. *Remote Sensing of Environment*, 172 :28–38.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, 91(3-4) :332–344.

Comité de la Sécurité Alimentaire Mondiale (2012). S'entendre sur la terminologie : Sécurité alimentaire, sécurité nutritionnelle, sécurité alimentaire et nutrition, sécurité alimentaire et nutritionnelle. page 17.

de Beurs, K. M. and Henebry, G. M. (2010). Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology. In *Phenological Research*, pages 177–208. Springer Netherlands, Dordrecht.

Delécolle, R., Maas, S. J., Guérif, M., and Baret, F. (1992). Remote sensing and crop production models : present trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 47(2-3) :145–161.

Diouf, O. (2001). La culture du mil (Pennisetum glaucum (L .) R . Br .) en zone siemi-aride : bases agrophysiologiques justificatives d ' une fertilisation azotée. page 26.

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., and Bargellini, P. (2012). Sentinel-2 : ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120 :25–36.

Duarte, L., Teodoro, A. C., Monteiro, A. T., Cunha, M., and Gonçalves, H. (2018). QPhenoMetrics : An open source software application to assess vegetation phenology metrics. *Computers and Electronics in Agriculture*, 148(October 2017) :82–94.

Eilers, P. H. (2003). A perfect smoother. *Analytical Chemistry*, 75(14) :3631–3636.

Eklundh, L. and Jönsson, P. (2017). TIMESAT 3.3 with seasonal trend decomposition and parallel processing Software Manual. *Lund and Malmo University, Sweden*, pages 1–92.

Espinoza-Dávalos, G. E., Bastiaanssen, W. G. M., Bett, B., and Cai, X. (2017). A Python Implementation of the Harmonic ANalysis of Time Series (HANTS) Algorithm for Geospatial Data.

FAO, FIDA, OMS, PAM, and UNICEF (2017). *L'état de la sécurité alimentaire et de la nutrition dans le monde 2017. Renforcer la résilience pour favoriser la paix et la sécurité alimentaire*. Rome, FAO.

Fieuzal, R., Marais Sicre, C., and Baup, F. (2017). Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, 57 :14–23.

Geng, L., Ma, M., Wang, X., Yu, W., Jia, S., and Wang, H. (2014). Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the heihe river basin, China. *Remote Sensing*, 6(3) :2024–2049.

Goulet, F. (2012). La notion d'intensification écologique et son succès auprès d'un certain monde agricole français. Une radiographie critique. *Le Courrier de l'Environnement de l'Inra*, 62 :19–30.

Helman, D. (2018). Land surface phenology : What do we really 'see' from space ? *Science of the Total Environment*, 618 :665–673.

Hermance, J. F. (2007). Stabilizing high-order, non-classical harmonic analysis of NDVI data for average annual models by damping model roughness. *International Journal of Remote Sensing*, 28(12) :2801–2819.

Hird, J. N. and McDermid, G. J. (2009). Noise reduction of NDVI time series : An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113(1) :248–258.

Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11) :1417–1434.

Ingram, K. T., Roncoli, M. C., and Kirshen, P. H. (2002). Opportunities and constraints for farmers of west Africa to use seasonal precipitation forecasts with Burkina Faso as a case study. *Agricultural Systems*, 74(3) :331–349.

Jakubauskas, M. E., Legates, D. R., and Kastens, J. H. (2001). Harmonic analysis of time - series AVHRR NDVI data. *Photogrammetric engineering and remote sensinghotogrammetric engineering and remote sensing*, 67(4) :461 – 470.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., and Wang, J. (2018). A review of data assimilation of remote sensing and crop models.

Jin, Z., Azzari, G., and Lobell, D. B. (2017). Improving the accuracy of satellite-based high-resolution yield estimation : A test of multiple scalable approaches. *Agricultural and Forest Meteorology*, 247(July) :207–220.

Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 141 :116–128.

Jönsson, P. and Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8) :1824–1832.

Jönsson, P. and Eklundh, L. (2004). TIMESAT - A program for analyzing time-series of satellite sensor data. *Computers and Geosciences*, 30(8) :833–845.

Julien, Y. and Sobrino, J. A. (2010). Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sensing of Environment*, 114(3) :618–625.

Julien, Y., Sobrino, J. A., and Verhoef, W. (2006). Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sensing of Environment*, 103(1) :43–55.

Kandasamy, S., Baret, F., Verger, A., Neveux, P., and Weiss, M. (2012). A comparison of methods for smoothing and gap filling time series of remote sensing observations : application to MODIS LAI products. *Biogeosciences Discussions*, 9(12) :17053–17097.

Kim, S.-R., Prasad, A. K., El-Askary, H., Lee, W.-K., Kwak, D.-A., Lee, S.-H., and Kafatos, M. (2014). Application of the Savitzky-Golay Filter to Land Cover Classification Using Temporal MODIS Vegetation Indices. *Photogrammetric Engineering & Remote Sensing*, 80(7) :675–685.

Kimball, J. S. (2014). Vegetation phenology. In *Encyclopedia of Earth Sciences Series*, pages 886–890. Springer, New York, NY.

Klisch, a., Royer, a., Lazar, C., Baruth, B., and Genovese, G. (2006). Extraction of phenological parameters from temporally smoothed vegetation indices. *Methods*, pages 91–96.

Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., and Lavrenyuk, A. (2013). Winter wheat yield forecasting in Ukraine based on Earth observation, meteorologicaldata and biophysical models. *International Journal of Applied Earth Observation and Geoinformation*, 23(1) :192–203.

Kogan, F. N. (1995). Droughts of the Late 1980s in the United States as Derived from NOAA Polar-Orbiting Satellite Data. *Bulletin of the American Meteorological Society*, 76(5) :655–668.

Lambert, M. J., Traoré, P. C., 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.

Leroux, L. (2015). Suivi et caractérisation des dynamiques de la production agricole en Afrique de l'Ouest par télédétection moyenne résolution spatiale. Thèse de d :292.

Leroux, L., Baron, C., Zoungrana, B., Traore, S. B., Lo Seen, D., and Begue, A. (2016). Crop Monitoring Using Vegetation and Thermal Indices for Yield Estimates : Case Study of a Rainfed Cereal in Semi-Arid West Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(1) :347–362.

Liu, R., Shang, R., Liu, Y., and Lu, X. (2017). Global evaluation of gap-filling approaches for seasonal NDVI with considering vegetation growth trajectory, protection of key point, noise resistance and curve stability. *Remote Sensing of Environment*, 189 :164–179.

Lu, X., Liu, R., Liu, J., and Liang, S. (2007). Removal of Noise by Wavelet Method to Generate High Quality Temporal Data of Terrestrial MODIS Products. *Photogrammetric Engineering & Remote Sensing*, 73(10) :1129–1139.

Lunetta, R. S., Knight, J. F., Ediriwickrema, J., Lyon, J. G., and Worthy, L. D. (2006). Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, 105(2) :142–154.

Marinho, E., Vancutsem, C., Fasbender, D., Kayitakire, F., Pini, G., and Pekel, J. F. (2014). From remotely sensed vegetation onset to sowing dates : Aggregating pixel-level detections into village-level sowing probabilities. *Remote Sensing*, 6(11) :10947–10965.

Marteau, R., Sultan, B., Moron, V., Alhassane, A., Baron, C., and Traoré, S. B. (2011). The onset of the rainy season and farmers' sowing strategy for pearl millet cultivation in Southwest Niger. *Agricultural and Forest Meteorology*, 151(10) :1356–1369.

Maselli, F., Romanelli, S., Bottai, L., and Maracchi, G. (2000). Processing of GAC NDVI data for yield forecasting in the Sahelian region. *International Journal of Remote Sensing*, 21(18) :3509–3523.

Menenti, M., Azzali, S., Verhoef, W., and Vanswol, R. (1993). Mapping agroecological zones and time-lag in vegetation growth by means of fourier-analysis of time-series of ndvi images. *Advances in Space Research*, 13(5) :233–237.

Misra, G., Buras, A., and Menzel, A. (2016). Effects of Different Methods on the Comparison between Land Surface and Ground Phenology—A Methodological Case Study from South-Western Germany. *Remote Sensing*, 8(9) :753.

Monteith, J. L. (1972). Solar Radiation and Productivity in Tropical Ecosystems. *Journal of Apllied Ecology*, 9(3) :747–766.

Moulin, S., Kergoat, L., Viovy, N., and Dedieu, G. (1997). Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. *Journal of Climate*, 10(6) :1154–1170.

Myneni, R. B. and Williams, D. L. (1994). On the relationship between FAPAR and NDVI. *Remote Sensing of Environment*, 49(3) :200–211.

Neumann, K., Verburg, P. H., Stehfest, E., and Müller, C. (2010). The yield gap of global grain production : A spatial analysis. *Agricultural Systems*, 103(5) :316–326.

Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G. A., Yan, J., and Liu, J. (2015). Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. *International Journal of Applied Earth Observation and Geoinformation*, 34 :188–197.

Prince, S. (1990). HIGH TEMPORAL FREQUENCY REMOTE SENSING OF PRIMARY PRODUCTION USING NOAA AVHRR. In *Applications of Remote Sensing in Agriculture*, pages 169–183. Butterworths.

Rasmussen, M. S. (1997). Operational yield forecast using AVHRR NDVI data : Reduction of environmental and inter-annual variability. *International Journal of Remote Sensing*, 18(5) :1059–1077.

Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, 5(4) :1704–1733.

Rodrigues, A., Marcal, A. R., and Cunha, M. (2013). Monitoring vegetation dynamics inferred by satellite data using the pheno sat tool. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4) :2096–2104.

Roerink, G. J., Menenti, M., and Verhoef, W. (2000). Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, 21(9) :1911–1917.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. (1973). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. *Progress Report RSC 1978-1*, page 112.

Savitzky, A. and Golay, M. J. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8) :1627–1639.

Shao, Y., Lunetta, R. S., Wheeler, B., Iiames, J. S., and Campbell, J. B. (2016). An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. *Remote Sensing of Environment*, 174 :258–265.

Sibley, A. M., Grassini, P., Thomas, N. E., Cassman, K. G., and Lobell, D. B. (2014). Testing remote sensing approaches for assessing yield variability among maize fields. *Agronomy Journal*, 106(1) :24–32.

Sivakumar, M. V. K. (1990). Exploiting rainy season potential from the onset of rains in the Sahelian zone of West Africa. *Agricultural and Forest Meteorology*, 51(3-4) :321–332.

Tateishi, R. and Ebata, M. (2004). Analysis of phenological change patterns using 1982-2000 Advanced Very High Resolution Radiometer (AVHRR) data. *International Journal of Remote Sensing*, 25(12) :2287–2300.

Tittonell, P. and Giller, K. E. (2013). When yield gaps are poverty traps : The paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research*, 143 :76–90.

Tucker, C. (1979). Red and photographic infarecd linear combinations for monitoring vegetation. *Remote sensing of Enviroment, 8*, 150 :127–150.

Tucker, C. J., Holben, B. N., Elgin, J. H., and McMurtrey, J. E. (1981). Remote sensing of total dry-matter accumulation in winter wheat. *Remote Sensing of Environment*, 11(C) :171–189.

Tucker, C. J. and Sellers, P. J. (1986). Satellite remote sensing of primary production. *International Journal of Remote Sensing*, 7(11) :1395–1416.

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.

Verhoef, W., Menenti, M., and Azzali, S. (1996). Cover a colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981-1992). *International Journal of Remote Sensing*, 17(2) :231–235.

Viovy, N., Arino, O., and Belward, A. S. (1992). The best index slope extraction (Bise) : A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing*, 13(8) :1585–1590.

Vrieling, A., Skidmore, A. K., Wang, T., Meroni, M., Ens, B. J., Oosterbeek, K., O'Connor, B., Darvishzadeh, R., Heurich, M., Shepherd, A., and Paganini, M. (2017). Spatially detailed retrievals of spring phenology from single-season high-resolution image time series. *International Journal of Applied Earth Observation and Geoinformation*, 59 :19–30.

White, M. A., Thornton, P. E., and Running, S. W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, 11(2) :217–234.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., and Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3) :471–475.

Zhou, J., Jia, L., Hu, G., and Menenti, M. (2012). Evaluation of Harmonic Analysis of Time Series (HANTS) : Impact of gaps on time series reconstruction. *Proceedings of the 2nd International Workshop on Earth Observation and Remote Sensing Applications, EORSA 2012*, 5(June 2012) :31–35.

Zhou, J., Jia, L., and Menenti, M. (2015). Reconstruction of global MODIS NDVI time series : Performance of Harmonic ANalysis of Time Series (HANTS). *Remote Sensing of Environment*, 163 :217–228.

Zhu, W., Pan, Y., He, H., Wang, L., Mou, M., and Liu, J. (2012). A changing-weight filter method for reconstructing a high-quality NDVI time series to preserve the integrity of vegetation phenology. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4) :1085–1094.