



ECOLE SUPERIEURE D'AGRICULTURES  
55, rue Rabelais - B.P. 748 ADRESSE  
49007 ANGERS CEDEX 01  
TEL. : 02.41.23.55.55



INRA Grignon – UMR Agronomie  
78850 THIVerval-GRIGNON  
TEL : +33 (0) 1 30 81 54 20

## **Télédétection du colza buissonnant à partir des images satellites Sentinel-2**

**Stage Recherche et Innovation 2019**

**Mots clefs : Télédétection, Sentinel-2, Colza buissonnant**

**Antoine BRASSET**

**118**



## **Remerciements**

En premier lieu, je tiens à remercier M. Corentin BARBU, chargé de recherche à l'UMR agronomie de l'INRA, pour le temps qu'il a consacré à mon encadrement. Il m'a guidé tout au long de mon étude et m'a apporté les connaissances nécessaires à sa réalisation.

Je tiens à remercier M. Thomas DELAUNE, ingénieur de recherche à l'UMR agronomie de l'INRA, pour son accompagnement et la transmission de son savoir. Je le remercie également pour ses conseils précieux lors de la rédaction du rapport de stage.

Un grand merci à Mme Céline ROBERT, chargée d'études ravageurs des cultures et faune auxiliaire chez Terre Inovia, pour les données d'épidémiosurveillance qu'elle nous a fournies.



## **NOTICE BIBLIOGRAPHIQUE**

AUTEUR : Antoine BRASSET

Promotion : 118

Identification du rapport : Télédétection du colza buissonnant à partir des images satellites Sentinel-2. 75 pages, 26 figures, 57 bibliographies.

Mot-clé : Télédétection, Sentinel-2, Colza buissonnant

## **RESUME D'AUTEUR**

La télédétection est employée dans de nombreux domaines d'applications, dont l'agriculture où elle est notamment utilisée pour caractériser les peuplements végétaux et surveiller l'impact des bio-agresseurs sur les cultures. En France, certains ravageurs ont développé des résistances aux pyréthrinoïdes. C'est le cas du charançon du bourgeon terminal du colza qui induit son buissonnement et engendre d'importantes pertes de rendements. Dans cette étude, nous allons utiliser les données satellites pour caractériser le peuplement végétal du colza et déterminer dans quelles mesures ces données sont pertinentes pour détecter les parcelles de colzas buissonnantes. Nous nous sommes basés sur les observations de 2017 de la base de données nationale d'épidémirosurveillance Vigicultures® et sur les données satellites Sentinel-2. Les résultats montrent qu'il existe un lien entre les données satellites et la présence de colzas buissonnantes dans les parcelles observées. Plus précisément, les longueurs d'ondes de l'infrarouge court ont la meilleure aptitude à discriminer les colzas buissonnantes. Ces longueurs d'ondes sont déjà utilisées pour estimer la quantité d'eau présente au sein d'un couvert végétal. Bien que notre étude ait mis en évidence un lien entre l'infrarouge court et les colzas buissonnantes, nos données ne nous ont pas permis de réaliser de modèle de prédiction fiable. A terme, associer la télédétection des colzas buissonnantes aux données sur les pratiques culturales des agriculteurs et aux facteurs environnementaux permettrait de conseiller sur les leviers à actionner pour limiter les dégâts du charançon du bourgeon terminal.



## BIBLIOGRAPHIC RECORD

Author: Antoine BRASSET

Years group: 2018

Report description: Remote sensing of bushy rapeseed using satellite images Sentinel-2.

75 pages, 26 figures, 57 bibliography.

Keywords: remote sensing, Sentinel-2, bushy rapeseed

## ABSTRACT

Remote sensing is used in many fields of application such as agriculture in order to characterize vegetation cover. It is used in particular to monitor the impact of pest on certain crops. In France, some pests have developed resistance to pyrethroids, such as rape stem weevil, which causes rapeseed to shrink and causes significant yield loss. In this study we will use satellite data to characterize the vegetation cover of rapeseed and determine how relevant these data are to detect bushy rapeseed patches. We used data from the national epidemiological surveillance database of 2017 (Vigicultures®) as well as Sentinel-2 satellite data. The results show that there is a clear link between satellite data and the presence of bushy canola in the observed plots. More precisely, it would appear that short infrared wavelengths can be used to discriminate bushy rapeseed. These wavelengths are more commonly used to estimate the amount of water present within the vegetation cover. Despite the fact that these explanatory variables are linked to the percentage of bushy rapeseed, we do not obtain a sufficiently reliable prediction model. Succeeding to remotely detecting bushy rapeseed plots and then going back to the practices and environment that caused this situation, would allow farmers to be advised on the levers to be used to circumvent the damage caused by these pests.



## **Signes et abréviations**

GPS : Global Positioning System

Ha : hectare

IFT : Indicateur de Fréquence de Traitements

ITB : Institut Technique de la Betterave

Km : kilomètre

m: mètre

mm : millimetre

OR : Odds Ratio

## **Table des matières**

Introduction .....	7
1 Institut National de la Recherche Agronomique .....	9
1.1 Historique .....	9
1.2 Activités.....	11
1.3 Organisation.....	13
2 Etude bibliographique .....	13
2.1 Projet de recherche .....	13
2.2 Le colza.....	15
2.2.1 Surface et utilisation.....	15
2.2.2 Une culture sensible aux ravageurs .....	15
2.3 Les données de télédétection .....	17
2.3.1 Concept général.....	17
2.3.2 Données Sentinel-2 .....	19
2.1 Utilisation de la télédétection en agriculture .....	21
2.1.1 Les indices de végétation .....	21
2.1.2 Caractérisation du peuplement végétal à partir d'images satellites .....	23
3 Problématique.....	25
4 Matériels et méthodes.....	27
4.1 Organisation générale .....	27
4.2 Base de données.....	27
4.2.1 Données d'épidémiosurveillance .....	27
4.2.1.1 R2D2 .....	27
4.2.1.2 Vigicultures®.....	27
4.2.1.3 Vigicultures® et R2D2 .....	29
4.2.2 Données Sentinel-2 .....	29
4.3 Création des variables de travail.....	31



4.3.1	Variables explicatives .....	31
4.3.1.1	Les bandes spectrales .....	31
4.3.1.2	Les indices de végétation .....	33
4.3.1.3	Variable explicative centrée réduite.....	35
4.3.2	Variables à expliquer.....	35
4.3.2.1	Pourcentage de colzas buissonnants .....	35
4.3.2.2	Colzas buissonnants ou non buissonnantes .....	35
4.4	Analyses statistiques.....	37
4.4.1	LASSO .....	37
4.4.2	Random forest .....	37
5	Résultats .....	39
5.1	Etude pilote : R2D2 .....	39
5.2	Vigicultures® .....	41
5.2.1	Analyse descriptive .....	41
5.2.2	Exploration de l'association des variables et de leurs importances .....	43
5.2.2.1	Modèle LASSO.....	43
5.2.2.2	Modèle Random Forest.....	43
5.2.2.3	Variables discriminant les colzas buissonnants .....	45
5.2.3	Prédiction avec random forest.....	45
5.2.3.1	Variable buissonnant binaire.....	45
5.2.3.2	Variable de pourcentage de plantes buissonnantes .....	47
5.2.3.3	Amélioration de la prédiction par ajout de variables .....	47
5.3	Analyse complémentaire (Vigicultures® et R2D2) .....	49
5.3.1	Analyse descriptive .....	49
5.3.2	Exploration des variables importantes et de la qualité de prédiction .....	49
6	Discussion .....	53
6.1	Discussion Vigicultures® .....	53



6.1.1	Résultat général de l'étude .....	53
6.1.2	Association et importance des variables .....	53
6.1.3	Prédictions avec random forest .....	55
6.2	Discussion de Vigicultures® associé à R2D2 .....	57
6.3	Limite de l'étude.....	59
Conclusion.....		63
Bibliographie.....		65
Table des figures .....		73



## **Introduction**

L'INRA, au cœur de la recherche agronomique française depuis plus de 70 ans, s'est adapté aux évolutions du monde agricole. Actuellement, de nombreuses problématiques liées à la gestion des ravageurs et à l'utilisation des produits phytosanitaires émergent. Certaines cultures, comme le colza, sont très sensibles aux ravageurs à cause des résistances à certaines matières actives présentes dans les insecticides. La problématique de la résistance est un sujet de préoccupation important, notamment concernant le charançon du bourgeon terminal qui induit une modification morphologique du colza qui génère son buissonnement et d'importantes pertes de rendements. Pour mieux lutter contre les ravageurs, il est important de bien connaître leurs habitats, leurs ressources mais aussi les auxiliaires qui sont en mesure de les réguler. La télédétection est l'un des moyens qui peut être mis en œuvre pour collecter des informations sur les dégâts causés par les ravageurs. Cartographier les dégâts des ravageurs aux champs est une première étape dans l'analyse des situations présentant des colzas buissonnantes. A terme, l'analyse des pratiques culturales des agriculteurs et des facteurs environnementaux qui ont contribué à ces dégâts détermineront les principaux leviers à actionner pour limiter les pertes de rendements.

Dans ce rapport, nous allons nous intéresser à la pertinence de l'utilisation de la télédétection pour détecter les parcelles de colzas buissonnantes. Dans un premier temps nous exposerons le contexte global qui entoure notre étude, puis nous allons mettre en place une méthodologie spécifique pour répondre à notre problématique. Enfin, nous analyserons nos résultats et nous les discuterons.



# **1 Institut National de la Recherche Agronomique**

## **1.1 Historique**

L’Institut National de la Recherche Agronomique (INRA), fut fondée en 1946 au lendemain de la seconde guerre mondiale. Son objectif était de trouver une solution pour nourrir la France au lendemain d’un conflit qui a ravagé tout le pays et où la nourriture se faisait rare. A la fin des années 1960, l’INRA réussit sa mission, la France produit suffisamment pour subvenir à ses besoins. L’INRA se lance alors dans un développement régional avec l’ouverture de pôles de recherches sur l’ensemble du territoire. Dès les années 70, le pays se retrouve avec des excédents alimentaires et la France devient exportatrice de produits agricoles. La mission de l’INRA va alors changer, elle va se focaliser sur la qualité de la production et la valeur ajoutée des produits agricoles. Elle va également se tourner vers la transformation de la matière première d’un point de vue qualitatif et sur les process appliqués à l’agroalimentaire. Dans les années 80, l’INRA prend le virage des biotechnologies. Lors de la décennie suivante, les consommateurs demandent une alimentation plus sécuritaire et qualitative, intégrant en outre les données environnementales. L’INRA va alors étendre son domaine de compétences. Ses recherches s’élargissent aussi à la préservation des ressources naturelles et à l’étude de l’impact des pratiques agricoles sur l’environnement. Durant les années 2000, l’INRA se tourne vers le développement durable et le changement climatique. Pour traiter au mieux ces problématiques, l’INRA crée des unités mixtes de recherches associant institut de recherche, université, enseignement agronomique et vétérinaire. La finalité est alors d’aboutir à un modèle agricole durable, économiquement viable et qui produit des aliments de qualité. En 2010, l’INRA change son calendrier et se projette sur une échelle décennale. Les problématiques mondiales sur lesquelles travaillent l’INRA demandent de la restructuration et l’apprentissage de “nouvelles” sciences. Les enjeux auxquels doit répondre l’INRA sont formulés au travers de huit programmes de recherches :

- Gestion intégrée de la santé des plantes,
- Gestion intégrée de la santé animale,
- Adaptation de l’agriculture et de la forêt au changement climatique,
- Méta-omiques des écosystèmes microbiens,
- Déterminants et impacts des régimes alimentaires,



- Sélection génomique,
- Services rendus par les écosystèmes,
- Transitions pour la sécurité alimentaire mondiale.

L'INRA s'est donc adapté aux différentes problématiques que le monde agricole a rencontrées depuis l'après-guerre.

## **1.2 Activités**

Les activités de l'INRA se divisent en six missions distinctes :

- *Produire et diffuser des connaissances scientifiques :*  
L'INRA produit de la connaissance scientifique qui est publiée en accès libre, avec plus de 4 000 publications par an, qui en fait le deuxième institut en sciences agronomiques.
- *Former à la recherche par la recherche :*  
L'INRA encadre notamment 1 800 doctorants dans 11 métadisciplines différentes. Il dispose de 130 équivalents temps-plein d'enseignant-chercheurs destinés à la formation.
- *Eclairer les décisions publiques :*  
L'INRA contribue à la mise en œuvre des politiques publiques aussi bien nationales qu'internationales. Elle apporte des arguments dans les débats qui opposent science et société et contribuent à l'identification des futures thématiques de recherche.
- *Contribuer à l'innovation :*  
L'INRA dispose de 17 domaines d'innovations qu'il valorise à travers trois structures, une société de transfert de l'innovation, une structure d'obtention variétale et une maison d'édition.
- *Elaborer la stratégie de recherche européenne et nationale :*  
Il contribue à l'élaboration de nouveaux axes de recherches internationales au travers d'alliances nationales et d'Eranet (programme européen pour la recherche et l'innovation).
- *Contribuer au dialogue entre science et société :*  
Il souhaite favoriser le débat et la réflexion critique autour de ses activités en mettant à contribution la société. L'INRA communique sous différentes plateformes pour atteindre différents publics.

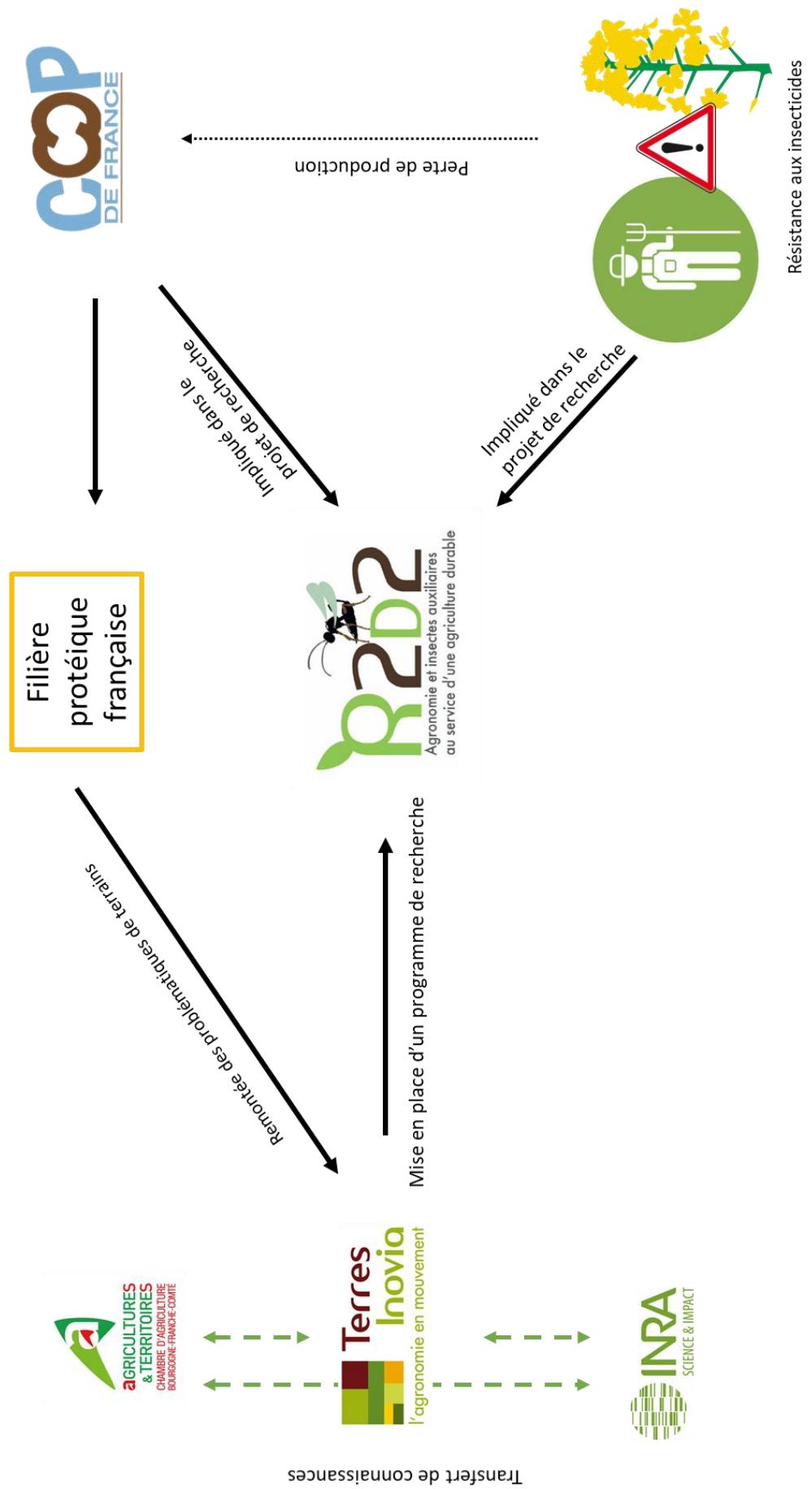


Figure 1 : Schéma des différents acteurs du projet R2D2

### **1.3 Organisation**

L'INRA est structuré en 17 centres de recherches régionaux qui font partie d'une vingtaine de pôle thématiques. Un centre-siège situé à Paris vient compléter cette organisation. L'INRA conduit ses recherches dans 13 départements scientifiques qui regroupent un ensemble de champs disciplinaires. Les questions de recherche s'articulent dans 8 métaprogrammes cités dans la partie 1.1. De plus, l'INRA possède de nombreux dispositifs scientifiques tels que des fermes expérimentales, des observatoires de recherche en environnement, des ressources génétiques etc. En accueillant des stagiaires, l'INRA remplit sa mission de formation via la recherche. Il permet aux étudiants de se former au sein de ses unités pour acquérir des compétences dans divers domaines. Le stagiaire est intégré à un programme de recherche au sein duquel il est en charge d'une mission précise. Les stagiaires contribuent ainsi à la production de connaissances en interne qui sont ensuite réutilisées au sein des programmes de recherche.

## **2 Etude bibliographique**

### **2.1 Projet de recherche**

Notre étude s'inscrit dans un projet de recherche conduit sur les plateaux de Bourgogne dans l'Yonne pour une durée de six ans. Ce projet est piloté par Terre Inovia (institut technique de la filière des huiles et protéines végétales) avec de nombreux partenaires tels que des coopératives locales, la chambre d'agriculture de l'Yonne, l'INRA (Figure 1). Il se concentre principalement sur la culture du colza qui sur ce territoire rencontre de nombreuses impasses techniques (Terre Inovia). Les ravageurs, comme l'altise d'hiver et le charançon, ont développé des résistances aux insecticides de la famille des pyréthrinoïdes et prolifèrent, entraînant des pertes économiques importantes (Terre Inovia). Le projet porte sur l'étude des systèmes de production de colza dans le territoire de l'Yonne. Il s'intitule : *restauration de la régulation naturelle et amélioration de la robustesse des cultures sur les plateaux de Bourgogne pour réduire durablement la dépendance aux insecticides (R2D2)*. Son objectif est de réduire l'utilisation des produits phytosanitaires en favorisant les processus de régulation naturelle. Cela passe notamment par une meilleure connaissance des auxiliaires, de leurs habitats et de leurs ressources alimentaires. Le stage s'inscrit dans ce projet avec pour objectif de déterminer la faisabilité d'utiliser les images satellites pour télédéetecter les colzas buissonnants. L'objectif à long terme de cette mission est de pouvoir remonter jusqu'aux pratiques des agriculteurs pour

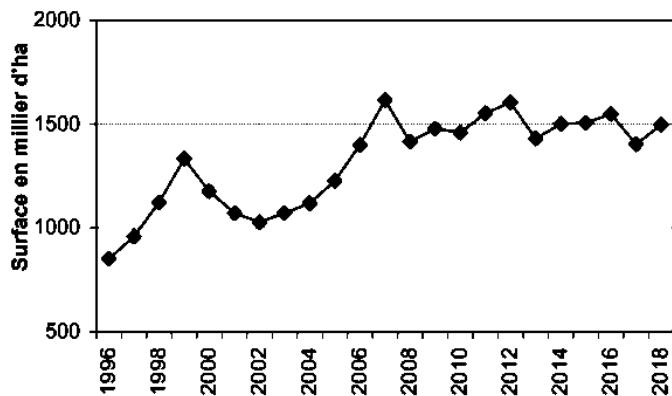


Figure 2: Evolution des surfaces de colza en France depuis 1996. (Source : Agreste Infos Rapides - Grandes cultures et fourrages - Février 2018 - n°2018 - 017).

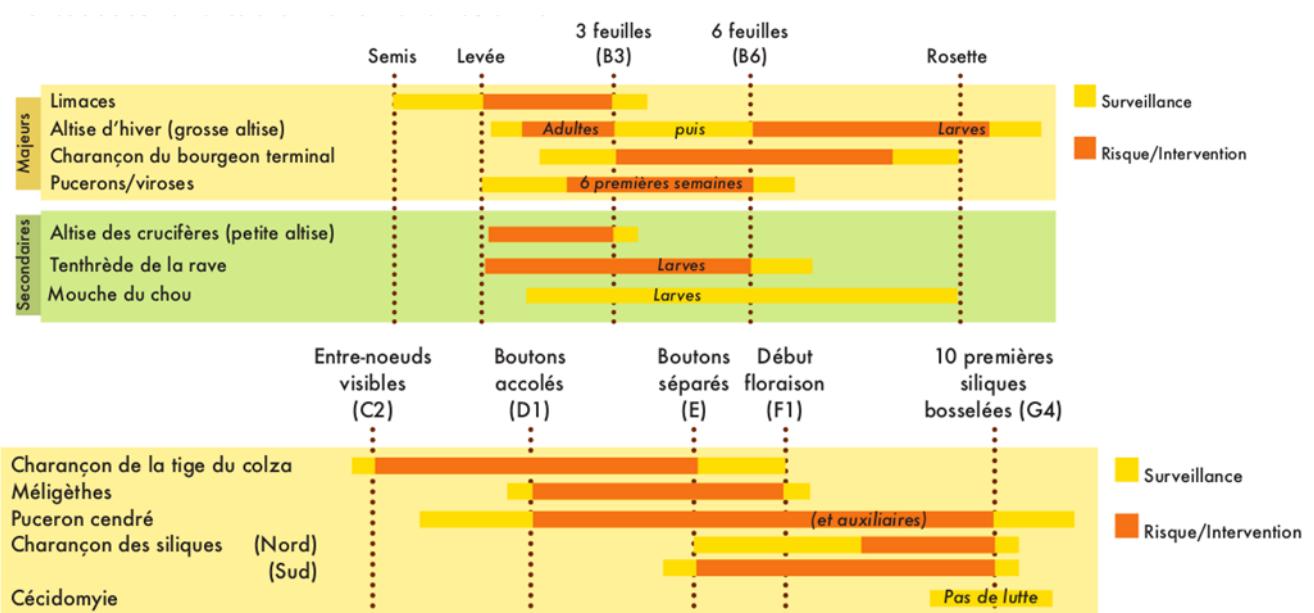


Figure 3 : Présentation des principaux ravageurs et des stades de sensibilités du colza. (Source : Guide de culture colza 2019)

comprendre les déterminants des attaques des ravageurs du colza. C'est une information essentielle pour être en mesure de conseiller les agriculteurs sur leurs pratiques, pour gérer aux mieux les ravageurs et réduire l'utilisation de produits phytosanitaires.

## 2.2 Le colza

### 2.2.1 Surface et utilisation

Le colza est une culture incontournable des plaines céréalières françaises. Depuis les années 90 les surfaces n'ont pas cessé d'augmenter (Figure 2, Agreste, 2018). Cependant, depuis quelques années, les surfaces semées diminuent. Sur la période 2014-2018 on assiste à une baisse de 18,8% des surfaces (Agreste, 2018). Selon les estimations de l'Agreste, il faut s'attendre à ce que cette tendance se confirme pour l'année 2019 avec une diminution des surfaces de plus de 28% par rapport à l'année 2018. Cette diminution est imputable aux mauvaises conditions de levées depuis quelques années liées au manque d'eau au moment des semis et à la pression des ravageurs tout au long du cycle de la culture (Agreste, 2019).

Le colza est utilisé dans la production d'huile alimentaire, de tourteaux à destination de l'élevage et de biocarburants (GNIS). Le colza peut aussi être une culture intéressante dans l'optique de diminuer la dépendance protéique française vis-à-vis des importations de soja. Le tourteau de colza possède une teneur en protéines de 39%, ce qui en fait un substitut potentiel aux tourteaux de soja utilisés pour l'alimentation animale (Brevault et al, 2016). Préserver les surfaces de colza est donc un enjeu stratégique pour la France qui arrive déjà, en partie grâce à ce dernier, à couvrir 55% des besoins en protéines végétales de ses élevages (Terre Inovia).

### 2.2.2 Une culture sensible aux ravageurs

Le colza est une culture sensible aux ravageurs, c'est ce qui explique le recours important aux produits phytosanitaires sur cette culture (IFT = 5.5). Il possède le plus haut indicateur de fréquence de traitement (IFT) en insecticides (Agreste, 2013), notamment en raison du grand nombre de ravageurs qui le menace tout au long de son cycle (Figure 3). En automne, les principaux risques proviennent des limaces entre le stade levée et le stade B3, puis des altises d'hiver de la levée au stade rosette. Le Charançon du Bourgeon Terminal (CBT) présente des risques du stade B3 au stade rosette. Ces deux derniers sont la cause principale du buissonnement des colzas. Au printemps, le charançon de la tige impacte le colza entre le stade C2 et le stade F1.



Les mélèthes sont à surveiller du stade D1 au stade F1 et le puceron cendré est nuisible du stade D1 au stade G4. Ces ravageurs constituent les principaux risques pour le colza (Terre Inovia). Cette recrudescence des ravageurs est due en partie aux rotations appliquées dans les bassins de productions céréaliers. Depuis les années 90, le colza a progressivement remplacé le pois et le tournesol dans les assolements français. Cette diminution du nombre d'espèces cultivées a engendré une simplification des systèmes de cultures, contribuant à l'augmentation de la pression des ravageurs (Schott, 2010).

Notre étude porte sur le port buissonnant du colza qui se caractérise par un changement morphologique de la plante. Le port buissonnant est majoritairement causé par le CBT qui pond ses œufs au cœur des tiges de colza de septembre à mars. Les larves vont se développer au cœur de la tige et se nourrir de cette dernière. Les pieds touchés par ces attaques deviennent jaunes et le port devient buissonnant (absence de bourgeon terminal). Les altises d'hiver, quant à elles, affaiblissent la plante, favorisant le buissonnement. Ensuite, le colza au port buissonnant présente un développement hétérogène et une fragilité accrue contre les autres ravageurs, ce qui engendre des pertes importantes de rendement. Le buissonnement du colza opère un changement morphologique très visible sur la plante ce qui implique qu'il est potentiellement détectable grâce aux images satellites.

## 2.3 Les données de télédétection

### 2.3.1 Concept général

La télédétection est utilisée en agriculture depuis de nombreuses années, son utilisation commence à se démocratiser grâce à la diminution du coût des acquisitions des données. Elle a été longtemps utilisée pour l'analyse et l'inventaire à grande échelle des cultures (Laguette et al, 1997). Maintenant, elle est de plus en plus utilisée à l'échelle de la parcelle pour diagnostiquer l'état du couvert végétal et y adapter les pratiques agricoles en fonction du couvert. La télédétection, associée aux données de positionnement et aux informations géographiques, permet de créer des données nécessaires à l'agriculture de précision (Huang et al, 2018) et d'adapter les interventions à l'intérieur d'une parcelle en fonction de son hétérogénéité (Seelan et al, 2003).

La télédétection consiste à collecter des informations sur un objet à distance à l'aide de capteurs optiques situés sur un satellite, un avion ou un drone. Ces capteurs optiques sont en mesure de capter les longueurs d'ondes émises par la terre, dans le domaine du visible et de

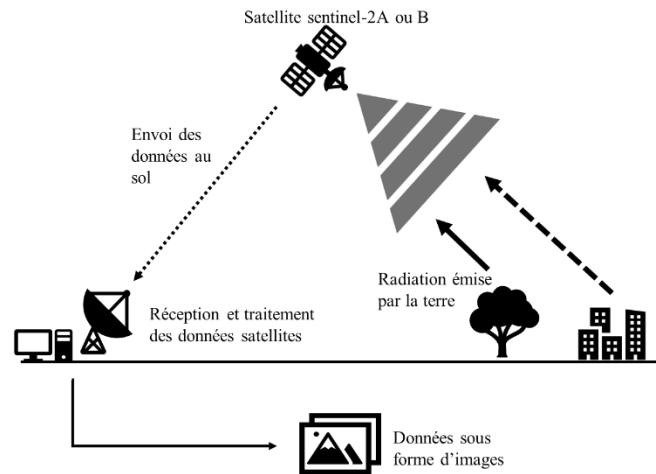


Figure 4: Schématisation des différents flux dans le cadre de la télédétection.

l'invisible. Ces mesures permettent de construire une image du paysage analysé (Richards, 2013). Chaque élément sur terre renvoie une longueur d'onde différente qui permet de l'identifier (Horning, 2008). La Figure 4 schématise le processus d'acquisition de données en télédétection.

Les données acquises par ces capteurs sont sous forme d'images qui sont accessibles à l'utilisateur sous format raster. Une image raster est composée de lignes et de colonnes qui forment une grille. Chaque case de cette grille est appelée pixel et contient une information. La résolution d'une image correspond au nombre de pixels que l'on trouve sur l'image, plus la résolution est élevée plus l'image possède d'informations. Les images possèdent plusieurs couches qui font référence à différentes "bandes" ou "canaux" qui eux-mêmes correspondent à différentes longueurs d'ondes. Le nombre de bandes dont dispose une image dépend du type de capteur utilisé par le satellite (Horning, 2008). La résolution des images et l'intervalle de temps entre la prise de deux images sur une même localisation se sont grandement améliorés ce qui fournit des données de plus en plus précises.

### 2.3.2 Données Sentinel-2

Les données sur lesquelles nous travaillons sont issues de la mission Sentinel-2. Cette mission fait partie du programme d'observation et de surveillance de la terre appelé Copernicus dirigé par l'agence spatiale européenne. La mission Sentinel-2 consiste en deux satellites conçus et produits par l'agence spatiale européenne, Sentinel-2a et Sentinel-2b. Ils sont en orbite autour de la terre à 786 km d'altitude et sont en mesure de prendre des images d'une largeur au sol de 290 km tous les 5 jours (CNES, 2017). La terre est divisée en carrés appelés tuiles de 110 km \* 110 km avec une superposition de 10 km avec les tuiles adjacentes. La résolution des images varie entre 10 et 60 mètres en fonction des bandes spectrales choisies. Les deux satellites embarquent un équipement multispectral (MSI) couvrant 13 bandes spectrales allant du visible au moyen infrarouge. Le MSI est constitué d'optique passive qui repose sur la réflexion de la lumière du soleil sur la surface de la Terre.

Les données fournies par la mission Sentinel-2 servent à cartographier l'occupation des sols, l'urbanisation, l'évolution des forêts, des zones humides et de la végétation. La mission Sentinel-2 s'articule autour de quatre grands thèmes :

- Changements des sols,
- Ressources hydriques,
- Cartographie d'urgence et des risques,
- Santé des plantes et phénologie.

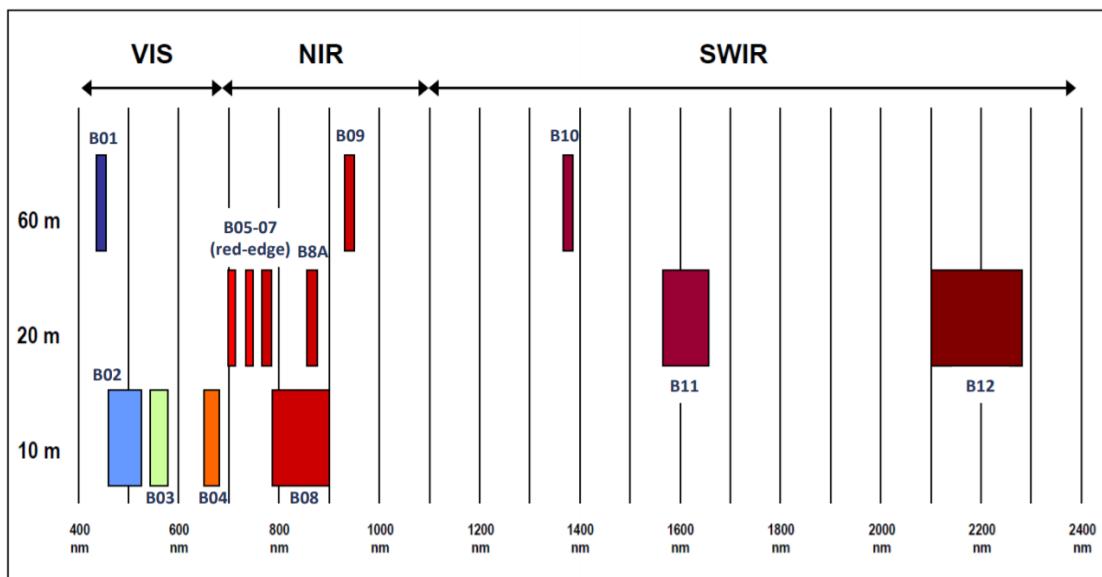


Figure 5 : Représentation des différentes bandes spectrales sentinel-2. VIS : Visible, NIR : proche infrarouge, SWIR : infrarouge cour. (Source : Gaetano, Le programme Copernicus et la mission Sentinel-2, 2018)

Notre étude s'inscrit dans ce dernier thème qui regroupe des données sur la végétation, la couverture humide, la chlorophylle acquise grâce aux fréquences rouges proche infrarouges (CNES, 2018). Les images sont prises avec un large spectre d'acquisition allant du visible au proche-infrarouge ce qui induit une résolution spatiale hétérogène. La Figure 5 nous montre les différentes bandes avec leurs longueurs d'ondes et leurs résolutions spatiales (Gaetano, 2018). Les bandes B1, B9 et B10 ont une résolution de 60 mètres. Elles servent à estimer les propriétés de l'atmosphère pour pouvoir ensuite corriger les images de ces mêmes effets qui, dans certaines conditions, bruent les données. (CESBIO). Les autres bandes sont en accès libre pour les utilisateurs.

## 2.1 Utilisation de la télédétection en agriculture

### 2.1.1 Les indices de végétation

En agriculture, l'utilisation de données satellites est le plus souvent couplée au calcul d'indices de végétation. Les indices de végétation sont des combinaisons de différentes bandes spectrales qui fournissent un renseignement sur le couvert végétal. Une multitude d'indices existent, ils permettent aussi bien de calculer la biomasse du couvert végétal que sa teneur en eau. L'indice le plus utilisé pour caractériser le couvert végétal est le Normalized Difference Vegetation Index (NDVI, Laguette et al, 1997). Le NDVI est une combinaison des réflectances dans les longueurs d'onde du Rouge et du Proche infrarouge. Cet indice est corrélé à la quantité de biomasse et à l'indice foliaire (Carlson et al, 1997). Il existe également le Green Normalized Difference Vegetation Index (GNDVI). Ce dernier est basé sur le NDVI, mais les longueurs d'onde du rouge y sont remplacées par les longueurs d'onde du vert (Gitelson, 1997). Le NDVI se trouve être sensible aux caractéristiques du sol et à l'effet de l'atmosphère (Rondeaux, 1996), c'est pour cela que certains indicateurs ont été développés pour pallier ces faiblesses : le Soil Adjusted Vegetation Index (SAVI, Huete et al, 1988), le Atmospherically Resistant Vegetation Index (ARVI, Kaufman and Tanre, 1992) ou encore le Soil and Atmospherically Resistant Vegetation Index (SARVI, Kaufman and Tanre, 1992) qui est la combinaison des deux précédents. Cependant le NDVI reste, encore à l'heure actuelle, l'indice le plus utilisé en télédétection, conséquence de la plus grande complexité des autres indices qui les rendent compliqués à mettre en œuvre. De plus leurs démonstrations ne sont pas suffisamment convaincantes ou n'améliorent pas réellement les résultats obtenus avec le NDVI (Rondeaux, 1996). D'autres indices tels que le Normalize Difference Water Index (NDWI) permettent de



renseigner sur la teneur en eau de la plante. Il se trouve être moins sensible aux variations atmosphériques que le NDVI. Le NDWI renseigne sur la teneur en eau de la plante. Il est moins sensible aux variations atmosphériques que le NDVI. Le NDWI se base sur les longueurs d'onde du rouge et du proche infrarouge. Le rouge correspond à la région d'absorption de la chlorophylle et le proche infrarouge à la réflectance du couvert végétal (Gao, 1996).

La liste des indices faite ici n'est pas exhaustive, elle a pour but de se familiariser avec les principaux indices et de montrer l'étendue des possibilités d'utilisation des indices de végétation.

### 2.1.2 Caractérisation du peuplement végétal à partir d'images satellites

La télédétection peut avoir une multitude d'applications dans le domaine agricole. Grâce aux images satellites, il est possible, entre autres, de cartographier l'état d'un peuplement végétal au travers d'indices de végétation. On peut notamment déterminer la sévérité d'un stress subi par les plantes ou l'état biochimique et biophysique du couvert végétal (Mulla et al, 2013). Grâce aux variations des spectres de la couverture végétale il est possible de détecter une infestation d'adventices sur une parcelle (Thorp, 2009). Les indices de végétation peuvent prédire l'évapotranspiration des cultures et ainsi permettre d'optimiser l'utilisation de l'irrigation (Glenn et al, 2011). Les indices, comme le GNDVI, peuvent aussi être utilisés pour estimer le rendement de maïs grain sous certaines conditions (Shanahan, 2001). Le NDVI peut même fournir des données sur les infestations des parcelles de blé par la mouche de Hesse à partir d'images satellites (Bhattarai et al, 2019). L'acquisition régulière de données satellites a un grand intérêt pour mener en temps réel le suivi des cultures et des bioagresseurs (Mulla et al, 2013). Sur le colza, le NDVI permet de mesurer la biomasse du colza qui est corrélée aux rendements finaux (Cowley, 2014 ; Huan et al, 2016). La télédétection est déjà largement utilisée pour prédire le rendement des céréales et du colza (Piekarczyk et al 2011, Serrano et al. 2000), mais elle l'est dans une moindre mesure pour la détermination des stades phénologiques. Une étude (Domínguez et al. 2015) s'est intéressée au NDVI pour prédire les stades phénologiques du blé d'hiver et du colza avec l'échelle Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH). Le NDVI peut aussi mettre en évidence les dommages des gelées sur le colza (She et al, 2015). La quantité de plantes touchées par le gel est obtenue en faisant la différence entre le NDVI avant et après les gelées. Il est aussi possible d'utiliser la télédétection pour prédire la pression des bio agresseurs sur un territoire (Sydha, 2018). La présence de maladies ou de ravageurs sur une culture entraîne des changements de pigmentation, de

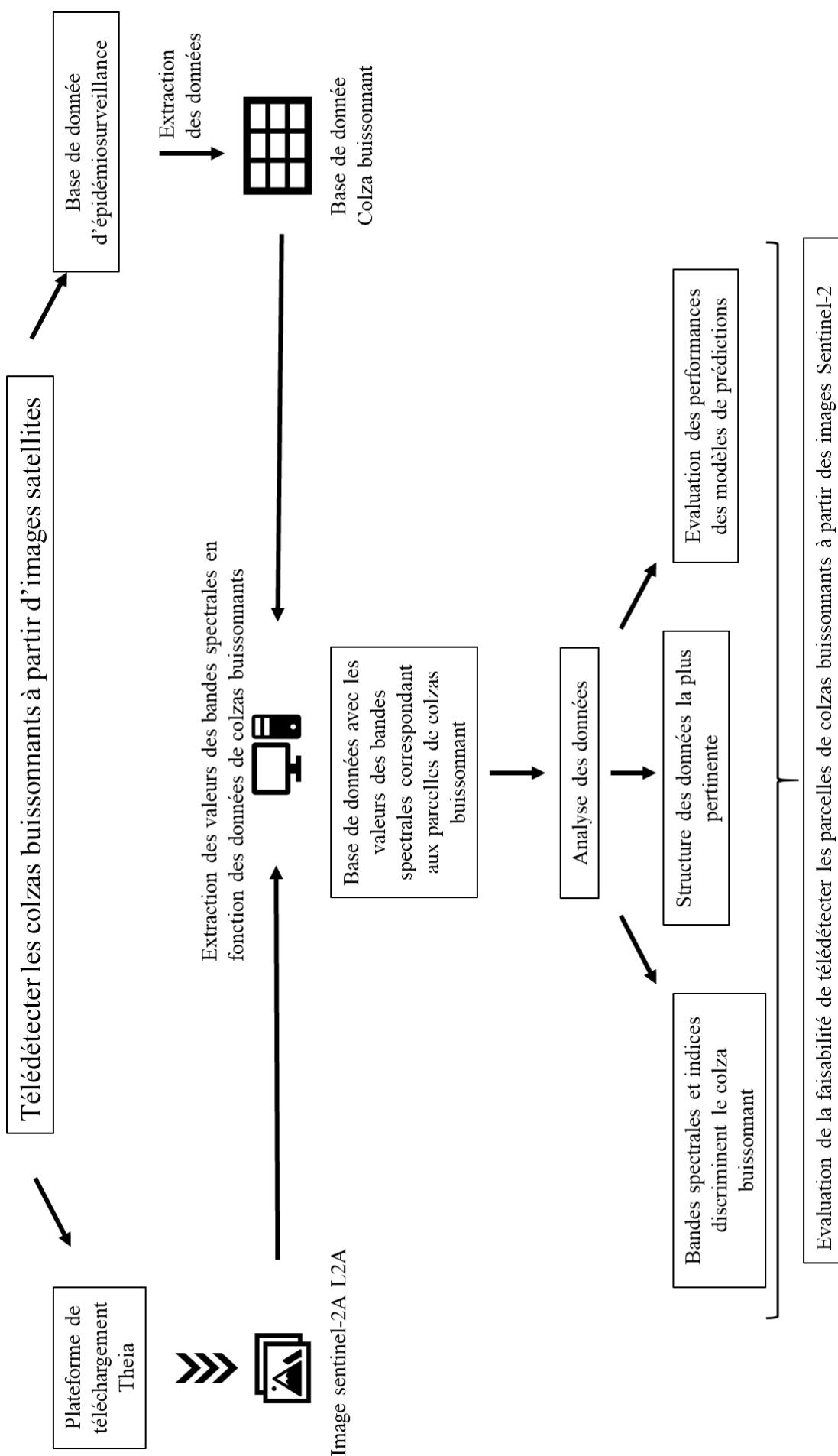


Figure 6 : Schéma de la démarche mise en œuvre pour répondre à la problématique

concentration en composés chimiques, modifie sa structure, l'absorption de l'eau et des nutriments (Huang, 2012). Tous ces changements se traduisent par des changements de couleur et de température du couvert, ce qui influence la réflectance émise par le couvert végétal. Ces changements de réflectance peuvent être perçus par la télédétection (Raikes and Burpee 1998). La télédétection est devenue, dans les pays développés, un moyen de surveillance des ravageurs pour pouvoir alerter en temps réel les agriculteurs pour certaines cultures (Acharya and Thapa, 2015). A ce jour, aucune étude ne mentionne la détection de colza buissonnant à partir d'images satellites.

### **3 Problématique**

Notre étude s'inscrit dans une démarche de plusieurs années qui a pour principal objectif la réduction de l'utilisation des produits phytosanitaires en grande culture. La télédétection est perçue comme un outil de collecte de données sur l'état sanitaire des cultures qui seront ensuite couplées aux pratiques des agriculteurs et aux contextes environnementaux. Ces informations pourront permettre de conseiller les agriculteurs sur les leviers à actionner pour contourner les dégâts de certains ravageurs.

Dans cette étude nous allons utiliser les données satellites pour caractériser le peuplement végétal du colza et déterminer dans quelle mesure les données sentinel-2 sont pertinentes pour détecter les parcelles de colzas buissonnantes.

Dans un premier temps nous étudierons les différents agencements possibles entre les bandes spectrales et les indices de végétation pour la télédétection des colzas buissonnantes. Il est fort probable qu'associer les bandes et les indices dans un même jeu de données permettent d'obtenir de meilleurs résultats grâce à un nombre plus important d'informations disponibles pour les modèles de prédiction. Dans un second temps nous observerons les variables explicatives qui semblent être en lien avec les colzas buissonnantes. Nous pouvons nous attendre à retrouver des indices de végétation en rapport avec la structure et la biomasse du couvert végétal. Ensuite nous effectuerons un travail de modélisation pour évaluer, dans l'état actuel de nos connaissances, les performances de prédiction auxquelles nous pouvons prétendre. Même si ces performances risquent d'être faibles, elles nous permettront d'évaluer la faisabilité de la télédétection des colzas buissonnantes. La méthodologie mise en œuvre pour répondre à cette problématique est schématisée dans la Figure 6.



## **4 Matériels et méthodes**

### **4.1 Organisation générale**

Premièrement nous réalisons une étude pilote sur un jeu de données de faible volume. L'objectif est de tester nos méthodes sur les bandes spectrales et les indices de végétation. Nous avons utilisé le jeu de données R2D2 comptant 16 points de données répartis dans l'Yonne. Une fois cette première étude réalisée, nous avons mis en œuvre les mêmes procédés sur la base de données Vigicultures® 2017. Nous disposons de 87 points répartis majoritairement dans le nord-est de la France. Ce sont sur ces données que porte notre étude.

Pour réaliser l'ensemble des opérations de traitement de données, allant du téléchargement à l'analyse en passant par la manipulation des données satellites, nous avons utilisé le logiciel R (3.4.4) ainsi qu'un serveur de calcul interne à l'UMR agronomie. Le logiciel R a été retenu pour traiter les données car il est maîtrisé au sein de l'unité. L'utilisation du serveur était indispensable pour la réalisation des calculs très volumineux.

### **4.2 Base de données**

#### **4.2.1 Données d'épidémosurveillance**

##### **4.2.1.1 *R2D2***

Les données du réseau R2D2 ont été fournies par Terre Inovia. Ce jeu de données comporte des comptages de plantes buissonnantes effectués le 18/04/2018 sur 16 parcelles du département de l'Yonne. Dans chacune des parcelles une zone précise a été choisie pour les relever, la zone de 50 m \* 20 m se trouve au minimum à 50 m du bord de la parcelle entre deux traces de pulvérisateur. Le comptage a été effectué sur 200 plantes prises aléatoirement dans la zone d'échantillonnage.

##### **4.2.1.2 *Vigicultures®***

Les données Vigicultures® regroupent des données d'épidémosurveillance sur l'ensemble de la France métropolitaine. Ce programme de surveillance a été mis en place par les instituts techniques (Arvalis, Terre Inovia, ITB). Son objectif est de surveiller la pression des ravageurs en temps réel. Il permet d'analyser les risques et de prévenir les agriculteurs au travers du Bulletin de Santé du végétal (BSV). Les BSV sont rédigés à 86% grâce aux données Vigicultures®, c'est donc une source d'information fiable alimentée depuis 2007 par des experts.



Figure 7 : Répartition spatiale des points Vigicultures®  
2017 pour le colza buissonnant.

Nous avons commencé par extraire de la base de données les observations concernant les colzas buissonnants. Ensuite, nous avons seulement sélectionné les données de l'année 2017 en ne gardant que les colonnes qui nous semblaient pertinentes pour notre analyse. Nous n'avions pas accès aux données 2018. Nous obtenons alors un nouveau jeu de données qui va nous servir de base pour notre travail. Nous avons accès au pourcentage de colzas buissonnants présents dans la zone d'échantillonnage. Nous disposons de 87 points, chacun possède des coordonnées GPS, le stade de la culture au moment de la mesure et la date de l'observation. Nos données sont situées majoritairement dans le nord-est de la France (Figure 7)

#### 4.2.1.3 *Vigicultures® et R2D2*

Pour pallier certains problèmes des données Vigicultures®, nous décidons de poursuivre nos investigations avec un jeu de données qui regroupe Vigicultures® et R2D2. Ce jeu de données contient davantage de points, notamment plus de parcelles comprenant des colzas buissonnants. Cependant pour R2D2 nous ne disposons que de données sur le mois de mai, nous ne gardons donc que les données de mai de Vigicultures® pour être cohérent. Nous obtenons alors 103 points de données avec pour chaque point le pourcentage de colzas buissonnants observé dans la parcelle.

#### 4.2.2 Données Sentinel-2

La mission sentinel-2 couvre l'ensemble des surfaces continentales et des littoraux avec une répétitivité de 5 jours, sans prendre en compte les jours nuageux. Les images satellites sont en accès libre sur différentes plateformes (Theia, Copernicus, peps). Nous avons fait le choix d'acquérir les images via la plateforme Theia qui propose un code python qui permet d'automatiser le téléchargement des données. Sur cette plateforme les images sont proposées au format L2A, ce qui inclut la détection des nuages et des ombres ainsi qu'une correction atmosphérique (CESBIO). Une fois téléchargé on obtient un fichier qui contient les 10 bandes spectrales (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12) au format raster (.tif). Ces 10 bandes sont données sous deux formes : une forme Surface Reflectance corrigée des effets atmosphériques et environnementaux (SRE\_Bx.tif), une autre forme, Flat Reflectance qui est en plus corrigé des effets de pentes (FRE\_Bx.tif). Nous allons travailler sur les données sentinel-2A L2A en utilisant les fichier raster FRE\_Bx.tif pour avoir des données les moins bruitées possible.



## **4.3 Création des variables de travail**

### **4.3.1 Variables explicatives**

#### **4.3.1.1 *Les bandes spectrales***

Pour récupérer les données qui nous intéressaient, nous avons mis en place une chaîne de traitement qui comporte trois étapes. La première étape consiste à récupérer pour chacun des points le nom de la tuile Sentinel-2 correspondante. Une fois le nom des tuiles connu nous les avons téléchargées sur une certaine période. Cette période correspond au moment où l'on peut observer les colzas buissonnantes. Pour les données R2D2 nous avons téléchargé la tuile de l'Yonne à la date où les relevés ont été effectués le 18/04/2018. Pour Vigicultures® les prises d'échantillons étant étalées sur une plus grande période, nous nous sommes centrés sur les mois d'avril et de mai. Dans un troisième temps, nous avons créé un polygone de 30 m \* 30 m autour de chaque point GPS pour Vigicultures® et de 50 m \* 20 m pour R2D2. Pour chaque polygone ainsi formé, nous avons extrait pour l'ensemble des bandes spectrales leurs valeurs moyennes sur la tuile correspondant au lieu échantillonné. Nous avons réalisé cette opération pour l'ensemble des tuiles de la période voulue. La valeur des bandes spectrales correspond à une intensité lumineuse réfléchie par le sol.

Les données récupérées sont ensuite nettoyées pour éviter au maximum de travailler sur des données bruitées, sachant que les tuiles téléchargées ont déjà subi un traitement contre certains bruits. La présence de nuages, de brouillard ou d'ombre fausse les valeurs, nous essayons de nous en débarrasser en refusant l'ensemble des valeurs négatives. Ce traitement n'est que partiel mais permet d'éliminer les valeurs les plus aberrantes. De plus une pré sélection a été effectuée au moment du téléchargement où nous avions seulement téléchargé les tuiles possédant moins de 50% de couverture nuageuse.

Pour simplifier l'exploitation des données nous décidons pour chaque point de moyennner par mois la valeur de ces bandes. Il y avait trop de valeurs manquantes pour pouvoir réaliser un découpage à la semaine ou aux 15 jours.

Nous obtenons à la fin pour chaque point R2D2 la valeur moyenne pour les différentes bandes spectrales sur la zone échantillonnée. Concernant Vigicultures®, chaque point possède une valeur par bande et par mois. Les valeurs des bandes correspondent aux moyennes de ces dernières sur les zones échantillonnées.



#### *4.3.1.2 Les indices de végétation*

Nous avons rajouté 5 indices de caractérisation de la végétation, ces derniers sont des combinaisons des bandes spectrales Sentinel-2 (Sentinel 2 EO products). Nous avons choisi les indices les plus couramment utilisés en agriculture et qui sont facilement calculables à partir des données Sentinel-2.

##### Normalized Difference Vegetation Index (NDVI)

Le NDVI est l'indice de végétation le plus connu et le plus utilisé pour quantifier la végétation. Il est calculé à partir des longueurs d'onde rouge et proche infrarouge. Il tient compte de la taille des feuilles et de la teneur en chlorophylle des plantes. Les valeurs de cet indice s'étendent de -1 à 1. Les valeurs aux alentours de -1 correspondent à de l'eau. Les valeurs proches de 0 correspondent à de la roche, un sol nu ou de la neige, et les valeurs proches de 1 représentent une zone de végétation dense.

Formule Sentinel-2 :

$$NDVI = (B8 - B4) / (B8 + B4)$$

##### Green Normalized Difference Vegetation Index (GNDVI)

Le GNDVI reprend les caractéristiques du NDVI mais a été modifié pour être plus sensible à la teneur en chlorophylle des plantes. Il est calculé avec les longueurs d'ondes du proche infrarouge et du vert dans le domaine du visible.

Formule Sentinel-2 :

$$GNDVI = (B8 - B3) / (B8 + B3)$$

##### Normalized Difference Water Index (NDWI)

Il possède deux formules qui permettent de mettre en avant des informations différentes. Le NDWI est un indice qui renseigne majoritairement sur la teneur en eau des plantes, il utilise les bandes proches infrarouge et infrarouge, tandis que le NDWI (2) est à l'origine utilisé pour se renseigner sur les masses d'eau de surface (lac, rivière...).

Formule Sentinel-2 :

$$NDWI = (B8 - B11) / (B8 + B11)$$

$$NDWI(2) = (B03 - B08) / (B03 + B08)$$

##### Moisture Index

Cet indice mesure l'humidité et se base en partie sur le NDWI, il utilise une plage de l'infrarouge plus restreinte.

Formule Sentinel-2 :



$$\text{Moisture Index} = (B8A - B11)/(B8A + B11)$$

Nous calculons ces indices pour le mois d'avril et le mois de mai, ils viennent enrichir notre base de données. Nous obtenons alors 15 variables explicatives pour R2D2 et 30 pour Vigicultures®.

#### *4.3.1.3 Variable explicative centrée réduite*

Sur le jeu de données regroupant Vigicultures® et R2D2, nous centrons et réduisons chaque bande spectrale selon la formule suivante :  $\frac{X-\mu}{\sigma}$ . La valeur des bandes spectrales correspond à l'intensité lumineuse réfléchie par la terre. Cette intensité peut être influencée par de nombreux paramètres. Pour niveler ces variations nous réalisons une opération sur l'ensemble des bandes. Nous moyennons pour chaque point l'intensité centrée réduite de l'ensemble des bandes spectrales. Nous obtenons l'intensité moyenne de chaque point que l'on appelle "lumin". Ensuite nous créons de nouvelles variables avec la nomenclature *lumin\_BX* (ou X correspond aux numéros d'une bande spectrale), pour chaque point nous retirons à la valeur de chacune de ces bandes centrées réduites, la luminosité moyenne centrée réduite :  $\text{lumin}_BX = BX - \text{lumin}$ . Nous obtenons alors 11 variables explicatives transformées : les 10 *Lumin\_BX* et la variable *Lumin*. Ces nouvelles variables prennent en compte les différences de luminosité qui peuvent exister entre deux points. En effet, on peut supposer que la luminosité est différente entre l'est et l'ouest de la France. Ainsi avec ces nouvelles variables nous pouvons comparer les différences d'intensité qui sont causées majoritairement par le couvert végétal, cette opération contribue à réduire le bruit des données. C'est une nouvelle manière d'observer nos données et cela va peut-être faire apparaître des liens importants avec le pourcentage de colzas buissonnantes.

### 4.3.2 Variables à expliquer

#### *4.3.2.1 Pourcentage de colzas buissonnants*

A partir des données de colzas buissonnants, nous avons obtenu une variable du pourcentage de colzas buissonnants sur les parcelles échantillonnées. Cette variable était déjà présente dans le jeu de données Vigicultures®. Pour R2D2, nous l'avons calculé à partir du nombre de plantes total échantillonnées et du nombre de plantes buissonnantes détectées.

#### *4.3.2.2 Colzas buissonnants ou non buissonnants*

Toujours à partir des données de colzas buissonnants nous avons créé une variable binaire, 1 pour buissonnant et 0 pour non buissonnant. Elle suit la règle de décision suivante :



si le nombre de plantes buissonnantes est supérieur à la médiane alors la variable prend la valeur 1 sinon elle prend la valeur 0.

## 4.4 Analyses statistiques

### 4.4.1 LASSO

La méthode LASSO est dérivée du modèle linéaire généralisé (GLM). C'est un algorithme de pénalisation qui sélectionne automatiquement les variables les plus pertinentes dans le cadre de la régression. Cette méthode est mise à disposition via le package R glmnet (V2.0). Il fournit deux coefficients, le lambda.min et le lambda.1se, nous privilégierons le second qui est plus sévère. Toutefois si ce dernier ne sélectionne aucune variable nous nous octroierons le droit de considérer le lambda.min. Ces coefficients déterminent le nombre de variables sélectionnées par le modèle et nous retourne leurs noms. Le fait que le LASSO sélectionne des variables signifie que les résultats de la régression ne sont pas aléatoires et qu'il existe bien un lien dans nos données. La validation croisée utilisée par le LASSO pour sélectionner les variables garantit la fiabilité de la sélection. Pour chaque lambda.1se ou lambda.min le LASSO fournit également un pourcentage de déviance. Plus ce pourcentage est faible plus le LASSO est juste. On peut utiliser le test de Fisher pour valider le lien entre valeurs prédites par le LASSO et les valeurs observées dans le cadre d'une classification binaire (matrice de confusion). La p-value du test nous fournit la significativité du lien ( $p\text{-value} < 0.05$ ). Le odds ratio du test de Fisher nous fournit l'importance de l'effet observé dans nos données, plus il est important plus l'effet est grand.

### 4.4.2 Random forest

La random forest est un algorithme rapide, robuste et stable. Il repose sur le principe des arbres de régressions et de classifications. Nous avons utilisé le package R randomForest (v 4.6). Le critère de segmentation de la random forest repose sur l'indice d'impureté de Gini. La Random Forest génère un grand nombre d'arbres de classifications de manière à réaliser une prédition. Une multitude de modèles peu fiables sont combinés pour aboutir à un modèle robuste. Lorsque l'on obtient un modèle suffisamment fiable, il est possible d'injecter de nouvelles données qui vont passer par tous les arbres, chaque arbre vote pour la classe de sortie ce qui aboutit à une prédition. La Random Forest fournit l'importance des variables, ce qui correspond à la diminution moyenne de l'impureté apportée par chaque variable.

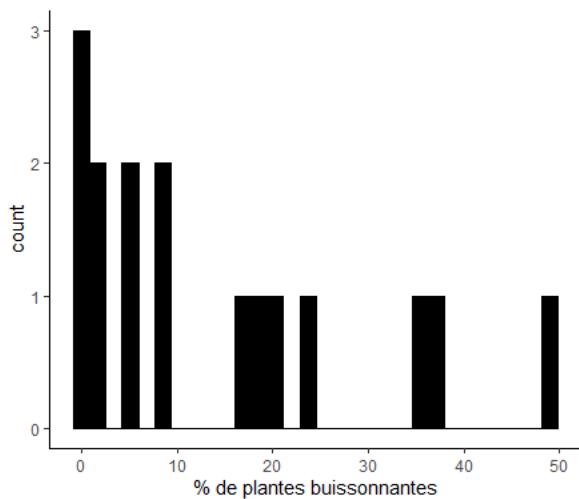


Figure 8 : Distribution du pourcentage de colza buissonnant des données R2D2 2018. La ligne pointillée symbolise la moyenne.

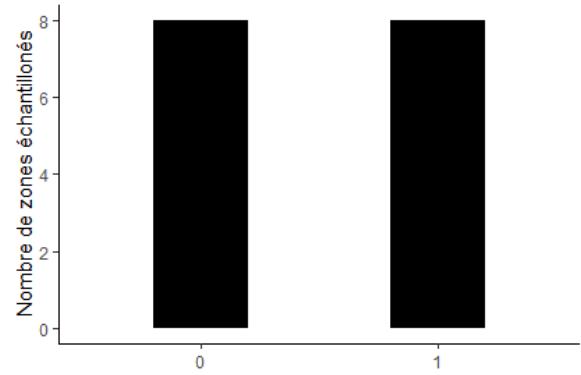


Figure 9 : Distribution du colza buissonnant. Le 0 signifie non buissonnant et le 1 buissonnant. Les données sont classées selon la médiane du pourcentage de plantes buissonnantes.

LASSO	Données	
	Bandes	Indices
Deviance	0.1759	0.2715
Df	2	3
Variable	B8 B11	NDWI GNDVI NDVI

Figure 10 : Résultats LASSO sur les bandes spectrales et les indices des données R2D2. Df signifie degrés de liberté (degree of freedom).

Random Forest	Données	
	Entrainement	
Kappa	0.25	
accuracy	0.63	
Variable	B11 NDWI B8	

Figure 11 : Résultats de la random forest avec arbre de classifications sur les bandes et les indices réunis.

L'impureté correspond aux éléments mal classés par le modèle. Cela permet de déterminer quelle variable contribue le plus au modèle.

Nous pouvons également déterminer des ratios tels que le Kappa et l'Accuracy pour évaluer la qualité de nos modèles. L'Accuracy correspond au pourcentage de valeur correctement classé par le modèle. Ce ratio est des plus utiles dans une classification binaire, cependant dans une classification à plusieurs classes il ne montre pas comment se décompose l'Accuracy pour chacune des classes. Le Kappa quant à lui reprend le principe de l'Accuracy mais prend en outre en compte le fait que certaines classifications sont prédites sur la base d'incertitudes (Brownlee, 2016). Plus il est proche de 1 plus le modèle réalise de bonnes prédictions.

## 5 Résultats

### 5.1 Etude pilote : R2D2

L'étude pilote est réalisée sur les données issues du projet R2D2, il comporte 16 observations. La figure 6 représente la distribution du pourcentage de plantes buissonnantes. En moyenne, il y a 14,5% de colzas buissonnants par zone échantillonnée. La majorité des zones échantillonnées comportent entre 0 et 15 % de colzas buissonnants (Figure8), il y a donc plus d'observations de parcelles à faible pourcentage de colzas buissonnants (0%). Ce pourcentage a été transformé en variable binaire, 0 pour non buissonnant (% de colza buissonnant < médiane) et 1 pour buissonnant (% de colza buissonnant > médiane) nous pouvons voir sa distribution sur la Figure 9. Cette dernière est parfaitement équilibrée entre buissonnant (1) et non buissonnant (0).

Nous avons couplé ces données à celles extraites de la tuile Sentinel-2 correspondant au jour de l'observation des parcelles. Ainsi nous avons 10 bandes spectrales et 5 indices calculés à partir de ces bandes. Les colzas buissonnants possèdent une structure particulière qui pourrait permettre sa détection grâce aux images satellites. Nous cherchons ici à mettre en évidence des bandes spectrales et des indices de végétations qui pourraient discriminer les colzas buissonnants. Pour ce faire, nous utilisons le LASSO avec un modèle binomial, les résultats sont présentés dans la Figure 10. Le LASSO effectué sur les bandes à une déviance plus faible que celui effectué sur les indices, le jeu de donné étant petit nous ne pouvons rien conclure. Cependant deux bandes spectrales sont sélectionnées, la bande B11 et B8 ainsi que trois indices, le NDWI, le GNDVI et le NDVI. Le LASSO sélectionnant certaines variables cela signifie que la télédétection est potentiellement en mesure de détecter les colzas buissonnants.

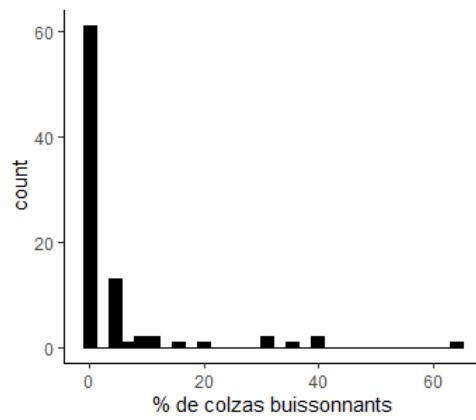


Figure 12 : Distribution du pourcentage de colzas buissonnantes. Graphique réalisé à partir des données Vigiculture 2017.

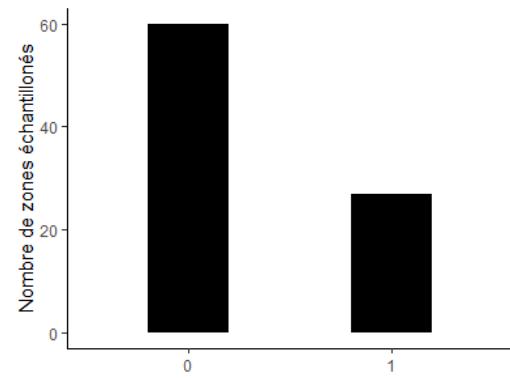


Figure 13 : Distribution du colza buissonnant. L'axe des abscisses prend la valeur 0 pour les colzas non buissonnantes et la valeur 1 pour les colzas buissonnantes. Les données sont classées selon la médiane du pourcentage de plantes buissonnantes. Graphique réalisé à partir des données Vigicultures® 2017.

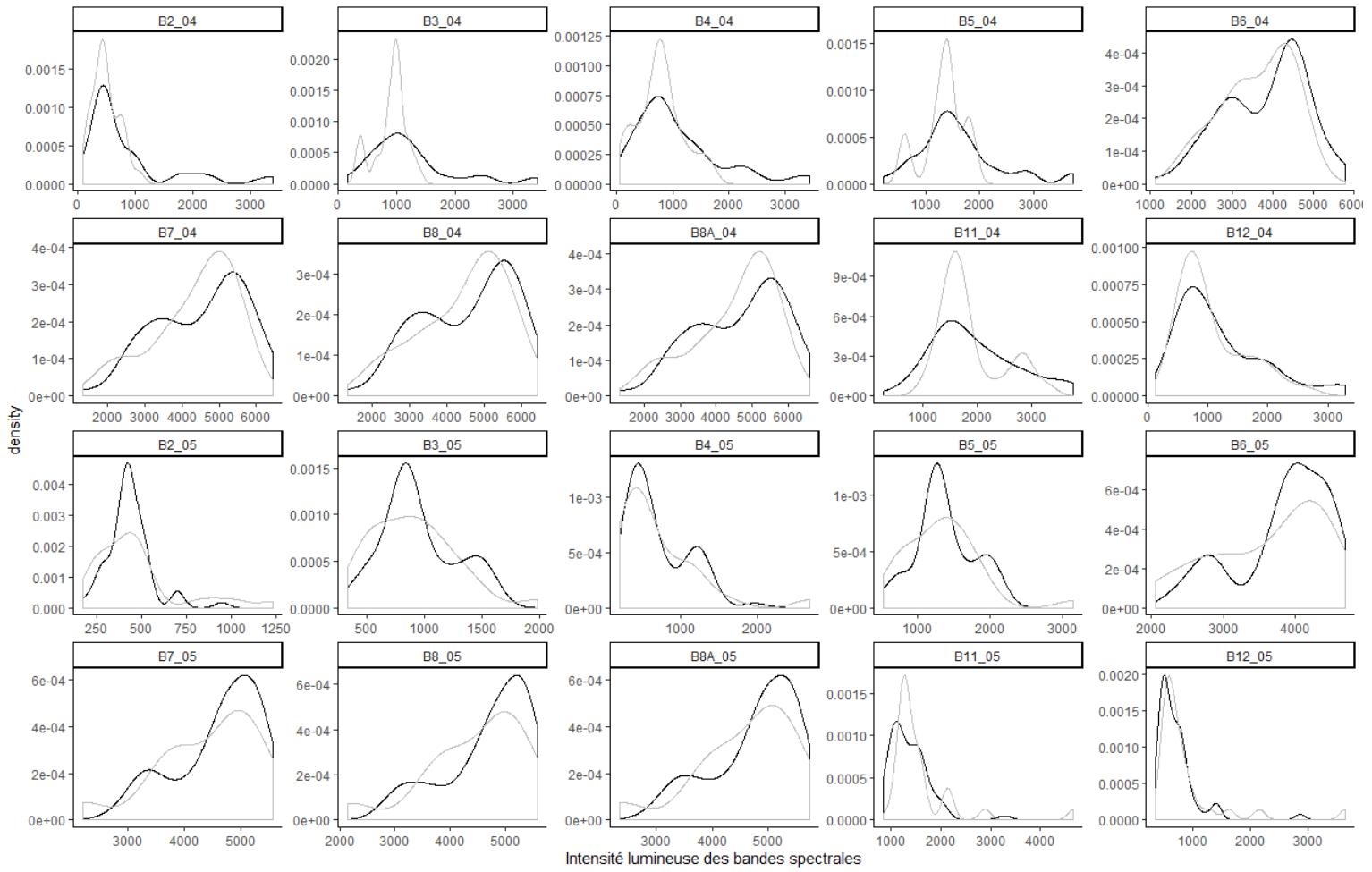


Figure 14 : Distribution des valeurs des bandes spectrales. Les courbes noires représentent les parcelles avec du colza non buissonnant. Les courbes grises correspondent aux parcelles avec du colza buissonnant. La nomenclature utilisée pour les bandes spectrales est Bxx\_mm, où xx correspond aux numéros de la bande et mm au mois d'acquisition des données satellites.

Notre objectif est d'aboutir à un modèle pour détecter les colzas buissonnantes, nous allons réaliser une Random Forest sur l'ensemble des 15 variables explicatives. La Figure 11 nous présente les résultats. L'accuracy est seulement de 0.63 et le kappa de 0.25. Les performances de la Random Forest restent faibles ce qui peut être imputé au faible nombre de données. Nous pouvons donc raisonnablement penser que lorsque l'on va passer sur les données Vigicultures® la prédiction du modèle sera améliorée. Lorsque l'on regarde les variables les plus importantes pour le modèle, on retrouve la bande B11, le NDWI et la bande B8. Ces résultats sont cohérents avec ceux obtenus ci-dessus.

## 5.2 Vigicultures®

### 5.2.1 Analyse descriptive

Les données Vigicultures® comptent 87 points répartis majoritairement dans le nord-est de la France. La variable à expliquer est le pourcentage de plantes buissonnantes dans la zone échantillonnée. La Figure 12 représente la distribution de cette variable. Nous observons que la majorité des parcelles ne présente aucune plante avec un port buissonnant. La moyenne se situe à 4% de plantes buissonnantes par zone échantillonnée avec un maximum à 64%. Notre jeu de données présente un certain déséquilibre en faveur des colzas non buissonnantes.

Nous avons associé au donnée Vigicultures® une variable binaire avec d'un côté les parcelles avec du colza buissonnant et de l'autre sans colza buissonnant. La Figure 13 nous montre sa distribution. Là encore nous observons un déséquilibre entre le nombre de parcelles avec port buissonnant et celles sans port buissonnant. Ce déséquilibre pourra potentiellement impacter nos futurs résultats.

Les données Vigicultures® ont été associées aux images Sentinel-2 qui sont composées de 10 bandes spectrales. La figure 14 représente la distribution des longueurs d'ondes de chaque bande pour les mois d'avril et de mai. Les courbes grises représentent les colzas buissonnantes et les courbes noires les colzas non buissonnantes. Nous observons de grosses différences pour les bandes : B3\_04 et 05, B4\_04, B5\_04 et 05, B11\_04 et 05 ainsi que pour B2\_05. Certaines bandes spectrales semblent plus pertinentes que d'autres pour télédétecenter les colzas buissonnantes.

Les indices de végétation sont souvent utilisés en télédétection. Pour les besoins de notre étude nous en avons utilisé cinq. Sur la Figure 15 nous comparons leurs valeurs entre les parcelles ayant du colza buissonnant (courbes grises) et celles n'en ayant pas (courbes noires).

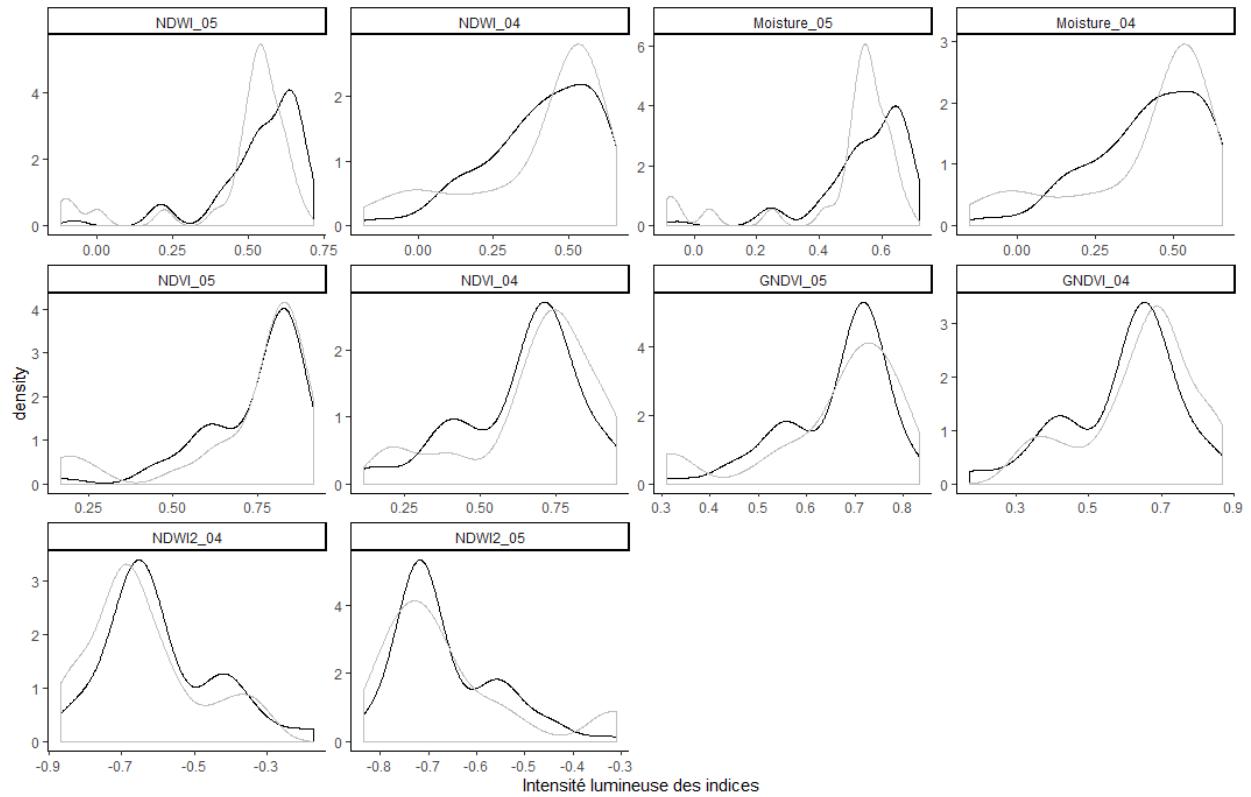


Figure 15 : Distribution des valeurs des indices de végétation. Les courbes noires représentent les parcelles avec du colza non buissonnant. Les courbes grises correspondent aux parcelles avec du colza buissonnant. La nomenclature utilisée pour les indices de végétation est Indice\_mm, où mm correspond au mois d'acquisition des données satellites qui ont servi à calculer l'indice.

LASSO	Association des données		
	Bandes	Indices	Bandes + Indices
Deviance	0.3168	0.4567	0.3907
Df	1	4	2
Variable	B11_05	NDWI_05 Moisture_04 GNDVI_04 NDWI2_04	B11_05 NDWI_05

Figure 16 : Tableau récapitulatif des LASSO effectués sur les 3 associations des variables explicatives. Df signifie degrés de liberté (degree of freedom). La ligne "variable" correspond aux variables sélectionnées.

Random Forest	Association des données		
	Bandes	Indices	Bandes + Indices
Kappa	0.35	0.35	0.44
accuracy	0.72	0.72	0.75
Class error Buissonnant	0.50	0.50	0.41
Class error non buissonnant	0.16	0.16	0.16
Variable	B11_05 B12_05	Moisture_05 Moisture_04 NDWI_5 NDWI_04	B11_05 NDWI_05 Moisture_05

Figure 17 : Tableau récapitulatif des Random Forest effectués sur les 3 associations des variables explicatives. Df signifie degrés de liberté (degree of freedom). La ligne "variable" correspond aux variables sélectionnées.

Nous remarquons que deux indices ressortent, le NDWI et le Moisture Index. Ces deux indices seront potentiellement utiles pour détecter les colzas buissonnants.

### **5.2.2 Exploration de l'association des variables et de leurs importances**

#### *5.2.2.1 Modèle LASSO*

Nos données sont composées de 10 bandes spectrales et de 5 indices de végétation. Nous allons dans un premier temps chercher à déterminer quel arrangement des variables explicatives entre elles est en mesure de discriminer au mieux les colzas buissonnants. Pour ce faire nous utilisons la méthode LASSO avec un modèle binomial. Le tableau de la Figure 16 présente les résultats des trois LASSO. Il s'avère que la déviance la plus faible est de 0.31 pour le LASSO qui utilise uniquement les bandes. L'association des bandes et des indices présente également un résultat correct (déviance = 0.39). Ce premier résultat nous conduit à penser que l'utilisation des bandes seules ou des bandes et des indices combinés peuvent permettre la détermination des colzas buissonnants.

Nous cherchons à déterminer quelles variables explicatives ont le plus d'importance pour la télédétection des colzas buissonnants. La dernière ligne du tableau de la Figure 16 nous présente les variables sélectionnées dans le cadre des LASSO. On remarque que l'association bande + indices sélectionne des variables qui étaient sélectionnées lors du LASSO sur les bandes et les indices seuls. On retrouve donc la bande B11\_05 et le NDWI\_05. Pour les Indices on retrouve également le Moisture\_04, le GNDVI\_04, et le NDWI2\_04. Pour résumer la bande B11\_05 semble relativement importante tout comme le NDWI\_05 car ils apparaissent deux fois dans les variables sélectionnées. Il semble donc que toute les bandes et les indices n'ont pas la même habilité à détecter le colza buissonnant.

#### *5.2.2.2 Modèle Random Forest*

Pour confirmer ces premiers résultats nous réalisons la même démarche avec la random forest. Nous cherchons à déterminer l'association des données qui permet les meilleurs résultats en comparant le résultat des random forest. La Figure 17, nous présente les résultats obtenus, l'utilisation des bandes ou des indices seuls donnent les plus mauvais kappa (0.35) et accuracy (0.72). Il semble qu'associer les bandes et les indices permettent, dans le cadre de la random forest, d'obtenir les meilleurs résultats avec un kappa de 0.44 et une accuracy de 0.75. Cela se confirme au vu des erreurs de classifications, notamment concernant les colzas buissonnants où l'on passe de 0.5 pour les bandes et indices seuls à 0.41 pour la combinaison des deux. En se basant sur la random forest, associer les bandes et les indices semble donc être pertinent.

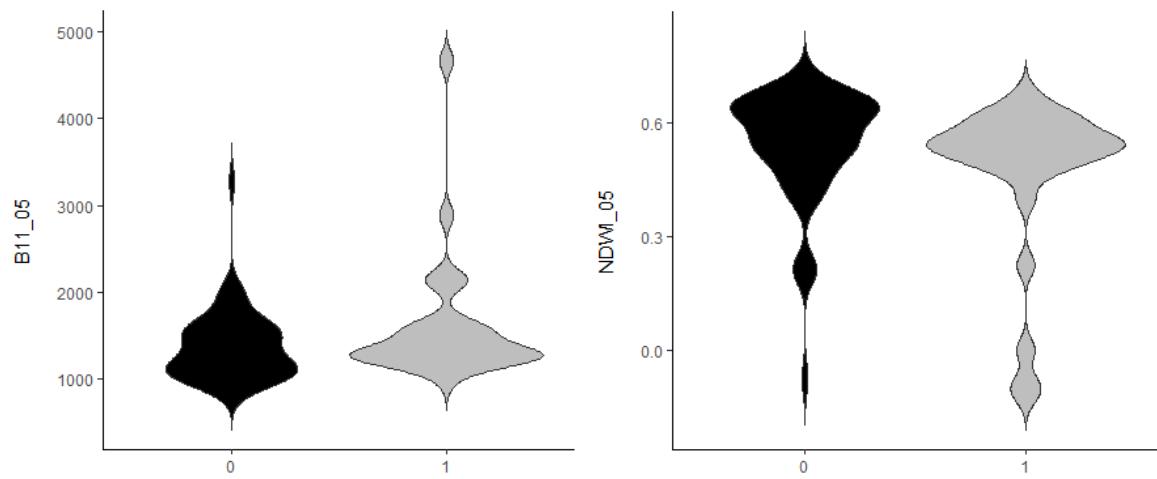


Figure 18 : Violine plot de la bande spectrale B11\_05 et du NDWI\_05. Sur l'axe des abscisses le 0 correspond au colza non buissonnant et le 1 au colza buissonnant.

Random Forest	Données	
	Entrainement	Validation
Kappa	0.44	0.30
accuracy	0.75	0.72
Class error Buissonnant	0.41	0.60
Class error non buissonnant	0.16	0.18
Variable	B11_05	-
	NDWI_05	-
	Moisture_05	-

Figure 19 : Résultat de la random forest à arbre de classifications sur les données Bandes + Indices.

Nous nous intéressons maintenant aux variables les plus importantes dans la construction des arbres de classifications des random forest. La Figure 17 nous présente les variables les plus importantes pour chaque jeu de données. On retrouve une nouvelle fois la bande B11\_05 et le NDWI\_05 qui apparaissent à deux reprises et notamment dans le meilleur modèle (Bandes + Indices). Nous remarquerons également l'apparition d'un nouvel indice, le Moisture Index qui apparaît deux fois dans les résultats. Il semble donc que la bande B11\_05 et le NDWI\_05 soient importants pour télédétecer les colzas buissonnantes. Il semble également que l'accumulation de variables explicatives permettent d'augmenter la prédition. L'ajout de variables supplémentaires sera sans doute nécessaire pour augmenter la précision des modèles.

#### *5.2.2.3 Variables discriminant les colzas buissonnantes*

Nous allons regarder plus en détail la bande B11 et le NDWI de mai, car ce sont les variables qui ressortent le plus au travers des analyses précédentes. La Figure 18 nous permet d'observer la distribution des valeurs de la bande B11 et du NDWI pour les colzas non buissonnantes (0) et buissonnantes (1). Visuellement on remarque que les distributions sont relativement similaires, même si pour les colzas buissonnantes les valeurs semblent plus centrées autour de la moyenne. La différence entre les colzas buissonnantes et non buissonnantes semble avoir un impact minime sur la bande B11 et le NDWI. Il va donc vraisemblablement falloir combiner plusieurs variables entre elles pour aboutir à un modèle le plus fiable possible.

### 5.2.3 Prédiction avec random forest

#### *5.2.3.1 Variable buissonnant binaire*

Nous avons maintenant une vision plus précise des variables discriminantes pour le colza buissonnant, nous allons pouvoir essayer de créer un modèle capable de détecter la présence de colza buissonnant. Pour réaliser cela, nous allons utiliser l'algorithme de la Random Forest utilisant des arbres de classifications. Les résultats sont présentés dans la Figure 19. L'accuracy et le kappa sont tous les deux de 0.44 et à 0.75 sur les données d'entraînement et diminuent pour passer à 0.3 et 0.72 lors de la validation. Par ailleurs on observe une nette augmentation de l'erreur de classification des colzas buissonnantes. La faible performance du modèle sur le jeu de validation est due au faible nombre de données qui le compose. On dispose d'un jeu de validation d'un total de 17 points correspondant à des colzas non buissonnantes et 5 points pour les colzas buissonnantes. Les performances de notre premier modèle ne permettent pas de détecter de manière fiable les colzas buissonnantes. L'ajout de variables autres que les bandes et les indices pourrait être une solution.

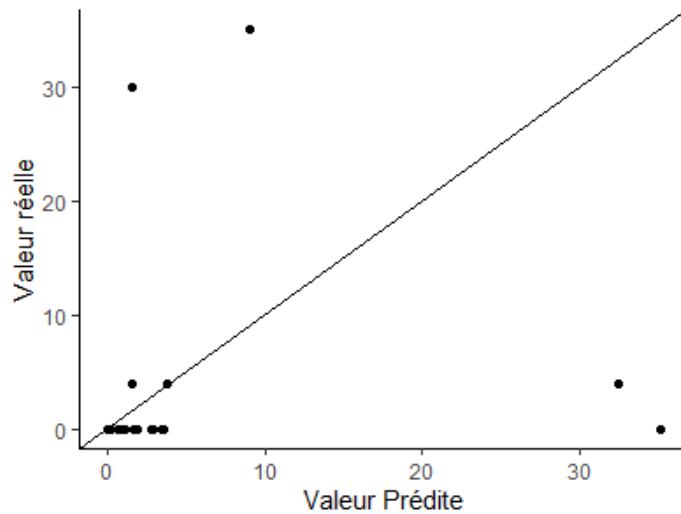


Figure 20 : Valeur réelle en fonction des valeurs prédites par la random forest avec arbre de régression. La droite symbolise la répartition des points dans le cadre d'un modèle parfait.

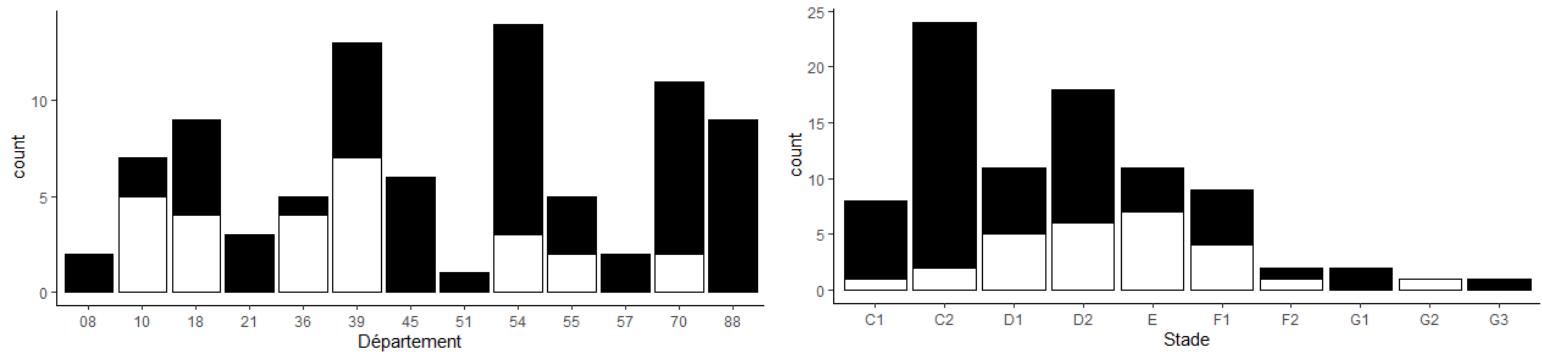


Figure 21 : Histogramme du nombre de parcelles buissonnantes par département et par stade au moment de l'observation. La partie noire de l'histogramme représente le colza non buissonnant et la blanche le colza buissonnant.

Random Forest	Données	
	Entrainement	Validation
Kappa	0.52	0.48
accuracy	0.77	0.82
Class error Buissonnant	0.32	0.40
Class error non buissonnant	0.16	0.12
Variable	Departement	-
	B11_05	-
	Stade	-
	Moisture_05	-
	NDWI_05	-

Figure 22 : Résultat de la random forest à arbres de classifications sur les données regroupant les bandes, les indices, les départements et les stades du colza.

### *5.2.3.2 Variable de pourcentage de plantes buissonnantes*

Dans le cadre du projet il serait aussi intéressant de pouvoir détecter un gradient de plantes buissonnantes pour à terme adapter au mieux les pratiques agronomiques. Nous allons chercher à déterminer le pourcentage de buissonnements du colza. Nous allons utiliser une random forest à arbre de régression sur les indices et les bandes. Nous obtenons un modèle qui explique 43% de la variance. La Figure 20 nous montre la différence entre les valeurs prédictives et les valeurs réelles. On observe une importante différence, dans le meilleur des cas nous devrions avoir une droite de la forme  $y = x$ . La prédiction du pourcentage du colza buissonnant semble donc peu évidente à partir des données sur lesquelles nous travaillons.

### *5.2.3.3 Amélioration de la prédiction par ajout de variables*

Notre objectif est maintenant d'améliorer notre premier modèle pour le rendre plus fiable. Il se trouve que nous disposons de 2 variables qualitatives qui sont potentiellement discriminantes pour la détection des colzas buissonnants : le département et le stade du colza au moment de l'observation. La Figure 21 nous représente le nombre de parcelles buissonnantes et non buissonnantes par département et par stade de la culture. Nous observons que tous les départements ne sont pas forcément concernés par les colzas buissonnants en 2017. Ajouter les départements aux données de la random forest peut donc permettre de discriminer un peu plus les colzas buissonnants. De même pour le stade (Figure 21), on s'aperçoit que la majorité des colzas buissonnants intervient entre le stade D et E du colza. Intégrer ces deux variables au modèle peut être une source d'amélioration de la prédiction.

Nous incorporons le département le stade du colza dans les données pour créer un nouveau modèle qu'on espère plus précis. Nous utilisons toujours un algorithme de random forest avec arbre de classification. Le résultat de ce nouveau modèle est présenté dans la Figure 22. Comparé au modèle de la figure 19, le kappa a été amélioré, il passe de 0.44 à 0.52. Il en est de même pour l'accuracy qui passe de 0.75 à 0.77, nous avons donc gagné en précision. Cela se ressent également au niveau des erreurs de classifications qui ont diminuées. Lors de la validation du modèle on retrouve également une bonne performance du modèle avec un kappa de 0.48 et une accuracy de 0.82. Ces résultats sont en parti faussés par la très bonne classification des colzas non buissonnantes (erreur de classification = 0.12). De manière générale la validation est aussi faussée par un jeu de données trop faible qui nous empêche de véritablement confirmer la fiabilité du modèle. Lorsque l'on regarde les variables importantes on retrouve le département ainsi que le stade, ces deux variables ont donc bien contribué à augmenter la précision du modèle.

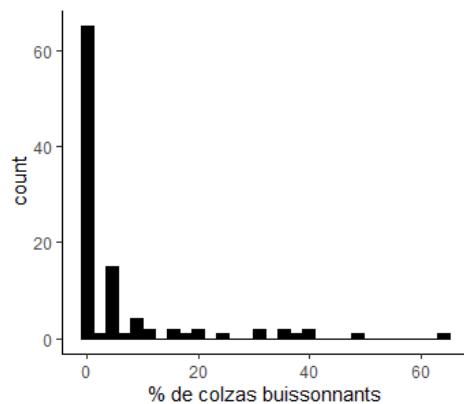


Figure 23 : Distribution du pourcentage de colzas buissonnants. Graphique réalisé à partir des données Vigicultures® 2017 et R2D2.

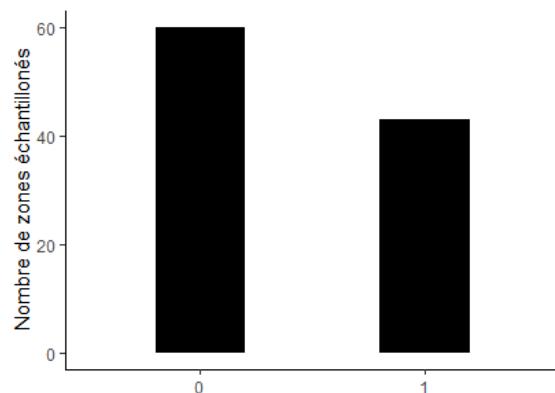


Figure 24 : Distribution du colza buissonnant. L'axe des abscisses prend la valeur 0 pour les colzas non buissonnantes et la valeur 1 pour les colzas buissonnantes. Les données sont classées selon la médiane du pourcentage de plantes buissonnantes. Graphique réalisé à parti de Vigicultures® et R2D2

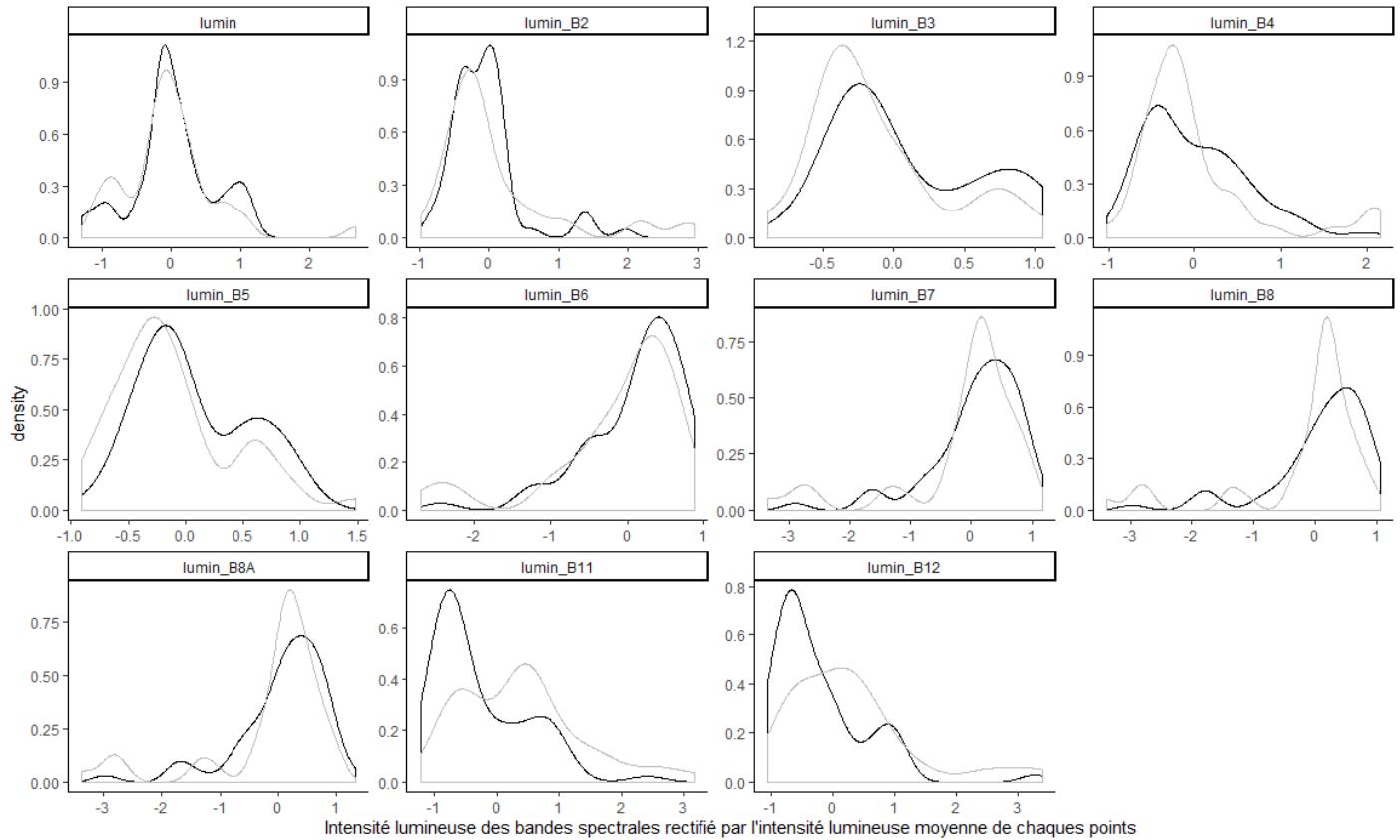


Figure 25 : Distribution des valeurs des bandes spectrales. Les courbes noires représentent les parcelles avec du colza non buissonnant. Les courbes grises correspondent aux parcelles avec du colza buissonnant. La nomenclature utilisée pour les bandes spectrales est  $\text{lumin}_Bx$  où  $x$  correspond aux numéros de la bande et mm au mois d'acquisition des données satellites.

L'analyse des données Vigicultures® nous a permis de déterminer le lien entre certaines bandes et le colza buissonnant.

### **5.3 Analyse complémentaire (Vigicultures® et R2D2)**

#### **5.3.1 Analyse descriptive**

Ce nouveau jeu de données comporte 103 points, il regroupe les 87 points de Vigicultures® et les 16 de R2D2. Il est composé de 21 variables explicatives, les 10 bandes spectrales du mois de mai, les 10 bandes spectrales rectifiées (lumin\_Bx) et de la moyenne de l'intensité lumineuse pour chaque point (lumin). La variable à expliquer est le pourcentage de plantes buissonnantes.

En additionnant les deux jeux de données, on obtient des données moins déséquilibrées. On observe sur la Figure 23 que l'on a plus de valeurs buissonnantes et que de valeur avec 0% de plantes buissonnantes est resté inchangé (comparaison à la Figure 12). On se rend plus compte de ce rééquilibrage lorsque l'on observe la Figure 24, l'écart entre plantes buissonnantes (1) et non buissonnantes (0) s'est fortement réduit.

Nous avons également récupéré les valeurs des bandes spectrales du mois de mai qui correspondent à ces données. Nous les avons centrées et réduites puis nous avons enlevé à chaque point sa valeur moyenne d'intensité lumineuse (lumin) sur l'ensemble de ces bandes. La Figure 25 nous présente les courbes de distribution de ces nouvelles variables. On remarque que les courbes sont relativement similaires. On note de réelles différences pour lumin\_B11 et lumin\_B12, on peut donc s'attendre à voir ressortir ces deux variables dans nos futures analyses.

#### **5.3.2 Exploration des variables importantes et de la qualité de prédiction**

Pour faire l'analyse des données de pourcentage de colzas buissonnants, nous avons essayé d'utiliser des modèles LASSO en supposant une distribution normale, binomiale (nombre de plans buissonnants sur 100 plans observés), de Poisson (adapté à des données de comptage) ou binomial négatif (étend le modèle poissonien en permettant un ajustement séparé de la variance). Les résultats sont représentés dans la Figure 26. L'application de la validation croisée à ces modèles fait systématiquement ressortir la bande B11 et/ou sa transformation par rapport à la « luminosité générale » (lumin\_B11). Nous avons aussi essayé d'utiliser ces modèles pour prédire des dépassements du seuil de 15 % de plants buissonnats. La plupart des modèles ne permettent pas une prédiction pertinente car tous les sites sont prédits au-dessus ou en dessous du seuil, suivant les modèles. Seul le modèle négatif binomial étale les prédictions

LASSO	Negative binomial		0 inflated negative binomial	
Donnée d'entraînement	% buiss	buiss > 0		seuil > 15%
Donnée de validation	=	buiss > 0	% buiss	=
OR	6.6	29.61		26.65
Interval de confiance	[1.24;+Inf]	[3.09;1519.46]		[4.87;196.6]
Kappa	0.22	0.59		0.53

Figure 26 : Résultats des LASSO réalisés sur le jeu de données Vigicultures® associé à R2D2. OR : odds ratio, intervalle de confiance calculé à 95%. Signification des abréviations pour les données : % buiss : jeux de données sur l'ensemble des valeurs (de 0 à 100), buiss > 0 : utilisation des valeurs strictement positive, seuil > 15% : utilisation d'un seuil de 15% de plantes buissonnantes, = : même type de donnée utilisé pour la validation que celles utilisées pour l'entraînement.

suffisamment pour réaliser des prédictions différentes suivant les sites, mais il ne prédit que 2 sites sur 14 supérieurs au seuil (Fisher-Test : OR 6.6 [1.24;+Inf], kappa=0.22).

La distribution des données avec un fort excès de zéros, associé à une forte dispersion des données non nulles suggère qu'en cas de faibles proportions de colzas buissonnantes les observateurs indiquent un taux de 0, augmentant artificiellement le nombre d'observations nulles. La distribution négative binomiale zéro-augmentée est particulièrement adaptée pour modéliser la variabilité de telles données. Nous avons donc aussi ajusté un modèle glm LASSO avec distribution négative binomiale à nos données. Ce modèle ajuste séparément les facteurs expliquant qu'une observation soit 0 parce qu'il n'y a pas vraiment de comptage et les facteurs déterminant la valeur de l'observation s'il y a observation. La validation croisée du modèle LASSO sélectionne bien des paramètres significatifs sur le modèle de 0 confirmant qu'il y a un excès de zéros. Cette procédure sélectionne *lumin\_B11* pour la prédiction des zéros et *B11* pour prédire les pourcentages de plants buissonnants si observés (généralement différents de 0). La corrélation est positive entre valeur de l'observation et la bande *B11* mais négative avec la bande similaire *lumin\_B11* confirmant qu'avec *B11*, la probabilité d'une observation réelle (non 0) augmente de manière similaire à la valeur observée s'il y a observation.

Si l'on utilise le modèle prédisant les pourcentages de plants buissonnants en cas d'observations pour prédire le dépassement du seuil de 15 % de plants buissonnants, le modèle est plus discriminant que le modèle binomial négatif simple : il prédit correctement 7 des 14 sites au-dessus du seuil mais au prix de 3 erreurs parmi les 89 sites en dessous du seuil (Fisher-Test : OR 26.65 [4.87;196.6] ; kappa = 0.53). On note que deux des trois erreurs sont des 0 qui pourraient être des faux négatifs, l'une d'elles estimée à 15.09 %, la troisième erreur est estimée à 4 % alors que la valeur estimée était 15.13 %. En revanche, les 7 sites estimés correctement au-dessus de 15 % ont des valeurs estimées supérieures à 16 % (moyenne à 23%, moyenne générale des prédictions à 12.31 %).

Pour compléter cette étude nous ajustons aussi un modèle supposant une distribution binomiale négative des données uniquement sur les données non nulles puisqu'il semble y avoir un grand nombre de 0 correspondant à un défaut d'observation. Ce modèle fait toujours ressortir *B11* (et à la marche *lumin\_B11*) et permet une assez bonne discrimination du dépassement du seuil de 15 % des plants buissonnants : 8 des 14 sites sont identifiés, 1 seul des 26 sites en dessous du seuil est mal classé (OR : 29.61[3.09;1519.46], kappa=0.59). Etant donné qu'il y a un lien d'après le modèle zéro augmenté entre non-observation (0) et faible pourcentage de plants buissonnants, on teste aussi l'efficacité de ce modèle sur l'ensemble du jeu de données



pour prédire le dépassement de seuil de 15 %. Le modèle se montre également très discriminant pour les sites à valeur nulle, n'attribuant que 2 sites sur 89 à une valeur au-dessus de 15 % résultant d'une bonne discrimination générale (OR 35.09 [6.5;260.9] ; kappa=0.59).

## 6 **Discussion**

### 6.1 **Discussion Vigicultures®**

#### 6.1.1 Résultat général de l'étude

Au terme de notre étude nous avons clairement identifié les bandes spectrales et les indices qui permettent de discriminer le colza buissonnant. Cependant nous n'avons pas réussi à obtenir un modèle suffisamment fiable pour télédétecter le port buissonnant des colzas.

Concernant l'agencement des données, le LASSO et la random forest font ressortir qu'il semble plus pertinent de combiner, au sein d'un même jeu de données, les bandes spectrales et les indices de végétation. La bande B11\_05 et le NDWI\_05 apparaissent comme discriminant pour les deux méthodes. Le premier modèle que nous obtenons avec uniquement les données de télédétection n'est pas suffisamment fiable. Nous y ajoutons le département et le stade de la culture, ce qui permet d'améliorer la prédiction et de valider l'importance de la bande B11 et du NDWI. Le faible jeu de données de validation rend difficile la validation du modèle. Enfin, notre modèle pour prédire le pourcentage de colza buissonnant s'est avéré incapable de réaliser une prédiction correcte.

#### 6.1.2 Association et importance des variables

La bande spectrale B11\_05 ressort dans la totalité de nos modèles. Elle correspond aux longueurs d'ondes de l'infrarouge court (SWIR). Ces longueurs d'ondes sont particulièrement appréciées pour quantifier l'humidité de la végétation et du sol (Ji et al, 2011). Le SWIR est sensible à la quantité d'eau contenue dans le feuillage mais ne peut pas être utilisé seul pour la quantifier. En effet la concentration en eau du feuillage est aussi associée à des paramètres de structure de la plante et de matière sèche (Ceccato, 2001), c'est pourquoi des indices ont été créés pour évaluer la teneur en eau d'un couvert végétal.

L'indice qui ressort le plus de nos analyses est le NDWI, un indice de la teneur en eau de la végétation. Le NDWI est calculé à partir des bandes spectrales B8 et B11(Sentinel 2 EO products). La bande B11 est sensible aux variations de la teneur en eau du couvert comme énoncé précédemment. La bande B8, correspond à la longueur d'onde proche infrarouge (NIR).



Le proche infrarouge permet de cartographier des différences dans la densité de végétation, il est extrêmement bien réfléchi par le feuillage. Il peut traverser les premières couches de feuilles et être réfléchi par les feuilles des étages inférieurs. Cette aptitude permet de prendre en compte la variabilité dans les étages inférieurs du couvert et ainsi faire transparaître la structure de la plante dans la densité du couvert végétal (Campbell, 2007). Les canaux NIR peuvent passer à travers 8 étages de feuilles distincts (Lillesaeter, 1982).

Il est aussi intéressant de noter qu'un autre indice semble être discriminant pour les colzas buissonnants, le moisture index, qui est lui aussi un indice d'humidité. Il utilise la bande B8A et la bande B11. La bande B8A fait partie du proche infrarouge mais a un spectre beaucoup plus restreint que celui de la bande B8.

Il semble donc que la teneur en eau du colza buissonnant soit différente de celle du colza non buissonnant. Cela peut être causé en partie par les larves de CBT qui se développent et se nourrissent au cœur des tiges. L'eau circule dans la plante via le Xylem et le Phloem qui passe au centre des tiges, s'ils servent de nourriture aux larves du charançon, l'alimentation de la plante se trouve compromise. Cette hypothèse pourrait expliquer la différence de concentration en eau entre les colzas buissonnantes et les colzas non buissonnantes.

Les variables sélectionnées par les random forest et les LASSO sont majoritairement celle du mois de mai. Les données satellites du mois d'avril semblent moins pertinentes. Cela est probablement dû au fait que la différence de morphologie entre le colza buissonnant et le colza non buissonnant apparaît au mois de mai.

### 6.1.3 Prédictions avec random forest

Pour réaliser nos modèles de prédiction nous avons utilisé la méthode de la random forest. Ce classificateur est souvent utilisé en télédétection pour la justesse des classifications qu'il fournit (Belgiu and Dragut, 2016). Les résultats obtenus par notre modèle sur les bandes et les indices pour la variable binaire ne sont cependant pas suffisamment fiables pour réaliser de la prédiction. On suppose que cela n'est pas dû à la méthode statistique utilisée mais au jeu de données. Notre jeu de données présente un déséquilibre en faveur des colzas non buissonnantes. Ce déséquilibre conduit à une plus mauvaise classification de la classe minoritaire, le colza buissonnant (Chao, 2004). On peut alors observer une très bonne accuracy, mais elle est biaisée par la bonne classification de la classe majoritaire. On notera que l'on retrouve ce problème dans tous nos modèles. De plus, le faible nombre de données qui compose les données de validation biaise la validation du modèle.



Notre second modèle intègre le département et le stade du colza au moment de l'échantillonnage, ces deux variables sont très liées au colza buissonnant. Certains départements sont très fortement touchés alors que d'autres pas du tout. Idem, l'apparition du buissonnement du colza est en lien avec le stade du colza. Avec ces nouvelles variables explicatives on remarque une amélioration générale du modèle. Cependant pour les raisons citées ci-dessus, il nous est impossible de réaliser une prédiction fiable. Toutefois le fait que l'on retrouve encore la bande B11 et le NDWI malgré la présence de ces variables explicatives témoigne d'un lien important entre la bande B11, le NDWI et le colza buissonnant. On observe donc une réelle cohérence dans l'ensemble de nos analyses.

Notre dernier modèle est une random forest réalisée sur le pourcentage de colzas buissonnantes. L'objectif est de pouvoir prédire un gradient de buissonnement et ne pas s'arrêter à buissonnant ou non buissonnant, ce qui présenterait plus de pertinences. Une parcelle avec 50% de colzas buissonnantes n'est pas conduite de la même manière qu'une parcelle avec 5%. Le modèle obtenu n'est pas fiable et ne permet pas de réaliser la prédiction. Nous supposons que cela est causé une fois de plus par le jeu de données déséquilibrées.

A la suite de l'étude de la base de données Vigicultures® couplée aux images Sentinel-2, il semble que le NDWI et la bande B11 soient en lien avec le colza buissonnant car ils ressortent tout au long de nos analyses. Il se peut toutefois que notre manière de regarder les données (utilisation des bandes brutes et des indices) ne soit pas la bonne. De plus un déséquilibre des données ne nous permet pas de réaliser des prédictions fiables. Pour tenter de résoudre une partie de ces problèmes nous poursuivons nos investigations en mutualisant les données Vigicultures® et R2D2.

## **6.2 Discussion de Vigicultures® associé à R2D2**

Ce nouveau jeu de données nous permet certes de rééquilibrer notre jeu de données mais il nous fait également perdre de l'information. Ne disposant pas des données de R2D2 pour le mois d'avril nous avons mis de côté les données Vigicultures® de ce mois. Nous perdons donc un mois entier de données ce qui impacte forcément la qualité de nos prédictions. Il faut également noter que les données R2D2 sont issues de la campagne 2017-2018 alors que celles de Vigicultures® sont issues de campagne 2016-2017. Ce mélange de périodes va nous permettre de voir si les bandes sélectionnées sont pertinentes sur différentes années. En effet, il est courant en télédétection que ce qui fonctionne une année ne marche pas l'année suivante et



qu'il soit nécessaire d'avoir les données de terrain relatives à l'année en cours pour réaliser des prédictions. Si nos résultats sont cohérents avec ceux trouvés précédemment cela voudrait dire que la télédétection pourrait être utilisée sans données de terrain pour n'importe quelle année pour la détection des parcelles de colzas buissonnants.

Les différents glm LASSO réalisés sur cette base de données confirme nos résultats précédents en sélectionnant la bande B11 et lumin\_B11. Cependant le NDWI n'est plus sélectionné ce qui est sans doute causé par le fait que les bandes soit centrées et réduites. Il existe un lien certain entre les colzas buissonnants et les images Sentinel-2, plus particulièrement avec la bande B11. Nous ne devons pas utiliser les données sous la bonne forme (bande seul, bande centrée réduite et Indice), une transformation supplémentaire est sûrement à envisager. Mais ce n'est pas la seule cause des "mauvaises" prédictions, le jeu de données influence aussi les modèles.

Notre jeu de données est moins déséquilibré mais comporte un excès de 0 qui biaise les modèles statistiques. Le fait que le modèle glm LASSO à zéro augmenté sélectionne une variable pour le modèle des 0 signifie que tous les zéros ne sont pas de vrais zéros. Certaines parcelles présentant un faible pourcentage de colzas buissonnants ont dû être notées 0 par l'observateur. Cette hypothèse est consolidée par le fait que le modèle sur la partie des données supérieur à 0% de colzas buissonnants obtient de bons résultats.

### **6.3 Limite de l'étude**

Notre étude comporte plusieurs points d'améliorations, à commencer par la base de données Vigicultures®. Elle comporte également des observations erronées comme nous avons pu le voir précédemment. Il est difficile de pallier les mauvaises notations des colzas buissonnants. Une des solutions serait de fonctionner par classe de pourcentage de plantes buissonnantes avec une amplitude de 10%. Ainsi les notations à 0% qui seraient mal notées seraient inclus au sein d'une classe et ne biaiserait plus les modèles. Cependant cette méthode ne rééquilibrera pas le jeu de données car la classe [0% - 10%] restera fortement majoritaire. Par ailleurs, l'utilisation d'un seuil au-delà duquel le pourcentage de colzas buissonnants induirait une perte économique significative pourrait être pertinent. Dans ce cas, on aboutirait à une classification binaire en fonction que la valeur soit en dessous ou au-dessus du seuil.

Il serait également possible de compléter notre base de données sur les colzas buissonnants en utilisant les données Epiphyt®. Ces données regroupent l'ensemble des données d'épidémiosurveillance à l'échelle de la France dont les données Vigicultures®. Il



serait également envisageable de regrouper différentes années de données d'épidémirosurveillance (associé à leurs données satellites) et de les regrouper au sein d'un même jeu de données. Ces deux options permettraient d'augmenter les informations fournies aux modèles et donc probablement leurs performances.

Les images sentinel-2 comportent également de nombreux bruits tels que les nuages et les ombres. Le traitement que nous leur avons appliqué (tuile téléchargée avec au maximum 50% de nuage et suppression des valeurs négatives) ne permet pas de s'affranchir de tous les bruits. Une des solutions serait d'utiliser la chaîne de traitement IOTA2 qui détecte et prend en compte les nuages. Par manque de temps et de ressources informatiques il nous a été impossible d'explorer cette option.

Certaines tuiles qui comportaient plus de 50% de nuages n'ont pas été téléchargées, ce qui a créé des trous sans données satellites pour certains points. Une des solutions pour éviter cette perte d'information serait d'effectuer une interpolation qui prendrait en compte la valeur du point sur l'image précédent et suivant la date de la tuile non téléchargée. Cette opération est également réalisée par IOTA2. Ceci permettrait d'obtenir une base de données plus précise ou au lieu d'avoir une moyenne par mois, on aurait une moyenne par semaine.



## **Conclusion**

Cette étude a permis d'associer les outils de télédétection aux données satellites Sentinel-2. L'utilisation d'images satellites pour télédétecer les colzas buissonnantes est envisageable. Plus particulièrement, il semble que l'infrarouge court soit en mesure de discriminer les parcelles buissonnantes des parcelles non buissonnantes. Ces longueurs d'ondes sont déjà utilisées en télédétection pour quantifier la quantité d'eau d'un couvert végétal.

Les résultats de notre étude sont encourageants, ils montrent la pertinence de l'utilisation des images Sentinel-2 pour collecter des données sur les dégâts subis par les parcelles de colzas. La télédétection est donc un outil potentiellement capable de cartographier à grande échelle les colzas buissonnantes. Cela constitue la première étape avant de pouvoir remonter aux pratiques des agriculteurs et aux facteurs environnementaux pour ensuite déterminer les leviers à actionner pour réduire les dégâts du charançon du bourgeon terminal.



## Bibliographie

- “2018\_017info rapide grandes cultures.Pdf.” Accessed September 4, 2019. [http://agreste.agriculture.gouv.fr/IMG/pdf/2018\\_017inforapgrandescultures.pdf](http://agreste.agriculture.gouv.fr/IMG/pdf/2018_017inforapgrandescultures.pdf).
- Acharya, M. C., and R.B. Thapa. “remote sensing and its application in agricultural pest management,” 2015.
- Agreste. “Agreste Infos Rapides – Grandes Cultures et Fourrages,” 2018.
- Agreste. “Agreste Infos Rapides – Grandes Cultures et Fourrages,” 2019.
- AGRESTE. “Les Indicateurs de Fréquencede Traitement (IFT) En 2011,” 2013. [http://agreste.agriculture.gouv.fr/IMG/pdf/dossier18\\_ift.pdf](http://agreste.agriculture.gouv.fr/IMG/pdf/dossier18_ift.pdf).
- Bégué, Agnés, Beatriz Bellon, Nadine Dessay, Stéphane Dupuy, Raffaele Gaetano, Jordi Inglada, Guerric Le Maire, et al. “Cartographie et caractérisation des systèmes agricoles,” n.d., 1.
- Belgiu, Mariana, and Lucian Drăguț. “Random Forest in Remote Sensing: A Review of Applications and Future Directions.” *ISPRS Journal of Photogrammetry and Remote Sensing* 114 (April 1, 2016): 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Bhattarai, Ganesh P., Ryan B. Schmid, and Brian P. McCornack. “Remote Sensing Data to Detect Hessian Fly Infestation in Commercial Wheat Fields.” *Scientific Reports* 9, no. 1 (April 16, 2019): 6109. <https://doi.org/10.1038/s41598-019-42620-0>.
- Bleiholder, Boom, Langelüddecke, and Stauss. “Einheitliche Codierung Der Phänologischen Stadien Bei Kultur Und Schadpflanzen.” *Gesunde Pflanzen*, no. 41 (1989): 381–84.
- Brownlee, Jason. “Machine Learning Evaluation Metrics in R,” February 29, 2016. <https://machinelearningmastery.com/machine-learning-evaluation-metrics-in-r/>.
- Campbell, J.B. *Introduction to Remote Sensing: Fourth Edition*. The Guilford Press: New York, NY, 2017.
- Carlson, Toby N., and David A. Ripley. “On the Relation between NDVI, Fractional Vegetation Cover, and Leaf Area Index.” *Remote Sensing of Environment* 62, no. 3 (December 1, 1997): 241–52. [https://doi.org/10.1016/S0034-4257\(97\)00104-1](https://doi.org/10.1016/S0034-4257(97)00104-1).
- Ceccato, Pietro, Stéphane Flasse, Stefano Tarantola, Stéphane Jacquemoud, and Jean-Marie Grégoire. “Detecting Vegetation Leaf Water Content Using Reflectance in the Optical Domain.” *Remote Sensing of Environment* 77, no. 1 (July 1, 2001): 22–33. [https://doi.org/10.1016/S0034-4257\(01\)00191-2](https://doi.org/10.1016/S0034-4257(01)00191-2).
- Chen, Chao. “Using Random Forest to Learn Imbalanced Data,” n.d., 12.
- CNES. “Copernicus.” Copernicus, 2015. <https://copernicus.cnes.fr/fr/copernicus>.
- CNES. “Sentinel-2.” Sentinel-2, March 6, 2017. <https://sentinel2.cnes.fr/fr/sentinel-2>.
- CNES. “Sentinel-2.” Sentinel-2, 2018. <https://sentinel2.cnes.fr/fr/sentinel-2>.
- Cowley, R. B., D. J. Luckett, J. S. Moroni, and S. Diffey. “Use of Remote Sensing to Determine the Relationship of Early Vigour to Grain Yield in Canola (*Brassica Napus L.*) Germplasm.” *Crop and Pasture Science* 65, no. 12 (December 10, 2014): 1288–99. <https://doi.org/10.1071/CP14055>.
- Domínguez, J. A., J. Kumhálová, and P. Novák. “Winter Oilseed Rape and Winter Wheat Growth Prediction Using Remote Sensing Methods.” *Plant, Soil and Environment* 61 (2015), no. No. 9 (September 14, 2015): 410–16. <https://doi.org/10.17221/412/2015-PSE>.
- “Ga09-Report.Pdf.” Accessed September 6, 2019. <http://pierre.gaillard.me/doc/Ga09-report.pdf>.
- Gaetano, Raffaele. “Le Programme Copernicus et La Mission Sentinel-2,” 2018. [http://www.modelia.org/moodle/pluginfile.php/2882/mod\\_resource/content/0/J1\\_02\\_Copernicus\\_gaetano\\_PayoteModelia\\_27juin.pdf](http://www.modelia.org/moodle/pluginfile.php/2882/mod_resource/content/0/J1_02_Copernicus_gaetano_PayoteModelia_27juin.pdf).



Gao, Bo-cai. "NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space." *Remote Sensing of Environment* 58, no. 3 (December 1, 1996): 257–66. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3).

Gitelson, Anatoly A., Yoram J. Kaufman, and Mark N. Merzlyak. "Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS." *Remote Sensing of Environment* 58, no. 3 (December 1, 1996): 289–98. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7).

Glenn, Edward P., Christopher M. U. Neale, Doug J. Hunsaker, and Pamela L. Nagler. "Vegetation Index-Based Crop Coefficients to Estimate Evapotranspiration by Remote Sensing in Agricultural and Natural Ecosystems: VEGETATION INDEX-BASED CROP COEFFICIENTS." *Hydrological Processes* 25, no. 26 (December 30, 2011): 4050–62. <https://doi.org/10.1002/hyp.8392>.

GNIS. "Colza : Origine et Physiologie," June 20, 2019. <https://www.gnis-pedagogie.org/colza-origine-physiologie-plante.html>.

Han, Jiahui, Chuanwen Wei, Yaoliang Chen, Weiwei Liu, Peilin Song, Dongdong Zhang, Anqi Wang, Xiaodong Song, Xiuzhen Wang, and Jingfeng Huang. "Mapping Above-Ground Biomass of Winter Oilseed Rape Using High Spatial Resolution Satellite Data at Parcel Scale under Waterlogging Conditions." *Remote Sensing* 9 (2017): 238. <https://doi.org/10.3390/rs9030238>.

Horning, N. "Remote Sensing." In *Encyclopedia of Ecology*, edited by Sven Erik Jørgensen and Brian D. Fath, 2986–94. Oxford: Academic Press, 2008. <https://doi.org/10.1016/B978-008045405-4.00237-8>.

Huang, Wenjiang, Juhua Luo, Jingcheng Zhang, Jinling Zhao, Chunjiang Zhao, Jihua Wang, Guijun Yang, Muyi Huang, Linsheng Huang, and Shizhou Du. "Crop Disease and Pest Monitoring by Remote Sensing," 2012. <https://doi.org/10.5772/35204>.

Huang, Yanbo, Zhong-xin Chen, Tao Yu, Xiang-zhi Huang, and Xing-fa Gu. "Agricultural Remote Sensing Big Data: Management and Applications." *Journal of Integrative Agriculture* 17, no. 9 (September 1, 2018): 1915–31. [https://doi.org/10.1016/S2095-3119\(17\)61859-8](https://doi.org/10.1016/S2095-3119(17)61859-8).

Huete, A. R. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25, no. 3 (August 1, 1988): 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).

Inglauda, Jordi. "La Première Carte d'occupation Des Sols 2016 de La France Avec Sentinel-2 | Séries Temporelles," March 30, 2017. <http://www.cesbio.ups-tlse.fr/multitemp/?p=10104>.

Ji, Lel, Li Zhang, Bruce Wylie, and Jennifer Rover. "On the Terminology of the Spectral Vegetation Index (NIR – SWIR)/(NIR + SWIR)," 2011. <https://pubs.er.usgs.gov/publication/70043293>.

Kaufman, Y. J., and D. Tanre. "Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS." *IEEE Transactions on Geoscience and Remote Sensing* 30, no. 2 (March 1992): 261–70. <https://doi.org/10.1109/36.134076>.

Laguette, S, A Vidal, and P Vossen. "Télédétection et estimation des rendements en blé en Europe," 1997, 17.

"Le Colza - Arysta LifeScience." Accessed September 4, 2019. <http://www.arystalifescience.fr/les-cultures/le-colza#inline-auto152>.

"Le suivi de l'état hydrique (NDWI) des cultures par les Séries Temporelles Sentinel-2 | Quadratic.be." Accessed August 23, 2019. <http://www.quadratic.be/le-suivi-de-letat-hydrique-ndwi-des-cultures-par-les-series-temporelles-sentinel-2/>.

Lillesaeter, O. "Spectral Reflectance of Partly Transmitting Leaves: Laboratory Measurements and Mathematical Modeling." *Spectral Reflectance of Partly Transmitting Leaves: Laboratory Measurements and Mathematical Modeling* 12 (1982): 247–54.

"MAJA's Native Sentinel-2 Format | Séries Temporelles." Accessed August 30, 2019. [http://www.cesbio.ups-tlse.fr/multitemp/?page\\_id=10464](http://www.cesbio.ups-tlse.fr/multitemp/?page_id=10464).



Mulla, David J. "Twenty Five Years of Remote Sensing in Precision Agriculture: Key Advances and Remaining Knowledge Gaps." *Biosystems Engineering*, Special Issue: Sensing Technologies for Sustainable Agriculture, 114, no. 4 (April 1, 2013): 358–71. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>.

Piekarczyk, Jan, Hanna Sulewska, and Grażyna Szymbańska. "Winter Oilseed-Rape Yield Estimates from Hyperspectral Radiometer Measurement." *Quaestiones Geographicae*, no. 30 (2011): 77–84. <https://doi.org/10.2478/v10117-011-0007-z>.

Piekarczyk, Jan, Hanna Sulewska, and Grażyna Szymbańska. "Winter Oilseed-Rape Yield Estimates from Hyperspectral Radiometer Measurements." *Quaestiones Geographicae* 30, no. 1 (January 1, 2011): 77–84. <https://doi.org/10.2478/v10117-011-0007-z>.

Raikes, C., and L. L. Burpee. "Use of Multispectral Radiometry for Assessment of Rhizoctonia Blight in Creeping Bentgrass." *Phytopathology* 88, no. 5 (May 1, 1998): 446–49. <https://doi.org/10.1094/PHYTO.1998.88.5.446>.

Recoules, E, and N Brevault. "L'autonomie protéique : état des lieux et voies d'amélioration pour l'alimentation des volailles," 2016, 11.

Richards, J. A. *Remote Sensing Digital Image Analysis: An Introduction*. Fifth edition. Berlin: Springer, 2013.

Rondeaux, Geneviève, Michael Steven, and Frédéric Baret. "Optimization of Soil-Adjusted Vegetation Indices." *Remote Sensing of Environment* 55, no. 2 (February 1, 1996): 95–107. [https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7).

Schott, Céline, Catherine Mignolet, and Jean-Marc Meynard. "Les oléoprotéagineux dans les systèmes de culture : évolution des assolements et des successions culturales depuis les années 1970 dans le bassin de la Seine." *Oléagineux, Corps gras, Lipides* 17, no. 5 (September 1, 2010): 276–91. <https://doi.org/10.1051/ocl.2010.0334>.

Seelan, Santhosh K, Soizik Laguette, Grant M Casady, and George A Seielstad. "Remote Sensing Applications for Precision Agriculture: A Learning Community Approach." *Remote Sensing of Environment*, IKONOS Fine Spatial Resolution Land Observation, 88, no. 1 (November 30, 2003): 157–69. <https://doi.org/10.1016/j.rse.2003.04.007>.

"Sentinel 2 EO Products | Sentinel Hub." Accessed August 28, 2019. [https://www.sentinel-hub.com/develop/documentation/eo\\_products/Sentinel2EOProducts](https://www.sentinel-hub.com/develop/documentation/eo_products/Sentinel2EOProducts).

Serrano, Lydia, Iolanda Filella, and Josep Peñuelas. "Remote Sensing of Biomass and Yield of Winter Wheat under Different Nitrogen Supplies." *Crop Science* 40, no. 3 (May 1, 2000): 723–31. <https://doi.org/10.2135/cropsci2000.403723x>.

Shanahan, John F., James S. Schepers, Dennis D. Francis, Gary E. Varvel, Wallace W. Wilhelm, James M. Tringe, Mike R. Schlemmer, and David J. Major. "Use of Remote-Sensing Imagery to Estimate Corn Grain Yield." *Agronomy Journal* 93, no. 3 (May 1, 2001): 583–89. <https://doi.org/10.2134/agronj2001.933583x>.

She, Bao, Jing-feng Huang, Rui-fang Guo, Hong-bin Wang, and Jing Wang. "Assessing Winter Oilseed Rape Freeze Injury Based on Chinese HJ Remote Sensing Data." *Journal of Zhejiang University-SCIENCE B* 16, no. 2 (February 1, 2015): 131–44. <https://doi.org/10.1631/jzus.B1400150>.

"Spatial - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online." Accessed August 30, 2019. <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>.

Sudha Rani, D., M.N. Venkatesh, Ch. Naga Satya Sri, and K. Anand Kumar. "Remote Sensing as Pest Forecasting Model in Agriculture." *International Journal of Current Microbiology and Applied Sciences* 7, no. 03 (March 20, 2018): 2680–89. <https://doi.org/10.20546/ijcmas.2018.703.310>.

Terre Inovia. "Cultiver Des Grandes Cultures sans Insecticides, Est-Ce Possible ?," n.d.



Terre Inovia. “Guide de Culture Du Colza 2018,” 2018.  
[https://www.terresinovia.fr/documents/20126/453413/guide\\_culture\\_colza2018.pdf/31678f7b-105f-3fce-82d4-c47c03f0f2af?t=1553979977748](https://www.terresinovia.fr/documents/20126/453413/guide_culture_colza2018.pdf/31678f7b-105f-3fce-82d4-c47c03f0f2af?t=1553979977748).

Thorp, K.R., and L.F. Tian. “A Review on Remote Sensing of Weeds in Agriculture.” *Precision Agriculture* 5, no. 5 (October 1, 2004): 477–508. <https://doi.org/10.1007/s11119-004-5321-1>.



## **Table des figures**

Figure 1 : Schémas des différents acteurs du projet R2D2 .....	12
Figure 3 : Présentation des principaux ravageurs et des stades de sensibilités du colza.....	13
Figure 2: Evolution des surfaces de colza en France depuis 1996.....	14
Figure 4: Schématisation des différents flux dans le cadre de la télédétection.....	18
Figure 5 : Représentation des différentes bandes spectrales sentinel-2 .....	20
Figure 6 : Schémas de la démarche mise en oeuvre pour répondre à la problématique .....	24
Figure 7 : Répartition spatiale des points Vigicultures 2017 pour le colza buissonnant. ....	28
Figure 11 : Résultats de la random forest avec arbre de classifications sur les bandes et les indices.....	38
Figure 10 : Résultats LASSO sur les bandes spectrales et les indices des données R2D2. ....	38
Figure 9 : Distribution du colza buissonnant. ....	38
Figure 8 : Distribution du pourcentage de colza buissonnant des données R2D2 2018. ....	38
Figure 12 : Distribution du pourcentage de colzas buissonnantes .....	40
Figure 14 : Distribution des valeurs des bandes spectrales.....	40
Figure 13 : Distribution du colza buissonnant.. ....	40
Figure 15 : Distribution des valeurs des indices de végétation.....	42
Figure 16 : Tableau récapitulatif des LASSO effectué sur les 3 associations des variables explicatives.....	42
Figure 17 : Tableau récapitulatif des Random Forest effectué sur les 3 associations des variables explicatives.....	42
Figure 18 : Violine plot de la bande spectrale B11_05 et du NDWI_05.. ....	44
Figure 19 : Résultat de la random forest à arbre de classification sur les données Bandes + Indices. ....	44
Figure 20 : Valeur réelle en fonction des valeurs prédites par la random forest avec arbre de régression. ....	46
Figure 21 : Histogramme du nombre de parcelles buissonnantes par département et par stade au moment de l'observation. ....	46
Figure 22 : Résultat de la random forest à arbre de classification sur les données regroupant les bandes, les indices, les départements et les stades du colza.....	46
Figure 23 : Distribution du pourcentage de colzas buissonnants.....	48
Figure 25 : Distribution des valeurs des bandes spectrales. on des données satellites. ....	48



Figure 24 : Distribution du colza buissonnant ..... 48

Figure 26 : Résultats des LASSO sur le jeu de données Vigicultures® associé à R2D2.....50