



**Université de Toulouse**

**MASTER 1 GEOMATIQUE**

« **ScIences Géomatiques en environneMent et AménageMent** » (**SIGMA**)

<http://sigma.univ-toulouse.fr>

**RAPPORT DE STAGE**

**Développement de méthodes de classification de l'occupation du sol robustes ne nécessitant pas de données d'apprentissage spécifiques de l'année**

**De CAUNES Robin**



**Centre d'Etudes Spatiales de la  
BIOsphère**



**Institut de Recherche pour le  
Développement**

**Maître de stage : SIMONNEAUX Vincent  
Enseignant-référent : SHEEREN David**

**26/08/2022**

# Résumé

L'occupation du sol est une variable spatialisée importante, car elle procure une connaissance de base pour de nombreux domaines d'application dont des modèles bio-physiques.

La plaine du Haouz est une région semi-aride étudiée par le CESBIO et le LMI TREMA dans le cadre du projet Altos. La demande de ressources en eau y augmente d'année en année sous les contraintes de sécheresse et d'agriculture intensive, malgré un gaspillage important dû à une gestion de la ressource inefficace. Pour répondre à cela, les laboratoires mettent en place des modèles bio-physiques alimentés par des données d'occupation du sol dans le but d'étudier le fonctionnement hydrologique de la zone afin de proposer des solutions. L'apparition de séries temporelles d'images à haute répétitivité (Sentinel-2) permet de suivre l'occupation du sol plus fréquemment qu'à l'aide de relevés terrain.

Dans ce contexte, le but du travail est de développer une méthode de classification de l'occupation du sol robuste et qui ne nécessite pas forcément de données d'apprentissage. Cela a été rendu possible à partir de séries temporelles d'indices de végétation, issues d'images Sentinel-2. Le travail a déjà été réalisé sur une petite zone d'étude, l'objectif du stage a été de porter ce travail à l'échelle de la plaine du Haouz.

# Abstract

Land cover is an important spatial variable, as it provides basic knowledge for many fields of application including bio-physical models.

The Haouz plain is a semi-arid region studied by the CESBIO and the LMI TREMA as part of the Altos project. There, the demand for water resources is increasing year after year under the constraints of drought and intensive agriculture, despite significant waste due to inefficient resource management. To respond to this, laboratories are setting up bio-physical models fed by land use data in order to study the hydrological functioning of the area and propose solutions. The appearance of time series of high repetition images (Sentinel-2) makes it possible to monitor land use more frequently than with field surveys.

In this context, the aim of the work is to develop a robust land cover classification method that does not necessarily require training data. This was made possible by using time series of vegetation indices from Sentinel-2 images. The work has already been carried out on a small study area, the aim of the internship was to extend this work to the scale of the Haouz plain.

# Remerciements

Tout d'abord, j'aimerais remercier l'ensemble de l'équipe du CESBIO de m'avoir accueilli pendant la durée de ce stage. Tout particulièrement Dominique TARISSE qui m'a sauvé dans les démarches du voyage au Maroc.

Je remercie chaleureusement Vincent SIMONNEAUX pour m'avoir encadré et guidé tout au long du stage, ainsi que Vincent RIVALLAND et Vincent THIERION pour leurs quelques dépanages.

Mes remerciements sincères s'adressent également aux stagiaires avec qui j'ai partagé un espace de travail très convivial et parfois animé. Merci Ainhoa, Serdar et Victor pour vos conseils. Merci Odile, Martin et Mathilda pour votre bonne humeur.

Merci également à Maryam CHOUKRI d'avoir été mon acolyte sur le terrain.

Pour finir, un profond merci à l'ensemble de l'équipe enseignante du Master SIGMA. Particulièrement David SHEEREN qui a été mon encadrant universitaire.

# Table des matières

Table des figures . . . . .	3
Liste des tableaux . . . . .	4
<b>1 Introduction</b>	<b>6</b>
1.1 Contexte général . . . . .	6
1.1.1 Structure d'accueil . . . . .	6
1.1.2 Le projet ALTOS et le LMI TREMA . . . . .	6
1.2 Problématique . . . . .	7
1.3 Objectifs . . . . .	7
<b>2 Matériels et méthodes</b>	<b>8</b>
2.1 Zone d'étude . . . . .	8
2.2 Indice de végétation : NDVI . . . . .	9
2.3 Classes à prédire . . . . .	9
2.4 Données d'entrée . . . . .	9
2.4.1 Données satellites . . . . .	9
2.4.2 Données terrain . . . . .	11
2.5 Outils de développement . . . . .	12
2.5.1 Python . . . . .	12
2.5.2 Scripts pré-existants . . . . .	13
2.5.3 Git . . . . .	13
2.5.4 Serveur de travail . . . . .	13
2.6 Méthodologie . . . . .	14
2.6.1 Téléchargement des données satellites . . . . .	14
2.6.2 Chaîne de pré-traitements . . . . .	14
2.6.3 Création des profils temporels . . . . .	16
2.6.4 Prédiction de l'occupation du sol . . . . .	16
2.6.5 Rassemblement de classes pour le calcul des matrices de confusion . . . . .	17
<b>3 Résultats</b>	<b>18</b>
3.1 Profils temporels . . . . .	18
3.1.1 Analyse des classes . . . . .	18
3.1.2 Comparaison des classes . . . . .	23
3.2 Prédiction de l'occupation du sol . . . . .	24
3.3 Arbre de décision retenu . . . . .	24
3.4 Cartographie de l'occupation du sol . . . . .	25
3.4.1 Cartes d'occupation du sol . . . . .	25
3.4.2 Validation de la méthode . . . . .	26
<b>4 Discussion</b>	<b>29</b>
<b>5 Conclusion et perspectives</b>	<b>31</b>
<b>6 Bilan personnel</b>	<b>32</b>
<b>7 Bibliographie</b>	<b>33</b>

# Table des figures

2.1	Localisation de la plaine du Haouz. . . . .	8
2.2	Signatures spectrales de trois types de surfaces . . . . .	10
2.3	Arbre de décision établi par Bouchra BARGAM lors de son stage en 2019. . . . .	14
2.4	Algorithme de calcul du NDVI et d'empilement des images par date . . . . .	15
2.5	Comparaison d'une même image Sentinel-2 masquée et interpolée . . . . .	16
3.1	Profils de NDVI pour la classe 1 en hiver et en été 2018 . . . . .	20
3.2	Profils de NDVI pour la classe 2 en hiver et en été 2018 . . . . .	20
3.3	Profils de NDVI pour la classe 3 en hiver et en été 2018 . . . . .	21
3.4	Profils de NDVI pour la classe 4 en hiver et en été 2018 . . . . .	21
3.5	Profils de NDVI pour la classe 5 en hiver et en été 2018 . . . . .	22
3.6	Profils de NDVI pour la classe 7 en hiver et en été 2018 . . . . .	22
3.7	Comparaison des profils moyens de NDVI pour chacune des classes d'occupation du sol. . . . .	24
3.8	Arbre de décision de 2019 modifié . . . . .	24
3.9	Classification de l'occupation du sol de la plaine du Haouz en hiver 2018. . . . .	25
3.10	Classification de l'occupation du sol de la plaine du Haouz en été 2018. . . . .	26
6.1	Diagramme de Gantt. . . . .	32

# Liste des tableaux

2.1	Classes prédites par l'arbre de décision. . . . .	10
2.2	Bandes spectrales captées par les satellites Sentinel-2. . . . .	11
2.3	Contenu d'un fichier Sentinel-2 L2A de Théia. . . . .	11
2.4	Correspondance entre les classes prédites et les classes relevées sur le terrain. . . . .	12
3.1	Matrice de confusion d'apprentissage en hiver 2018 et son rapport . . . . .	26
3.2	Matrice de confusion de validation en hiver 2018 et son rapport . . . . .	27
3.3	Matrice de confusion d'apprentissage en été 2018 et son rapport . . . . .	27
3.4	Matrice de confusion de validation en été 2018 et son rapport . . . . .	28

# Introduction

En 1998, l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (ONUAA) définit l'Occupation du Sol (OS) comme « la couverture (bio-)physique de la surface des terres émergées ». Il s'agit d'une variable spatialisée importante, car elle procure une connaissance de base pour de nombreux domaines d'application comme des suivis de végétation, des productions statistiques, la planification, les études environnementales et climatiques, la modélisation bio-physique ou encore l'étude de l'activité humaine. Ce vaste champ de domaines d'application explique les travaux importants de cartographie d'occupation du sol réalisés dans la plupart des régions du monde. Seulement sur, le territoire français, plusieurs cartographies de l'occupation du sol sont disponibles (OSO, Corine Land Cover, OCS GE, ...).

## 1.1 Contexte général

### 1.1.1 Structure d'accueil

Le Centre d'Etudes Spatiales de la Biosphère (CESBIO) est une Unité Mixte de Recherche (UMR) créée en 1995 qui est sous la tutelle de plusieurs structures de recherche. Ces tutelles sont :

- le Centre National d'Etudes Spatiales (CNES)
- le Centre National de Recherche Scientifique (CNRS)
- l'Institut de Recherche pour le Développement (IRD)
- l'Université Paul Sabatier (Toulouse 3)

L'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) est également sous contrat avec l'UMR.

L'Observatoire Midi-Pyrénées (OMP) fédère les laboratoires midi-pyrénéens axés sur la recherche pour les sciences de l'univers, de la planète et de l'environnement, dont le CESBIO. Le laboratoire a pour vocation le développement des connaissances sur le fonctionnement de la biosphère continentale et sa relation avec le climat et les activités humaines via une approche principale : la télédétection spatiale. Pour cela plusieurs disciplines scientifiques sont mobilisées, comme l'écologie appliquée à l'agriculture, les sciences de l'environnement ou la physique.

Pour répondre à ses missions, le CESBIO possède deux équipes pour un total de 65 employés permanents et environ le même nombre d'employés en contrat à durée déterminée. La première équipe travaille sur la création de modèles qui décrivent le fonctionnement des surfaces continentales. La deuxième équipe est spécialisée en télédétection et ses missions comptabilisent la cartographie de la végétation, le support de l'équipe de modélisation et des études climatiques et environnementales (CESBIO, 2022).

### 1.1.2 Le projet ALTOS et le LMI TREMA

Les territoires du bassin méditerranéen subissent, en plus des risques grandissants liés aux changements climatiques, de fortes pressions sur leurs ressources en eau. Dans le cas d'application de l'agriculture et pour répondre à la demande croissante d'eau et de nourriture, des solutions doivent être trouvées pour adapter la gestion de l'eau et des terres. Une modification des pratiques d'utilisation des terres agricoles est nécessaire, de manière à obtenir un compromis entre les besoins à divers niveaux. Le projet ALTOS rassemble des chercheurs de l'ensemble du pourtour méditerranéen, dont des chercheurs du CESBIO, dans le but d'améliorer des modèles de gestion de l'eau en agriculture pluviale et irriguée dans cette même zone (Projet Altos, 2022).

Le Laboratoire Mixte International de Télédétection et Ressources en Eau en Méditerranée semi-Aride (LMI TREMA) est un laboratoire structuré autour de la recherche sur le fonctionnement hydrologique des zones semi-arides et sur les indicateurs régionaux du cycle de l'eau. Il travaille en collaboration avec le CESBIO et sa zone d'étude principale est le bassin-versant de Tensift, au centre du Maroc. Son approche scientifique est de combiner la modélisation de processus biophysiques avec de l'observation satellite et des mesures réalisées sur le terrain (LMI TREMA, 2022).

## 1.2 Problématique

Les zones géographiques soumises à un climat aride ou semi-aride font face à des pénuries d'eau de plus en plus fréquente. La région du centre du Maroc est affectée par ce phénomène. Cette situation s'explique par des sécheresses de plus en plus fréquentes et par une demande de ressources en eau croissante. La ville de Marrakech illustre bien ce cas de figure, elle est la ville la plus peuplée et touristique du Maroc. De ce fait, elle utilise des quantités d'eau importantes de manière directe, de par le tourisme et l'usage domestique, mais aussi indirecte, de par son besoin important de nourriture issue de l'agriculture. De plus, le secteur agricole représente plus 40% des emplois du royaume (AgriMaroc, 2021) qui consacre 80% de ses ressources en eau à l'irrigation. Ces ressources en eau sont gaspillées en grande partie par une gestion des ressources inefficace (Benhadj, 2008). Un moyen permettant de proposer des solutions à ce problème est l'étude de fonctionnement hydrologique de ces zones à l'aide de modèles bio-physiques alimentés par des données d'occupation du sol.

Les méthodes de classification de l'occupation du sol les plus performantes actuellement (RF, IA) sont efficaces dans les régions où le volume de données d'apprentissage est important et que son apport est constant. Dans la région centrale du Maroc, ces données n'ont pas la garantie d'être disponible régulièrement. L'occupation du sol étant une variable dynamique dans le temps, une solution doit être trouvée pour faire fonctionner les modèles biophysiques correctement. L'arrivée de satellites d'acquisition d'images terrestres à une fréquence temporelle élevée permet de suivre les changements d'occupation du sol à l'échelle saisonnière. Ces séries temporelles d'images permettraient d'actualiser plus fréquemment les données d'occupation du sol.

## 1.3 Objectifs

Dans ce contexte, le but du travail est de développer une méthode de classification de l'occupation du sol robuste et qui ne nécessite pas forcément de données d'apprentissage. A partir de séries temporelles d'indice de végétation, l'algorithme doit être capable de générer des paramètres de classification optimisés et réutilisables sur des périodes différentes.

Dans un premier temps, le travail a consisté en l'appropriation et le débogage de scripts Python créés et améliorés lors des générations précédentes de stages. Ensuite l'objectif sera de réussir à obtenir une classification sur l'ensemble de la zone d'étude, les stagiaires précédents n'ayant réalisé des tests de classification que sur des zones réduites.

La ligne directrice du stage peut être résumée de la manière suivante : le portage du projet vers l'échelle spatiale supérieure.

# Matériels et méthodes

## 2.1 Zone d'étude

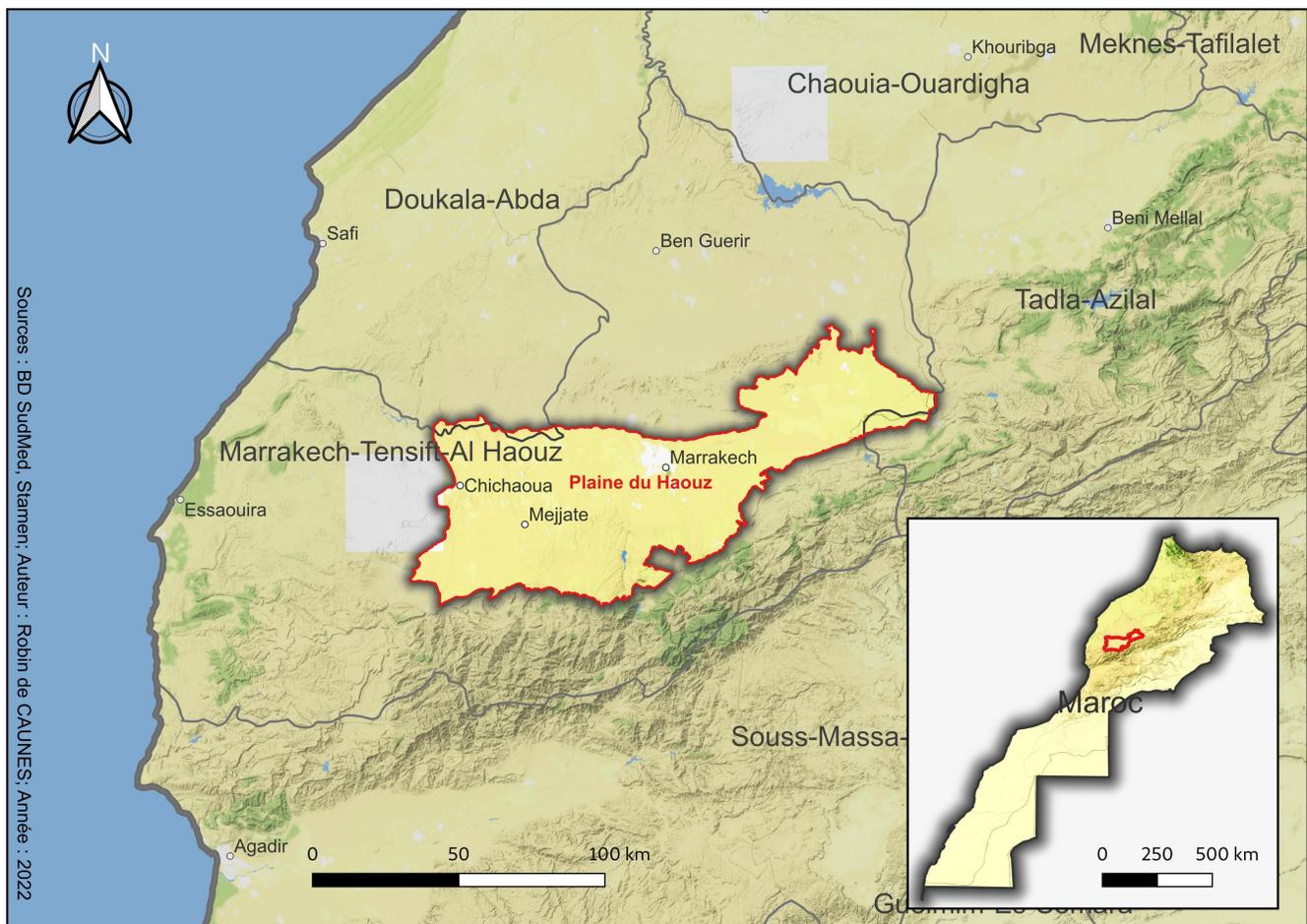


Figure 2.1 – Localisation de la plaine du Haouz.

L'étude a été menée sur la plaine du Haouz qui est située sur le bassin versant du Tensift, au centre-sud du Maroc, entre le petit massif de Jbilet au Nord et le Haut-Atlas au Sud. Elle couvre une surface d'environ 6000 km<sup>2</sup> pour une longueur d'environ 150 km de l'Oued Chichaoua à l'Ouest jusqu'à sa limite Est marquée par l'Oued Tassaout. La commune de Marrakech, située en son centre, est le chef-lieu de la région.

Le relief est caractérisé par un point culminant à une altitude de 1000m sur les contreforts du Haut-Atlas, baissant très progressivement en direction du Nord pour atteindre 300m d'altitude à hauteur du Tensift.

Le paysage de la plaine est considéré comme semi-aride. La pluviométrie annuelle y est faible avec des précipitations inférieures à 300mm/an. Le climat y est chaud et continental, et est marqué par des contrastes de température important au cours de l'année.

Les cultures irriguées, traditionnellement alimentées par un réseau de canaux ouverts appelés "seguia" eux-même alimenté par les oueds en provenance du Haut-Atlas, ont vu leur surface croître au cours du temps ; notamment grâce à un apport d'eau issue d'un canal artificiel acheminant cette dernière depuis plusieurs barrages construits en amont. La plaine étant régulièrement soumise à des épisodes de sécheresse importants, les zones de culture irriguée ne sont pas toujours alimentées par les Oueds. En complément, l'eau est puisée directement dans la nappe phréatique grâce à des forages.

## 2.2 Indice de végétation : NDVI

Le NDVI (Normalized Difference Vegetation Index) est un indice spectral permettant de quantifier la végétation à la surface du sol par télédétection. Il repose sur le contraste des caractéristiques biophysiques de deux canaux issus du spectre lumineux : l'absorption des pigments chlorophylliens dans le rouge et la haute réflectance de la végétation dans le proche infrarouge. Ces canaux sont acquis par des capteurs satellites ou embarqués sur drone. La valeur du NDVI se calcule grâce à l'équation suivante :

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

où  $PIR$  est la valeur de réflectance dans le proche infrarouge et  $R$  est la valeur de réflectance dans le rouge.

Les valeurs obtenues sont comprises entre -1 et 1. Les valeurs négatives correspondent généralement à de l'eau, de la neige ou des nuages. Les valeurs positives faibles (en dessous de 0,1) correspondent à de la roche, du sable ou du sol nu. Les valeurs intermédiaires (comprises entre 0,2 et 0,3) correspondent à une végétation faible de type adventice, arbuste ou prairie. Les valeurs élevées (comprises entre 0,6 et 0,8) correspondent à une végétation dense de type forêt.

Plusieurs raisons nous ont dirigés vers l'utilisation de cet indicateur pour l'algorithme de classification. D'abord, il est fiable et robuste de par sa conception, les seuils entre les valeurs sont clairement définis et interprétables. Ensuite, il est l'indice le plus couramment utilisé en agriculture pour mesurer la santé et la densité de végétation. Cela ajouté à son calcul simple, rend son interprétation compréhensible et accessible au plus grand nombre. Une des limites du NDVI est la saturation de l'indice lorsque la couverture végétale est trop importante. Cette limite a été négligée pour notre cas d'application, car son impact est faible.

## 2.3 Classes à prédire

L'algorithme de classification est conçu dans l'objectif de pouvoir prédire 7 classes d'occupation du sol. Il s'agit du sol nu, des jachères, des cultures d'été, des cultures d'hiver, des arbres sempervirents, des arbres caducs et des arbres avec des cultures. La description de ces classes est consultable dans le tableau 2.1.

## 2.4 Données d'entrée

### 2.4.1 Données satellites

#### Imagerie satellite et télédétection

L'imagerie satellite permet d'acquérir des informations sur les objets terrestres sans contact physique, depuis l'espace. La télédétection passive est une technique d'acquisition d'image qui utilise des capteurs photosensibles pour mesurer la réflectance des surfaces terrestres. Le calcul de la réflectance  $R$  se fait grâce à l'équation suivante :

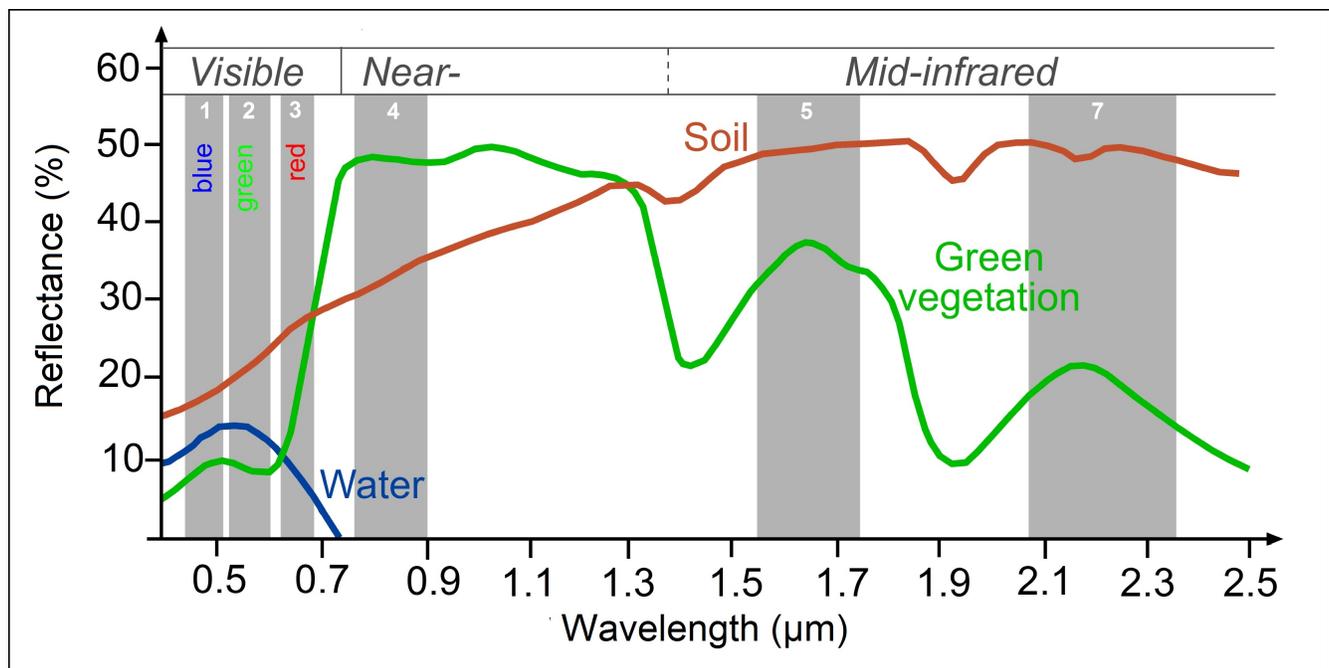
$$R = \frac{\Phi_r}{\Phi_i}$$

il s'agit du rapport entre le flux lumineux réfléchi  $\Phi_r$  et le flux lumineux incident  $\Phi_i$ . La réflectance peut être mesurée pour des longueurs d'onde variées, cela dans le but d'obtenir une signature spectrale pour différentes surfaces terrestres (Figure 2.2). La discrimination de ces signatures est au centre des algorithmes de classification et permet

Identifiant	Nom	Description	Exemples
1	Sol nu	Surface non agricole, dépourvue de végétation.	-
2	Jachère	Sol agricole mis au repos. Absence de végétation ou présence d'adventices mais aucune culture visible.	-
3	Culture d'été	Culture annuelle estivale. Les graines sont plantées à la fin de l'hiver ou au début de l'été. Les plants à maturité sont visibles de mi-avril à mi-octobre.	Aubergine, courgette, melon, pastèque, tomate
4	Culture d'hiver	Culture annuelle hivernale. Les graines sont plantées en début d'hiver. Les plants à maturité sont visibles de mi-février à mi-avril.	Blé, fève, maïs, orge, petit-pois
5	Arbre sempervirent	Surface agricole. Culture annuelle d'arbres. Présence d'arbres fruitiers à feuilles sempervirentes.	Agrume, olivier
6	Arbre avec culture	Surfaces agricole. Il s'agit d'une classe intermédiaire où il a présence d'arbre et de cultures.	-
7	Arbre caduc	Surface agricole. Culture saisonnière en période estivale. Présence d'arbres fruitiers à feuilles caducs.	Abricotier, poirier, pommier, nectarinier, vigne

Table 2.1 – Classes prédites par l'arbre de décision.

de générer des cartes d'occupation du sol.



Source : [https://seos-project.eu/classification/images/spectral\\_signatures\\_landsat.jpg](https://seos-project.eu/classification/images/spectral_signatures_landsat.jpg)

Figure 2.2 – Signatures spectrales de trois types de surfaces : l'eau (bleu), le sol nu (marron) et la végétation (vert) dans le spectre du visible et du proche et moyen infrarouge et canaux 1, 2, 3, 4, 5, 7 des satellites Landsat TM.

### Mission Sentinel-2

La mission Sentinel-2 fait partie du programme Copernicus menée par l'Agence Spatiale Européenne (ESA) et l'Union Européenne. Son objectif est de fournir des clichés de la surface dans 13 bandes spectrales tous les 5 jours. Pour cela l'ESA et ses partenaires ont développé et mis en orbite 2 satellites, Sentinel-2A et Sentinel-2B, respectivement le 23 juin 2015 et le 7 mars 2017 (CNES, 2017) pour une durée de vie de 7 ans. Les bandes captées sont dans le domaine du visible et du proche et moyen infrarouge, la résolution spatiale des images est comprise entre 10m et 60m pour une fauchée de 290km, les spécificités sont visibles dans la table 2.2. L'orbite de

ces satellites est héliosynchrone. Les images Sentinel-2 sont les plus adaptées pour l'étude de la zone d'étude, elles offrent le meilleur compromis de résolution et de temps de revisite.

Bande	Longueur d'onde centrale (nm)	Résolution spatiale (m)	Description
B1	442	60	Aérosol côtier
B2	492	10	Bleu
B3	559	10	Vert
B4	665	10	Rouge
B5	704	20	Végétation red edge
B6	740	20	Végétation red edge
B7	781	20	Végétation red edge
B8	833	10	Proche infrarouge (PIR)
B8a	864	20	Proche infrarouge (PIR) étroit
B9	944	60	Vapeur d'eau
B10	1375	60	Infrarouge court (SWIR) - Cirrus
B11	1612	20	Infrarouge court (SWIR)
B12	2194	20	Infrarouge court (SWIR)

Table 2.2 – Bandes spectrales captées par les satellites Sentinel-2.

### Sélection et obtention des images

Les données satellites utilisées sont des images L2A produites par Théia à partir à partir d'images L1C distribuées par Copernicus. Les images L1C de Copernicus fournissent une réflectance orthorectifiée de la couche supérieure de l'atmosphère (Copernicus, 2022). Elles sont traitées à nouveau par Théia qui applique des corrections atmosphériques et génère des masques pour les nuages et leurs ombres ainsi que l'eau et la neige (Théia, 2022). Les données Théia sont récupérées sous forme d'un fichier ZIP dont l'architecture est visible dans la table 2.3.

		Nombre de bandes	Description
<b>Images</b>	<b>SRE</b>	10	Réflectance du sol sans correction des effets de pente
	<b>FRE</b>	10	Réflectance du sol avec correction des effets de pente
	<b>ATB</b>	2	Paramètres atmosphériques et biophysiques
<b>Dossier masques</b>	<b>CLM</b>	2	Masques des nuages (1 par résolution)
	<b>EDG</b>	2	Masques des bordures (1 par résolution)
	<b>SAT</b>	2	Masques de saturation (1 par résolution)
	<b>MG2</b>	2	Masques géophysique de niveau 2 (1 par résolution)
	<b>IAO</b>	2	Masques de pixels AOT interpolés (1 par résolution)
<b>Métadonnées</b>	<b>MTD</b>	-	-
<b>Fichier quicklook</b>	<b>QKL</b>	-	-

Table 2.3 – Contenu d'un fichier Sentinel-2 L2A de Théia.

Pour réaliser des séries temporelles de NDVI sur la zone d'étude, seules les bandes FRE B4 (rouge) et B8 (PIR) doivent être conservées. Théia offre la possibilité définir un seuil de couverture nuageuse lors de la recherche des images. De manière à obtenir le plus de données possible, aucun filtre de couverture nuageuse n'a été appliqué. Les masques de nuages, de saturation et de bordures ont été également conservés. Un script python développé par Olivier HAGOLLE permet de télécharger automatiquement les données Sentinel-2 à l'aide d'un requête dans un terminal de commande.

### 2.4.2 Données terrain

Depuis 2016, des échantillons d'occupation du sol sont relevés sur la zone d'étude à une fréquence de deux séries par an en moyenne. Une série a lieu au début de la période de culture estivale ce qui permet de relever les cultures du début de l'été ainsi que les cultures héritées de l'hiver. Les relevés se font par observation directe ou par déduction grâce à l'observation de résidus de cultures présents sur les parcelles ou encore grâce à l'interaction directe avec les agriculteurs présents sur site. De la même manière, une autre série de relevés a lieu au début de

la période hivernale, permettant d’observer les cultures du début d’hiver et d’observer ou de déduire les cultures de fin d’été. Il est important de noter que les relevés ne correspondent pas aux classes que l’on cherche à prédire. Les observations notées sur le terrain sont volontairement plus précises. On note notamment les essences d’arbres, l’espace entre ces derniers, le type de plante cultivée ou encore l’âge estimé des jachères. Au retour de mission, ces informations sont utilisées pour déterminer à quelle classe d’occupation du sol le relevé appartient (Tableau 2.4).

Classe relevée	Classe prédite
Sol nu	Sol nu
Jachère	Jachère
Aubergine, courgette, melon, pastèque, tomate	Culture d’été
Blé, fève, orge, petit-pois	Culture d’hiver
Abricot, poire, pomme, maïs, nectarine, vigne	Arbre caduc
Agrume, olivier	Arbre sempervirent
Classe non relevée mais prédite	Arbre sempervirent et culture
Luzerne	Classe relevée mais non prédite

Table 2.4 – Correspondance entre les classes prédites et les classes relevées sur le terrain.

Historiquement, les polygones sont relevés et répertoriés sur une carte papier avec une image Sentinel-2 en une composition colorée de fausses couleurs (R, NIR, SWIR) au format A1, puis reportée dans un SIG en retour de mission. Depuis 2022, les polygones sont relevés directement dans un SIG grâce à une application Android open-source, QField. Cette méthode a pour avantage d’accroître significativement la rapidité des relevés et leur nombre grâce à un suivi GPS en temps réel et une simplification du remplissage de la table attributaire. Elle permet également un gain de temps significatif lors du retour de mission, puisque la couche de relevés peut être aussitôt transférée sur machine.

## 2.5 Outils de développement

### 2.5.1 Python

Le programme de classification est historiquement écrit en langage Python. Ce langage a donc été choisi pour la reprise du travail. La distribution Python utilisée est Anaconda, elle a l’avantage de faciliter l’exportation d’environnements Python d’une machine à l’autre, notamment entre une machine utilisateur et un serveur de travail. L’environnement de développement choisi est *Scientific Python Development EnviRonment* (Spyder), il s’agit d’un environnement de développement Python gratuit et open-source à destination des scientifiques, ingénieurs et analystes de données. Dans le but de réaliser les objectifs du projet, le programme fait appel à plusieurs bibliothèques clés qui sont décrites dans les paragraphes suivants.

#### NumPy

*Numerical Python* (NumPy) est une bibliothèque Python fondamentale pour les scientifiques et les ingénieurs. Son utilisation est basée sur son propre format de structure de données, le *ndarray*. Il s’agit de tableaux multidimensionnels, non dynamiques et qui ne peuvent contenir qu’un type de données homogènes. Les fonctions apportées par NumPy permettent la création et la manipulation des *ndarrays*. Le langage C, utilisé en partie pour écrire la bibliothèque, en fait un outil puissant pour l’application de fonctions mathématiques et de calculs matriciels. Il permet notamment une réduction significative des temps de calcul en comparaison des fonctions basiques offertes par Python. Numpy est également à la base du fonctionnement de plusieurs autres bibliothèques Python comme GDAL, SciPy et scikit-learn qui seront utilisées au cours de ce projet.

#### GDAL

*Geospatial Data Abstraction Library* (GDAL) est une bibliothèque libre maintenue par l’*Open Source Geospatial Foundation* (OSGeo) disponible dans plusieurs langages de programmation, dont Python. Elle permet historiquement de lire et de traiter des images géoréférencées. Désormais, GDAL comprend également la bibliothèque

*OpenGIS Simple Features Reference Implementation* (OGR) qui permet de lire et de traiter des données vectorielles.

### **rasterstats et Rasterio**

Comme son nom l'indique, *rasterstats* est une bibliothèque Python qui vise à calculer des statistiques sur des images géoréférencées. Elle fonctionne de paire avec Rasterio qui est une bibliothèque Python basée sur GDAL. Son objectif est de fournir les mêmes services que GDAL en utilisant plus de caractéristiques du langage Python, GDAL, à l'inverse, repose majoritairement sur le langage C.

### **SciPy**

*Scientific Python* (SciPy) est une bibliothèque Python à destination des scientifiques. Basée sur NumPy, ses fonctions profitent également de performances excellentes grâce à l'utilisation du langage C. Elle contient divers outils pour l'algèbre linéaire, le traitement du signal et de l'image, l'intégration, et plus particulièrement l'interpolation et l'optimisation. Ces deux derniers seront ceux utilisés au cours du projet.

### **scikit-learn**

*scikit-learn* est une bibliothèque Python d'apprentissage machine conçue pour fonctionner en harmonie avec NumPy et SciPy. Elle présente des outils pour l'analyse prédictive de données tels que des algorithmes de classification, de réduction, de régression ou de regroupement. Seules les fonctions de calcul de matrice de confusion et de leur rapport seront utilisées au cours du projet.

## **2.5.2 Scripts pré-existants**

Le projet repose sur des scripts Python générés au cours de plusieurs stages. La première version des scripts a été générée en 2018 par Antoine RENAUDIN, cette version définit l'architecture générale des programmes ; une chaîne de pré-traitements et un algorithme d'optimisation et de classification robuste. Cette première version a été pensée pour fonctionner sur une année complète.

La chaîne de pré-traitement crée des images NDVI à partir des bandes 4 et 8 des images Sentinel-2 et applique de masques pour les nuages et les pixels saturés. Les images sont ensuite empilées pour former un bloc de données ordonné par dates d'acquisition. Suite à cela, une interpolation temporelle du bloc de données est réalisée de manière à obtenir une image par date, grâce à une fonction d'interpolation linéaire. La prédiction de l'occupation du sol est effectuée à partir du bloc de données interpolé à l'aide d'un arbre de décision robuste qui utilise des critères d'experts. En amont de la classification, une optimisation des paramètres d'entrée peut être réalisée à l'aide de données terrain.

Ces scripts ont été repris et améliorés en 2019 par Bouchra BARGAM, la figure 2.3 montre le nouvel arbre de décision pour la prédiction de l'occupation du sol qui a été proposé pour améliorer la classification. Cette nouvelle version a été pensée pour réaliser des classifications à partir de bloc de données par saison agricole (hiver ou été).

## **2.5.3 Git**

Git est un logiciel libre et gratuit de gestion de versions décentralisé. Son principe de fonctionnement repose sur le stockage chronologique de fichiers en conservant la chronologie des modifications effectuées en son sein. Il agit sur une arborescence de fichiers en comparant les ajouts et les délétions entre les mises à jour de ces derniers. Son utilisation la plus commune consiste à faciliter le travail en groupe. Dans le cadre d'application au projet, l'intérêt de Git est simplement de garder un historique de modifications avec les commentaires associés.

## **2.5.4 Serveur de travail**

L'emprise de la zone d'étude est située à l'intersection de 4 tuiles Sentinel-2. Après découpage, elle peut être estimée à une taille de 18500x10000 pixels environ. Avec une résolution temporelle de 1 jour, la taille estimée d'un bloc de données interpolé est de 90Go. Pour permettre de traiter un volume de données aussi conséquent, un serveur de travail a été mis à disposition. Il s'agit d'une machine virtuelle sous Linux disposant de 32Go de

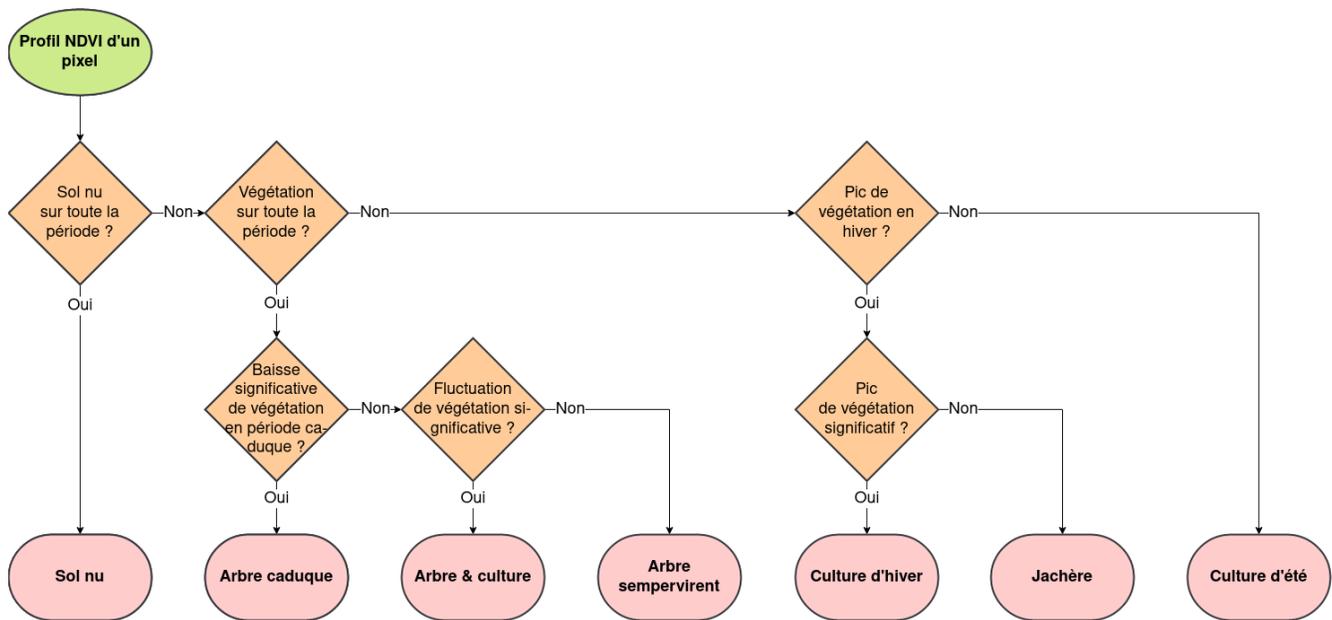


Figure 2.3 – Arbre de décision établi par Bouchra BARGAM lors de son stage en 2019.

mémoire vive ainsi que de plusieurs To de stockage de données à disposition pour les tuiles Sentinel-2 et les blocs de données générés. L'accès à ce serveur a été permis grâce à un protocole SSH et le logiciel d'émulation de terminal MobaXterm.

## 2.6 Méthodologie

L'approche du projet a été inspirée des travaux réalisés sur le même sujet par les stagiaires des années précédentes. A l'aide d'un arbre de décision basé sur des critères de classification robustes et de blocs de données qui contiennent les valeurs de NDVI sur les périodes étudiées, des cartes d'occupation du sol ont été générées sur l'emprise de la zone d'étude.

### 2.6.1 Téléchargement des données satellites

Les données Sentinel-2 pour les 4 tuiles sur la zone d'étude pour l'année 2017-2018 ont été téléchargées sur la plateforme Théia grâce au script de téléchargement automatisé de Olivier HAGOLLE. Les données au format ZIP sont extraites des archives et sont stockées sur le serveur de travail dans 4 répertoires associés aux nom des tuiles en question : T29RNQ, T29RPQ, T29SNR et T29SPR.

### 2.6.2 Chaîne de pré-traitements

Après le téléchargement des données, une phase de pré-traitements est nécessaire avant de pouvoir lancer un algorithme de classification. Cette phase peut être divisée en deux parties. La première permet de calculer le NDVI pour chaque image Sentinel-2. La deuxième partie est une interpolation temporelle de ces images de NDVI.

#### Calcul du NDVI, application des masques et empilement des images

La première partie vise dans un premier temps à calculer des images de NDVI à partir des bandes 4 et 8 des images Sentinel-2. L'indice de NDVI est compris entre -1 et 1. Une variable adaptée pour le stocker serait un *float*. Cependant, par souci d'espace de stockage, il est plus intéressant de le stocker en entier 8-bit positif (*entier compris entre 0 et 255*). Pour faire cela, une fonction permet de transformer la valeur de l'indice de NDVI classique :

$$NDVI_{int} = (NDVI + 1) \times 100$$

Chaque jour où une image est prise, des algorithmes génèrent des masques de données pour les images Sentinel-2. Ces masques se présentent sous la forme de raster. En faisant des opérations sur les images, il est possible de retirer les valeurs aberrantes des images de NDVI. Le masquage a été appliqué pour chacune toutes les dates récupérées. En plus des masques de nuages et de saturation, le masque de bordure a été rajouté à l’algorithme (Figure 2.4).

Une fois le NDVI calculé et les masques appliqués, les valeurs sont triées par chronologie et empilées dans une image géoréférencée. Pour permettre de retrouver les dates par la suite, elles sont inscrites dans la description des bandes. Ces blocs de données empilées seront désignés par *stack*.

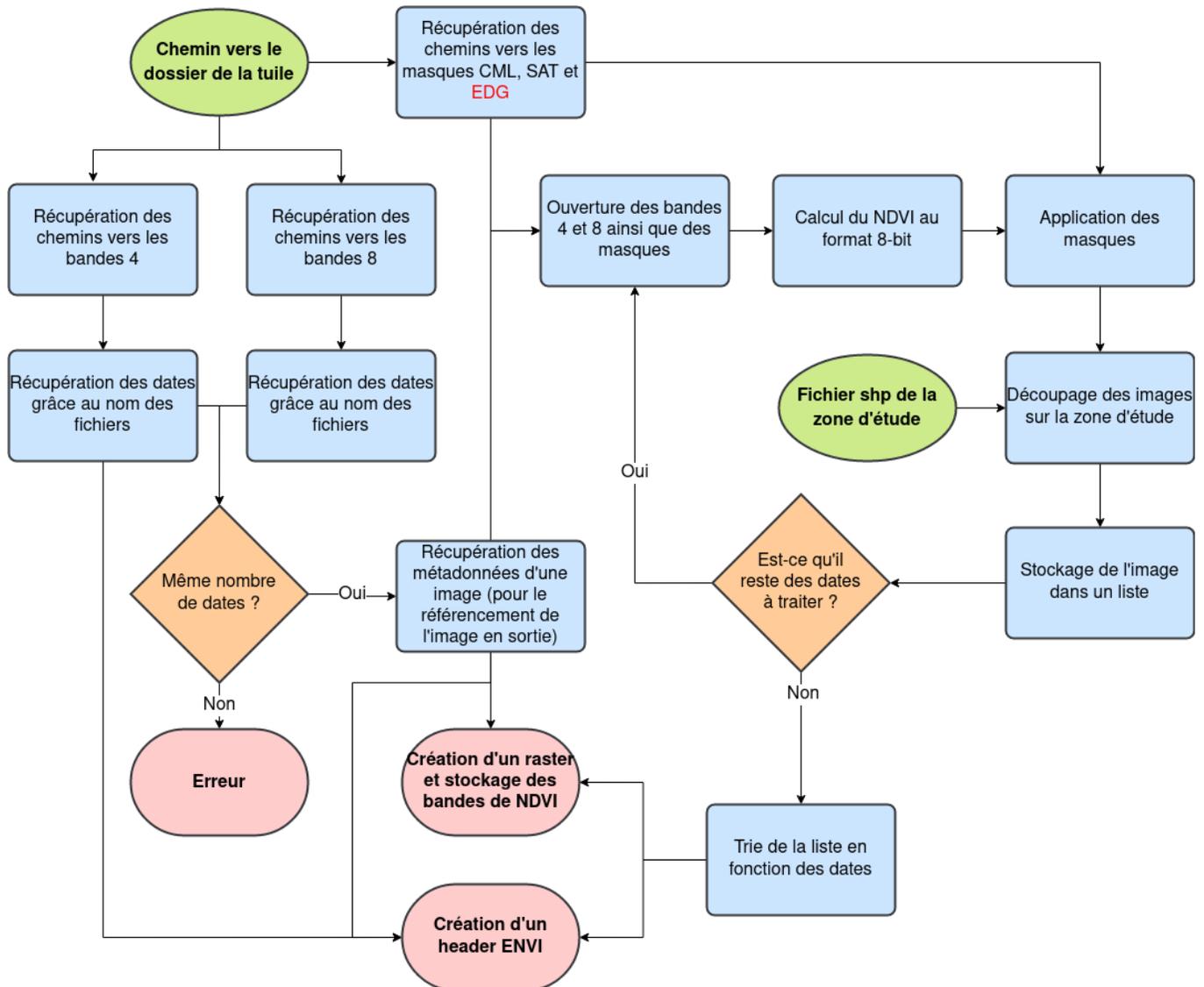


Figure 2.4 – Algorithme de calcul du NDVI et d’empilement des images par date. Le texte en rouge indique les modifications effectuées au cours du projet.

Après la création des *stacks*, il est important de regarder son profil spectral à plusieurs endroits pour déceler d’éventuelles aberrations causées lors de la prise de l’image par les capteurs. Les jours qui présentent des données aberrantes se repèrent à une variation de NDVI forte et ponctuelle. Ces jours devront être retirés. Cette opération a permis de déceler un problème d’échange de dates lors de la génération des *stacks*. Ce problème était dû au téléchargement de fichiers en double. Il peut aussi survenir dans le cas exceptionnel où les deux satellites prennent une image le même jour. Ce problème a été corrigé par la sélection aléatoire d’une seule image

dans le cas où plusieurs seraient disponibles pour la même date.

### Interpolation temporelle journalière

Après avoir généré les *stacks* de NDVI pour chaque tuile Sentinel-2, les valeurs de NDVI sont interpolées sur l'axe temporel à l'aide d'un second script. L'image obtenue est un cube de données avec une résolution temporelle de 1 jour au lieu de 5. L'intérêt de cette interpolation est principalement de remplir les vides laissés par les valeurs masquées comme le montre la figure 2.5.

Cette interpolation est réalisée grâce à la fonction `interpolate.interp1d` de la bibliothèque *Scipy*.

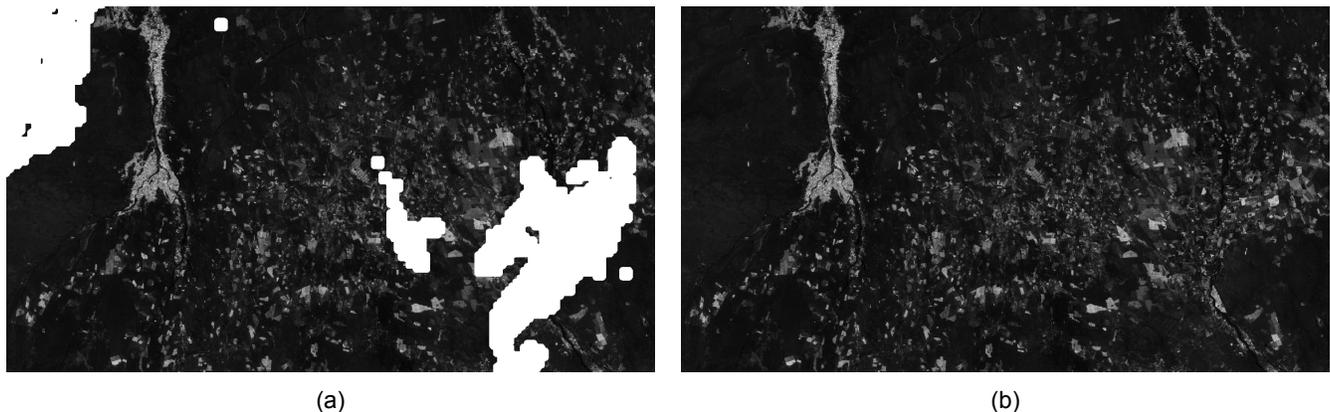


Figure 2.5 – Comparaison d'une même image Sentinel-2 masquée (a) et interpolée (b).

Plusieurs méthodes ont été testées pour créer un cube interpolé sur l'ensemble de la zone d'étude. Le problème réside dans la taille des fichiers à ouvrir et à manipuler (un cube sur 1 tuile Sentinel-2 a une taille d'environ 10 à 15Go), la plupart des ordinateurs ne possédant pas la mémoire vive nécessaire pour ouvrir ces fichiers. La méthode qui a été retenue consiste à créer des cubes interpolés pour les 4 tuiles Sentinel-2 (notés cubes TS2). A l'aide de la bibliothèque *GDAL* et de la fonction *BuildVRT* une mosaïque virtuelle est créée à partir des cubes TS2. Une image virtuelle peut accéder aux données d'une image de référence à l'aide de liens XML. Sur ce principe, une mosaïque virtuelle permet de souder les cubes TS2 sans en lire le contenu directement. De cette façon, il est possible de lire les bandes de l'image virtuelle et de les écrire une à une dans une image physique. Deux blocs sont générés de cette manière, un sur la période hivernale et l'autre sur la période estivale.

### 2.6.3 Création des profils temporels

La lecture des profils temporels par classes d'occupation du sol permet d'observer le comportement de ces classes pour de déterminer des règles de classifications robustes. Pour générer de tels profils à partir des cubes interpolés, la fonction *zonal\_stats* de la librairie *rasterstats* calcule la moyenne des pixels par parcelles d'une classe d'occupation du sol à partir d'une couche vecteur de relevés terrain.

### 2.6.4 Prédiction de l'occupation du sol

Pour prédire l'occupation du sol, l'arbre de décision du stage de 2019 a été repris et modifié. Les détails de la modification sont consultables dans la partie 3.2.

Des modifications sur les fonctions de création des matrices de confusion ont été réalisés, dans le but de permettre à l'utilisateur des réunir ou d'écarter certaines classes d'occupation du sol des résultats de la classification.

Une boucle de tuilage du cube de données d'entrée a été ajoutée au script de classification. L'intérêt du tuilage est de permettre l'ouverture et le traitement d'images trop volumineuses pour la mémoire vive de la machine. Son fonctionnement repose sur l'ouverture des cubes par fenêtres. Ces fenêtres se déplacent sur l'image de manière à récupérer les valeurs du raster progressivement. La fonction de calcul de l'emprise des fenêtres est une version modifiée du script de l'utilisateur « Jon » sur la plateforme StackOverflow.

Un autre problème rencontré lors des premières exécutions du programme de classification livré a été la durée excessive des temps de calculs. Ce problème est dû au mode de calcul utilisé pour la lecture des pixels de l'image. Pour chaque opération, une boucle parcourait la totalité de l'image, pixel par pixel. En réponse à cela, toutes les boucles de calcul ont été remplacées par de fonctions de calculs matriciels issuent de la bibliothèque *NumPy*.

### **Optimisation des critères de classification**

La boucle d'optimisation est une fonctionnalité optionnelle de l'algorithme de classification qui, à l'aide d'un jeu de données terrain et d'une fonction d'optimisation, permet de déterminer les paramètres d'entrée optimales permettant d'obtenir les moins d'erreurs de classement dans la matrice d'apprentissage. Des modifications et du débogage ont été effectués sur son code pour prendre en compte le nouveau système de classification et permettre le traitement de classes sélectionnées par l'utilisateur.

### **2.6.5 Rassemblement de classes pour le calcul des matrices de confusion**

Dans certains cas, il a été demandé à l'algorithme de rassembler certaines classes d'occupation du sol pour calculer les matrices de confusion. Il s'agit des classes pour lesquelles il n'y a pas de données terrain qui sont similaires à une autre classe sur la saison étudiée. Pour toutes les matrices, les arbres avec cultures (classe 6) ont été réunis avec les arbres sempervirents. Dans le cas particulier de l'été, toutes les classes d'arbre (classes 5, 6 et 7) ont été réunies en une seule classe. Il en va de même pour les jachères (classe 2) qui ont été réunies avec les sols nus (classe 1). Malgré l'absence de relevés terrains pour les cultures de saisons opposées aux périodes étudiées, les cultures n'ont pas été réunies car leurs profils d'indice sont bien différents.

# Résultats

## 3.1 Profils temporels

### 3.1.1 Analyse des classes

#### Sol nu

Le profil temporel de NDVI pour la classe de sol nu est globalement linéaire. Ses valeurs sont faibles, avec une moyenne proche de 0,1 et des parcelles pour la majorité en dessous 0,15 (Figure 3.1).

#### Jachère

Le profil temporel de NDVI moyen pour la classe de jachère présente une augmentation modérée en fin d'hiver. Cette augmentation de la moyenne atteint presque une valeur de NDVI de 0,4. Le reste de l'année, la valeur observée stagne entre 0,1 et 0,2. Les courbes des parcelles individuelles sont comprises entre 0,05 et 0,2 en période basse pour atteindre un pic de NDVI variable, compris entre 0,10 et 0,5. Certaines parcelles présentent un profil atypique avec des valeurs qui dépassent 0,5 et qui ne suivent pas la même tendance que les autres parcelles. Ces dernières ont d'ailleurs tendance à voir leur NDVI augmenter sur toute la période hivernale (Figure 3.2).

#### Culture d'été

Les cultures d'été présentent une augmentation de NDVI moyen importante et clairement distincte en début de la période estivale, de début mai à début juillet. Au cours de cette période, l'ensemble des parcelles dépasse une valeur de NDVI de 0,4. En été comme en hiver, le NDVI minimum se situe entre 0,05 et 0,2 en fonction des parcelles. Bien que les tracés des parcelles soient relativement uniformes en période estivale, ce n'est pas le cas en période hivernale. Dans ce dernier cas de figure, on observe une diversité importante dans les profils temporels, avec des profils non seulement linéaires et stables en fonction des saisons, mais aussi des profils dont le NDVI fluctue fortement (Figure 3.3).

#### Culture d'hiver

Les cultures d'hiver présentent une augmentation de NDVI moyen importante et clairement distinctive à la fin de la période hivernale, entre début février et fin avril. Au cours de cette période, une grande majorité des parcelles dépassent une valeur de NDVI de 0,3. Deux périodes sont caractérisées par un NDVI faible, compris entre 0,05 et 0,15, il s'agit de la période de juillet et celle de novembre à janvier. Une faible augmentation de la moyenne de NDVI est visible entre août et novembre, si l'on regarde les tracés des parcelles, on observe deux tendances. Avec d'une part une forte augmentation de NDVI (aux alentours de 0,8) et d'autre part une valeur de NDVI faible et constante comprise entre 0,05 et 0,15 (Figure 3.4).

#### Arbre sempervirent

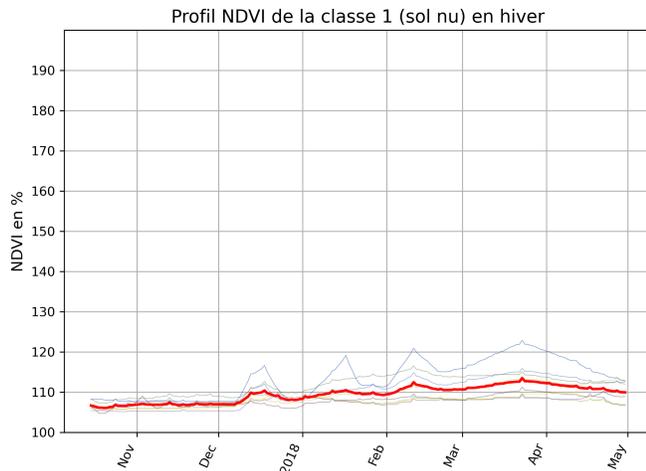
On observe une tendance dans les courbes des profils NDVI parcellaires des arbres sempervirents. Cependant, on observe également une forte variabilité inter-classe. Il y a une différence constante entre les valeurs extrêmes, de l'ordre de 0,6, sur toute la période étudiée. La moyenne de NDVI est relativement constante et toujours supérieure à 0,4. De même, la valeur maximale des profils parcellaires est pratiquement toujours supérieure à 0,4 (Figure 3.5).

### **Arbre caduc**

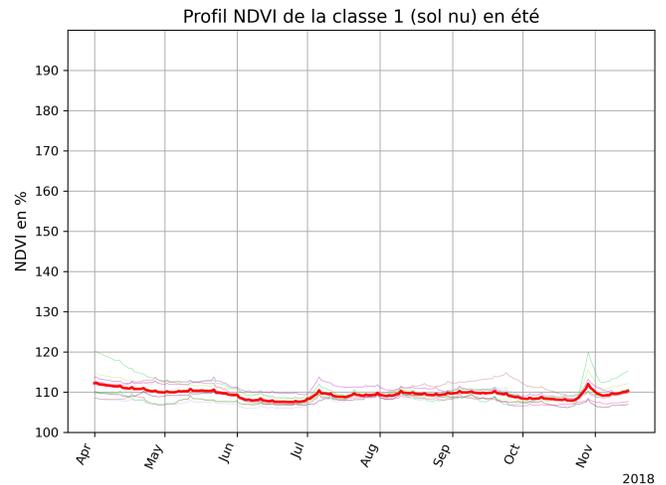
En moyenne, les arbres caducs sont caractérisés par une chute du NDVI en période hivernale. Chacun des profils parcellaires descendent sous la valeur approximative de 0,35 en hiver. La valeur la plus basse est toujours située entre janvier et mi-mars. Hors de la période hivernale, le NDVI est globalement supérieur à 0,4 (Figure 3.6).

### **Arbre et culture**

La classe arbre et culture n'a pas été retenue dans les polygones de vérité terrain. En effet, en mesurant le taux de recouvrement des arbres, les parcelles ont pu être classées soit en culture, soit en arbre. Il n'y a donc pas de résultats visibles pour cette classe.

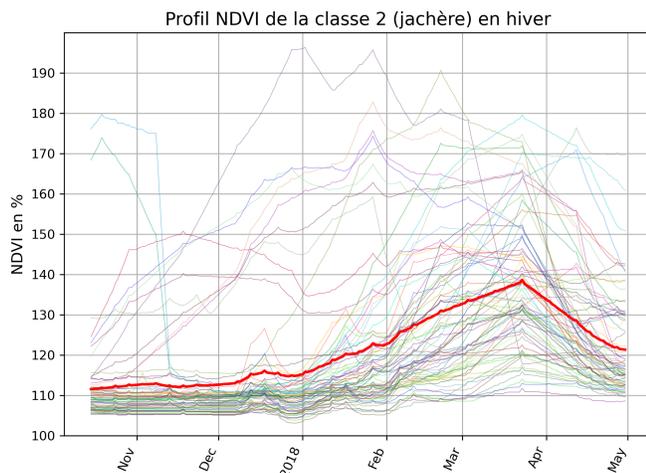


(a)

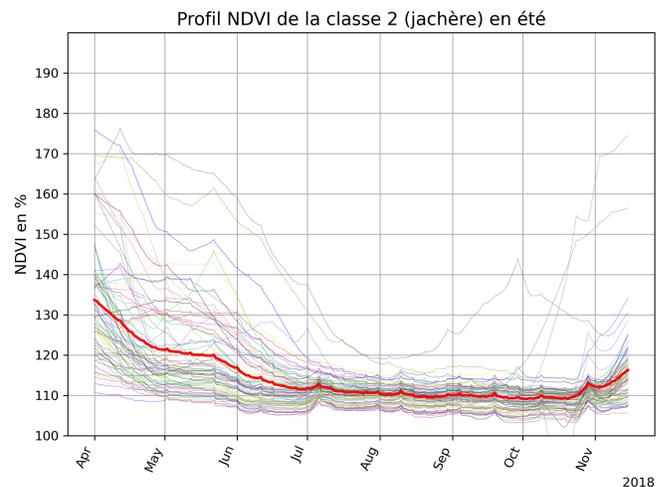


(b)

Figure 3.1 – Profils de NDVI pour la classe 1 en hiver et en été 2018. La ligne épaisse rouge désigne la moyenne de NDVI pour toutes les parcelles observées sur le terrain pour cette classe. Les courbes plus fines représentent les valeurs de NDVI de chaque parcelle observée pour la même classe. Les valeurs de NDVI sont comprises entre 0 et 200%, les valeurs en dessous de 100% correspondent aux valeurs négatives de l'indice.

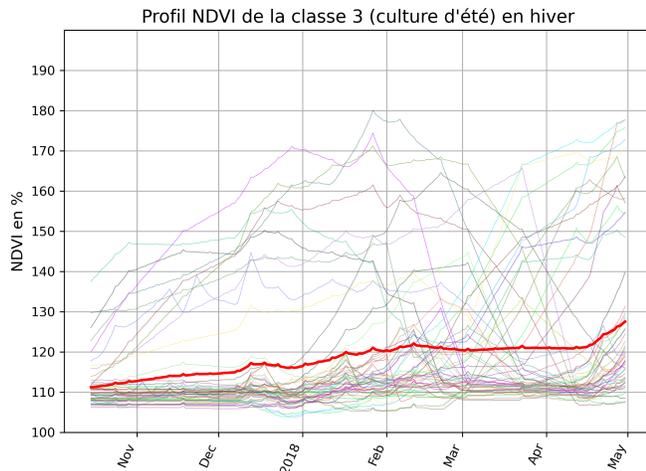


(a)

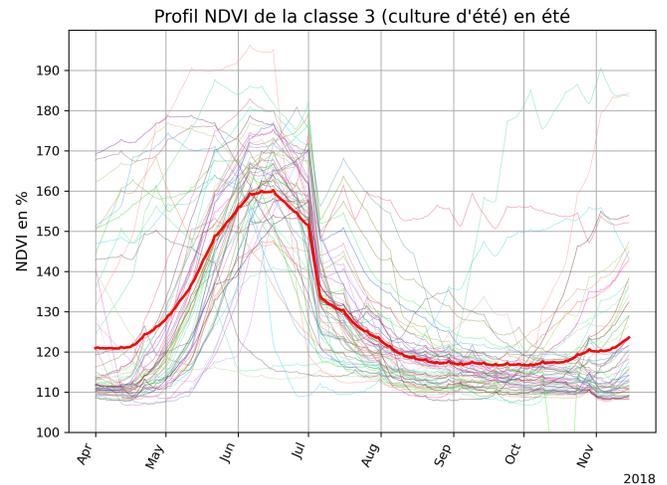


(b)

Figure 3.2 – Profils de NDVI pour la classe 2 en hiver et en été 2018. La ligne épaisse rouge désigne la moyenne de NDVI pour toutes les parcelles observées sur le terrain pour cette classe. Les courbes plus fines représentent les valeurs de NDVI de chaque parcelle observée pour la même classe. Les valeurs de NDVI sont comprises entre 0 et 200%, les valeurs en dessous de 100% correspondent aux valeurs négatives de l'indice.

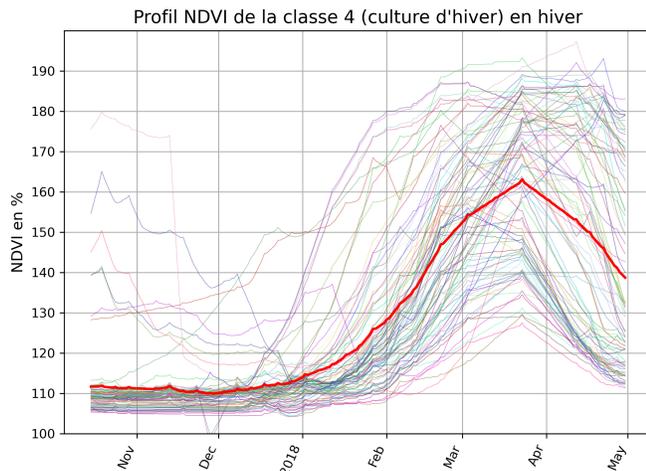


(a)

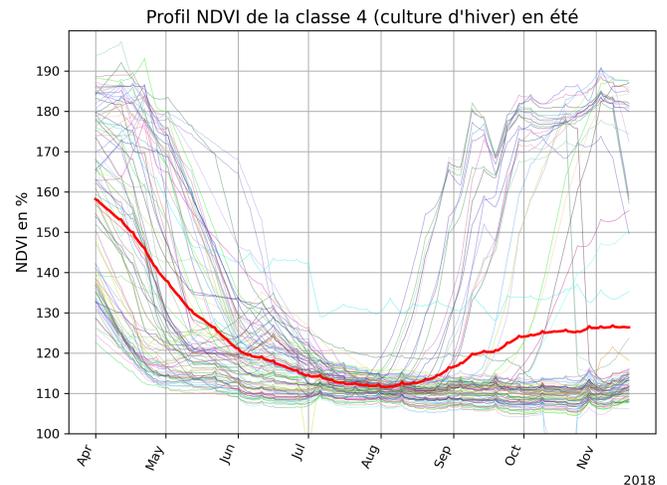


(b)

Figure 3.3 – Profils de NDVI pour la classe 3 en hiver et en été 2018. La ligne épaisse rouge désigne la moyenne de NDVI pour toutes les parcelles observées sur le terrain pour cette classe. Les courbes plus fines représentent les valeurs de NDVI de chaque parcelle observée pour la même classe. Les valeurs de NDVI sont comprises entre 0 et 200%, les valeurs en dessous de 100% correspondent aux valeurs négatives de l'indice.

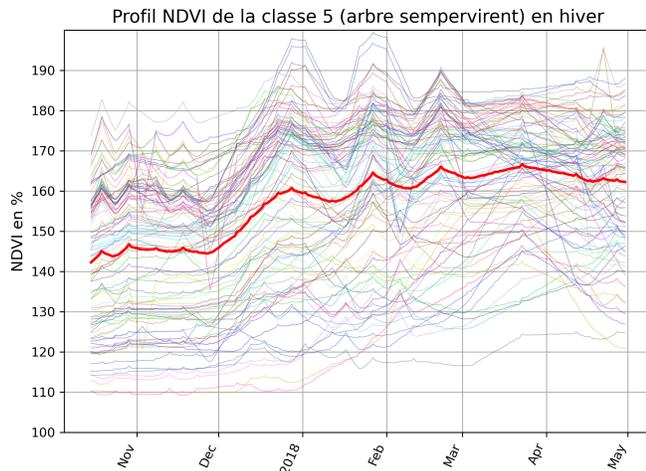


(a)

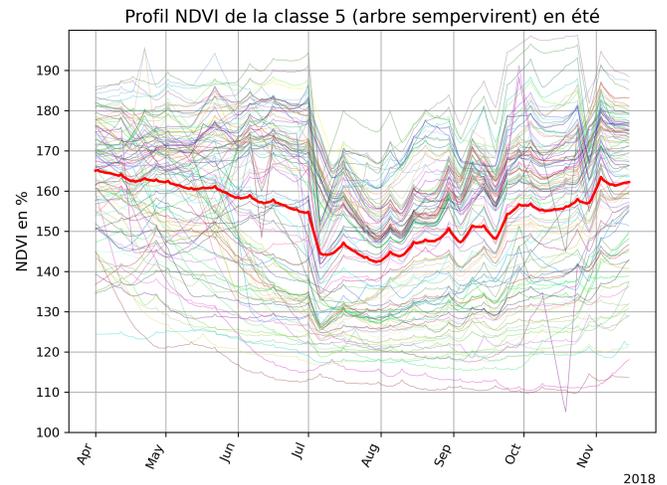


(b)

Figure 3.4 – Profils de NDVI pour la classe 4 en hiver et en été 2018. La ligne épaisse rouge désigne la moyenne de NDVI pour toutes les parcelles observées sur le terrain pour cette classe. Les courbes plus fines représentent les valeurs de NDVI de chaque parcelle observée pour la même classe. Les valeurs de NDVI sont comprises entre 0 et 200%, les valeurs en dessous de 100% correspondent aux valeurs négatives de l'indice.

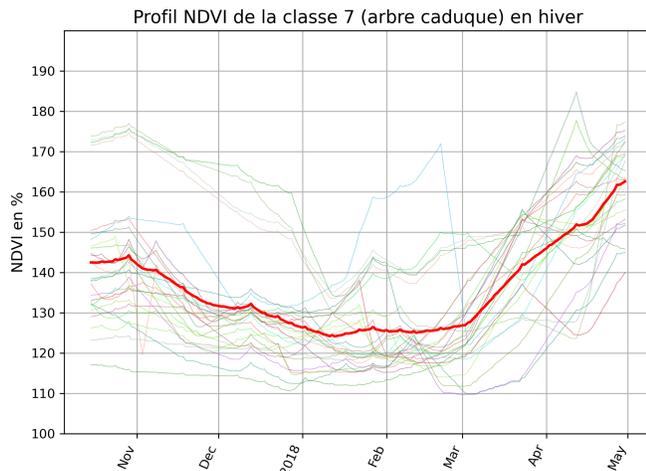


(a)

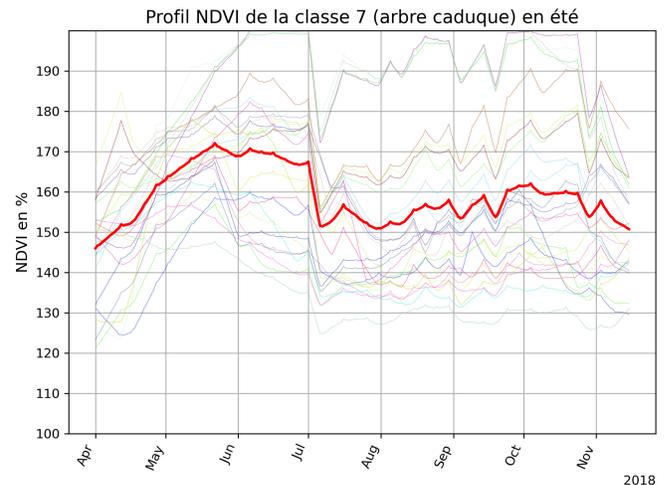


(b)

Figure 3.5 – Profils de NDVI pour la classe 5 en hiver et en été 2018. La ligne épaisse rouge désigne la moyenne de NDVI pour toutes les parcelles observées sur le terrain pour cette classe. Les courbes plus fines représentent les valeurs de NDVI de chaque parcelle observée pour la même classe. Les valeurs de NDVI sont comprises entre 0 et 200%, les valeurs en dessous de 100% correspondent aux valeurs négatives de l'indice.



(a)



(b)

Figure 3.6 – Profils de NDVI pour la classe 7 en hiver et en été 2018. La ligne épaisse rouge désigne la moyenne de NDVI pour toutes les parcelles observées sur le terrain pour cette classe. Les courbes plus fines représentent les valeurs de NDVI de chaque parcelle observée pour la même classe. Les valeurs de NDVI sont comprises entre 0 et 200%, les valeurs en dessous de 100% correspondent aux valeurs négatives de l'indice.

### **3.1.2 Comparaison des classes**

#### **Classes arbre et classes culture**

En période hivernale, il est possible de distinguer les cultures d'hiver des arbres en observant les variations de valeurs de NDVI moyen. Les cultures d'hiver présentent une variation positive marquée par une augmentation de NDVI moyen à la fin de la période (d'environ 0,15 à 0,6). A l'inverse, les profils des arbres caducs subissent une baisse de cette valeur sur l'ensemble de la période (d'environ 0,5 à 0,25). Le comportement spectral des arbres sempervirents est relativement stable tout au long de la période hivernale avec une valeur moyenne de NDVI forte comprise entre 0,45 et 0,7.

En période estivale, ces variations permettent de distinguer les cultures d'été et les classes d'arbres. Les arbres sont identifiables par un profil marqué par des oscillations faibles dans une gamme de valeurs constante comprise entre des valeurs de NDVI moyen allant de 0,4 à 0,7. Tandis que les cultures d'été sont globalement situées entre 0,15 et 0,25, à l'exception de leur pic de végétation (Figure 3.7).

#### **Arbre sempervirent et arbre caduc**

En période hivernale, les arbres sempervirents présentent une moyenne de NDVI forte et constante entre 0,45 et 0,7. Tandis que les arbres caducs présentent une baisse de végétation avec une valeur moyenne de NDVI presque toujours en dessous de 0,4. En période estivale, les profils temporels des deux types d'arbres ne sont pas distinguables, ils présentent une évolution linéaire comprise entre 0,4 et 0,7 (Figure 3.7).

#### **Culture d'hiver et culture d'été**

Les deux types de cultures présentent une augmentation importante et ponctuelle du NDVI pour une période de 2 à 4 mois sur l'année. En période basse leur NDVI est compris en moyenne en 0,1 et 0,2. Lors du pic de végétation, le NDVI moyen dépasse la valeur de 0,5. C'est la date à laquelle le pic survient qui permet de différencier les classes. Pour les cultures d'hiver, le pic a lieu entre mars et mi-avril. Dans le cas des cultures d'été, il apparaît entre mi-mai et mi-juillet (Figure 3.7).

#### **Classe culture et classes jachère et sol nu**

Le profil temporel des jachères partage une tendance avec celui des cultures d'hiver. Ils partagent un pic de végétation à la fin de l'hiver qui n'est pas de la même intensité. Le pic moyen des cultures d'hiver dépasse une valeur de 0,6, tandis que le pic moyen des jachères atteint frôle une valeur de 0,4. En période basse, les profils sont compris entre 0,1 et 0,2. En été, les jachères le profil des jachères reste linéaire tandis que les cultures d'été voient leur NDVI grimper de manière importante entre mi-mai et mi-juillet (Figure 3.7).

#### **Jachère et sol nu**

La classe jachère et la classe sol nu partagent un profil de NDVI moyen similaire en été. Il est linéaire et compris entre 0,05 et 0,15. Une distinction entre les classes est possible en hiver car, à l'inverse du sol nu qui reste linéaire et compris dans la même gamme de valeurs, les jachères voient une augmentation modérée de leur moyenne de NDVI en fin de période hivernale (Figure 3.7).

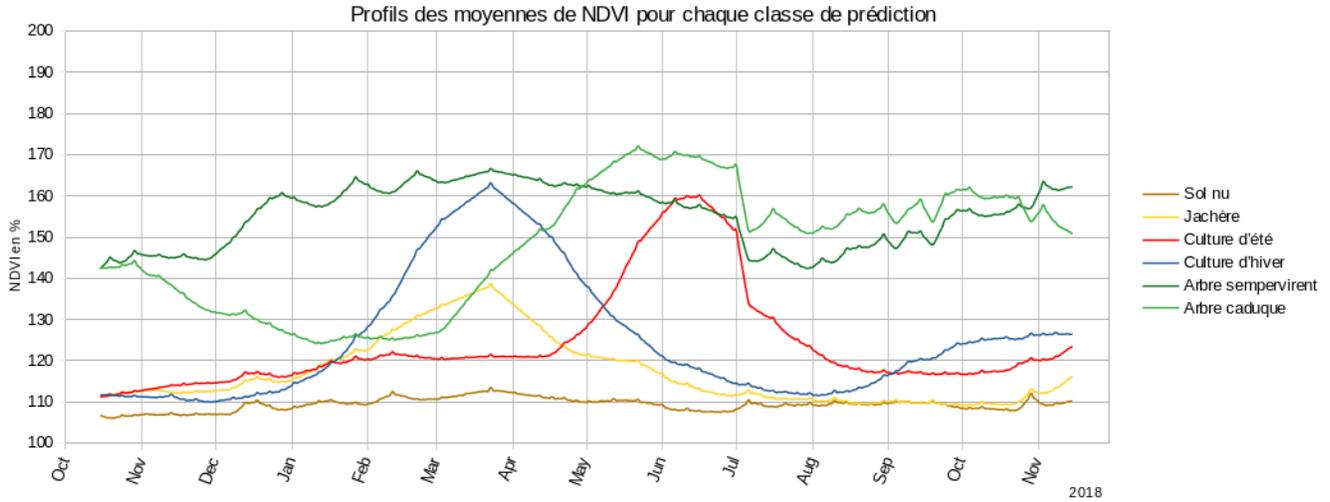


Figure 3.7 – Comparaison des profils moyens de NDVI pour chacune des classes d’occupation du sol.

### 3.2 Prédiction de l’occupation du sol

Les modifications apportées au script de classification de l’occupation du sol ont permis une réduction du temps de calcul d’un facteur 15. Le temps pour obtenir une classification avec des paramètres optimisés en utilisant le script non modifié était de 36 heures. Le passage de bouclage pixel par pixel aux calculs matriciels ont permis de réduire ce temps à 2 heures et 30 minutes environ.

### 3.3 Arbre de décision retenu

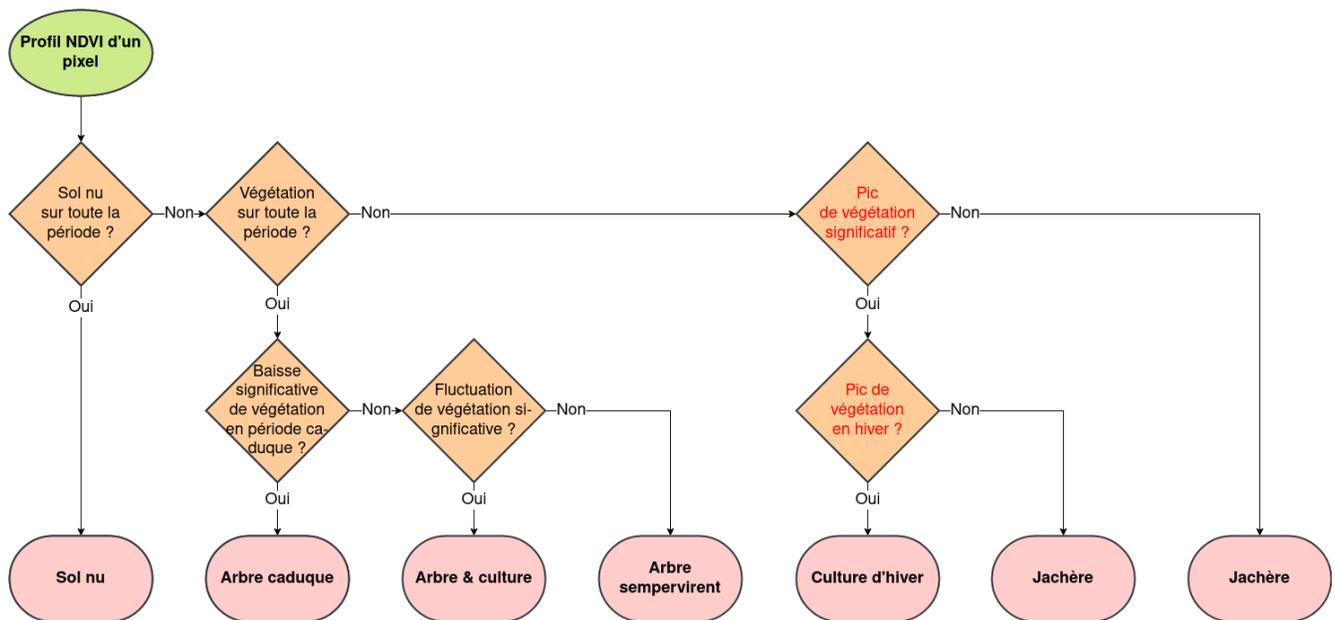


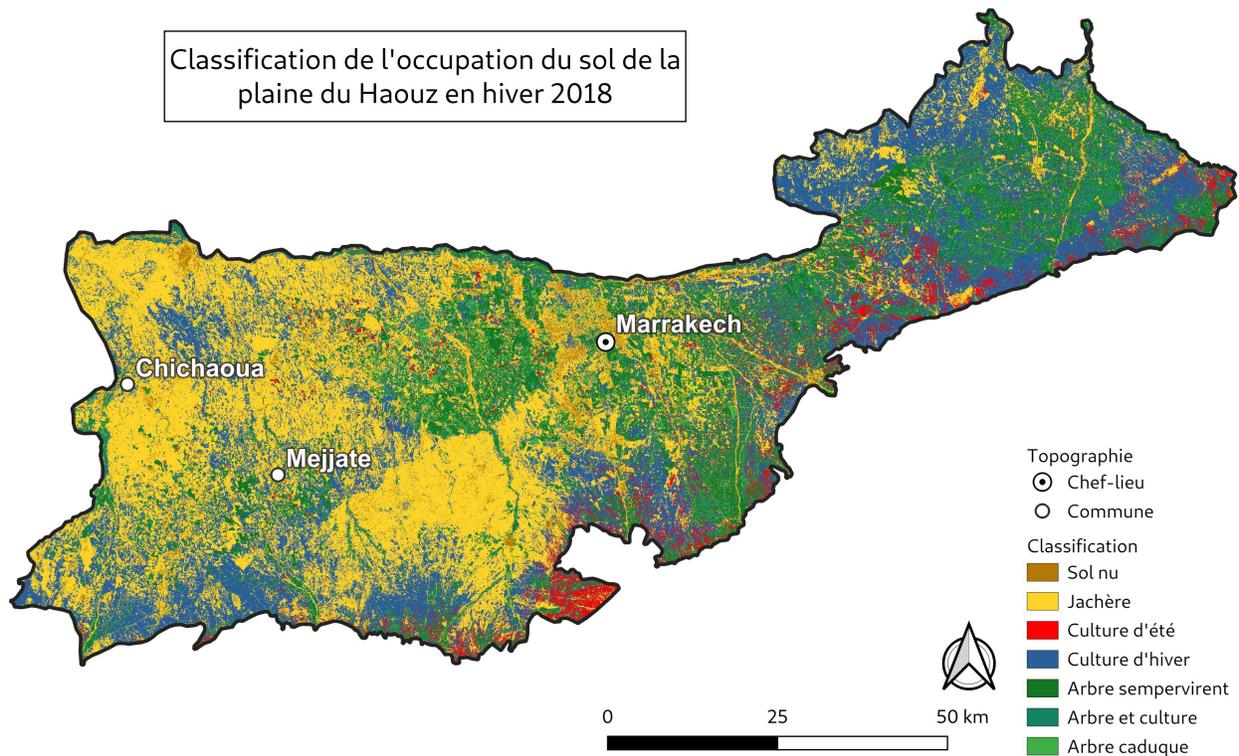
Figure 3.8 – Arbre de décision de 2019 modifié. Les modifications apportées sont visibles en rouge.

Des suites à un problème de sur-représentation de la classe culture d'été lors de la période estivale. L'arbre de décision de 2019 a été modifié de manière tester la significativité des pics de végétation avant de regarder la période à laquelle le pic à lieu.  
De manière générale, l'arbre de décision fonctionne à l'aide de tests logiques sur le profil NDVI des pixels du cube de données. Pour cela, des seuils d'indice et des périodes lui sont fournis en entrée de l'algorithme.

### 3.4 Cartographie de l'occupation du sol

#### 3.4.1 Cartes d'occupation du sol

Les cartes d'occupation du sol consultables sur les figures 3.9 et 3.10, permettent clairement de faire une distinction entre la partie Ouest et la partie Est de la zone d'étude. L'Est est marqué pour une forte présence des classes arborées pour les deux saisons. On observe le même phénomène avec la classe jachère sur la partie Ouest de la carte. La période estivale est marquée par une plus grande présence de la classe de culture d'été par rapport à l'hiver, surtout dans le secteur ouest de Marrakech.



Source : BD SudMed, Auteur : Robin de CAUNES, Année : 2022

Figure 3.9 – Classification de l'occupation du sol de la plaine du Haouz en hiver 2018.

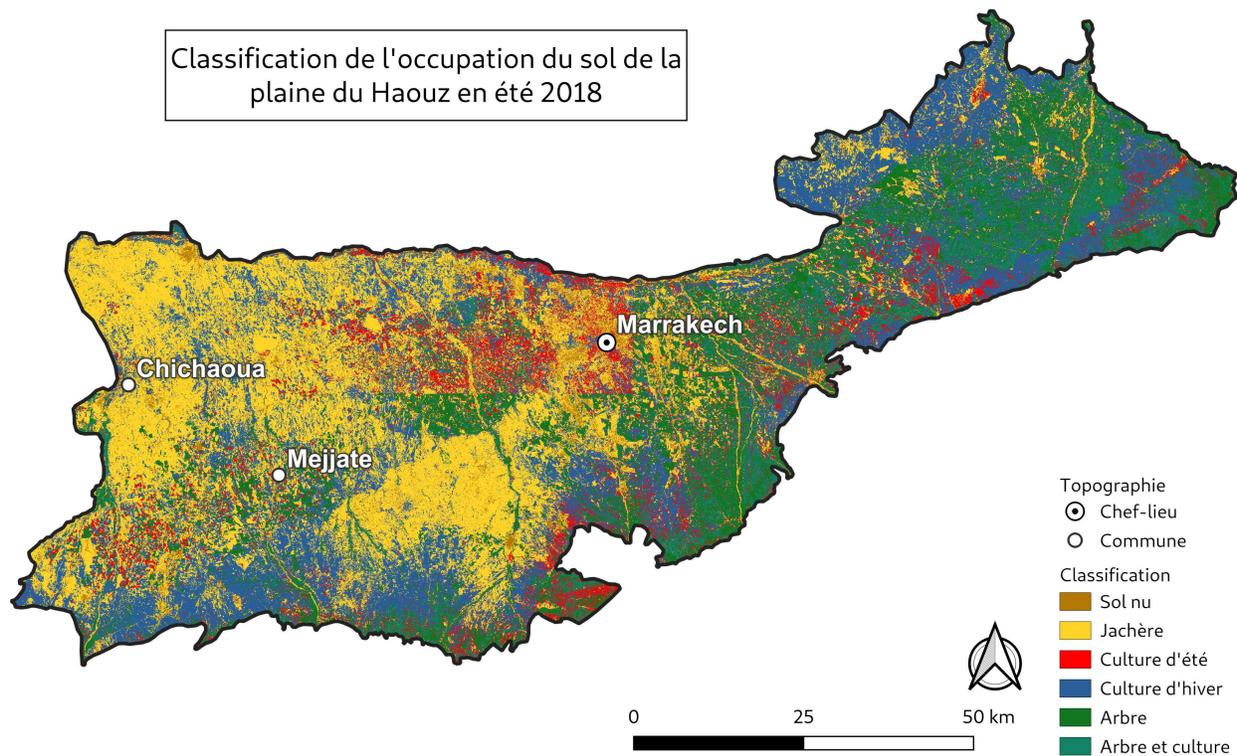


Figure 3.10 – Classification de l'occupation du sol de la plaine du Haouz en été 2018.

### 3.4.2 Validation de la méthode

#### Apprentissage hiver

Apprentissage hiver		Prédiction						
		1	2	3	4	5	7	
Réalité terrain	1	6	5	0	0	0	0	
	2	1	23	0	11	2	1	
	3	0	0	0	0	0	0	
	4	0	6	6	33	1	0	
	5	0	0	0	2	47	2	
	7	0	0	2	1	1	9	

(a)

	precision	recall	f1-score	support
1	0.86	0.55	0.67	11
2	0.68	0.61	0.64	38
3	NaN	NaN	NaN	NaN
4	0.7	0.72	0.71	46
5	0.92	0.92	0.92	51
7	0.75	0.69	0.72	13

<b>accuracy</b>	0.74	<b>support</b>	159
-----------------	------	----------------	-----

(b)

Table 3.1 – Matrice de confusion d'apprentissage en hiver 2018 (a) et son rapport (b). Correspondance des classes : 1 = sol nu, 2 = jachère, 3 = culture d'été, 4 = culture d'hiver, 5 = arbre sempervirent ou arbre avec culture, 7 = arbre caduc.

La matrice d'apprentissage en hiver (figure 3.1) présente un accord global de 0,74 pour 159 parcelles étudiés. La classes 5 présente un F-score de 0.91 tandis que les autres classes se situent entre 0,64 et 0,72. La classe 4 n'a pas été observée sur le terrain mais a été prédite.

## Validation hiver

Validation hiver		Prédiction					
		1	2	3	4	5	7
Réalité terrain	1	810	4990	0	13	1	0
	2	457	9690	66	2149	386	118
	3	0	0	0	0	0	0
	4	0	1020	291	3623	22	49
	5	0	0	2	129	8540	479
	7	0	0	340	18	50	1816

(a)

	precision	recall	f1-score	support
1	0.64	0.14	0.23	5814
2	0.62	0.75	0.68	12866
3	NaN	NaN	NaN	NaN
4	0.61	0.72	0.66	5005
5	0.95	0.93	0.94	9150
7	0.74	0.82	0.78	2224

<b>accuracy</b>	0.7	<b>support</b>	35059
-----------------	-----	----------------	-------

(b)

Table 3.2 – Matrice de confusion de validation en hiver 2018 (a) et son rapport (b). Correspondance des classes : 1 = sol nu, 2 = jachère, 3 = culture d'été, 4 = culture d'hiver, 5 = arbre sempervirent ou arbre avec culture, 7 = arbre caduc.

La matrice d'apprentissage en hiver (figure 3.1) présente un accord global de 0,7 pour 35059 pixels étudiés. La classe 5 présente un F-score de 0,94, la classe 1 présente un F-score de 0,23 tandis que les autres classes se situent entre 0,68 et 0,78. La classe 4 n'a pas été observée sur le terrain mais a été prédite. Le rappel de la classe 1 est extrêmement faible, les pixels ont majoritairement été classés dans la classe 2. Les classes 2 et 5 présentent un nombre de pixels traités bien plus important que les autres classes.

## Apprentissage été

Apprentissage été		Prédiction			
		1	3	4	5
Réalité terrain	1	7	0	0	0
	3	2	25	0	1
	4	0	0	0	0
	5	1	0	2	48

(a)

	precision	recall	f1-score	support
1	0.7	1	0.82	7
3	1	0.89	0.94	28
4	NaN	NaN	NaN	NaN
5	0.98	0.94	0.96	51

<b>accuracy</b>	0.93	<b>support</b>	86
-----------------	------	----------------	----

(b)

Table 3.3 – Matrice de confusion d'apprentissage en été 2018 (a) et son rapport (b). Correspondance des classes : 1 = sol nu ou jachère, 3 = culture d'été, 4 = culture d'hiver, 5 = arbre ou arbre avec culture.

La matrice d'apprentissage en été (figure 3.3) présente un accord global de 0,7 pour 86 parcelles étudiés. Les classes présentent un F-score supérieur à 0,82 tandis que la classe 4 n'a pas été observée sur le terrain mais a été prédite.

## Validation été

Validation été		Prédiction			
		1	3	4	5
Réalité terrain	1	8788	12	690	0
	3	128	6806	241	680
	4	0	0	0	0
	5	53	195	256	21114

(a)

	precision	recall	f1-score	support
1	0.98	0.93	0.95	9490
3	0.97	0.87	0.92	7855
4	NaN	NaN	NaN	NaN
5	0.97	0.98	0.97	21618

<b>accuracy</b>	0.94	<b>support</b>	38963
-----------------	------	----------------	-------

(b)

Table 3.4 – Matrice de confusion de validation en été 2018 (a) et son rapport (b). Correspondance des classes : 1 = sol nu ou jachère, 3 = culture d’été, 4 = culture d’hiver, 5 = arbre ou arbre avec culture.

La matrice de validation en été (figure 3.4) présente un accord global de 0,94 pour 38863 pixels étudiés. Les classes présentent un F-score supérieur à 0,92 tandis que la classe 4 n’a pas été observée sur le terrain mais a été prédite. La classe 5 présentent un nombre de pixels traités bien plus important que les autres classes.

# Discussion

Les profils spectraux atypiques de certaines parcelles de cultures d'hiver en été (figure 3.4b) et de certaines parcelles de cultures d'été en hiver (figure 3.3a) sont des données inévitables car invisibles lors des relevés terrain. Ces périodes pour ces classes sont donc à éviter pour tout critères de classification. Les pics d'indice observés sur ces cultures sont par contre la manifestation d'une forte activité chlorophyllienne qui correspond à la maturation des cultures associées.

Les jachères présentent une hausse d'activité chlorophyllienne étalée du milieu de l'hiver jusqu'à une partie de l'été (figure 3.2). La prise en compte de cette valeur maximale dans le cadre d'une classification uniquement sur la partie estivale doit être prise en compte pour éviter les confusions avec les cultures d'été. La différenciation entre ces deux classes étant possible en comparant l'intensité du pic de végétation sur la période étudiée (figure 3.3b). La variabilité importante de l'amplitude des profils d'arbres intra-classes (figures 3.5 et 3.6) est explicable par le taux de recouvrement des arbres. En effet, sur le terrain, il est difficile depuis la fenêtre du véhicule d'estimer la distance entre les arbres à l'arrière des parcelles. De ce fait, certaines parcelles n'auraient pas dû être relevées du fait de leur non-représentativité de la classe. Autrement, la chute de NDVI hivernale associées aux arbres caducs est associée à l'expression de l'apoptose des feuilles. Ces dernières perdent leur activité chlorophyllienne et finissent par tomber.

En comparant, les profils estivaux pour les arbres (figures 3.5b et 3.6b) ont peut prévoir une confusion dans la classification des deux classes d'arbres en été. Il en va de même pour la classe de sol nu et celle de jachère pour la même période (figures 3.1b et 3.2b).

Les cartographies d'occupation du sol (figures 3.9 et 3.10) présentent une forte démarcation entre l'Est et l'Ouest. Cela illustre parfaitement le caractère semi-aride de la région. En effet, la présence importante des classes d'arbres sur la partie Est s'explique pour l'irrigation bien plus intensive dans ce secteur. En effet, les barrages présents plus à l'Est en Atlas permettent l'acheminement d'eau par un canal artificiel. Plusieurs autres hypothèses peuvent être tirées concernant l'acheminement de l'eau dans ce secteur, comme des Oueds qui ont tendance à moins s'assécher ou encore la présence de forages qui puisent l'eau dans la nappe phréatique. A l'inverse, le secteur Ouest de la zone d'étude est beaucoup plus sec. Les cultures y sont moins irriguées et il est plus difficile d'y faire pousser les cultures. La raison du manque d'irrigation peut être l'épuisement de la nappe phréatique, et l'assèchement des Oueds en été. Une raison qui peut être minime, mais qui mérite d'être relevée a été l'observation de cultures sous serres et le recouvrement des vignes par des filets dans cette zone. Il se peut que le plastique réduise l'intensité de la réflectance captée par les satellites.

Les matrices de confusion pour l'hiver (figures 3.1 et 3.2) montrent des résultats encourageants pour la suite du projet. Les classes arbres sont très bien détectées et différenciées. La détection de la classe culture est satisfaisante 60% des pixels sont bien classés. Les 40% restants ont été classés comme jachère. Ce résultat n'est pas surprenant, car sur ce type d'occupation, la photosynthèse des adventices peut fortement augmenter sur la même période que les cultures hivernales. L'utilisation d'un autre indice pourrait permettre une meilleure distinction entre les classes. La classe jachère est d'ailleurs fortement sur-représentée, car elle récupère également des pixels de sol nu. Un biais que l'on peut relever dans la méthode est la sur-représentation de la classe jachère dans les échantillons et la faible représentation de la classe sol nu. Réduire le nombre d'échantillons de manière aléatoire permettrait peut-être d'améliorer la détection de la classe sol nu.

Les matrices de confusion pour l'été (figures 3.3 et 3.4) montrent des résultats plus qu'encourageants. Des statistiques aussi bonnes s'expliquent par la réunion de classes « problématiques » quand on essaye de les détecter individuellement, mais très bien détectées par l'algorithme en tant que classe unique. L'analyse des profils temporels a laissé supposer l'arbre de décision ne pourrait pas différencier les classes d'arbres caducs et sempervirents

en été ; c'est bien le cas, car elle n'apparaît pas sur l'image de classification. L'algorithme sait très bien détecter les arbres en général, et il en va de même pour la classe sol nu et la classe jachère qui partagent le même profil temporel en été.

# Conclusion et perspectives

Le sujet du stage portait sur le développement d'une méthode de classification de l'occupation du sol robuste, capable de fonctionner sans données d'apprentissage si l'utilisateur le souhaite. L'outil développé était déjà capable de déterminer l'occupation du sol à partir d'un indice de NDVI calculé à l'aide d'images Sentinel-2. L'objectif principal était de déboguer le code et de porter le projet à l'échelle spatiale supérieure.

L'analyse des profils temporels de NDVI a permis de renforcer l'arbre de décision établi en 2019. De plus, la classification de l'occupation du sol a été réalisée à l'échelle désirée, sur l'ensemble de la zone d'étude. Ensuite, la validation a montré des résultats optimistes pour la suite du projet.

Les cartes d'occupation du sol ont été générées pour l'année 2018, il serait intéressant de reproduire l'opération pour les années où l'on peut entraîner l'algorithme, de 2016 à 2022. De même, il serait intéressant de chercher d'autres méthodes de classification qui permettraient de faire plus de discrimination entre les classes. Voire même de distinguer plus des classes d'occupation du sol, au sein des cultures, par exemple. Pour cela, l'utilisation de classifier random forest pourrait se révéler utile.

La suite du projet va porter sur un changement de résolution temporelle, l'objectif sera de faire une classification sur plusieurs années, pour savoir s'il possible obtenir des paramètres de classification valables pour chaque année. De cette façon, la classification d'occupation du sol en milieu semi-aride ne reposera plus uniquement sur des données relevées sur le terrain et les modèles bio-physiques du CESBIO et du LMI TREMA pourront être alimentés en données d'occupation du sol.

# Bilan personnel

Ce stage au sein du CESBIO m'a permis de travailler dans un environnement de recherche et développement que j'ai beaucoup apprécié. Le sujet a été intéressant et m'a passionnée, car il touche à des problématiques que je trouve subjectivement très importantes. J'ai beaucoup appris sur la méthodologie et l'approche d'un sujet dans un contexte de recherche, même si mes missions portaient sur du technique. J'avais des appréhensions sur mes capacités avant de débiter le stage et je suis fier d'avoir trouvé les solutions aux problèmes que j'ai rencontrés par moi-même.

J'ai pu mettre en application des notions qui remontent à ma licence de biologie-écologie, et à mon parcours universitaire dans la géomatique, notamment dans le master SIGMA ; comme la mobilisation de connaissances en calcul d'indices spectraux, la classification supervisée et non supervisée et l'utilisation de langage de programmation dans la réalisation de tâches géographiques. J'ai enfin pu découvrir le fonctionnement de bibliothèques Python fondamentales, comme NumPy, GDAL, SciPY et scikit-learn. J'ai également pu me familiariser avec l'environnement Linux et les serveurs de travail à distance. La rédaction du rapport a également été l'opportunité d'apprendre le langage Latex qui ne m'était pas familier par avant.

Ce stage a aussi été l'occasion de travailler dans des conditions de travail exceptionnelles, flexibles et toujours dans la bonne humeur. Le prêt de matériel de pointe a été très appréciable. Travailler au CESBIO a été pour moi l'occasion de travailler dans une structure importante de la télédétection en France. Cette expérience m'a permis de confirmer mon souhait de faire de la télédétection ma profession.

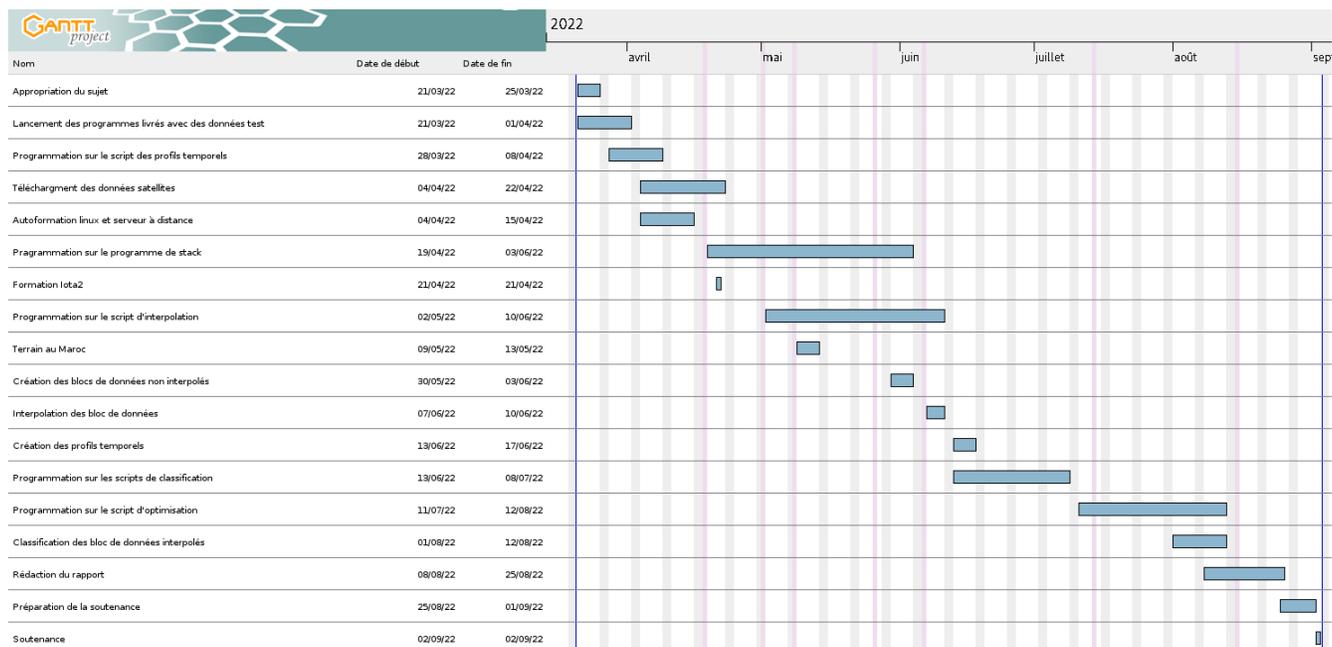


Figure 6.1 – Diagramme de Gantt.

# Bibliographie

- CESBIO | Centre d'Etudes Spatiales de la Biosphère. url : <https://www.cesbio.cnrs.fr/> (visité le 24/08/2022)
- Altos | Home. url : <https://www.cesbio.cnrs.fr/> (visité le 24/08/2022)
- LMI TREMA | LMI. url : <https://www.lmi-trema.ma/> (visité le 24/08/2022)
- AgriMaroc | Le secteur agricole au Maroc. 2021. url : <https://www.agrimaroc.ma/secteur-agricole-au-maroc/> (visité le 24/08/2022)
- Iskander Benhadj. "Observation spatiale de l'irrigation d'agrosystèmes semi-arides et gestion durable de la ressource en eau en plaine de Marrakech". These de doctorat. Toulouse 3, 1er jan. 2008
- CNES | Sentinel-2. url : <https://sentinel2.cnes.fr/fr/> (visité le 24/08/2022)
- CNES | Sentinel-2A L2A products description. url : [https://www.theia-land.fr/wp-content/uploads/2018/12/SENTINEL-2A\\_L2A\\_Products\\_Description.pdf](https://www.theia-land.fr/wp-content/uploads/2018/12/SENTINEL-2A_L2A_Products_Description.pdf) (visité le 24/08/2022)
- THEIA | Sentinel-2 Surface Reflectance. url : <https://www.theia-land.fr/en/product/sentinel-2-surface-reflectance/> (visité le 24/08/2022)
- Copernicus | Sentinel-2 MSI Technical Guide. url : <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi> (visité le 24/08/2022)
- V. Simonneaux, B. Duchemin, D. Helson, S. Er-Raki, A. Olioso & A. G. Chehbouni (2008) The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco, International Journal of Remote Sensing, 29 :1, 95-116, DOI : 10.1080/01431160701250390