

Faculté des bioingénieurs

Prédiction des cultures de printemps dans le contexte de la création d'un système d'alerte précoce pour les inondations boueuses

Auteur : Houdmont Pierre

Promoteurs : Prof. Charles Bielders (UCLouvain/ELI)
Prof. Pierre Defourny (UCLouvain/ELI)

Lectrices : Prof. Aurore Degré (Uliège-Gembloux ABT)
Sophie Bontemps (UCLouvain/ELI)

Année académique 2019-2020

Mémoire de fin d'études présenté en vue de l'obtention du diplôme de
Bioingénieur : sciences et technologies de l'environnement

Remerciement

En premier lieu, je tiens à remercier chaleureusement mon promoteur Charles Bielders ainsi que mon co-promoteur Pierre Defourny, qui m'ont éclairé de leurs connaissances tout au long de ce mémoire, et qui ont rendu possible sa réalisation sur une période réduite. Charles Bielders a assuré un suivi de l'avancement de ce mémoire, de façon exemplaire, par de nombreuses réunions passées à décortiquer les améliorations possibles des modèles de prédictions. Son expertise au sujet des inondations boueuses ainsi que ses connaissances agronomiques m'ont aussi été d'une grande aide. Pierre Defourny a toujours su prendre du temps quand l'occasion se présentait pour me remettre sur le droit chemin et m'aiguiller vers certaines méthodes. L'avoir comme co-promoteur m'a été très utile.

Je souhaite remercier également les chercheurs des pôles GERU et géomatique, qui m'ont aidé de près comme de loin lors de la réalisation de ce mémoire. Ils ont ainsi pu me guider sur certaines questions et, surtout, m'ont fourni l'accès à certaines données essentielles à mon mémoire. Je tiens à remercier en particulier Diane Heymans, pour tout le travail supplémentaire qu'elle a dû effectuer, ainsi que pour sa patience lors de la réalisation de mes tests sur le processeur SEN4CAP.

De manière plus symbolique, je voudrais remercier tous les différents sites internet et autres forums que j'ai pu consulter lors de ces 6 derniers mois et qui m'ont permis de réaliser toute une partie de ce mémoire de façon « autodidacte ».

Plus personnellement, je voudrais remercier tous les gens que j'ai pu rencontrer lors de mon cursus universitaire, à Louvain-la-Neuve comme à Bologna, de l'agro aux KAP, toutes ces personnes qui ont permis de rendre ces années magnifiques. Je tiens à remercier également ma famille, mes amis et ma copine qui m'ont soutenu tout au long de ces années dans les bons moments comme les plus difficiles. Dans le cadre de ce mémoire je tiens d'ailleurs à remercier particulièrement mes parents et Manon pour leur relecture.

Enfin, je remercie d'avance mes lecteurs, pour m'accorder de leurs temps sur mon mémoire. Je vous souhaite une bonne lecture.

Table des matières

1. Introduction	1
2. Synthèse Bibliographique	3
2.1. Inondations boueuses	3
2.1.1. Estimation des inondations boueuses	6
2.2. Cultures en Belgique centrale	7
2.3. Rotations.....	9
2.4. Méthode de prédictions des cultures par la rotation	11
2.5. Télédétection.....	13
2.5.1. Missions spatiales comme sources de données	13
2.5.2. Classification de cultures par télédétection.....	16
2.6. Conclusion.....	18
3. Objectif.....	19
4. Méthodologie	20
4.1. Zone d'étude.....	21
4.2. Sources de données et outils.....	22
4.2.1. Données utilisées	22
4.2.2. Logiciels utilisés	22
4.3. Classification	23
4.3.1. Regroupement des cultures pour une typologie hiérarchique.....	23
4.3.2. Evolution des répartitions de cultures	24
4.4. Prédiction par la méthode des rotations.....	26
4.4.1. Préparation des données de parcellaire	26
4.4.2. Choix de la méthode Random Forest	29
4.4.3. Types de modèles.....	29
4.4.4. Calibration des modèles.....	30
4.4.5. Prédiction 2019, validation des modèles	38
4.5. Détection des couverts hivernaux.....	39
4.5.1. Fonctionnement du processeur	39

4.5.2.	Calibration	44
4.5.3.	Validation du modèle	47
4.5.4.	Amélioration du modèle	49
5.	Résultats.....	51
5.1.	Prédiction par la méthode des rotations.....	51
5.1.1.	Comparaison des parcelles	51
5.1.2.	Calibration	53
5.1.3.	Validation des modèles	64
5.2.	Détection des couverts hivernaux.....	67
5.2.1.	Calibration	67
5.2.2.	Validation	72
5.2.3.	Amélioration du modèle	72
6.	Discussion générale et perspectives d'améliorations	76
7.	Conclusion	80
8.	Bibliographie	81

Liste des figures

Figure 1 : Fréquence inondations boueuses dans les communes de Belgique centrales sur une période de 10 ans (1991-2000 pour Wallonie et 1995-2004 pour la Flandre) [Evrard et al., 2007]	4
Figure 2 : Principe du SAR. Les points en haut représentent le satellite, les carrés en bas représentent le sol. On observe sur la figure que le satellite va prendre des données à plusieurs reprises sur les mêmes points du sol [Defourny, 2017]	14
Figure 3 : Organigramme représentant la construction des modèles de prédictions développés dans la méthodologie.....	20
Figure 4 : Localisation et occupation du sol de la zone d'étude en 2007 [Géoportail de la Wallonie] ..	21
Figure 5 : Illustration des principaux problèmes rencontrés dans l'association de cultures de deux années successives. Soit les cas 1 : la division des parcelles, 2 : la fusion de parcelles, 3 : l'agrandissement d'une parcelle au détriment d'une autre, 4 : la disparition de parcelles et 5 : l'apparition de parcelles.	27
Figure 6 : Illustration des fonctions « union » et « intersection » de polygone	27
Figure 7 : Schéma de la méthode des centroïdes dans les polygones à problèmes intersectés	28
Figure 8 : Schéma de l'utilisation des trois types de modèle développés dans la prédiction par les rotations.	29
Figure 9 : Illustration de la création du tableau de calibration de 3 ans de rotation (H= Hiver, Pt = Printemps, Pr = Prairie).....	31
Figure 10 : : Illustration du tampon interne et de son utilité dans le traitement des images	40
Figure 11 : Exemple de l'évolution des trois métriques d'une parcelle dans laquelle le processeur détecte un couvert hivernal [SEN4CAP].....	43
Figure 12 : Exemple du comportement du NDVI, et évolution des images en fausses couleurs, d'une parcelle ayant accueillie un couvert en 2019. (S2 = Sentinel 2 et L8 = Landsat 8) [ASAP].....	48
Figure 13 : Illustration de la prédiction des cultures par le modèle combiné 1	49
Figure 14 : Illustration de la prédiction des cultures par le modèle combiné 2	50
Figure 15 : a) Exemple de la différenciation symétrique entre les parcellaires 2018 et 2019 sur une partie de la zone d'étude. b) Idem sans les parcelles de moins de 1000 m ²	52
Figure 16 : Comparaison des prédictions des quatre classes pour l'année 2019 obtenues par le modèle 4 classes et le modèle 10 classes direct (regroupement des sept sous-classes de printemps)	65
Figure 17 : Comparaison des prédictions sur les cultures de printemps, obtenues par les méthodes indirecte (modèles 4 classes puis 10 classes indirect) et directe (modèle 10 classes direct). Les bonnes prédictions en 4 classes représentent les parcelles mal prédites en dix classes mais bien prédites par le modèle 4 classes ou par le rassemblement des sept sous-classes de printemps (10 classes direct).	66
Figure 18 : Comparaison des prédictions des cultures de printemps de 2019 par la détection des couverts hivernaux dans les périodes du 30/09 au 10/11 et du 30/09 au 15/11	71
Figure 19 : Confrontation de la détection des couverts dans la période du 30/09 au 10/11 avec des parcelles photo-interprétées comme des couverts.	72

Figure 20 : Représentation cartographiques des modèles de prédictions de cultures de printemps obtenus par la méthode de rotation, de détection des couverts et les modèles combinant les deux.....	75
Figure A-1 : Etat des lieux des activités de la cellule GISER (07/04/17) [Biolders et al., 2017]	II
Figure A-2 : Communes de Wallonie, et zone vulnérable du PGDA [Protect'eau-ZoneVulnérable]	II

Liste des graphes

Graphe 1 : Evolution du nombre de parcelle (a) et des superficies (b) des quatre classes de cultures entre 2014 et 2018 pour la zone d'étude	24
Graphe 2 : Evolution de la précision et de la sensibilité de la classe printemps du modèle 4 classes utilisé en prenant compte différentes durées de rotation, et écart-type autour de l'estimation de l'exactitude moyenne.	53
Graphe 3 : Comparaison de la précision de la classe « Printemps » des modèles 4 classes basé sur trois années de rotations, prenant, ou ne prenant pas en compte les parcelles de type « Autre »	54
Graphe 4 : Comparaison des précisions et sensibilités moyennes de la classe « Printemps » par suite des différents tests effectués sur le modèle 4 classes basé sur des rotations de trois ans, et écart-type autour de l'estimation de l'exactitude moyenne. (Modif = test d'évolution des parcelles, SPP= test des parcelles de petites tailles)	55
Graphe 5 : Comparaison des précisions et sensibilités de la classe « Printemps » par suite des différents tests effectués sur le modèle 4 classes basé sur des rotations de quatre ans, et écart-type autour de l'estimation de l'exactitude moyenne. (Modif = test d'évolution des parcelles, SPP= test des parcelles de petites tailles)	56
Graphe 6 : Comparaison des précisions et sensibilités de la classe « Printemps » des deux modèles 4 classes retenus et écart-type autour de l'estimation de l'exactitude moyenne.	56
Graphe 7 : Evolution de l'exactitude globale moyenne des modèles 10 classes indirect de base, prenant en compte un nombre différent d'années de rotations et écart-type autour de l'estimation de l'exactitude moyenne.	57
Graphe 8 : Exactitudes globales moyennes en nombre de parcelles et en aires obtenus suite aux différents tests effectués sur le modèle de 10 classes indirect basé sur cinq années de rotation dont les classes faibles ont été suréchantillonnées, et écart-type autour de l'estimation de l'exactitude moyenne. (Modif = test d'évolution des parcelles, SPP= test des parcelles de petites tailles, flow acc=test de l'accumulation de flux)	59
Graphe 9 : Effet de ROSE sur les exactitudes en aire et en nombre de parcelles des modèles sans et avec les variables retenues du modèle 10 classes indirect basé sur cinq années de rotations, et écart-type autour de l'estimation de l'exactitude moyenne.	60
Graphe 10 : Comparaison des prédictions (précision et sensibilité calculées sur base du nombre de parcelles) des sous-classes regroupées ou de la classe de printemps obtenues avec le modèle 4 classes retenu (droite) et le modèle 10 classes utilisé par la méthode directe (gauche) et écart-type autour de l'estimation de l'exactitude moyenne.....	61
Graphe 11 : Comparaison des exactitudes sur les classes de printemps obtenues avec les modèles de 10 classes direct (gauche) et indirect (droite).....	62
Graphe 12 : Comparaison de l'exactitude globale moyenne, de la précision et de la sensibilité de la classe printemps des prédictions sur l'année 2019 par le modèle 4 classes et le modèle 10 classes direct (regroupement des sept sous-classes de printemps)	64

Graphe 13 : Précisions et sensibilités des prédictions de cultures de printemps de 2019 obtenues par les modèles utilisant la rotation (10 classes direct), la détection des couverts (entre le 30/09 et le 10/11) et les modèles combinant les deux méthodes.	74
Graphe A-1 : Effet de la suppression des petites parcelles sur les performances de prédictions du modèle 4 classes basé sur trois années de rotations	VIII
Graphe A-2 : Effet de la suppression des petites parcelles sur les performances de prédictions du modèle 4 classes basé sur quatre années de rotations	VIII
Graphe A-3 : Effet de la suppression des petites parcelles sur les performances de prédictions du modèle 10 classes indirect basé sur cinq années de rotations	IX
Graphe A-4 : Effet de la suppression des parcelles dont les limites ont été modifiées selon des degrés différents dans le modèle 4 classes basé sur trois années de rotations.	IX
Graphe A-5 : Effet de la suppression des parcelles dont les limites ont été modifiées selon des degrés différents dans le modèle 4 classes basé sur quatre années de rotations.	X
Graphe A-6 : Effet de la suppression des parcelles dont les limites ont été modifiées selon des degrés différents dans le modèle 10 classes basé sur cinq années de rotations.	X
Graphes A-7 : Exemple de l'évolution du NDVI de parcelles détectées comme CIPAN dans la période du 30/09 au 10/11. Avec ; A : une parcelle contenant bien un couvert, B : une culture d'orge d'hiver détectée comme couvert, C : une culture de colza d'hiver et D : une culture « autre fourrage » détectés comme des couverts. Le segment horizontal vert montre la limite à laquelle le marqueur 6 du processeur détecte de la végétation sur base du NDVI (500).	XIV
Graphes A-8 : Exemple de l'évolution du NDVI de parcelles de couverts détectées (A) et non détectées (B, C et D) par le processeur dans la période du 30/09 au 10/11. Le segment horizontal vert montre la limite à laquelle le marqueur 6 du processeur détecte de la végétation sur base du NDVI (500). L'intervalle en vert clair représente la période de détection.	XIV

Liste des tableaux

Tableau 1 : Calendrier cultural des principales cultures retrouvées dans la région limoneuse (Biolders et al., 2011).....	7
Tableau 2 : Calendrier des deux types (classique et courte) de couvertures imposées dans la zone vulnérable.....	9
Tableau 3 : Principales rotations identifiées sur trois années en Wallonie (Biolders et al, 2011).....	10
Tableau 4 : Evolution du nombre de parcelles et de la superficie des 10 classes de cultures entre 2014 et 2018 sur la zone d'étude	25
Tableau 5 : Nombre de parcelles dans les tables de calibration regroupant les données issues des intersections des parcellaires sur 2, 3, 4 et 5 années de rotations.	30
Tableau 6 : Exemple d'une matrice de confusion.....	31
Tableau 7 : Exemple des variables inhérentes au modèle de 4 classes de rotation de 3 ans	33
Tableau 8 : Illustration de l'astuce trouvée pour contrer l'algorithme RF de R qui n'accepte que 53 classes.....	34
Tableau 9 : Périodes de détection des couverts hivernaux testés.....	46
Tableau 10 : Exemple des différents cas de figures qui apporte une différence entre les deux méthodes de prédictions combinant les prédictions par rotation et détection de couvert.	50
Tableau 11 : Pourcentage des cultures (en aire et en nombre de parcelles) dont l'aire intégrée (issue de l'intersection entre les parcellaires de différentes années) est similaire à plus de 80% avec les aires des parcelles entières de chacune des années.....	51
Tableaux 12 : Matrices de confusions issues d'une prédiction sur les parcelles tests de calibration avec (gauche) et sans (droite) l'augmentation de la proportion des classes faiblement représentées dans les parcelles d'entraînement.	58
Tableaux 13 : Matrices de confusions issues d'une prédiction sur les parcelles tests de calibration en utilisant les paramètres retenus, avec (dessus) et sans (dessous) l'augmentation de la proportion de classe faible dans les parcelles d'entraînement.....	60
Tableau 14 : Matrices de confusions obtenues en utilisant le modèle de 10 classes en méthode direct, en diminuant dans les parcelles d'entraînement la proportion de cultures d'hiver à 25%.....	61
Tableaux 15 : Comparaison des matrices de confusions obtenues avec le modèle de 4 classes (droite) et le modèle de 10 classes utilisé par la méthode directe (gauche).	62
Tableaux 16 : Comparaison des matrices de confusions des cultures de printemps obtenues avec les modèles de 10 classes indirect (droite) et le modèle de 10 classes direct (gauche).	63
Tableau 17 : Effet du choix de la date de début d'intervalle dans la performance de détection des couverts.....	67
Tableau 18 : Effet du choix de la date de fin d'intervalle dans la performance de détection des couverts	68
Tableau 19 : Effet de la durée et de la situation de la période dans l'année sur la détection des couverts.....	69
Tableau 20 : Matrice de confusion obtenues par la prédiction des cultures de maïs (précédées par une autre culture de maïs) de 2019 en utilisant les parcelles de monocultures de maïs sur la période 2016-2018	73

Tableau 21 : Evolution de la proportion de la superficie des cultures de printemps précédées par des cultures d'hiver entre 2015 et 2019 sur la zone d'étude	79
Tableau A-1 : Cutures considérées comme permanente et non prises en compte dans les classifications 4 et 10.	III
Tableau A-2 : Cutures classées dans la classe « prairie » des classifications 4 et 10.	III
Tableau A-3 : Cutures classées dans la classe « autre » des classifications 4 et 10.	IV
Tableau A-4 : Cutures classées dans la classe « hiver » des classifications 4 et 10.	IV
Tableau A-5 : Cutures classées dans la classe « printemps » des classifications en 4 classes et composition des 7 sous-classes de printemps de la classification en 10 classes.	V
Tableau A-6 : : Occurrence des rotations observées dans la zone d'études entre 2014 et 2018, classification en 4 classes. (Hiv = hiver, Pri = Printemps, Aut = Autre et Pra = Prairie)	VI
Tableau A-7 : Occurrence des rotations observées dans la zone d'études entre 2014 et 2018, classification en 10 sous-classes. (Hiv = hiver, Bet = Betterave, Maï = Maïs, PDT = Pomme de terre, Lin = Lin textile, Poi = Pois, AuP = Autre Printemps, CéP = Céréale de printemps, Aut = Autre et Pra = Prairie)	VII
Tableau A-8 : 15 classes retenues dans le test basé sur le type de sol issu du CNSW	XI
Tableau A-9 : Seuils entrés dans le processeur L4C utilisé par les différents marqueurs. Ces seuils ont été repris de ceux optimisés pour les Pays-Bas.....	XII
Tableau A-10 : Période de récoltes calibrées sur les récoltes des cultures de l'année 2018	XIII
Tableaux A-11 : Matrices de confusions obtenues par le croisement des détections de CIPAN et les cultures de 2019 pour les deux périodes retenues [ha].	XIII

Liste des abréviations

CIPAN	Culture Intermédiaire Piège à Nitrates
CR	Coefficient de Rétrodiffusion
NDVI	Indice de Végétation par Différence Normalisé
ESA	Agence spatiale européenne
L4C	SEN4CAP <i>Agricultural monitoring produces</i>
LPIS	<i>Land Parcel Identification System</i>
PAA	Parcellaire Agricole Anonyme
PAC	Politique Agricole Commune
RF	<i>Random Forest</i>
RN	Réseau de neurone
ROSE	<i>Random Over-Sampling Examples</i>
SEN4CAP	<i>Sentinels for Common Agricultural Policy</i>

1. Introduction

Les inondations boueuses sont des événements récurrents dans la partie nord de la Wallonie. Elles sont définies par des inondations d'eau fortement chargées en sédiments, issues du ruissellement de l'eau provenant des précipitations dans les parcelles agricoles. Leurs dégâts causés sur les zones urbanisées et sur l'environnement poussent les communes à chercher des solutions afin de diminuer l'apparition et l'impact de ce processus érosif. Dans ce contexte la Région wallonne a formé, en collaboration avec deux universités (UCLouvain et ULiège-Gembloux ABT) un pôle de recherche et d'information sur le ruissellement et l'érosion des terres agricoles : GISER (Gestion Intégrée Sol-Erosion-Ruissellement). GISER intervient notamment auprès des communes impactées pour diagnostiquer la problématique et y trouver des solutions. Beaucoup d'entre elles consistent à la mise en place d'aménagements permettant la rétention des sédiments ou la protection des habitations et des voiries (fascines, noues, bandes enherbées, etc.) [GISER]. Ces actions sont actuellement menées *a posteriori*, après survenance d'une inondation boueuse. L'établissement d'un modèle permettant de prédire les zones à risque d'apparition d'inondations boueuses, pourrait être un outil idéal d'aide à la décision concernant la mise en place de ces aménagements avant leur occurrence.

Des études ont montré que l'implantation des cultures de printemps constitue l'une des causes principales de ces inondations en Belgique. D'une part car elles sont semées durant la période où les pluies sont les plus érosives en Belgique, et d'autre part car la plupart de ces cultures sont des cultures dites « sarclées » [Boardman et al., 1994 ; Evrard et al., 2007 ; Evrard et al., 2008 ; etc.]. Prédire la répartition des cultures constitue dès lors une information capitale dans la prédiction des inondations.

À ce jour, aucun modèle ne permet de prédire les cultures à l'échelle de la parcelle de manière satisfaisante. Cependant, deux approches de prédictions de cultures de printemps ont été conceptualisées par des étudiants de l'UCLouvain dans le cadre de leurs mémoires en 2018.

La première approche est une méthode de prédictions à l'échelle de la parcelles, basée sur les cultures principales des années antécédentes. Les agriculteurs ont tendance à répéter des schémas de successions de cultures afin d'augmenter leurs rendements agricoles, sinon d'éviter des pertes de rendements dues principalement aux adventices et aux maladies s'installant dans les cultures. Ces successions peuvent alors être utilisées afin de prédire les cultures suivantes.

La seconde approche utilise la détection des couverts hivernaux sur base d'imagerie satellite afin de prédire les cultures de printemps. L'installation d'un couvert hivernal est obligatoire dans certaines zones, dont les régions les plus sujettes aux inondations boueuses afin d'éviter la lixiviation des nitrates vers les nappes phréatiques pendant la période de jachère (entre une

Introduction

culture d'hiver et de printemps). Un couvert y précède donc généralement une culture de printemps.

L'ambition de ce mémoire est de développer un modèle de prédictions des cultures de printemps en se basant sur, et en améliorant, les deux approches préexistantes. Le mémoire est subdivisé en 3 grandes parties. En premier lieu, la synthèse bibliographique présente l'état de l'art associé à la problématique des inondations boueuses ainsi qu'aux cultures présentes en Wallonie, aux rotations entre celles-ci, aux outils de prédictions et à la détection des cultures par imagerie satellite. Ensuite la méthodologie résume, étape par étape, le chemin effectué et les méthodes employées afin de réaliser ce travail. Ensuite, les différents résultats sont présentés ensemble dans une rubrique contenant également une appréciation de ceux-ci. Enfin, une discussion générale aborde, entre autres, des améliorations éventuelles applicables aux différentes méthodes développées.

2. Synthèse Bibliographique

La synthèse bibliographique aborde premièrement la problématique des inondations boueuses, ses causes et ses conséquences. Elle traite ensuite des cultures retrouvées dans la zone problématique de ces inondations. Enfin, elle dresse l'état de l'art associé aux rotations de cultures en Région wallonne ainsi qu'à la télédétection, qui pourraient tout deux aider à la prédiction des couverts hivernaux.

2.1. Inondations boueuses

Une inondation boueuse est une problématique issue du processus d'érosion due à l'écoulement de l'eau sur les sols, provoquant un déplacement de ses particules sous forme de sédiments en suspension ou en charge de fond [Boardman et al., 2006b].

Causes

Deux grands facteurs de risques peuvent amener à ces déplacements de sédiments : la vulnérabilité du paysage et les conditions climatiques. Le premier dépend de la géomorphologie (relief, type de sol, etc.), de l'occupation du sol et des pratiques culturales. Le second est relié aux précipitations qui occupent un rôle important et dans une moindre mesure au vent [Boardman et al., 1994].

Conséquences

Le mouvement des sédiments amène avec lui une série de conséquences pouvant être néfastes sur plusieurs niveaux et qui nécessite donc une gestion de ces inondations. Du point de vue agronomique, on observera une perte de sol accompagnée de fertilisants et pesticides *in situ* dans les parcelles cultivées et donc un appauvrissement de celles-ci [Boardman, 2006]. Les sédiments peuvent également ensevelir des portions de cultures et induire une diminution des rendements. Mais les conséquences se font surtout ressentir *ex-situ*. Les dégâts matériels issus des inondations boueuses dans les zones urbanisées peuvent mener à des coûts importants. À titre d'exemple, le coût des dégâts sur les habitations situées dans la ceinture loessique en Belgique est estimé de 1,6 à 16,5 millions d'euros par an [Evrard et al., 2007]. De plus, les inondations boueuses peuvent avoir une influence néfaste sur l'environnement à la suite du déplacement des particules de sol dans les eaux de ruissellements qui amènent avec elles, des fertilisants et pesticides qui induiront des problèmes hors site. À titre d'exemple, l'azote et les phosphates contribuent au phénomène d'eutrophisation dans les cours d'eau, les pesticides vont polluer ces mêmes cours d'eau [Evans, 2002] et les nitrates vont indirectement polluer les nappes aquifères [Boardman et al., 2003].

En Belgique

On observe fréquemment cette problématique en Europe continentale de l'ouest dans la « ceinture de lœss » [Boardman et al., 1994 ; Evrard et al., 2007]. Une région principalement composée de sols limoneux et sablo-limoneux déposés durant la dernière époque glaciaire [Van den haute et al., 2002]. La Belgique est traversée en son centre par cette ceinture où l'on répertorie de nombreux phénomènes d'inondations boueuses chaque année. 79% des communes belges de cette zone ont subi des inondations boueuses en 10 ans (1996-2006) et 22% d'entre-elles ont été impactées plus de 10 fois (Figure 1) [Evrard et al., 2007]. Comme vu précédemment elles induisent des dégâts sur les habitations, mais elles causent aussi des coûts pour les communes qui doivent se débarrasser des boues sur les voies publiques [Evrard et al., 2007].

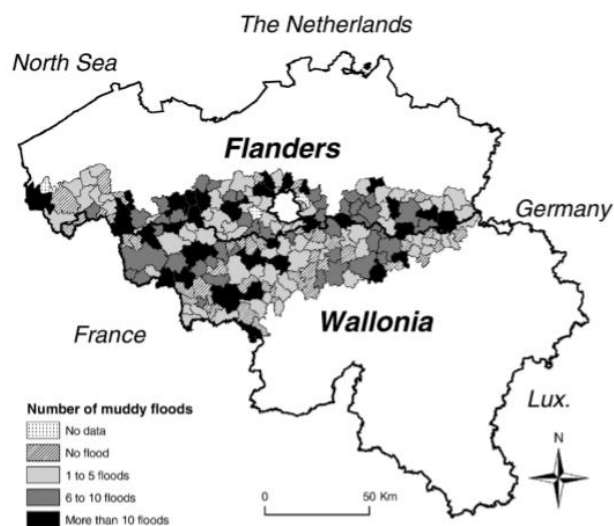


Figure 1 : Fréquence inondations boueuses dans les communes de Belgique centrales sur une période de dix ans (1991-2000 pour Wallonie et 1995-2004 pour la Flandre) [Evrard et al., 2007].

Les inondations boueuses ne touchent cependant pas seulement la région du Lœss en Wallonie. Des phénomènes peuvent se dérouler dans d'autres régions agricoles wallonnes. La carte recueillant les différentes activités répertoriées par la cellule GISER (Annexe A, Figure A-21) montre notamment que les cultures de la région agricole du Condroz sont fortement sensibles à l'érosion.

On distingue une différence temporelle dans l'occurrence de ces événements. En effet, les interventions des pompiers en relation aux inondations boueuses recensées dans la base de données de Saint-Trond entre 1977 et 2001 sont quasi exclusivement effectuées entre mai et septembre [Evrard et al., 2007]. Dans certains bassins versants, 60 à 70% des pertes de sédiments se dérouleraient durant les pluies intenses de printemps et d'été [Vandaele et al., 1995]. Et selon Evrard et al. (2008), 85% des inondations boueuses sont enregistrées lors de la période mai-août en Belgique centrale.

Deux grands facteurs expliquent cette diversité temporelle. Du point de vue climatique, les précipitations sont réparties tout au long de l'année mais l'érosivité de celles-ci montre un pic entre mai et septembre [Verstraeten et al., 2006].

Le deuxième facteur est lié aux cultures semées à cette période : les cultures de printemps. Ces dernières ont une faible couverture du sol durant les orages intenses de mai et juin, ce qui facilite l'érosion hydrique [Evrard et al., 2007]. De plus, une grande partie des cultures de printemps est composée de cultures dites « sarclées », comme le maïs, la betterave ou la pomme de terre, bien plus sensibles à l'érosion que les cultures dites « non sarclées » comme les céréales d'hiver (15 à 20 fois plus sensibles) [Boardman et al., 2003 ; Van der Helm, 1987]. Cette sensibilité à l'érosion est à relier à l'écartement important entre les lignes de semis qui expose fortement le sol aux pluies et favorise la concentration du ruissellement entre les lignes, phénomène encore accentué par la culture en butte dans le cas des pommes de terre. Environ 50 % de la ceinture en Belgique était occupée par des cultures de printemps en 2003 [Biielders et al., 2003].

Augmentation du risque

Les fréquences d'observation d'inondations boueuses en Belgique centrale ont sensiblement augmenté ces dernières décennies [Evrard et al., 2007]. Plusieurs causes peuvent expliquer cette évolution.

D'une part, l'augmentation de la taille des exploitations agricoles et des surfaces des parcelles a fait disparaître peu à peu les limites entre les champs qui produisaient un barrage au ruissellement. La spécialisation de l'agriculture et la disparition des cultures « mixtes », ou associées ont amené à cultiver de nombreuses terres d'un même paysage avec les mêmes cultures permettant donc un ruissellement ininterrompu entre ces cultures. L'apparition de machines plus puissantes a permis de cultiver des terrains situés sur des pentes plus inclinées [Boardman et al. 2003]. Enfin, la conversion de prairies en champs cultivés et l'augmentation des surfaces cultivées en cultures de printemps aux dépens des céréales d'hiver explique aussi cette augmentation [Biielders et al, 2003 ; Evrard et al, 2007].

D'autre part, une urbanisation du territoire avec le phénomène de l'étalement urbain (*urban sprawl*) apporte une augmentation dans le recensement des inondations boueuses et une augmentation des coûts liés aux dégâts de celles-ci [Denis, 1992]. Enfin, une autre cause d'augmentation des risques à prendre en compte dans le futur sera due à la modification des conditions climatiques et donc des pluies. Ce risque varie en fonction des modèles de prédictions climatique choisis, qui induisent des prévisions de précipitations très variables, mais dans la majorité des cas on observerait une augmentation de l'occurrence d'inondations boueuses [Mullan et al., 2016 ; Mullan et al., 2019].

2.1.1. Estimation des inondations boueuses

Afin d'estimer les pertes de sols, diverses approches co-existent.

- *Modèles empiriques, type USLE/RUSLE*

L'équation USLE (*Universal Soil Loss Equation*) et sa version révisée (*Revised Universal Soil Loss Equation*), sont des modèles mathématiques de calcul de perte de sol développés par l'*U.S. Department of Agriculture* (USDA).

Ils permettent d'estimer la perte de sol moyenne annuelle (A) [tonnes/ha.an] en tonnes par hectare sur le long terme, en multipliant différents facteurs. Ils prennent en compte l'érosivité de la pluie (R) [-], l'érodibilité du sol (K) [tonnes/hectare], un facteur tenant compte de la pente et de sa longueur (LS) [-], la culture (C) [-] et enfin les pratiques culturales (P) [-]. L'équation (1) se présente sous la forme énoncée ci-dessous [Stone et al., 2012 ; Renard et al., 2000].

$$A = R * K * LS * C * P \left[\frac{\text{tonnes}}{\text{ha} * \text{an}} \right]$$

1

Ce sont les modèles d'érosions les plus utilisés dans le monde. Ils peuvent être utilisés à une échelle régionale comme à l'échelle du champ. Il a en revanche le défaut de prédire les pertes de sol sur le long terme et non sur les événements courts [Renard et al., 2000 ; Tiwari et al., 2000]. De plus, ils ne prennent pas en compte l'érosion par ravinement, très souvent à l'origine des inondations boueuses. Enfin, les utilisateurs de ce modèle se basent sur des tables, préalablement calibrées sur des régions spécifiques. Les performances de ce modèle sont donc liées à l'adéquation géographique avec ces régions.

- *Modèles déterministes dont WEPP*

WEPP (*Water Erosion Prediction Project*) est un autre modèle développé par l'USDA, en 1985. Il a été créé dans le but de prédire des pertes de sol à l'échelle d'événements pluvieux [Kinnell et al., 2017].

C'est un modèle physique, continu, qui prédit les pertes et les dépôts de sol par l'érosion due à l'eau. Il est capable de prédire des distributions spatiales et temporelles de pertes de sol à toute période de temps [Tiwari et al., 2000 ; Amore et al., 2004].

- *Prédictions de de Walque et al. (2017) et indices de connectivités*

En ce qui concerne les prédictions d'inondations boueuses en Belgique, de Walque (2017) a développé un modèle de prédictions, construit sur base d'événements passés dans la ceinture loessique de Belgique centrale. Il a établi ses prédictions grâce à l'indice de connectivité des sédiments imaginé par Borselli et al. (2008).

L'indice de connectivité de Borselli est calculable à l'aide d'un modèle numérique de terrain (MNT) dans un système d'information géographique (GIS). C'est un indice adimensionnel

donnant une idée sur la probabilité d'un pixel d'amener des sédiments vers un puit qui peut être un cours d'eau ou dans le cas des inondations boueuses une zone urbanisée. Cet indice prend en compte l'aire contributive arrivant au pixel et le chemin que parcourront les sédiments de la source jusqu'au puit.

De Walque (2017) a révisé cet indice en utilisant l'équation USLE pour quantifier la production de sédiments. Pour le facteur C, il a donné à l'ensemble des surfaces agricoles la valeur d'un sol nu, ceci afin d'imiter les pires conditions possibles. Il a prouvé qu'avec cet indice révisé, il pouvait prédire les zones d'inondations boueuses avec une sensibilité et une spécificité de respectivement 0.812 et de 0.792.

Le fonctionnement de ces modèles requiert la connaissance, au préalable, des cultures disposées sur les parcelles (et notamment les cultures de printemps), afin de pouvoir prédire une éventuelle zone à risque d'inondations boueuses pour envisager la mise en place de mesures le plus tôt possible.

2.2. Cultures en Belgique centrale

En 2011, les cultures principales des régions limoneuses et sablo-limoneuses (régions où l'on retrouve principalement les problèmes d'inondations boueuses), sont la betterave, la pomme de terre, le maïs, le lin, le froment et l'escourgeon (Tableau 1) [Biielders et al., 2011]. Les quatre premières sont considérées comme cultures de printemps, semées au printemps et récoltées en automne. Les deux autres sont des cultures d'hiver, semées en automne, les dates de semis changeant en fonction des cultures précédentes, et récoltées en été.

Tableau 1 : Calendrier cultural des principales cultures retrouvées dans la région limoneuse (Biielders et al., 2011).

	Oct	Nov	Dec	Jan	Fév	Mar	Avr	Mai	Jun	Jul	Aou	Sep	Oct	Nov
Betterave														
Patate														
Maïs														
Lin														
Froment														
Escourgeon														

	Semis
	Culture en place
	Récolte

En plus des cultures principales, il arrive souvent qu'une interculture (couvert hivernal), soit implantée entre une culture d'hiver et une culture de printemps. En effet, on observe sur le calendrier du Tableau 1 qu'une longue période de jachère existe entre une culture d'hiver, récoltée en juillet-août, et une culture de printemps (environ 7 mois entre l'escourgeon et la betterave).

Ces couverts hivernaux comportent de nombreux effets positifs :

- Ils permettent une production de biomasse supplémentaire (comme une production fourragère par exemple) pour l'agriculteur.
- Ils vont induire une diminution de l'érosion des sols en maintenant une couverture végétale sur les parcelles, à la place d'un sol nu [Biielders et al., 2003].
- Ils peuvent aussi servir d'engrais vert. En effet, lorsqu'une légumineuse est utilisée dans l'interculture, elle va capter l'azote de l'aire et le restituer au sol. [Burket et al., 1997]. Ce processus d'enrichissement du sol par les couverts n'est cependant pas réservé qu'aux légumineuses. La biomasse de toutes les intercultures contient des nutriments. Leur destruction va alors permettre un apport de matières organiques, très bénéfique pour le sol en stimulant l'activité biologique.
- Surtout, ils permettent d'immobiliser des nutriments mobiles dans le sol (comme les nitrates) et ainsi d'éviter une lixiviation de ces nutriments dans les nappes, qui peuvent causer des problèmes de contamination ou d'eutrophisation [Destain et al., 2010 ; Protect'eau].

Pour cette dernière raison, les agriculteurs wallons sont désormais poussés à installer des CIPAN (culture intermédiaire piège à nitrate) par la région wallonne et son programme de gestion durable de l'azote (PGDA) [Destain et al., 2010]. Dans certaines conditions, ils sont dans le devoir d'en implanter. Ces conditions sont :

- Après un apport de matières organiques réalisé entre le 1^{er} juillet et le 15 septembre.
- Sur les parcelles en pente, après une culture non sarclée.
- Dans les zones vulnérables (risque de pollution des aquifères), soit après une culture récoltée avant le 1^{er} septembre, soit après une légumineuse [Protect'eau].

La mesure concernant les zones vulnérables (Annexe A, Figure A-22) a de l'intérêt pour ce travail, car on y retrouve les parties wallonnes des régions agricoles limoneuses et sablo-limoneuses. Cette zone couvre l'ensemble des territoires dont l'eau des nappes souterraines contient une teneur en nitrate dépassant ou risquant de dépasser les 50 mg/L. L'implantation d'une CIPAN avant le 15 septembre et couvrant au moins 90% de la surface y est obligatoire après une culture récoltée avant le 1^{er} septembre, si elle est suivie d'une culture implantée après le 1^{er} janvier [Protect'eau].

Autrement dit, les agriculteurs de la ceinture de lœss en Wallonie sont obligés d'implanter des couverts hivernaux, entre les cultures d'hiver, ou des cultures de printemps récoltées tôt (comme le lin), et les cultures de printemps. À noter qu'il existe aussi des intercultures courtes, implantées avant une culture d'hiver notamment dans les zones vulnérables après une légumineuse. Celles-ci peuvent alors être détruites après le 1^{er} octobre afin de permettre le semis des cultures hivernales [Protect'eau]. Le Tableau 2, ci-dessous présente le calendrier cultural des deux types de couverts abordés.

Synthèse Bibliographique

Tableau 2 : Calendrier des deux types (classique et courte) de couverts imposés dans la zone vulnérable.

	Jun	Jul	Aou	Sep	Oct	Nov	Dec	Jan	Fév	Mar	Avr	Mai	Jun	Jul
CIPAN classique														
Interculture courte														

Semis du CIPAN
 Croissance du CIPAN
 Destruction du CIPAN

Culture d'hiver
 Culture de printemps

2.3. Rotations

Un moyen de prédire les cultures serait d'utiliser les rotations. Celles-ci sont des successions de cultures différentes qui suivent une certaine logique permettant d'augmenter les rendements de productions des cultures [Bullock et al., 1992]. Ainsi, une suite prédéfinie de cultures principales se succèdent chaque année sur une parcelle. Cette suite peut être fixe (une culture possible par année de rotation) ou flexible (plusieurs cultures possibles pour certaines années de rotation) [Castellazzi et al., 2008]. La décision finale du choix des séquences est un arbitrage entre les objectifs financiers, agricoles (p.ex., le rendement), et environnementaux (p.ex. l'utilisation de moins de pesticides) [Castellazzi et al., 2008])

L'augmentation des rendements due aux successions n'est pas encore tout à fait comprise. L'apport d'azote par l'utilisation de légumineuses dans les rotations (cultures captant l'azote atmosphérique), la diminution des adventices (généralement liées aux cultures), la réduction d'insectes et des maladies sont des facteurs qui expliquent l'augmentation des rendements. Cependant, ces facteurs n'expliquent pas l'entièreté de cette hausse de rendement. En effet, un apport de fertilisants chimiques et de pesticides dans les monocultures ne permet pas d'égaliser les rendements apportés par les rotations [Bullock et al., 1992].

Des études sur les parcellaires ont permis de connaître les probabilités de retrouver différentes rotations à l'échelle régionale.

- En Wallonie, les successions de cultures ont été étudiées par Leteinturier et al. (2007) en superposant les LPIS¹ (*Land Parcel Identification System*) de plusieurs années (1993-2007) afin d'obtenir des données et traiter les successions de cultures de la Région wallonne. De cette manière, ils ont réussi à isoler les probabilités des rotations de cultures pour chacune des régions agricoles de Wallonie.
- Dans le cadre de la *Convention de Recherche d'Intérêt Général Gestion Intégrée Sol Erosion Ruissellement (GISER)*, Biolders et al. (2011) ont identifié les principales

¹ Le LPIS est un parcellaire contenant les informations sur cultures principale des années. (4.2.1)

Synthèse Bibliographique

rotations de Wallonie sur 3 ans en superposant les LPIS de trois années (2006-2007-2008) et reprises dans le Tableau 3.

Tableau 3 : Principales rotations identifiées sur trois années en Wallonie (Bielders et al, 2011).

Rotations	Monoculture	Maïs
		Froment
		Escourgeon
	Succession de deux cultures	Froment-Betterave
		Froment-Maïs
		Froment- Pomme de terre
		Froment-Lin
		Froment-Colza
		Froment-Escourgeon
		Escourgeon-Pomme de terre
		Escourgeon-Lin
		Escourgeon-Colza
		Epeautre-Betterave
		Epeautre-Maïs
		Epeautre-Pomme de terre
	Succession de trois cultures	Betterave-Froment-Pomme de terre
		Betterave-Froment-Maïs
		Betterave-Froment-Lin
		Maïs-Froment-Pomme de terre
		Pomme de terre-Escourgeon-Betterave *
		Lin-Escourgeon-Betterave*
		Pomme de terre-Escourgeon-Maïs*
		Escourgeon-Pomme de terre-Froment
		Escourgeon-Lin-Froment
		Escourgeon-Colza-Froment
		Froment-Escourgeon-Maïs
		Froment-Escourgeon-Betterave

* : ne peut être lu que tel qu'écrit

Le Tableau 3 montre que la plupart des rotations forment un cycle entre des cultures d'hiver et de printemps, même si des monocultures sont possibles comme pour le maïs ou le froment par exemple.

Cette approche de superposition des LPIS a été reprise dans le développement d'un modèle : « *Crop Rota* » est un modèle qui permet de définir les rotations principales retrouvées à l'échelle d'une exploitation ou d'une région. Pour se faire, il se base sur 7 années de rotation et des critères agronomiques. Il a été validé sur 570 exploitations en Autriche [Schönhart et al., 2011].

2.4. Méthode de prédictions des cultures par la rotation

Les probabilités de rencontrer certaines rotations (2.3) peuvent être utiles dans la prédiction des cultures à l'échelle de la parcelle. Les différentes cultures successives peuvent être utilisées comme variables dans un algorithme de *machine learning* (apprentissage automatique) afin d'obtenir des modèles de prédictions. Les *machine learning* sont des formes d'intelligences artificielles utilisées dans la classification supervisée. C'est un ensemble d'algorithmes ayant la capacité de s'entraîner de manière autonome sur des données afin d'obtenir des classifications. Il en existe de toutes sortes, sous différentes formes. Les plus utilisés sont les arbres de décisions, les forêts aléatoires (*random forest* ou RF) et les réseaux de neurones.

Les arbres de décisions forment des « arbres » à la manière de clés dichotomiques utilisées en biologie. Chaque « nœud » ou « embranchement » de l'arbre représente une décision permettant de discriminer au mieux les données en vue d'une classification. Les arbres de décisions ont la particularité de se surajuster aux données utilisées pour entraîner le modèle. Ce qui peut apporter des pertes de performances du modèle de classification, lorsqu'il est utilisé sur d'autres données.

Pour pallier ce problème de surajustement, les RF ont été créés. Il s'agit de forêts d'arbres décisionnels (ensemble de plusieurs arbres de décisions) construits sur base des mêmes données d'entraînement. Afin de ne pas trop s'ajuster aux données, les arbres ne sont pas entraînés avec l'ensemble des données, et chaque arbre n'utilise qu'une partie des variables disponibles dans les données. La classification peut ensuite être effectuée par un système de vote. Ainsi la classe attribuée à une certaine donnée est celle attribuée par le plus grand nombre d'arbres.

Les réseaux de neurones (RN) ont été créés en s'inspirant des neurones biologiques. Les modèles sont composés d'une suite de neurones liés les uns aux autres sous différentes formes selon le choix de l'utilisateur et de l'algorithme utilisé. Ces neurones vont représenter un calcul, ils sont entraînés sur des données de manière régressive afin de pouvoir obtenir une classification. Comme chaque neurone est un calcul, les variables catégoriques doivent préalablement être transformées en variables linéaires (transformation en variables binaires).

Dans le cadre de prédictions, aussi bien les RF que les RN sont utilisés. Il n'existe pas de consensus dans le choix d'une méthode ou d'une autre, les performances des modèles de prédictions changeant en fonction du type de données. En général, pour le traitement de variables catégoriques peu nombreuses, le RF va être préféré car il est plus facile à utiliser. De nombreuses comparaisons entre les deux types d'algorithmes ont été effectuées dans la littérature.

- Were et al. (2014) ont voulu prédire et développer une carte du stock de carbone organique présent dans les sols de la réserve « *Eastern Mau Forest* » au Kenya. Leur

étude avait pour but d'évaluer les performances de prédictions d'une RF, d'un RN et d'un autre algorithme de machine learning (*Support Vector Regression* (SVR)). Les variables qu'ils traitaient étaient numériques. Dans leur cas, l'algorithme permettant d'obtenir la meilleure performance était le SVR avec une erreur moyenne (ME) de $-4.4 \text{ Mg C ha}^{-1}$, suivi du RN avec une ME de $-4.7 \text{ Mg C ha}^{-1}$. La RF a obtenu la moins bonne performance avec une ME de $-6.5 \text{ Mg C ha}^{-1}$.

- Hsieh et al. (2011) ont comparés les performances de différents algorithmes de *machine learning* dans le cadre du diagnostic d'appendicites aiguës sur base de variables catégoriques (binarisées) et avec peu de variables numériques continues. Dans ce contexte, RF a obtenu une meilleure performance que RN, avec des exactitudes de respectivement 0.96 et 0.91.
- Dans le cadre de son mémoire sur la prédiction de type de cultures à l'échelle de la parcelle, Urdiain-Arraiza (2018) a revisité la méthode de Leteinturier et al. (2007) pour obtenir les données des cultures successives. Il a ensuite effectué les prédictions grâce aux chaînes de Markov et aux réseaux de neurones. Il a montré que les chaînes de Markov étaient plus efficaces que les réseaux de neurones, mais les résultats n'étaient pas suffisants pour les utiliser dans la prédiction des inondations boueuses.

2.5. Télédétection

On a vu précédemment (2.2) qu'un couvert hivernal est imposé entre une culture d'hiver et une culture de printemps. Par ailleurs, les principales successions de cultures en Wallonie ont montrés que la majorité des cultures de printemps sont précédées d'une culture d'hiver [Leteinturier et al., 2007 ; Biielders et al., 2011]. Théoriquement, une détection des couverts hivernaux permettrait, donc, de prédire la plupart des cultures de printemps. Les informations sur les couverts hivernaux pourraient être prélevées *in situ*, ce qui amènerait alors à un travail considérable pour détecter l'ensemble des couverts des régions limoneuses et sablo-limoneuses. Une autre méthode bien plus rapide, *ex situ* cette fois, serait d'utiliser la télédétection, soit la détection de ces couverts en utilisant des images satellites. En effet, la télédétection est de plus en plus utilisée dans le monde de l'agriculture de précision car elle permet de couvrir de larges surfaces, plus rapidement et de manière moins couteuse que les inspections sur le terrain [Contiu et Groza, 2016].

Différents programmes d'observations de la Terre par satellites capturant des images utilisées dans la classification de cultures ont été lancés et se sont perfectionnés au cours du temps. Ceux-ci sont situés sur une orbite héliosynchrone. Ce sont des orbites basse altitude (300-1500 km) décrivant un tour de la Terre sur un plan orbital quasi polaire et de périodicité courte (environ 1h30). L'altitude et l'inclinaison de l'orbite héliosynchrone sont choisies de manière à obtenir un angle quasi constant entre le plan de l'orbite et le Soleil. De cette manière l'intensité du rayonnement sur une latitude reste constante, ce qui représente une propriété intéressante pour les satellites à imagerie optique qui dépend de cette intensité [Defourny, 2017].

2.5.1. Missions spatiales comme sources de données

Copernicus

Copernicus est l'ambitieux programme européen d'observation et de surveillance de la Terre développé récemment par l'ESA (Agence Spatiale Européenne). Une des composantes de ce projet est le lancement de satellites à imagerie optique, Sentinel-2, s'inscrivant dans la continuité des programmes Landsat et SPOT.

La mission Sentinel-2 est composée de deux satellites disposés sur une orbite héliosynchrone, située à 786 km d'altitude, de manière à repasser sur une même zone tous les 5 jours. Ils sont tous deux composés d'un imageur multispectral composé de 13 bandes allant du visible à l'infrarouge avec une résolution spatiale comprise entre 10 m et 60 m selon la bande. Le renouvellement des images, au minimum tous les 5 jours (à l'équateur), a apporté de nouvelles opportunités comme un meilleur suivi des cultures permettant de s'inscrire dans un projet d'agriculture de précision. Dans une zone comme la Belgique, située à une latitude d'environ 50°, les images sentinel-2 sont renouvelées tous les 3-4 jours.

Une autre composante du programme est le lancement de satellite à image radar à synthèse d'ouvertures (SAR), Sentinel-1. Les SAR sont des radars synthétisant une antenne virtuelle gigantesque, en capturant plusieurs images d'une même zone à partir de positions différentes (Figure 2). Les signaux émis par les SAR peuvent se situer dans certaines bandes de fréquence

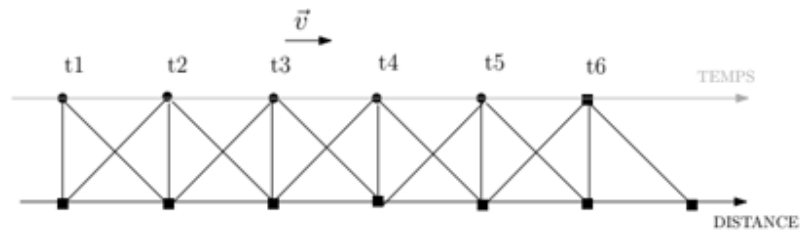


Figure 2 : Principe du SAR. Les points en haut représentent le satellite, les carrés en bas représentent le sol. On observe sur la figure que le satellite va prendre des données à plusieurs reprises sur les mêmes points du sol [Defourny, 2017].

permettant une large gamme d'applications car elles permettent une pénétration différente de la canopée ou du sol, par exemple. Les signaux sont en général polarisés linéairement, transmis verticalement ou horizontalement et leurs retours peuvent être réceptionnés, soit verticalement (VV si émetteur et récepteur en parallèle, HV si croisé), soit horizontalement (HH si émetteur et récepteur en parallèle, VH si croisé). Ce principe de polarisation permet d'obtenir des informations supplémentaires sur les surfaces, comme sur la végétation par exemple, les végétaux ayant tendance à dépolariser les signaux. Les signaux radars réceptionnés forment des images complexes (sens mathématique) caractérisés par une amplitude et une phase qui peuvent mener à une multitude d'applications. Par exemple, il est possible de recréer la topographie grâce au déphasage entre deux images prises de positions différentes capturant une même zone (interférométrie). De plus, ils sont aussi utiles dans des applications similaires à celles rendues possibles grâce à l'imagerie optique, comme les indices de végétations.

Contrairement aux imageurs optiques, les imageurs radars ont l'avantage de ne pas dépendre de l'illumination du soleil ou des conditions climatiques (la présence de nuage n'influence pas les signaux), ce qui permet d'obtenir des images en tout temps [Defourny, 2017 ; Glaude, 2018]

Les signaux émis par les SAR Sentinel-1 se situent dans la bande de fréquence C (4 à 8 GHz). Comme pour Sentinel-2, la mission Sentinel-1 est composée de deux satellites disposés sur une orbite héliosynchrone, à une altitude de 693 km. La résolution temporelle d'un satellite est de 12 jours (à l'équateur), mais disposés de part et d'autre de la terre, à 180° l'un de l'autre, la fréquence de passage est divisée par deux et passe à 6 jours. La résolution spatiale de leurs imageurs peut descendre jusqu'à 10 m. [ESA Copernicus ; Defourny, 2017]

Précurseurs du programme Copernicus

Landsat

Landsat est un programme d'observation de la terre développé par la NASA et l'USGS, dont le premier satellite, Landsat 1, fut lancé en 1972. Aujourd'hui, 2 satellites sont opérationnels. Le dernier lancé en 2013, Landsat 8, est composé de 8 bandes spectrales à 30 m de résolution, 1 bande panchromatique à 15 m de résolution, ainsi que deux bandes spectrales dans l'infrarouge moyen de 100 m de résolution. Sa résolution temporelle est quant à elle de 16 jours [Defourny, 2017]

SPOT

SPOT (Système Probatoire d'Observation de la Terre), est un programme français, développé en collaboration avec la Belgique et la Suède, destiné à l'observation de la terre. Les images obtenues grâce aux satellites SPOT permettent d'obtenir de meilleures résolutions spatiales que celles de Landsat (5 m pour la bande panchromatique de SPOT 5) avec une résolution temporelle de 26 jours. Les satellites SPOT 4 et 5 ont aussi été équipés d'un capteur à large fauchée avec une résolution spatiale de 1 km, SPOT VEGETATION, permettant d'obtenir une couverture globale de la Terre mise à jour quotidiennement. Cet instrument, non réinstallé sur les satellites SPOT nouvelles générations, a été posé sur un microsatellite de l'agence spatiale européenne (ESA), PROBA-V pour assurer la continuité des données.

Les principaux inconvénients des images SPOT est, d'une part, la résolution temporelle largement supérieure à celle des programmes Landsat et plus tard de Copernicus. Cela induit un moins bon suivi des cultures dans le cadre d'une agriculture de précision par exemple. D'autre part, le coût des images est assez important en comparaison à ceux des images issues du programme Landsat ou Copernicus qui sont aujourd'hui mises gratuitement à la disposition du public. [Defourny, 2017 ; ProbaV]

ERS

ERS (*European Remote-Sensing*) était un programme de satellites équipés de radiomètre imageur de type radar (SAR). Le projet, développé par l'ESA prit fin en 2011. Les deux satellites faisant partie du programme étaient situés sur une orbite héliosynchrone à 785 km d'altitude. Le radar émettait dans la bande C (4 à 8 GHz) et la résolution spatiale de leurs images pouvait aller jusque 6 m au nadir [ESA earth online]

ENVISAT

ENVISAT (*Environment Satellite*) est un satellite à imagerie radar développé par l'ESA pour prendre la suite des satellites ERS. Il était notamment constitué d'un radar SAR émettant en bande C. Il était situé sur une orbite héliosynchrone à 800 km d'altitude. Ses images SAR étaient caractérisées par une résolution spatiale de 30 à 150m. Sa mission s'est achevée en 2012 [ESA earth online]

2.5.2. Classification de cultures par télédétection

Les images obtenues via les différents satellites peuvent être transformées en différentes variables que l'on nomme métriques. Certaines de ces métriques, notamment les indices de végétation, sont utilisées pour effectuer les classifications de cultures. L'une d'elles souvent utilisée est le NDVI (*Normalized Difference Vegetation Index*) (équation 2), un indice de végétation utilisant la réflectance, captée par les imageurs optiques de la bande spectrale du rouge (R) (les végétaux absorbant l'énergie de ce spectre) et du proche infrarouge (pIR) (les végétaux réfléchissant l'énergie dans le proche infrarouge).

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

2

Le résultat de l'équation 2 est compris entre -1 et 1, -1 correspondant à une surface d'eau, les pixels se situant aux environs de 0 correspondant à un sol nu et 1 aux forêts tropicales.

En effectuant des statistiques sur les métriques, isolées par parcelles du LPIS, il est possible d'effectuer une classification des parcelles et d'identifier les différentes cultures. Plusieurs méthodes existent, pouvant prendre en compte plusieurs métriques et/ou des séquences temporelles de ces métriques, la classification pouvant ensuite se faire avec des outils statistiques comme le maximum de vraisemblance, k-moyenne, les classifications hiérarchiques ou l'utilisation de *machine learning*.

Blaes et al. (2004), ont comparé les classifications de cultures obtenues par imagerie optique et imagerie radar (SAR) sur une zone du nord-est de la Wallonie contenant 6751 parcelles. Pour ce faire, ils ont utilisé l'algorithme de maximum de vraisemblance. Les images utilisées provenaient de SPOT et de Landsat pour les images optiques, et de ERS et RADARSAT (programme canadien de satellite à imagerie SAR) pour les images radars. En imagerie optique, les meilleurs résultats ont été obtenus par la combinaison de trois images de manière à obtenir une série temporelle répartie sur une année de cultures (mars, juin et août). Ainsi, leur modèle a pu atteindre une exactitude globale d'environ 75 %. La meilleure exactitude globale obtenue par un modèle n'utilisant qu'une seule image optique (celle d'août) était de 64%. Aucun modèle utilisant les images radar (ni ERS ni RADARSAT) n'ont permis d'obtenir de meilleurs résultats que ceux obtenus pour la série temporelle des images optiques. La meilleure exactitude globale atteinte par les modèles utilisant les images radar a été obtenue grâce à une série temporelle d'image ERS capturées entre le 4 février et le 6 juillet et était de 65%. Blaes et al. ont de plus montré qu'il était possible d'augmenter l'exactitude globale par la combinaison des images SAR et optique. Ainsi, en utilisant les séries temporelles des images optiques et ERS, elle a atteint 80%.

Lors de son mémoire, A. Danzin (2018) a essayé d'identifier les couverts hivernaux à partir d'une série temporelle d'images optiques obtenues grâce à Sentinel-2. Elle a utilisé le

maximum de vraisemblance comme méthode de classification. Malheureusement, un seul des deux Sentinel-2 était disponible à l'époque et les mauvaises conditions climatiques n'ont pas permises d'obtenir assez de données lors de la croissance des couverts hivernaux. En utilisant une série temporelle lors de la croissance des cultures hivernales après l'hiver, elle a malgré tout réussi à distinguer environ 96% des couverts hivernaux rassemblés en une classe, des cultures hivernales séparées en deux classes (froment et escourgeon). Elle n'a néanmoins pas pu utiliser son modèle sur une année différente afin de valider son modèle. De plus, son modèle ne permet pas de détecter les couverts hivernaux assez tôt afin de prédire les futures cultures de printemps. Désormais, il serait possible de les détecter plus tôt grâce à l'amélioration de la résolution temporelle des images optiques de Sentinel-2 (avec l'apparition du deuxième satellite) et l'utilisation de l'imagerie radar qui ne dépend pas des conditions nuageuses.

Projet SEN4CAP

Le projet SEN4CAP (*Sentinels for Common Agricultural Policy*) est un projet d'observation de la terre par satellite dans l'objectif de suivre l'application, sur le terrain, des pratiques agricoles éligibles pour la Politique Agricole Commune (PAC) de l'Union Européenne. Le projet est financé par l'agence spatiale européenne et coordonné par l'équipe de recherche en géomatique de l'UCLouvain. Le projet est actuellement en fin de phase test sur 15 millions de parcelles, de 6 pays pilotes parmi lesquels la Belgique ne figure pas.

Le système SEN4CAP est séparé en plusieurs processeurs permettant chacun d'observer ou de contrôler des composantes différentes de la PAC, en utilisant les images issues de Sentinel-1 et Sentinel-2. Un de ceux-ci, *Cultivated crop type map*, permet d'identifier les types de cultures en temps réels. Il permettrait donc de détecter des couverts hivernaux au même titre que les autres cultures.

Un autre processeur qui peut être utile dans le cadre de ce mémoire est celui permettant de suivre les pratiques culturales (*Agricultural practices monitoring product*), comme les CIPAN par exemple. Ce processeur fonctionne comme un modèle logique en détectant une récolte et en observant si une pratique culturale (sélectionnée par l'utilisateur) a bien été effectuée sur la parcelle. *Agricultural practices monitoring product* a l'avantage de n'utiliser que 3 métriques. Ce qui va permettre au processeur d'être plus rapide que celui de *Cultivated crop type map* dans l'objectif de détecter les couverts hivernaux [SEN4CAP].

Les trois métriques utilisées par le processeur sont le NDVI (équation 2) obtenu par imagerie optique, le coefficient de rétrodiffusion (CR) et la cohérence, tous deux obtenus par imagerie radar.

- Le CR représente la portion du signal radar émit par l'antenne du satellite qui est directement réfléchi par la cible vers l'antenne radar [Sentinel online]. Le CR n'est donc fonction que de l'amplitude et non de la phase perçue par le récepteur. Dans le cadre

de son application sur un sol agricole, une haute valeur de CR est reliée à un sol nu qui réfléchit bien le signal. A l'inverse, un sol couvert par de la végétation renvoie une valeur de CR faible car le signal est absorbé ou réfléchi de manière anisotrope par la végétation.

- La cohérence est une corrélation complexe (dans le sens mathématique) entre deux images radars représentant une même zone mais capturée d'une position différente. Elle est donnée par la formule (équation 3) :

$$coherence = \frac{|\sum_w M S|}{\sqrt{\sum_w |M|^2 * \sum_w |S|^2}}, \in [0,1]$$

3

Où M est dite l'image « maitresse » et vaut $A_M e^{j\varphi_M}$ et S, l'image « esclave » et vaut $A_S e^{j\varphi_S}$ dans laquelle A représente l'amplitude du signal reçu par le satellite et φ son déphasage par rapport à la phase du signal dans l'émetteur. W représente une fenêtre autour du pixel. Son utilisation permet de moyenner les pixels et ainsi diminuer le bruit [Glaude, 2018]. Dans le cadre de SEN4CAP, la cohérence est calculée par pixel (et pas par fenêtre de pixels). Les moyennes de cohérences sont effectuées par parcelle, le bruit est donc diminué de cette manière.

Le processeur *Agricultural practices monitoring product* utilise la cohérence entre les images obtenues tous les 6 jours par Sentinel-1. Les images utilisées sont plus précisément issues des signaux polarisés linéairement transmis et reçu verticalement (VV). Cette polarisation est utilisée car les végétaux ont tendance à dépolariser les signaux. Une valeur de cohérence faible, proche de zéro, correspond à des conditions dynamiques sur le pixel, qui, dans le cadre de zones cultivées est signe d'une présence de végétation. A l'inverse, un sol nu sera caractérisé par une cohérence plutôt forte, plus proche de 1.

2.6. Conclusion

Le point majeur à retenir de cette section est qu'il n'existe pas de modèle de prédictions de cultures à l'échelle de la parcelle. Deux méthodes, la prédiction grâce aux rotations de cultures et la prédiction des cultures de printemps par la détection des couverts hivernaux, ont toutefois été conceptualisées. Elles n'ont pas permis d'aboutir aux résultats escomptés. Cependant, de nouveaux modèles basés sur ces deux mêmes concepts permettraient, possiblement, d'obtenir de meilleures prédictions.

3. Objectif

Ce mémoire s'inscrit dans le cadre de la création d'un outil destiné aux communes wallonnes leur permettant de cibler, le plus tôt possible dans l'année, les zones à risques d'inondations boueuses. Il permettrait de pouvoir prendre des mesures préventives afin de limiter les pertes de sol ainsi que les dégâts liés à celles-ci grâce à différents aménagements.

La clé du ciblage de ces zones à risques se trouvant dans la connaissance préalable des parcelles de cultures de printemps, **l'objectif général du mémoire est de développer un modèle de prédictions des cultures de printemps, à l'échelle des parcelles, utilisable le plus tôt possible dans l'année.**

Pour ce faire, deux méthodes distinctes seront étudiées, l'une fondée sur les rotations de cultures, l'autre sur la télédétection des couverts hivernaux, amenant à plusieurs sous-objectifs :

1. Construction et validation d'un modèle utilisant les logiques de rotations et certaines données inhérentes aux parcelles (types de sol, pente, etc.) comme prédicteurs des cultures grâce à l'algorithme Random Forest.
2. Détection des couverts hivernaux sur base des images satellites (Sentinel 1 et 2), grâce aux marqueurs de *Sen4CAP-Agricultural practices monitoring product* et validation de ce modèle de détection pour prédire les cultures de printemps.
3. Comparaison des deux méthodes et combinaison de celles-ci dans le but d'améliorer les prédictions.

4. Méthodologie

La méthodologie s'articule en cinq sous-parties. En préalable, la zone d'étude (4.1) et les outils utilisés (4.2) sont présentés. Ensuite, les sections sont exposées selon l'organisation du mémoire (Figure 3). En premier lieu, l'étude apporte une description de la classification (4.3), une étape préalable aux prédictions. Ensuite sont développées, les deux méthodes de prédictions utilisées : la rotation (4.4) et la détection des couverts hivernaux (4.5). Les deux méthodes seront calibrées et validées. Finalement, la sous-section « amélioration du modèle de détection » (4.5.4), présente une méthode combinant les deux modèles dans l'objectif d'augmenter les performances de prédictions.

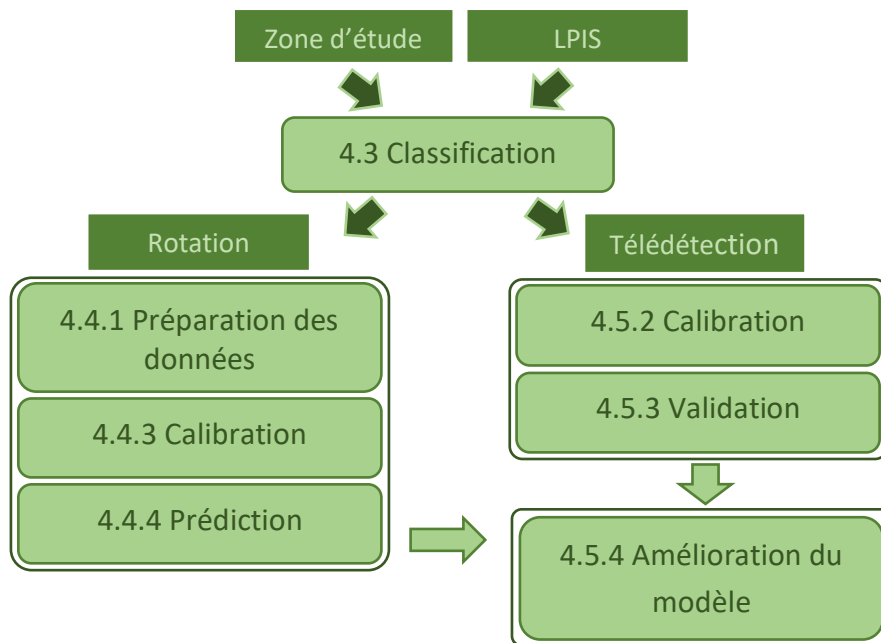


Figure 3 : Organigramme représentant la construction des modèles de prédictions développés dans la méthodologie.

4.1. Zone d'étude

La zone d'étude utilisée dans ce travail est assez représentative de la région touchée par les inondations boueuses. Elle est située dans la ceinture loessique de Belgique centrale, à la frontière entre les provinces du Brabant Wallon, de Liège et de Namur. Elle s'étend sur une superficie de 156.25 km² comprenant 8 communes ; Héléchine (BW), Jodoigne (BW), Orp-Jauche (BW), Ramillies (BW), Eghezée (N), Hannut (L), Lincet (L) et Wasseige (L). L'ensemble de la zone se situe dans la région agricole limoneuse et dans la zone agro-géographique de la Hesbaye. C'est la même zone que celle sur laquelle avaient travaillé M. Urdian-Arraiza (2018) et A. Danzin (2018) ce qui offre la possibilité de comparer les résultats avec les leurs.

En plus d'avoir un type de sol propice aux événements d'inondations boueuses, la zone d'étude est majoritairement agricole. En effet, 84.6% de la zone était occupée par des terres agricoles en 2007, un pourcentage bien supérieur à la moyenne belge qui était de 50% en 2009 [Chiffres clés de l'agriculture 2011, Statbel]. On retrouve en deuxième position les territoires artificialisés (12.1%) sur lesquels peuvent se répercuter des coûts importants liés aux inondations de boues. Ensuite les forêts (2.9%) et les surfaces en eaux (0.4%) sont peu présentes [COSW 2007, Geoportail de Wallonie]. 4320 parcelles issues du parcellaire agricole anonyme du SIGEC étaient recensées dans la zone pour l'année 2018.

Occupation du sol de la zone d'étude

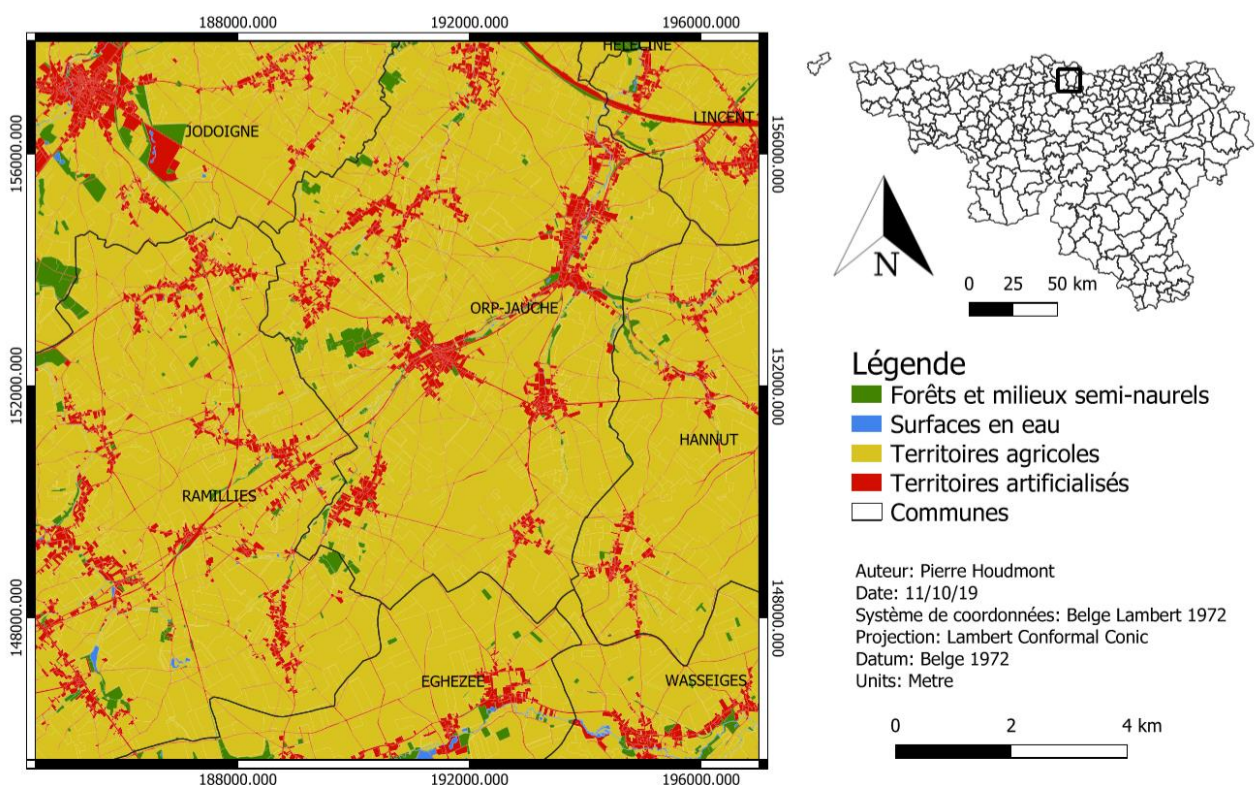


Figure 4 : Localisation et occupation du sol de la zone d'étude en 2007 [Géoportail de la Wallonie].

4.2. Sources de données et outils

4.2.1. Données utilisées

- **Parcellaire agricole anonyme du SIGEC** [Geoportail de Wallonie]

Le parcellaire agricole anonyme (PAA) [Geoportail de Wallonie] est le LPIS wallon. C'est une donnée renouvelée chaque année, référençant la culture principale présente ainsi que les limites des parcelles. Les cultures sont notées dans le PAA selon un code, choisi parmi une multitude de codes référencés. Les agriculteurs sont dans le devoir de le remplir eux-mêmes dans le cadre de la Politique Agricole Commune. Le parcellaire est ensuite mis à la disposition du public l'année suivante via le Géoportail de Wallonie. Ces données ont permis d'identifier les cultures principales de chaque année pour chaque parcelle afin de connaître les logiques de rotation. Lors de ce travail, une série de six années consécutives (2014-2019) de PAA a été utilisée.

- **Carte numérique des sols de Wallonie (CNSW)** [Geoportail de Wallonie]
- **Modèle numérique de terrain lissé (Lidaxes)** [Geoportail de Wallonie]
- **Métriques de télédétection des NDVI (Sentinel 2, Landsat 8), des CR et cohérences (Sentinel 1)**

4.2.2. Logiciels utilisés

- **QGIS** [QGIS]

Le prétraitement des données de la première approche « *Prédiction par la méthode des rotations* » s'est entièrement effectuée grâce au système d'information géographique QGIS. Il a ensuite été utilisé pour représenter les prédictions dans l'espace. Le logiciel a de nouveau servi lors de la deuxième approche « *Détection des couverts hivernaux* » pour comparer les résultats au parcellaire de 2019.

- **R, RStudio** [RStudio]

La calibration des modèles et leurs prédictions de la première approche « *Prédiction par la méthode des rotations* » ainsi que les graphes de la section « résultats », ont été réalisés grâce au logiciel RStudio.

- **Excel**

- **Processeur SEN4CAP utilisant Python et R**

Lors de la partie de télédétection, la détection des couverts hivernaux s'est effectuée grâce au processeur SEN4CAP *Agricultural practices monitoring product*. Celui-ci est composé de deux parties, la première utilisant le logiciel Python, et la deuxième utilisant le logiciel R. Plus d'informations sont apportées au point 4.5

4.3. Classification

4.3.1. Regroupement des cultures pour une typologie hiérarchique

Le grand nombre de cultures recensées (88 codes cultures différents dans le parcellaire de 2016) conduit à un nombre extrêmement élevé de rotations possibles. Pour cette raison, il semblait judicieux, dans un premier temps, d'effectuer les prédictions sur une classification plus « simpliste » des cultures avant d'essayer de prédire les cultures de façon plus « fine ». Dès lors, deux jeux de données ont été constitués de manière à obtenir une typologie de cultures hiérarchique. Ainsi, le deuxième jeu de données, subdivise les classes du premier. Les inondations de boues trouvant leur origine de manière générale dans les cultures de printemps, il s'agissait entre autres de mettre en évidence les cultures de printemps par rapport aux cultures d'hiver.

Premièrement, ne suivant pas une logique de rotation et ne présentant pas un intérêt pour la prédiction des cultures de printemps, les cultures permanentes n'ont pas été prises en compte. Il s'agit des prairies permanentes, des cultures fruitières, des vignes, des sapinières, etc¹.

La première classification, est constituée de quatre classes de cultures, elle a donc été nommée « 4 Classes ». Tout d'abord, les cultures de printemps, regroupant l'ensemble des cultures de printemps semées entre mars et juin et récoltées en été ou en automne (maïs, betterave, carotte, pomme de terre, pois, avoine de printemps, etc.¹) et qui offrent un faible degré de couverture du sol au printemps (de mars à mai-juin). Les cultures d'hiver regroupent les cultures semées avant l'hiver (navette, froment, orge d'hiver, etc.¹). Elles assurent une bonne couverture du sol au printemps. La classe prairie regroupe les prairies non permanentes, ainsi que les cultures semblables à ces dernières en termes d'effet sur le ruissellement (luzerne, couvert à finalité environnementale, couvert favorisant la faune, etc.¹). En effet, ces cultures sont pérennes et assurent une couverture permanente pendant l'année. Enfin la dernière classe est composée de cultures notées comme « Autre » dans le PAA, et de cultures inclassables entre la classe printemps et hiver.

En ce qui concerne la classification plus fine des cultures, la classe « Printemps » a été divisée en sept sous-classes. Une sous-classe pour chaque culture de printemps importante en nombre et en superficie, soit le maïs, la betterave, la pomme de terre, le pois et le lin textile. Une sous-classe regroupant l'ensemble des céréales de printemps (avoine, orge de printemps, etc.¹) et enfin une sous-classe regroupant toutes les autres cultures de printemps moins importantes¹. Afin de garder un nombre de classes acceptable, l'hypothèse que le type de culture d'hiver n'aurait pas d'influence sur la culture suivante a été émise. Les trois autres classes « Hiver », « Prairie » et « Autre » ont donc été gardées telles quelles. Ce choix aurait du sens agronomiquement car le type de culture d'hiver n'a pas d'influence sur le choix de la

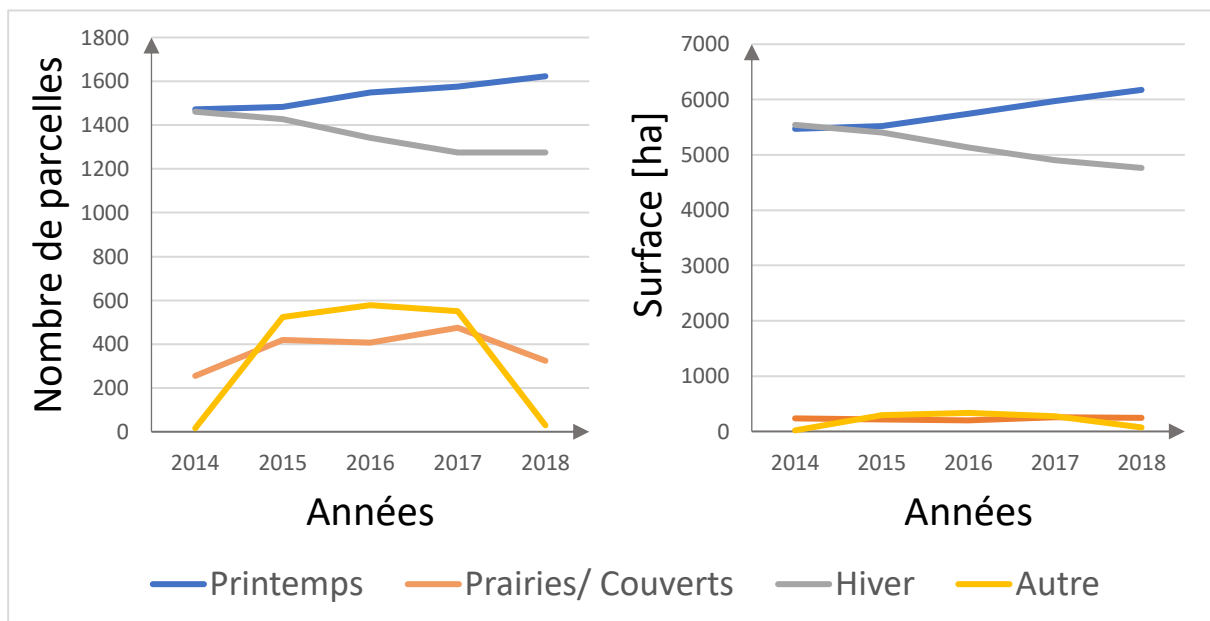
¹ Le détail des classes est donné en Annexe B, Tableau A-22 à Tableau A-26.

culture de printemps suivante. Cependant il pourrait apporter des erreurs au modèle car une culture d'hiver peut avoir une influence sur une future culture d'hiver. Finalement 10 sous-classes sont répertoriées dans cette classification, que l'on a nommé « 10 Classes ».

4.3.2. Evolution des répartitions de cultures

4 classes

Comme vu précédemment dans la revue de littérature, il s'observe une diminution progressive des cultures d'hiver au profit des cultures de printemps entre 2014 et 2018 (Graphe 1). Les cultures de printemps dans la zone d'étude sont même plus importantes que les cultures d'hiver aussi bien en termes de superficie qu'en nombre de parcelles. La superficie de la classe « Prairie » semble assez constante sur les cinq années, bien que le nombre de parcelles fluctue un peu plus. Enfin, la classe « Autre » n'est quasi pas existante en 2014 et 2018 par rapport aux autres années, la culture « Autre » n'étant pas présente dans le SIGEC de ces deux années. Ces manques de stationnarité dans les classes pourraient apporter des erreurs dans une prédiction par rotation, c'est la raison pour laquelle effectuer une validation du modèle sur une autre année est importante.



Graph 1 : Evolution du nombre de parcelle (a) et des superficies (b) des quatre classes de cultures entre 2014 et 2018 pour la zone d'étude.

10 classes

Pour la classification en 10 sous-classes, trois cultures ressortent du lot en termes de nombre de parcelles : le maïs, la betterave et la pomme de terre (Tableau 4). Ce sont toutes trois des cultures sarclées, cultures qui posent le plus de problème d'inondations boueuses. Elles représentent la majorité des cultures de printemps (70% en nombre et 64.3% en aire, pour l'année 2018). Arriver à prédire les cultures de printemps pourrait déjà être un bon pas en avant dans la prédiction des zones à risques.

Tableau 4 : Evolution du nombre de parcelles et de la superficie des 10 classes de cultures entre 2014 et 2018 sur la zone d'étude.

	2014		2015		2016		2017		2018	
	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]	Aire [%]
1 Maïs	11,67	4,15	10,74	5,20	10,06	4,86	9,78	4,31	12,52	4,83
2 Betterave	12,02	14,88	7,99	12,13	8,98	12,67	9,81	14,44	11,90	15,10
3 Pomme de terre	9,11	12,45	7,11	12,20	7,89	12,80	8,90	13,21	10,55	14,80
4 Lin textile	2,87	4,16	2,91	5,26	4,05	6,82	3,38	6,64	3,81	5,85
5 Pois	2,93	4,93	3,79	6,79	3,30	6,09	3,12	6,12	3,72	6,34
6 Autre printemps	6,05	7,64	4,70	6,34	4,64	6,78	4,41	7,26	6,46	7,74
7 Céréales printemps	1,28	0,33	1,25	0,35	1,06	0,30	1,24	0,38	0,95	0,18
8 Prairies/ Couverts	7,96	2,09	10,85	1,89	10,47	1,75	12,26	2,25	9,97	2,19
9 Hiver	45,60	49,19	37,06	47,31	34,64	45,00	32,90	42,96	39,22	42,30
10 Autre	0,50	0,17	13,60	2,53	14,91	2,93	14,19	2,43	0,89	0,68

On peut noter que la classe « Maïs », importante en nombre de parcelles, ne l'est plus autant en superficie. On peut en déduire que la plupart des parcelles sont plus petites pour cette culture, et l'utilisation de la taille de la parcelle comme paramètre du modèle de prédictions pourrait donc avoir du sens. On retrouve la même observation, plus accentuée encore, dans la classe de « Prairie ».

On observe une légère diminution du pourcentage en nombre de la plupart des cultures entre 2015 et 2017, du fait de l'apparition des cultures « autres ». Cette diminution est moins marquée dans les pourcentages surfaciques, les cultures « autres » ne dépassant pas les 3 % en superficie. Les pourcentages surfaciques restent stables pour la plupart des cultures avec de surcroît, une légère augmentation des cultures de printemps (à l'exception des céréales) en parallèle à une diminution des cultures d'hiver.

4.4. Prédiction par la méthode des rotations

Cette partie a pour objectif d'obtenir des modèles de prédictions sur base des successions de cultures. Elle est subdivisée en cinq points. Premièrement, la préparation des données est développée afin de connaître les successions de cultures sur chaque parcelle. Ensuite le choix de l'algorithme de classification est présenté, il s'agit de RF. Après, les différents types de modèles développés sont expliqués. La grosse étape de calibration des modèles est alors abordée. Enfin les modèles sont utilisés pour prédire l'année 2019 en tant que validation.

4.4.1. Préparation des données de parcellaire

Comme l'avaient effectué Leteinturier et al. (2007), une superposition des parcellaires de plusieurs années était nécessaire pour connaître les différentes rotations. Afin de prédire les cultures présentes en 2019, des combinaisons de deux à cinq années de successions de cultures ont été étudiées. C'est pourquoi les données nécessaires à la calibration du modèle sont, pour rappel, les cultures de l'année 2014 à l'année 2018 comprise.

Premièrement, les parcelles situées en bordure de la zone d'étude ont été supprimées. N'apportant pas toutes les informations (comme la surface par exemple), elles étaient une source potentielle d'erreur.

Il a ensuite fallu connaître les successions de cultures sur chaque parcelle afin de pouvoir caractériser la logique de rotation. Le problème majeur de cette étape est que les parcelles peuvent changer de taille et de forme au gré des années, ce qui rend plus compliqué la détermination des successions de cultures sur les parcelles. Plusieurs causes peuvent expliquer ces modifications de parcelles, qui sont souvent dues à des facteurs agricoles ou économiques. Par exemple, dans le cadre de l'implantation de certaines cultures comme la betterave et la pomme de terre, pour respecter leurs contrats, les agriculteurs peuvent agrandir ou diminuer la taille de leurs parcelles.

En observant l'évolution des parcellaires, plusieurs modifications reviennent régulièrement (Figure 5) : la division des parcelles (cas 1), le regroupement de plusieurs parcelles en une seule entité (cas 2), l'agrandissement d'une parcelle au détriment d'une autre (cas 3), ainsi que la disparition (cas 4) et l'apparition de nouvelles parcelles (cas 5).

Méthodologie

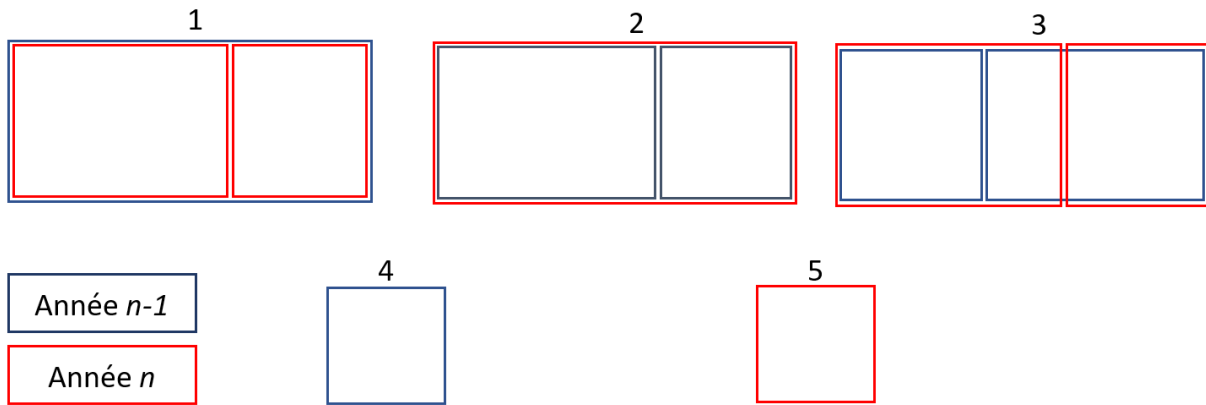


Figure 5 : Illustration des principaux problèmes rencontrés dans l'association de cultures de deux années successives. Soit les cas 1 : la division des parcelles, 2 : la fusion de parcelles, 3 : l'agrandissement d'une parcelle au détriment d'une autre, 4 : la disparition de parcelles et 5 : l'apparition de parcelles.

Deux méthodes de fusionnement de parcelles sont disponibles dans les logiciels SIG, l'union et l'intersection (Figure 6).

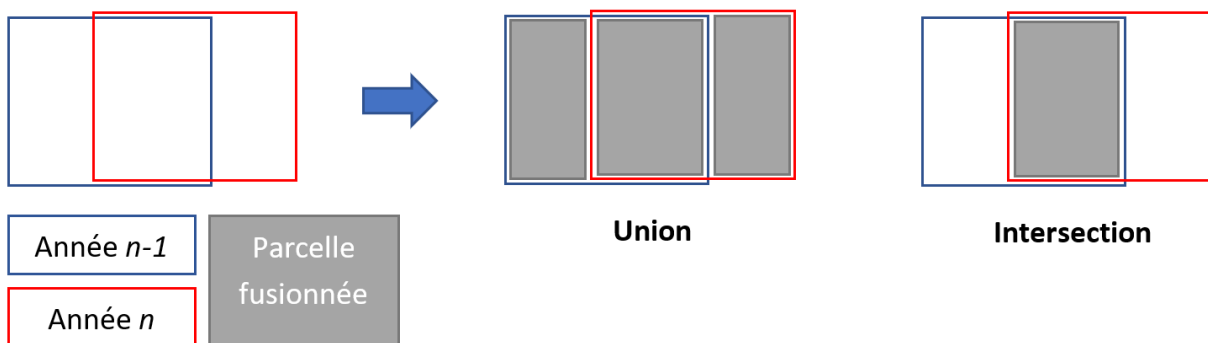


Figure 6 : Illustration des fonctions « union » et « intersection » de polygone.

L'outil « union » n'a pas été retenu car il crée des polygones manquant d'information (la Figure 6 montre que des parcelles créées par cet outil ne contiennent de l'information que sur l'une des deux années). Les parcelles des cas 4 et 5 (Figure 5) auraient été gardées, malgré le fait qu'elles ne contiennent pas d'information sur la rotation. Un autre outil, « intersection », permet lui de s'affranchir de ces erreurs. Ce dernier a donc été préféré.

Néanmoins, effectuer une simple intersection entre les années engendre encore des problèmes. Une première source d'erreur se trouve au niveau des petits morceaux de parcelles intersectés issus des cas 1, 2 et 3 (Figure 5), la logique de rotation pouvant ne pas y être respectée. Mais surtout, n'ayant pas accès au parcellaire de l'année à prédire ($n+1$) en avance, les résultats de la prédiction sont projetés sur le parcellaire de l'année n . Il est alors préférable de n'avoir qu'une seule succession de cultures par parcelle (et donc d'éviter le cas 2) pour n'obtenir finalement qu'une prédiction par parcelle.

Il faut aussi souligner que la suppression des parcelles des cas 4 et 5 (Figure 5), avec l'intersection, vont amener irrémédiablement à des pertes de données et finalement à des parcelles non prédites en année $n+1$.

Méthodologie

La méthode des centroïdes, déjà utilisée par Urdiain-Arraiza (2018), permet de palier partiellement à certains de ces problèmes. Après avoir effectué les intersections entre les différentes années voulues pour les rotations (de deux à cinq années selon le cas), les centroïdes des parcelles de l'année n (2018 dans notre cas) peuvent être calculés et utilisés pour n'établir qu'une seule information de rotation pour chaque parcelle. Celle-ci peut alors être utilisée pour prédire la culture de l'année $n+1$ pour la parcelle telle que délimitée pour l'année n . Afin d'éviter d'avoir des centroïdes hors des parcelles pour les polygones de forme concave par exemple, l'option « centroïde à l'intérieur des polygone » a été choisie.

Sur la Figure 7 ci-dessous, on peut observer que les parcelles de cas 1, donc celles s'étant divisées l'année n , seront gardées, à l'inverse de celles des cas 2 et 3. De plus, malgré la présence d'un centroïde sur la parcelle de type 5, aucune information n'est récoltée (car pas de culture en année $n-1$) et cette parcelle n'est donc pas prédictible pour l'année $n+1$.

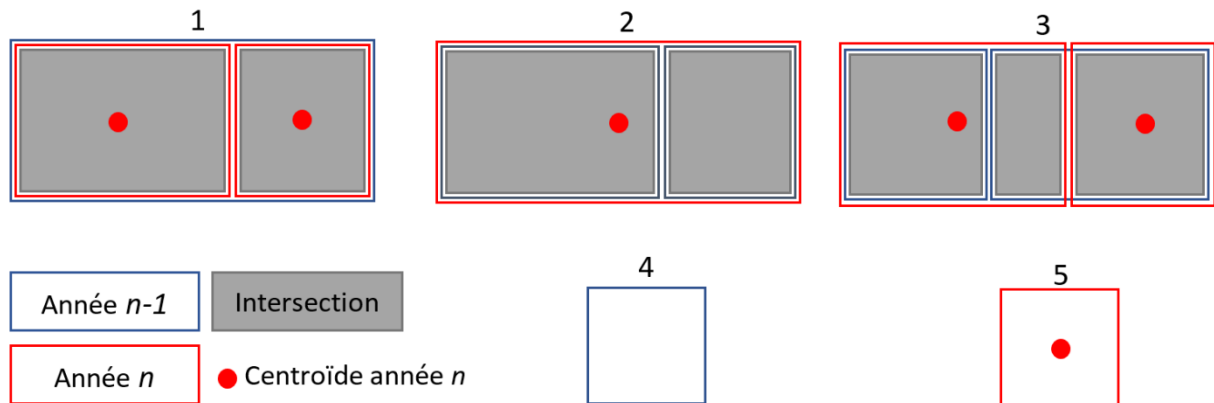


Figure 7 : Schéma de la méthode des centroïdes dans les polygones à problèmes intersectés.

Les difficultés liées à l'intersection ont été illustrées pour deux années de rotation, mais les différents cas cités précédemment se multiplient avec le nombre d'années retenues pour la rotation. Notamment au niveau des pertes d'informations dues aux cas 5 qui pourraient influencer le choix du nombre d'années de rotation à garder dans le modèle de prédictions. Des cas 1 et 2 peuvent se succéder sur une parcelle, etc.

Pour avoir une idée de la proportion de parcelles variant au cours du temps (prenant en compte les cas 1, 2 et 3), et donc pour voir s'il est judicieux d'utiliser la méthode des centroïdes, une quantification a été effectuée. En effet, si la proportion des parcelles à problèmes est vraiment importante, la méthode des centroïdes n'est sans doute pas appropriée car elle induirait trop d'approximations.

Les choix méthodologiques de la préparation des données sont fort similaires à ce qu'avaient effectués Leteinturier et al. (2007), à l'exception qu'ils n'ont pas utilisé la méthode des centroïdes. Cette méthode était cependant requise dans le cadre d'une prédiction pour les raisons abordées plus tôt.

4.4.2. Choix de la méthode Random Forest

Comme les variables sont peu nombreuses, et pour la plupart catégoriques, l'algorithme choisi pour les modèles de prédictions de la culture est le RF. Il se base sur la création d'une multitude d'arbre décisionnel. Pour éviter un surajustement aux données d'entraînement, un échantillon d'environ 2/3 d'entre elles va être tiré aléatoirement pour la construction de chaque arbre. Cette étape se nomme : le *bagging*. De plus, chaque arbre est entraîné en n'utilisant qu'un échantillon de prédicteurs (variables), également tirés aléatoirement. En classification, le nombre de prédicteurs choisi vaut, par défaut, la racine carrée du nombre total de variables. Les arbres ne contiennent donc pas toutes les variables.

4.4.3. Types de modèles

A partir des données de successions de cultures déterminées au point 4.4.1, le développement de plusieurs modèles peut être réalisé.

Le **modèle 4 classes** est le modèle le plus simple. Il prédit les 4 classes de la classification « simplifiée » à partir des données de successions de cultures classées en 4 classes.

Concernant les prédictions des 10 classes de cultures, deux méthodes sont possibles :

- La *méthode directe*

Le **modèle 10 classes direct** fonctionne de la même manière que le modèle 4 classes, en utilisant, cette fois, les données de successions classées en 10 sous-classes. Il permet ainsi la prédiction des parcelles classées en 10 sous-classes. En outre, les classifications étant hiérarchiques, les 10 sous-classes peuvent être regroupées en 4, les performances du modèle pouvant alors être comparées avec le modèle 4 classes.

- La *méthode indirecte*

La méthode indirecte, consiste à utiliser le modèle prédisant les 10 sous-classes sur les parcelles préalablement prédites en « printemps » par le modèle 4 classes. Il est nommé **modèle 10 classes indirect**.

Pour récapituler, ces trois modèles mènent à quatre prédictions (Figure 8) : prédiction en 4 classes à partir du modèle 4 classes ou du modèle 10 classes direct, et prédictions en 10 sous-classes à partir des modèles 10 classes indirect ou direct

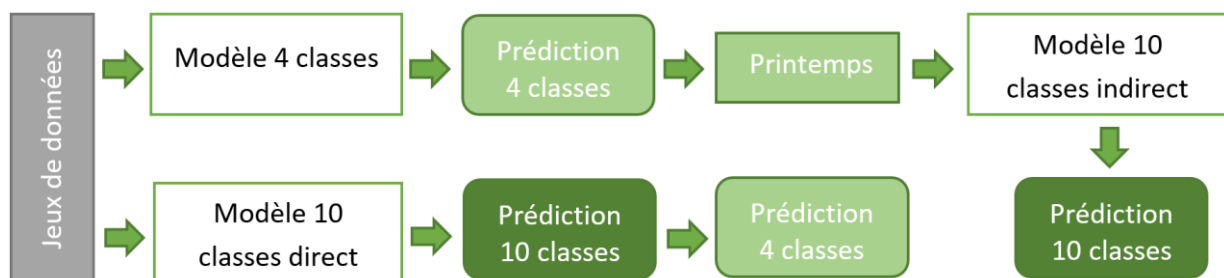


Figure 8 : Schéma de l'utilisation des trois types de modèle développés dans la prédiction par les rotations.

4.4.4. Calibration des modèles

La calibration des modèles est la composante essentielle de la prédiction par rotations. Les points suivants énumèrent la manière avec laquelle les modèles ont pu être créés. Dans un premier temps, la formation des tables de calibration pour entraîner les modèles est présentée. Le point suivant énonce les variables maximisant les performances de prédictions des modèles qui ont été testées. Ensuite, l'élaboration du modèle de 10 classes direct est expliquée.

4.4.4.1. Construction des tables de calibration

La préparation des données (4.4.1) a montré comment obtenir les successions de cultures sur les parcelles. Les successions de cultures entre deux, trois, quatre et cinq années successives vont pouvoir être utilisées pour calibrer le modèle. Afin d'obtenir une prédiction plus robuste, un maximum de données était nécessaire pour effectuer la calibration. Ainsi, pour la calibration d'un modèle sur base des rotations sur deux années, une table de calibration a été créée par la compilation des intersections de 2014-2015, 2015-2016, 2016-2017 et 2017-2018. De même, le modèle de rotation sur trois années se base sur une table contenant les intersections de 2014-2015-2016, 2015-2016-2017 et 2016-2017-2018 (Figure 9), et ainsi de suite jusqu'à la rotation sur cinq ans qui ne se basera que sur l'intersection de 2014-2015-2016-2017-2018. Plus le nombre d'années pris en compte dans la rotation est élevé, moins les données utiles à la calibration sont nombreuses (Tableau 5). Le modèle issu de la calibration sur une rotation de cinq ans en sort donc moins robuste.

Tableau 5 : Nombre de parcelles dans les tables de calibration regroupant les données issues des intersections des parcellaires sur 2, 3, 4 et 5 années de rotations.

<i>Nombre d'années de rotations</i>	2 ans	3 ans	4 ans	5 ans
<i>Nombre de parcelles</i>	13111	9545	5870	2802

L'occurrence de rencontrer chaque succession de cultures a donc pu être calculée à partir des tables de calibration pour les deux classifications. Elles sont disponibles en annexe (Annexe C, Tableau A-27 et Tableau A-28).

Pour déterminer les variables à utiliser afin de créer le modèle, une série de tests ont été effectués. Afin de valider ou non la prise en compte d'une variable dans le modèle, une mesure de l'exactitude du modèle devait être effectuée sur une partie des données non utilisées dans la création du modèle. La méthode de construction de RF prévoit une validation interne (*Out of Bag Error*). Afin d'avoir la maîtrise des parcelles de validations et de l'exploitation des résultats (créer des matrices de confusions, calculer les exactitudes, précision, etc.), une validation externe a été faite. En pratique, les tables de calibration construites ci-dessus sont séparées de manière aléatoire : 2/3 de la table sert de données pour l'entraînement du

modèle (calibration), l'autre tiers est ensuite utilisé pour évaluer l'exactitude de la prédiction (validation) (Figure 9).

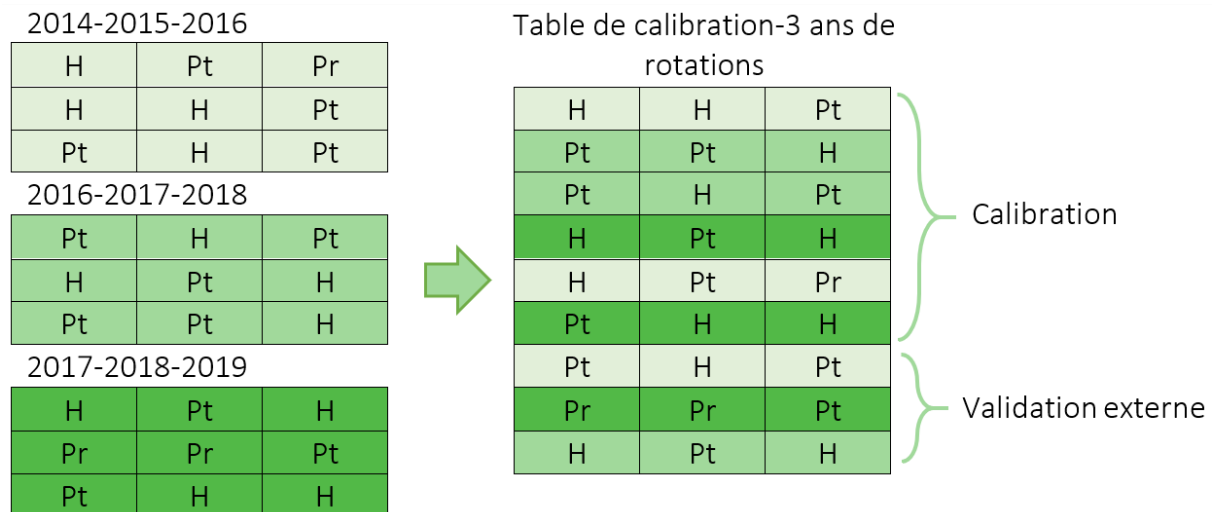


Figure 9 : Illustration de la création du tableau de calibration de 3 ans de rotation (H= Hiver, Pt = Printemps, Pr = Prairie).

4.4.4.2. Evaluation des prédictions

L'évaluation de l'exactitude des prédictions se base sur des matrices de confusion. Les matrices de confusion sont des tableaux croisés (référence/classification) confrontant les résultats d'une classification (ou d'une prédiction) à la réalité observée (référence). Elles permettent de quantifier les éventuelles erreurs de classifications sur chacune des classes. Le Tableau 6 montre le type de matrice de confusion utilisée afin d'analyser les résultats des différentes prédictions.

Tableau 6 : Exemple d'une matrice de confusion.

		Référence					
		A	B	C	Total _p	Cont.	Er.C.
Prédiction	A	A _r A _p	B _r A _p	C _r A _p	A _r A _p +B _r A _p +C _r A _p	B _r A _p +C _r A _p	Cont. _A /Total _{pA}
	B	A _r B _p	B _r B _p	C _r B _p	A _r B _p +B _r B _p +C _r B _p	A _r B _p +C _r B _p	Cont. _B /Total _{pB}
	C	A _r C _p	B _r C _p	C _r C _p	A _r C _p +B _r C _p +C _r C _p	A _r C _p +B _r C _p	Cont. _C /Total _{pAC}
	Total _r	A _r A _p +A _r B _p +A _r C _p	B _r A _p +B _r B _p +B _r C _p	C _r A _p +C _r B _p +C _r C _p	Total		
	Om.	A _r B _p +A _r C _p	B _r A _p +B _r C _p	C _r A _p +C _r B _p			
	Er. O.	Om. _A /Total _{rA}	Om. _B /Total _{rB}	Om. _C /Total _{rC}			A _r A _p +B _r B _p +C _r C _p

Dans laquelle :

- X_rY_p représente le nombre d'objets classés (ou prédits) en classe Y qui en réalité font partie de la classe X.
- Total_{rX} est le total d'objets appartenant réellement à la classe X et Total_{pX} est le nombre d'objets totaux classés (ou prédits) dans la classe X.

- $Om.X$ est l'omission d'une classe, soit le nombre d'objets de la classe X réelle qui ne sont pas classés (prédits) comme tel. $Er. O_X$ est l'erreur d'omission, soit le nombre d'objets de la classe réelle X non prédits dans la classe X sur le nombre total d'objets de la classe réelle X.
- $Cont.X$ est la contamination, soit le nombre d'objets prédits dans la classe X mais n'appartenant pas à X dans la réalité. $Er.C.X$ est l'erreur de contamination, soit la somme des objets mal prédits en classe X sur le total des objets prédits en X.

De cette matrice de confusion, il est possible de retrouver les trois indicateurs permettant d'apprécier la qualité des prédictions : l'exactitude globale moyenne, la sensibilité d'une classe et la précision d'une classe (équation 4, 5 et 6 ; [Parikh et al., 2008]). Comme l'intérêt se porte plus sur le fait d'obtenir une bonne prédiction sur une plus grande superficie et non pas sur le plus grand nombre de parcelles, le calcul de ces indices de qualité s'est fait en utilisant l'aire des parcelles et non pas le nombre.

$$Exactitude\ globale = \frac{\sum \text{aires des objets correctement prédits}}{\text{Superficie totale d'objets}}$$

4

$$Précision\ d'une\ classe = \frac{\sum \text{aires des objets de la classe correctement prédits}}{\sum \text{aires des objets prédits dans cette classe}}$$

5

$$Sensibilité\ d'une\ classe = \frac{\sum \text{aires des objets de la classe correctement prédits}}{\sum \text{aires des objets réels de cette classe}}$$

6

Dans le cas de la matrice de confusion de l'exemple, l'exactitude globale vaudrait $(A_r A_p + B_r B_p + C_r C_p) / Total$. La sensibilité d'une classe X vaudrait $X_r X_p / Total_{rX}$ ou $1 - Er. O_X$. Et la précision de cette classe vaudrait $X_r X_p / Total_{pX}$ ou $1 - Er. C_X$.

Lors de la calibration, chaque test a été effectué à plusieurs reprises, au minimum cinq fois, en réitérant à chaque fois la séparation aléatoire (2/3-1/3) du jeu de données global afin de pouvoir calculer une moyenne et un écart type à l'exactitude et ainsi connaître la stabilité et l'amélioration éventuelle que pourrait apporter le modèle testé.

En ce qui concerne les méthodes de 4 classes, la classe d'intérêt principal pour une prédiction optimale est la classe « Printemps ». L'appréciation des améliorations éventuelles des modèles testés s'est donc faite sur base de la précision et de la sensibilité de cette classe au lieu de l'exactitude globale.

Dans le cadre de la prédiction des cultures de printemps avec l'objectif de prédire les zones à risques d'inondations boueuses, il vaut mieux maximiser la sensibilité que la précision. En

effet, dans une logique de « principe de précaution », il vaut mieux surestimer le nombre de cultures de printemps que le sous-estimer afin de ne pas sous-estimer le risque d'inondations. Cela étant dit, une mauvaise précision va induire des coûts de protection contre les pertes de sol inutile. Il faut donc trouver un bon équilibre entre les deux indices de qualité.

4.4.4.3. Choix des variables

Plusieurs tests ont été effectués afin de connaître le nombre d'année de rotation, les variables du modèle, et le type de parcelles d'entraînement à prendre en compte dans les modèles pour obtenir les meilleures performances de prédictions. Les tests ont été effectués pour les modèles 4 classes et 10 classes indirect. Le modèle 10 classes direct a ensuite été développé en se basant sur les variables retenues pour le modèle 10 classes indirect.

Comme la méthode indirecte ne se base que sur les cultures préalablement prédites en culture de printemps (4.4.3), la calibration des modèles ne devait se faire que sur les parcelles dont l'année n est une culture de printemps. Les parcelles dont l'année n est une culture d'hiver, autre ou une prairie, ont donc été enlevées des tables de calibration de ce modèle.

Les différentes variables (et suppressions de parcelles) testées dans la création des modèles de 4 et 10 classes indirect sont répertoriées ci-dessous.

Nombre d'année de rotation

Les premières variables du modèle sont bien évidemment les cultures de chaque année prises en compte dans la rotation voulue. Les modèles de rotation sur deux années n'auront que la variable de l'année $n-1$ pour prédire l'année n , les rotations sur trois années en comprendront deux (Tableau 7) et ainsi de suite. Le nombre d'années de rotation permettant d'obtenir la meilleure exactitude globale (dans le cas du modèle de 10 classes) où la meilleure balance entre précision et sensibilité (dans le cas du modèle de 4 classes) va être gardée.

Tableau 7 : Exemple des variables inhérentes au modèle de 4 classes de rotation de 3 ans.

Variables			Classes à prédire
<i>Culture n-2</i>	<i>Culture n-1</i>		<i>Culture n</i>
Printemps	Hiver		Printemps
Hiver	Printemps		Hiver
Prairie	Prairie		Printemps

Ensuite, afin d'améliorer l'exactitude du modèle, des variables supplémentaires où la suppression de certaines parcelles (pouvant apporter de l'erreur dans les prédictions) dans les jeux de données d'entraînements, ont été testées une par une. Les variables d'intérêts ont ensuite été repérées, grâce à leurs améliorations sur la performance du modèle. Des tests combinant ces variables d'intérêts dans le modèle ont ensuite été effectués afin d'observer si l'apport de ces variables associées est bénéfique ou non sur les performances du modèle. Le

modèle testé apportant la meilleure performance est finalement le modèle choisi pour la prédiction finale.

- **Séquences de cultures**

La méthode de RF n'utilisant pas nécessairement toutes les variables comme prédicteurs de ces arbres (4.4.2), elle ne tient pas compte de la séquence de ces cultures. Or, les successions de cultures peuvent apporter un intérêt en tant que prédicteur. L'ajout de nouvelles variables rassemblant les cultures des années précédentes permet d'y remédier. Par exemple, pour un modèle utilisant cinq années de rotation, trois variables sont ajoutées, une prenant la séquence des deux années antécédentes (exemple : Printemps-Hiver), une la séquence des trois années antécédentes (ex : Hiver-Printemps-Hiver) et enfin celles prenant en compte les quatre années (ex : Autre-Hiver-Printemps-Hiver).

Un problème est survenu : la fonction RF de R ne permettant pas qu'une variable possède plus de 53 classes. Or, pour la prédiction des rotations de cinq années, le nombre de classes théoriques en prenant en compte la séquence des quatre années précédentes est de $4^4=256$, ce qui est largement supérieur aux 53 classes possibles. Le même problème est observé dans les modèles de 10 classes pour les rotations de plus de deux ans. Afin de pouvoir utiliser l'information sur les séquences, chaque variable de séquence comprenant plus de 53 classes a été transformée en une multitude de variables. Chaque nouvelle variable correspond à une séquence (exemple : Autre-Hiver-Printemps-Hiver) et est exprimée de manière binaire, 1 quand elle est présente dans la donnée, 0 quand elle ne l'est pas (Tableau 8). Les variables sous-représentées (moins de cinq observations) ne sont pas utilisées, afin d'éviter d'avoir des variables non représentées dans les parcelles d'entraînement mais qui le sont dans les parcelles de validations, et ainsi de prévenir des problèmes dans les prédictions.

Tableau 8 : Illustration de l'astuce trouvée pour contrer l'algorithme RF de R qui n'accepte que 53 classes.

Séquences		Print-Hiver-Print	Hiver-Hiver-Print	Prair-Prair-Print
Print-Hiver-Print		1	0	0
Hiver-Hiver-Print		0	1	0
Prair-Prair-Print		0	0	1

Une vérification sur un modèle de 4 classes de quatre ans de rotation a montré que cette astuce induit une variation qui reste cependant très légère. Ainsi, tenir compte de la séquence en une seule variable amène à une exactitude globale de 68.5% contre 68.64% pour le même modèle utilisant une variable par séquences présentes.

- **Superficies**

Les superficies des parcelles pourraient avoir de l'importance dans le choix des cultures. En effet, on a remarqué précédemment que les superficies des prairies (couvertures, fourrages, etc.) ou des parcelles de maïs sont en moyenne plus petites. À l'inverse, les pommes de terre

et betteraves occupent une surface plus importante pour un même nombre de parcelle. La prise en compte des aires des parcelles comme variable a donc été testée.

- **Parcelles évoluant au cours du temps**

Il a été souligné lors de la préparation des données que certaines parcelles à problèmes (cas 1, 2 et 3 (4.4.1, Figure 5)), pourraient apporter une certaine imprécision aux prédictions. Exclure ces parcelles des données d'entraînement permettrait alors de s'affranchir d'erreurs éventuelles.

Une manière de supprimer ces parcelles se modifiant au cours du temps est de comparer l'aire de l'intersection avec l'aire de la parcelle de chacune des années. Les parcelles dont le quotient de l'aire intersectée sur l'aire de la parcelle est inférieur à un certain seuil (par exemple 80 %) sont alors supprimées du jeu de données de calibration. Plusieurs valeurs seuil ont été testées afin d'optimiser la prédiction du modèle.

- **Parcelles de type « Autre »**

Puisque le nombre de parcelles de type « Autre » varie fortement au cours du temps, un test a été réalisé pour évaluer l'erreur de l'inclusion de ces parcelles dans les prédictions. Pour ce faire, un modèle a été construit en excluant l'ensemble des parcelles ayant fait partie de la classe « Autre » au moins sur une année.

- **Parcelles de petite taille**

Les petites parcelles en superficie pourraient apporter du bruit en ne suivant pas le schéma classique de rotation de la zone. En effet, les agriculteurs effectuent plus souvent des tests de cultures sur ce genre de parcelles en comparaison avec les plus grandes. De plus, prédire ce type de parcelles est moins important compte tenu de l'objectif de ce mémoire, celles-ci ayant un effet moindre sur les inondations boueuses en raison de leur faible taille et de la faible superficie totale occupée. Différents modèles ont donc été construits, ne prenant pas en compte dans les parcelles d'entraînements, les parcelles dont l'aire est inférieure à un seuil choisi (par exemple 1 hectare).

Variables complémentaires pour les modèles à 10 classes

Par manque de résultats satisfaisants pour les modèles de 10 classes, des tests additionnels ont été effectués :

- **Sur-représentation des classes faibles**

Au regard des prédictions obtenues dans les autres tests, les classes de cultures non sarclées (les plus faibles en termes de parcelles) sont sous représentées et mal prédites. En effet, l'algorithme RF a tendance à prédire la plupart des données dans les trois classes, importante en nombre, de cultures sarclées (représentant environ 70% des cultures de printemps) afin d'augmenter l'exactitude globale du modèle. Or, l'intérêt d'obtenir une prédiction sur base d'une classification plus fine, était de discriminer ces cultures non sarclées, des sarclées qui posent plus de problèmes d'inondations boueuses. Il était donc utile d'augmenter le poids des

classes de cultures non-sarclées dans les parcelles d'entraînement et de comparer les exactitudes, avant et après la modification.

Pour augmenter le nombre d'observations des classes faibles, le logiciel R permet de suréchantillonner les données par la fonction ROSE (*Random Over-Sampling Examples*). Cette fonction permet entre-autres de reproduire, de manière aléatoire, les données d'une variable mal équilibrée. Les classes de lins textiles, de pois et de céréales d'hiver de la variable représentant les cultures de l'année à prédire ($n+1$) des parcelles d'entraînement ont ainsi été suréchantillonnées. La proportion de chacune des classes a été augmentée à hauteur de 10% des parcelles d'entraînement. Cela permettait d'avoir, dans les parcelles d'entraînement, plus ou moins 20% de pomme de terre, 20% de betterave, 20% de maïs, 10% de lins textiles, 10% de pois, 10% de céréales d'hiver et 10% d'autres cultures de printemps.

Des variables supplémentaires, comme le type de sol et le relief des parcelles, peuvent aussi influencer le choix de la culture. Cela est d'autant plus vrai lorsqu'on affine la classification. C'est pourquoi ces tests n'ont pas été effectués lors de la calibration de type 4.

- **Type de sol**

Le type de sol pourrait influencer le choix de cultures de l'agriculteur. En effet, les différentes propriétés des sols (rétention et teneur en eau, propriétés chimiques, charge caillouteuse, etc.) jouent un rôle majeur dans les caractéristiques agronomiques du sol et du choix de la culture. Pour obtenir le type de sol à l'échelle des parcelles, la carte numérique des sols de Wallonie (CNSW) a été utilisée. La CNSW (disponible sur le Géoportail de Wallonie) est une numérisation de la carte des sols de Wallonie. Elle regroupe 6000 types de sols sur base de sa nomenclature propre en trois lettres. La première lettre exprime la texture du sol, la deuxième caractérise le drainage naturel du sol et la troisième la présence d'un horizon diagnostique classant le sol [Géoportail Wallonie].

Les types de sol n'étant pas uniformes sur les parcelles, celui se trouvant à l'intersection avec le centroïde des parcelles a été retenu comme type de sol de la parcelle.

Un premier test utilise la classification du CNSW comme variable du modèle. La classification comportant plus de 53 classes, certaines classes aux propriétés similaires ont été regroupées entre-elles (comme par exemple les sables tous ensembles) de manière à obtenir finalement 15 classes (le détail des classes gardées est disponible en Annexe C, Tableau A-29). Deux autres tests de nouvelles variables ont aussi été fait, un d'eux n'utilisant plus que la texture, et l'autre recourant au drainage naturel.

- **Relief**

Le relief des parcelles et la localisation de celles-ci dans le relief peuvent aussi influencer le choix des cultures. Certaines cultures se cultivent plus difficilement sur de fortes pentes, c'est

pourquoi on retrouve souvent des prairies dans ce genre de parcelles ainsi que du maïs ensilage qui a remplacé beaucoup de prairies.

Grâce au modèle numérique de terrain (MNT) lissé (LIDAXES) (disponible sur le Géoportail de Wallonie) de la région wallonne, la pente est calculable sur un système d'information géographique. Un test a été effectué en prenant en compte la pente moyenne des parcelles comme variable du modèle.

La localisation même des parcelles dans le relief a aussi un intérêt. En effet, le ruissellement des précipitations dans le relief va rendre des zones plus ou moins susceptibles de recevoir de l'eau par ruissellement. Des zones réceptrices de nombreux volumes d'écoulement seraient donc moins aptes à accueillir des cultures souffrant de l'excès d'eau. Le MNT a permis de calculer l'accumulation de flux, qui montre pour chaque pixel le nombre de pixel contributeur à l'écoulement de l'eau sur ce pixel. Un modèle a été testé en prenant en compte la médiane de l'accumulation de flux sur chaque parcelle comme variable. La médiane est davantage pertinente que la moyenne dans ce cas-ci, car la disparité des valeurs entre un grand axe d'écoulements et un pixel « lambda » est importante et, de plus, dans ce même axe d'écoulement une grande différence de valeur existe entre l'amont et l'aval d'un bassin. La médiane permet de s'affranchir de ces grandes valeurs.

Une fois les choix de variables pour les modèles de 4 classes et de 10 classes (méthode indirecte) fait, le modèle de prédictions de 10 classes par la méthode directe peut être construit.

4.4.4.4. Modèle de 10 classes, méthode directe

Les variables maximisant les performances du modèle 10 classes indirect, ont été reprises pour la méthode directe. Pour rendre les résultats de la méthode directe comparable à ceux obtenus par la méthode indirecte, il a fallu diminuer le poids de la classe « Hiver », qui est beaucoup trop important (à peu près 40 % des cultures se trouvant dans cette classe). En effet, le modèle a tendance à prédire toutes les cultures en « Hiver » afin d'obtenir une meilleure exactitude.

Il est possible d'y remédier en diminuant la proportion des cultures d'hiver avec l'outil ROSE décrit précédemment. Afin de connaître cette proportion, un regroupement des cultures de printemps a été effectué pour être comparé au modèle de 4 classes retenu. Après plusieurs essais, c'est la proportion de 25% de cultures de la classe « hiver » qui permet d'obtenir des performances de prédictions de la classe hiver, similaires à celles du modèle de 4 classes.

Les prédictions ainsi obtenues ont ensuite été comparées à celles des modèles 10 classes indirect et 4 classes (4.4.3).

4.4.4.5. Calibration finale

Une fois les trois modèles développés, n'ayant plus besoin de parcelles de validation dans la phase finale des prédictions, la séparation des tables de calibration en 2/3-1/3 (4.4.4.1) n'a plus beaucoup de sens. Les entraînements des modèles finaux utilisés pour la prédiction de 2019 vont alors s'effectuer sur l'ensemble des tables de calibrations (les 3/3) afin de les rendre plus robustes. Les modèles d'intérêt vont ensuite pouvoir être validés sur l'année 2019.

4.4.5. Prédiction 2019, validation des modèles

Il s'agit de prédire les cultures de printemps de 2019. Ne disposant pas anticipativement du parcellaire de 2019, ce sont sur les limites de parcelles de 2018 qu'ont été projetées les prédictions. Cela peut donc créer des erreurs là où le parcellaire 2019 n'est pas identique au parcellaire 2018. Une comparaison de l'évolution des limites de parcelles entre 2018 et 2019 a été faite pour se rendre compte de ces erreurs induites par l'évolution du parcellaire.

Afin de ne pas perdre la trace de la localisation géographique des parcelles en envoyant les données du SIG vers R, un identifiant a été donné à chaque parcelle de 2018, avant les intersections expliquées dans l'étape du prétraitement. Cet identifiant permet d'associer aisément les cultures prédites aux parcelles du parcellaire après la prédiction.

Le/les modèle(s) retenu(s) sont ensuite appliqué(s) sur les parcelles intersectées des années de rotation correspondantes au(x) modèle(s). Par exemple, si un modèle retenu est de quatre années de rotations, les parcelles sur lesquelles vont être appliquées le modèle seront celles de l'intersection entre les années 2016, 2017 et 2018.

Enfin, une fois les prédictions faites dans le logiciel R, elles ont été projetées sur les parcelles de 2018 et comparées aux cultures réelles du SIGEC de 2019. Cette comparaison entre les deux peut être effectuée par une union ou une intersection des deux couches. L'intersection ne permet pas de garder les parcelles apparues en 2019. Elle permet en revanche d'avoir une idée de l'exactitude des/du modèle choisi pour la prédiction contrairement à une union qui garde des parcelles impossibles à prédire pour le modèle.

4.5. Détection des couverts hivernaux

Comme expliqué précédemment, la détection des couverts hivernaux par imagerie satellite pourrait permettre de prédire une majorité des cultures de printemps. Pour ce faire, le processeur SEN4CAP spécifique à la détection de pratiques culturales, *L4C-Agricultural practices monitoring product processors* (L4C), a été utilisé. Il s'agit d'un modèle logique utilisé pour détecter des pratiques déclarées dans le cadre de la PAC, comme la déclaration des couverts hivernaux par exemple. Le fait qu'il s'agisse d'un ensemble de règles logiques fonctionnant sur une série temporelle de métriques lui donne l'avantage d'être utilisable d'années en années contrairement aux classifications utilisant une séquence d'images. En effet, dans ce cas, ces séries d'images sans nuages sont fonction de la météo, et évoluent différemment chaque année (semis et récoltes plus précoces ou tardives en fonction de facteurs externes).

Cette section aborde en premier lieu le fonctionnement du processeur. Ensuite, les phases de calibration et de validation sont expliquées. Le processeur a été dans un premier temps calibré sur l'année 2018, pour détecter les couverts et les différencier des cultures hivernales. Par la suite, le modèle de détection a été validé par photo-interprétation sur l'année 2019. Enfin différentes méthodes d'améliorations des prédictions par la combinaison, avec les modèles de rotations, sont abordées.

Données d'entrée utilisées

- Images prétraitées de cohérences et CR dérivées de Sentinel-1, et de NDVI dérivées de Sentinel-2 et Landsat 8.
- Parcellaires Agricoles anonymes 2018 et 2019.
- Périodes attendues de végétation et de récoltes des cultures principales présentes sur chaque parcelle des parcellaires utilisés, ainsi que la période sur laquelle on s'attend à détecter les cultures hivernales.
- Différents paramètres, définissant les seuils de NDVI, cohérence et CR permettant de détecter une récolte et un couvert hivernal.

Logiciels utilisés

Le processeur SEN4CAP L4C est séparé en deux grandes étapes : l'extraction et l'exploitation des données. L'extraction des métriques s'effectue avec un code Python tandis que le modèle logique de décision s'effectue avec un code R.

4.5.1. Fonctionnement du processeur

4.5.1.1. Extraction des données

Le processeur est séparé en deux grandes étapes. La première se déroule avec l'aide du logiciel Python. Elle consiste en l'extraction des métriques calculées par le processeur des images satellites SEN4CAP. Les métriques utilisées par le processeur sont le NDVI, le CR et la cohérence. Elles proviennent d'images prétraitées obtenues par les satellites à imagerie radar, Sentinel-1 (CR et cohérence), et à imagerie optique Sentinel-2 et Landsat 8 (NDVI).

L'extraction se fait sur l'ensemble d'images prises lors de la période de végétation et de récolte de la culture principale ainsi que sur la période de détection du couvert, préalablement définie à cette étape.

Le processeur va utiliser les limites des parcelles issues d'un parcellaire (LPIS) afin d'obtenir des statistiques (moyenne, déviation standard, minimum et maximum) sur les métriques pour chaque parcelle. Un tampon interne (Figure 10) est préalablement appliqué aux parcelles, afin d'éviter la prise en compte de pixels chevauchant les limites de parcelles, source potentielle d'erreur. La distance entre le tampon et la limite de la parcelle vaut la moitié de la taille des pixels de l'image. De cette manière, tous les pixels compris dans le tampon font bien partie de la parcelle. Les distances choisies sont de 5 m pour les images issues de Sentinel-2 et de 10 m pour celles de Sentinel-1 et Landsat-8. Lorsque les statistiques montrent une mauvaise uniformité des résultats dans une parcelle, les données ne sont pas prises en compte. A l'inverse, lorsqu'elles sont uniformes, les statistiques sont sauvegardées, accompagnées de la date de capture de l'image, dans un fichier intermédiaire qui va être utilisé dans la deuxième étape. Pour les images optiques, un masque de nuage est appliqué préalablement au calcul des statistiques, pour s'assurer qu'au moins 10% des pixels de la parcelles renvoient des valeurs.

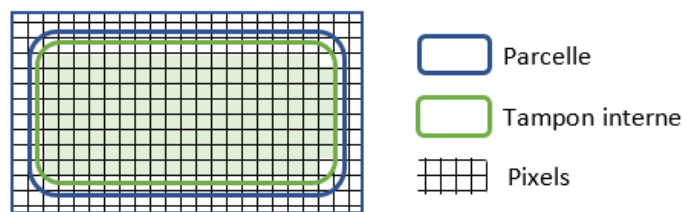


Figure 10 : : Illustration du tampon interne et de son utilité dans le traitement des images.

4.5.1.2. Exploitation des données

La deuxième partie du processeur SEN4CAP mobilise un code R, dérivant un modèle logique de décision utilisant des marqueurs successifs pour détecter la présence éventuelle d'une certaine pratique agricole (*catch crop* (couvert hivernal) dans le cadre de ce mémoire) sur chaque parcelle. Un marqueur est un indicateur qui évalue la valeur d'une métrique en la comparant à une valeur seuil. Il retourne une appréciation sous forme binaire (OUI / NON) signifiant si le seuil est atteint ou non. Afin de détecter les couverts hivernaux, le modèle va d'abord s'assurer de la présence d'une récolte en utilisant cinq marqueurs dans une période spécifique à la récolte. Seulement lorsque celle-ci est décelée, cinq autres marqueurs vont permettre de déceler la présence ou l'absence d'une pratique culturale durant une période de détection choisie.

Fonctionnement des marqueurs

Les marqueurs sont testés de manière hebdomadaire. Les données extraites au point 4.5.1.1 doivent donc être transformées pour être utilisables dans les marqueurs. Pour chaque semaine (commençant le lundi), les données extraites sont calculées de manière à obtenir les

moyennes du CR et du NDVI, les valeurs maximales de CR et de cohérence observées dans la semaine, ainsi que l'augmentation maximale de cohérence dans la semaine. De plus, la moyenne du CR sur trois semaines, la différence entre les CR moyens d'une semaine et la moyenne des trois semaines précédentes vont être calculées.

Tous les marqueurs sont composés de valeurs seuils permettant, par exemple, de discriminer certaines pratiques de cultures. Dans le cadre de ce mémoire, les valeurs seuils du modèle calculé pour les Pays-Bas, ont été utilisées. Celles-ci sont disponibles en annexe (Annexe D, Tableau A-30).

Les 5 premiers marqueurs, utilisés pour détecter une éventuelle récolte sur la parcelle sont :

- *Marqueur 1 : Présence de végétation sur la parcelle évaluée grâce aux valeurs de NDVI.*

Ce marqueur permet de s'assurer de la présence d'une culture principale sur la parcelle avant la récolte. Les moyennes hebdomadaires de NDVI sont comparées à une valeur seuil. Lorsque la valeur moyenne du NDVI sur une semaine est supérieure à la valeur seuil, le marqueur 1 est activé (noté TRUE). Si aucune des valeurs moyenne de NDVI ne dépasse le seuil dans la période de végétation, le modèle estime qu'il n'y a pas de culture implantée sur la parcelle en question. Le marqueur 1 est noté FALSE, et le processeur ne cherche pas à détecter de date de récolte ni de pratique culturale sur la parcelle grâce aux autres marqueurs.

- *Marqueur 2 : Perte de végétation dans la parcelle évaluée grâce aux valeurs de NDVI.*

Pour chaque semaine de la période de récolte, le processeur compare la valeur de NDVI de la semaine précédente et suivante. Lorsque la perte de NDVI est supérieure à un seuil (calculé grâce à la valeur moyenne de NDVI sur la semaine concernée et les différents seuils du marqueur 2 (Annexe D, Tableau A-30)), ou que la valeur de NDVI descend sous la valeur seuil minimum de NDVI, le marqueur détecte une perte de végétation.

- *Marqueur 3 : Perte de végétation dans la parcelle évaluée grâce aux valeurs de CR.*

Ce deuxième marqueur évaluant une perte de végétation, compare chaque semaine la différence de CR entre la moyenne de celui-ci dans la semaine concernée, et la moyenne des trois semaines précédentes. La perte est alors détectée lorsque cette différence dépasse le seuil de perte de CR calculé sur base de la déviation standard d'une série temporelle choisie. Le marqueur est évalué sur l'ensemble des semaines encodées comme période de récolte pour la culture en place sur la parcelle.

- *Marqueur 4 : Absence ou présence faible de végétation sur la parcelle évaluée grâce aux valeurs de CR.*

Chaque semaine de la période de récolte, le ratio maximal de CR est comparé à un seuil d'absence de végétation calculé sur base de la moyenne et de la déviation standard d'une

série temporelle choisie. Lorsque ce-dernier est dépassé, le marqueur détecte la présence d'un sol nu ou faiblement couvert.

- *Marqueur 5 : Absence ou présence faible de végétation dans la parcelle évaluée grâce aux valeurs de cohérences.*

Chaque semaine de la période de récolte, l'augmentation maximum de cohérence, ainsi que la valeur maximum de cohérence, sont comparées aux seuils du marqueur 5. Lorsque les seuils sont dépassés, le marqueur détecte une faible couverture végétale.

Une fois les cinq marqueurs détectés, une récolte est présumée sur la parcelle. La semaine dans laquelle la récolte est observée est enregistrée sur le fichier de sortie. Si le processeur ne détecte pas de récolte sur la parcelle, ou s'il n'avait pas détecté de culture avec le marqueur 1, la recherche d'une éventuelle pratique culturale sur la parcelle n'est pas effectuée.

Les cinq marqueurs suivant, utilisés pour détecter une pratique culturale dans une période définie, sont :

- *Marqueur 6 : Présence de végétation sur la parcelle évaluée grâce aux valeurs de NDVI.*

Chaque semaine de la période de détection, les moyennes hebdomadaires des NDVI obtenus sont comparées à un seuil. Lorsque le seuil est dépassé, une végétation est détectée sur la parcelle.

- *Marqueur 7 : Croissance de végétation sur la parcelle, évaluée grâce aux valeurs de NDVI.*

Chaque semaine de la période de détection, un seuil est calculé sur base de la moyenne du NDVI et des valeurs limites implémentées pour le marqueur 7. Les valeurs moyennes de NDVI doivent rester supérieures à ce seuil calculé ainsi qu'à un seuil minimum pour qu'une croissance de végétation soit diagnostiquée.

- *Marqueur 8 : Pas de perte de végétation sur la parcelle, évaluée grâce aux valeurs de NDVI*

Chaque semaine, la moyenne hebdomadaire de NDVI est comparée au seuil calculé pour le marqueur 7. Elle doit rester supérieure à cette valeur sur l'ensemble de la période de détection pour que le marqueur soit noté *TRUE*.

- *Marqueur 9 : Pas de perte de végétation sur la parcelle, évaluée grâce aux valeurs de CR*

Une régression linéaire est effectuée entre les données de CR entre la deuxième et l'avant-dernière semaine de la période de détection. La pente et la p-valeur de la régression est ensuite calculée. Lorsque cette pente est inférieure à une pente seuil et que la p-valeur est plus faible que 0.05, le marqueur est noté *TRUE*, et le processeur ne détecte pas de perte de végétation sur la parcelle.

- *Marqueur 10 : Présence de végétation sur la parcelle évaluée grâce aux valeurs de cohérence*

Chaque semaine, l'augmentation maximale de la cohérence, ainsi que la valeur maximale des cohérences sont comparées à des seuils. Si l'une d'elles dépasse les seuils, une perte de végétation est détectée sur la parcelle.

La Figure 11 ci-dessous présente l'évolution des trois métriques prises en compte dans le processeur sur une parcelle accueillant un couvert.

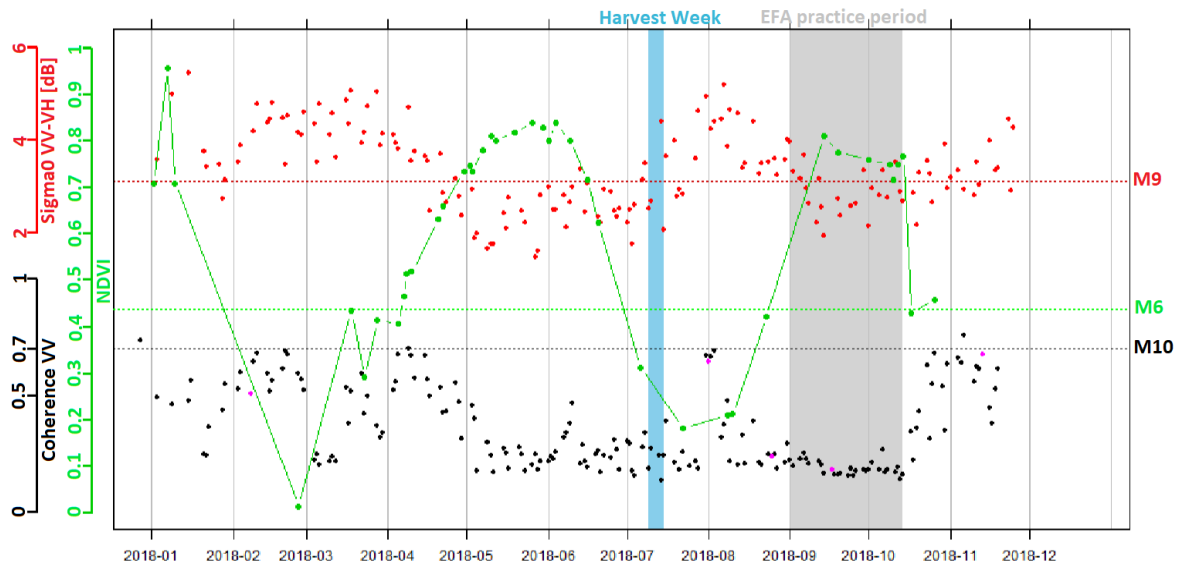


Figure 11 : Exemple de l'évolution des trois métriques d'une parcelle dans laquelle le processeur détecte un couvert hivernal [SEN4CAP].

Lorsque la pratique culturale à décoder est un couvert hivernal, l'appréciation de sa présence éventuelle se réalise comme ceci :

- Lorsque les cinq marqueurs donnent « *TRUE* » :
→ Confiance **forte** en la présence d'un couvert hivernal sur la parcelle.
- Lorsque M6 (marqueur 6) donne « *TRUE* » et qu'un seul des autres marqueurs donne « *FALSE* » :
→ Confiance **modérée** en la présence d'un couvert hivernal sur la parcelle.
- Lorsque M6 donne « *TRUE* » et que deux autres marqueurs donnent « *FALSE* » :
→ Confiance **faible** en la présence d'un couvert hivernal sur la parcelle.
- Dans le cas où les valeurs de NDVI ne sont pas utilisables dans l'intervalle de temps (M6, M7 et M8 donnent « *NA* »). Lorsque M9 et M10 donnent « *TRUE* » :
→ Confiance **modérée** en la présence d'un couvert hivernal sur la parcelle.
- Dans le cas où M6, M7 et M8 donnent « *NA* ». Lorsque seulement l'un des deux autres marqueurs ne donne « *TRUE* » :
→ Confiance **faible** en la présence d'un couvert hivernal sur la parcelle.
- Dans tous les autres cas :
→ Peu de confiance (**pauvre**) dans la présence d'un couvert hivernal sur la parcelle.

Finalement, un fichier de sortie recueille les résultats des différents marqueurs (noté TRUE, FALSE ou NA) pour chacune des parcelles. Ce fichier donne aussi une information sur la détection d'une récolte et la semaine dans laquelle elle a été détectée (par Sentinel-1 et 2), l'appréciation de la présence d'un couvert, ainsi qu'une information sur la qualité (nombre de pixels sur la parcelle) et la quantité (nombre d'images dans la période de détection) des données sentinel-1 utilisées.

4.5.2. Calibration

Le but de la calibration était de pouvoir détecter un maximum de couverts hivernaux, tout en essayant de maintenir une contamination par des cultures hivernales raisonnable en utilisant le processeur. Les seuils des marqueurs calculés pour le processeur aux Pays-Bas n'ont pas été modifiés. La calibration a plutôt été effectuée en modifiant la période de détection de la pratique culturale.

Choix des parcelles à tester

Dans le fichier d'entrée contenant les parcelles à examiner par le processeur, il n'était pas nécessaire d'ajouter les parcelles dans lesquelles les cultures principales étaient récoltées lors de la période de détection. Premièrement, car le processeur n'y détecterait pas un couvert (les marqueurs de pertes de végétation seraient notés FALSE). Deuxièmement, parce que ces parcelles sont peu susceptibles d'accueillir un couvert dans cette période, étant récoltées après le 1^{er} septembre, l'implantation de CIPAN n'y est pas obligatoire. Les parcelles contenant les grandes cultures de printemps sarclées (pomme de terre, betterave, maïs) ainsi que les carottes, oignons, chicons, et autres cultures récoltées tardivement de 2018, n'ont donc pas été utilisées pour la détection. De plus, il y a très peu de chance de voir apparaître une culture de printemps après certaines cultures classées en «Prairie» («prairie temporaire», «couvert favorisant la faune» et «couvert à finalité environnementale»). Entre les parcelles de 2018 et de 2019, seule une parcelle de culture de printemps (maïs) a succédé à ces cultures. Elles n'ont également pas été utilisées pour la détection afin d'éviter d'éventuelles contaminations.

Choix des périodes de récoltes des cultures

Les récoltes de chaque culture ont, en premier lieu, été déterminées sur des périodes assez larges, par connaissance agronomique. Après un premier essai de prédictions, elles ont été réajustées. En effet, le fichier de sortie donne, pour chaque parcelle, la semaine dans laquelle la récolte y est détectée. Cela a permis de connaître de manière plus précise, les périodes de récoltes par culture et ainsi, d'améliorer les performances de détection des couverts. Pour rappel, le processeur cherche à détecter les couverts, si et seulement s'il a détecté préalablement une récolte sur la parcelle. Les périodes de récoltes des cultures, calibrées sur l'année 2018 sont disponibles en annexe (Annexe D, Tableau A-31).

Les périodes de récoltes ainsi choisies, sont encore assez larges, ce qui en théorie, leur permettraient d'être utilisables d'année en année. De plus, comme la détection se fait après la récolte, il serait toujours possible d'ajuster la période de détection chaque année afin de maximiser les performances du modèle de détection.

Choix de la période de détection des couverts

Le processeur demande une période dans laquelle il va chercher à détecter les couverts. Le choix de cet intervalle est important car il va influencer fortement la performance de la détection. Les principales erreurs attendues dans la détection des couverts étaient de détecter les intercultures courtes (ne précédant pas les cultures de printemps) et de confondre les cultures d'hiver avec des couverts. Plus la date de fin de la période sera reculée dans l'année, plus il y a de chance de confondre les cultures hivernales avec les couverts et d'augmenter la contamination. Plus elle se situera tôt, plus les indices de végétation des couverts à la fin de la période seront faibles, et plus le risque de ne pas détecter certains couverts est élevé. De plus, si la date de fin se situe proche du 1^{er} octobre, des intercultures courtes pas encore détruites pourraient être détectées. Dans ce contexte, le choix de mettre comme fin de période, une date proche du 15 novembre a été fait. Pour rappel, les CIPAN peuvent être détruites à partir de cette date (2.2). Elle permet de prendre en compte les couverts à leurs maximums de végétation sans risquer de se situer dans la période de destruction.

La date de début de la période est aussi importante. Si la récolte sur la parcelle est détectée plus tard que la date de début de la pratique, le processeur ne va pas chercher à détecter la pratique. Si la date de début de la pratique choisie est située avant le semis du couvert, des données de la culture précédente comme la perte de végétation due à la récolte vont être prises en compte et fausser les résultats.

Jouer sur la période de détection lors de la phase de calibration du modèle a permis de maximiser les détections des couverts. Une gamme de choix de dates a été testée (Tableau 9) afin de pouvoir comparer les effets des dates de début et de fin ainsi que de la durée de la période sur les prédictions.

Méthodologie

Tableau 9 : Périodes de détection des couverts hivernaux testés.

Début de la période	Fin de la période	Durée de la période
15/08	15/11	3 mois
01/09	15/11	2 mois et demi
10/09	10/11	2 mois
15/09	15/11	2 mois
20/09	30/10	2 mois et 10 jours
20/09	10/11	1 mois et 20 jours
20/09	15/11	1 mois et 25 jours
25/09	30/10	1 mois et 5 jours
25/09	10/11	1 mois et 15 jours
25/09	15/11	1 mois et 20 jours
30/09	30/10	1 mois
30/09	10/11	1 mois et 10 jours
30/09	15/11	1 mois et 15 jours
30/09	30/11	2 mois
05/10	10/11	1 mois et 5 jours
05/10	15/11	1 mois et 10 jours

Analyse des données sortant du processeur

Le processeur donnant quatre niveaux de confiance, il a été retenu que les parcelles sur lesquelles le processeur donne une confiance forte et modérée en la présence d'un couvert sont classées en tant que couverts hivernaux. Les parcelles caractérisées par les deux autres niveaux de confiance (faible et pauvre) sont considérées comme des parcelles sans couverts.

Génération des données de référence pour l'analyse des détections

N'ayant pas accès à un parcellaire référençant les couverts hivernaux effectivement présents sur le terrain, une astuce a dû être trouvée afin de pouvoir vérifier les résultats obtenus avec le modèle. En effectuant une intersection entre les parcellaires de l'année 2018 et de 2019, classés en 4 classes (4.3.1), il est possible d'isoler les parcelles ayant accueillies une culture d'hiver en 2018 et une culture de printemps en 2019. Sachant que l'implantation d'une CIPAN est obligatoire dans la zone d'étude entre ces deux cultures, il est possible de connaître, avec une certitude assez élevée, une bonne partie des couverts hivernaux. Ces parcelles peuvent permettre d'obtenir la sensibilité du modèle de détection des couverts hivernaux. Cette méthode de validation ne s'établit donc pas sur des parcelles vérifiées. Des erreurs peuvent exister si les agriculteurs ne respectent pas leurs déclarations de cultures principales, ou n'implantent pas de CIPAN malgré l'obligation. Dans la partie résultats, cet indice de performance est dénommé « sensibilité sur les CIPAN présumées ».

Une manière permettant de se rendre compte de la précision de la détection a aussi dû être trouvée. Il a déjà été dit que le plus gros risque de contamination provenait des cultures d'hiver semées tôt (environ dans la même période que les couverts hivernaux) et des

intercultures courtes. Ces dernières précédant nécessairement une culture d'hiver, il est possible de se rendre compte de la proportion de ces deux sources de contaminations en croisant les parcelles sur lesquelles un couvert a été détecté et les parcelles classées en « hiver » du parcellaire de 2019 classé en 4 classes (4.3.1). Cette erreur de contamination peut ensuite être utilisée comme une indication de la précision. Dans la partie « Résultats », cet indice de performance est dénommé « contamination par des parcelles en culture d'hiver ».

Enfin, dans ce modèle, les parcelles sur lesquelles les couverts sont détectés sont considérées comme des futures cultures de printemps. A contrario, les parcelles sur lesquelles le processeur ne détecte pas de couverts sont considérées comme des parcelles accueillant un autre type de culture. Il est donc possible de créer une matrice de confusion en croisant les résultats de la détection avec les futures cultures sur la parcelle classés en 4 classes (4.3.1). La précision et la sensibilité de la classification peuvent en être déduites. Elles apportent aussi une aide dans l'appréciation de la performance du modèle de prédictions. Dans la partie résultats, ces indices de performance sont dénommés « précision et sensibilité sur les cultures de printemps 2019 prédictibles par EO¹ ». Il est important de retenir ici que seuls les parcelles sur lesquelles l'algorithme a été tourné, et pour lesquelles l'algorithme a pu donner un avis de confiance (forte, modérée, faible et pauvre) sont reprises dans cette matrice de confusion. La précision et la sensibilité ainsi calculée ne prend donc pas en compte les parcelles sur lesquelles une culture de printemps était précédée d'une autre culture de printemps récoltée tardivement.

Comme pour les prédictions sur base de rotation, l'ensembles des indicateurs permettant l'évaluation de la performance du modèle (sensibilité de la prédiction des CIPAN présumées, erreur de contamination par les cultures d'hiver, et précision et sensibilité de la prédiction de cultures de printemps prédictibles par EO) ont été calculés sur base de l'aire et non du nombre de parcelles.

Dans un second temps, une fois la période maximisant la détection des couverts choisie, la précision et la sensibilité de la classe printemps de 2019 ont été recalculées, en tenant compte cette fois de toutes les parcelles de printemps de 2019 (incluant cette fois les parcelles non prédictibles). Une représentation cartographique des prédictions a enfin été faite à la manière de celle obtenue par les modèles de rotations.

4.5.3. Validation du modèle

Initialement, la calibration du modèle devait s'effectuer sur la détection des couverts installés entre les cultures principales de 2017 et de 2018. La validation aurait alors pu se faire sur les couverts installés en 2018 pour prédire les cultures de printemps de 2019. Cependant, un des deux satellites sentinel-2 a été lancé en 2017 et était en phase de calibration durant ce qui

¹ EO signifie *Earth Observation* ou observation de la terre

auraient été les périodes de récolte et de détection des couverts. La calibration du modèle de détection du processeur a donc été réalisée sur l'année 2018, ce qui a amené une difficulté pour la validation. En effet, à l'heure de la rédaction de ce mémoire, les cultures de printemps n'ont pas encore été semées, et le parcellaire 2020 n'est naturellement pas encore disponible. Les méthodes d'analyse de prédictions utilisées lors de la calibration (4.5.2) ne sont donc pas reproductibles dans ce cas.

Un moyen de valider le modèle de prédictions était donc d'utiliser directement les séries temporelles d'images satellites. L'évolution des métriques des couverts étant fort caractéristiques (Figure 12), il a été possible d'identifier des parcelles comme des couverts hivernaux (photo-interprétation). 50 parcelles ont ainsi été retenues comme données de référence répartie dans la zone d'étude pour la validation.

Pour ce faire, le site internet « *ASAP - Anomaly Hotspots of Agricultural Production* » de la commission européenne a été utilisé [ASAP]. Il permet d'obtenir des séries temporelles d'images et de métriques issues du programme COPENICUS et de ses précurseurs encore en service. La Figure 12 suivante, montre l'évolution du NDVI caractéristique des couverts hivernaux, ainsi que les images en fausses couleurs obtenues avant la croissance du couvert, pendant, et après la destruction du couvert.

Malheureusement, les couverts n'étant pas encore tous détruits au moment de la rédaction du mémoire, il est difficile de distinguer formellement un couvert d'une culture d'hiver par interprétation d'images satellites. Les erreurs de contaminations par les cultures d'hiver dans les prédictions n'ont donc pas pu être évaluées.

Les périodes de récolte et de détection des couverts calibrées sur 2018 ont ensuite été utilisées pour prédire les couverts de 2019 grâce au L4C. L'appréciation de la performance de détection a ensuite été obtenue par la confrontation des résultats aux 50 parcelles de couverts identifiées par photo-interprétation.

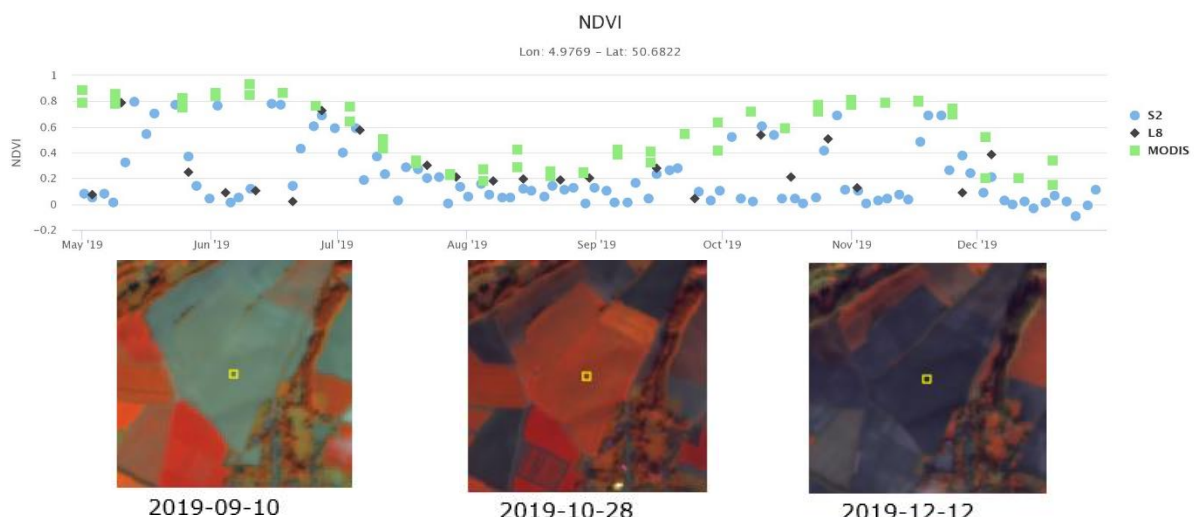


Figure 12 : Exemple du comportement du NDVI, et évolution des images en fausses couleurs, d'une parcelle ayant accueillie un couvert en 2019. (S2 = Sentinel 2 et L8 = Landsat 8) [ASAP].

4.5.4. Amélioration du modèle

Une partie des cultures de printemps n'est pas précédée par des couverts hivernaux. En effet, lorsqu'elles sont précédées d'une culture de printemps se récoltant tardivement (après le 1^{er} septembre), l'agriculteur n'est pas obligé de semer un couvert. Ces parcelles ne seront donc pas prédites en culture de printemps, ce qui rend impossible la prédiction totale des cultures de printemps. Ces cultures pourraient néanmoins être prédites par les probabilités issues des successions de cultures.

Probabilité simple de succession de cultures

Dans les successions de cultures principales retrouvées en Wallonie (2.3), les seules rotations entre cultures de printemps recensées sont les monocultures de maïs. Dans les occurrences des successions calculées grâce aux tables de calibration (Annexe C, Tableau A-28), c'est également la seule succession entre deux cultures de printemps dont l'occurrence des observations dépasse 1% (4.04%).

Afin d'obtenir une connaissance de la part que représente ces monocultures de maïs dans l'ensemble des parcelles de maïs, des calculs de probabilités ont été réalisés grâce aux intersections créées lors de la calibration des modèles de prédictions obtenus par rotation (4.4.3). S'il est possible d'isoler les parcelles sur lesquelles s'effectuent les monocultures de maïs, il serait possible de connaître une partie des cultures de printemps qui seraient oubliées dans le modèle de prédictions par la détection des couverts hivernaux.

En plus de se retrouver sous forme de monoculture, il s'observe dans les intersections que le maïs précède de manière récurrente d'autres cultures de printemps comme la pomme de terre ou la betterave. Un calcul de la probabilité de retrouver une culture de printemps (maïs, pomme de terre, etc.) après un maïs a également été réalisé. Si cette probabilité est importante, il pourrait être envisagé d'utiliser les cultures de maïs comme prédicteur de cultures de printemps pour les parcelles n'ayant pas accueillies de couverts.

Association avec le modèle de prédictions par rotation

Une autre manière d'améliorer la prédiction par la détection des couverts hivernaux était de le combiner avec le modèle de prédictions par les rotations. Comme la calibration de la méthode de prédictions s'est effectuée sur l'année de 2018 pour prédire l'année 2019, le modèle de rotation utilisé pour la combinaison a été celui obtenu dans la validation (4.4.5). Deux méthodes ont été implémentées pour associer les deux modèles.

Le **modèle combiné 1** (Figure 13) reprend les parcelles prédites en cultures de printemps par la détection des couverts hivernaux. Il les combine avec les résultats du modèle de prédictions par rotations sur toutes les autres parcelles afin de prédire le reste des cultures de printemps.

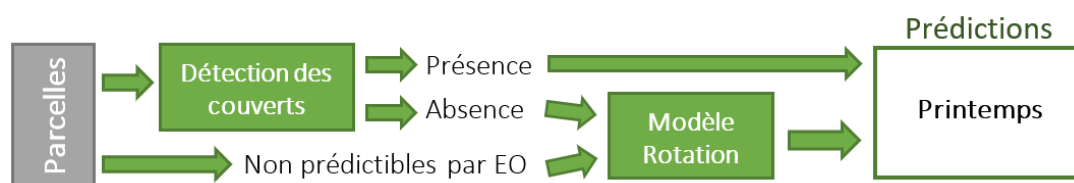


Figure 13 : Illustration de la prédiction des cultures par le modèle combiné 1.

Méthodologie

Afin de prédire les cultures de printemps, le **modèle combiné 2** (Figure 14) reprend les parcelles prédites en cultures de printemps par la détection des couverts hivernaux. Il les combine ensuite seulement avec les résultats du modèle de prédictions par rotations des parcelles récoltées après le 1^{er} septembre (non prédictibles par EO) et non plus ceux des parcelles sur lesquelles une couverture n'a pas été détectée. A noter que certaines parcelles devant être prédites par la méthode de détection de couverts ne pourront pas être détectées pour des raisons expliquées dans la partie résultat (5.2.1). Ces parcelles, peu nombreuses, ont été prédites par la méthode de rotation également.

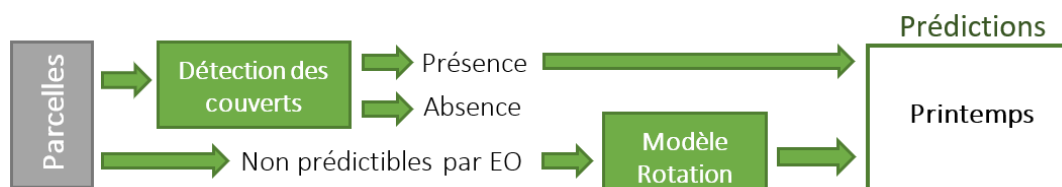


Figure 14 : Illustration de la prédiction des cultures par le modèle combiné 2.

Ces deux méthodes vont avoir un impact différent sur les prédictions. Le Tableau 10 montre que les prédictions du modèle combiné 1 souffriront plus d'omission alors que celles du modèle combiné 2 souffriront, elles, de plus de contamination. Ces différences de prédictions vont jouer sur la précision et la sensibilité du modèle de prédictions. Dans les autres cas de figures que ceux montrés sur le Tableau 10, les deux modèles vont prédire les mêmes résultats.

Tableau 10 : Exemple des différents cas de figures qui apportent une différence entre les deux méthodes de prédictions combinant les prédictions par rotations et détection de couverts.

	Culture 2019	Détection CIPAN	Rotation	Combiné 1	Combiné 2
Parcelle X ₁	Printemps	Non-printemps	Printemps	Printemps	Non-printemps
Parcelle X ₂	Hiver	Non-printemps	Printemps	Printemps	Non-printemps

5. Résultats

5.1. Prédiction par la méthode des rotations

5.1.1. Comparaison des parcellaires

Cette section présente le résultat des comparaisons des parcellaires, utile à la justification de l'utilisation de la méthode des centroïdes (4.4.1) et à la connaissance des différences entre les parcellaires de 2018 et de 2019 (4.4.5).

Test des centroïdes

Ce test avait pour objectif de caractériser les (di-)similarités de la taille et de la forme des parcelles au gré des années, et donc l'occurrence des cas 1, 2 et 3 (4.4.1). Il permet de pouvoir décider si l'utilisation de la méthode des centroïdes était appropriée. Le tableau suivant montre le pourcentage de parcelles dont l'aire de l'intersection vaut au moins 80% de leurs aires chaque année.

Tableau 11 : Pourcentage des cultures (en aire et en nombre de parcelles) dont l'aire intégrée (issue de l'intersection entre les parcellaires de différentes années) est similaire à plus de 80% avec les aires des parcelles entières de chacune des années.

	2015		2016		2017		2018	
	Aire [%]	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]	Aire [%]	Nbr parcelles [%]
2 années	83,77	77,72	94,53	95,77	87,34	87,12	89,44	87,42
3 années	\	\	77,85	71,32	81,24	79,49	81,55	79,11
4 années	\	\	\	\	73,32	66,31	76,82	73,93
5 années	\	\	\	\	\	\	69,99	65,63

On observe, comme on pouvait s'y attendre, que le nombre de parcelles modifiées augmente avec le nombre d'années pris en compte. Dans le pire des cas, l'intersection des parcelles sur cinq années montre qu'environ 65 % d'entre elles ont été peu modifiées lors de ces années. L'utilisation du centroïde peut donc être justifiée car une majorité assez satisfaisante de parcelles restent stables. Une partie significative des données va néanmoins être perdue, et il faut rester conscient que 35% des parcelles ont subi des modifications de type « cas 1, 2 et 3 ». De plus il faut retenir que dans ce test, les parcelles disparues à la suite des problèmes de type « cas 4 et 5 » ne sont pas employées dans les calculs car elles ne sont pas reprises dans les intersections, ce qui rajoute encore des pertes de données.

Différence entre les parcellaires de 2018 et 2019

Il était intéressant de connaître les erreurs apportées par l'utilisation du parcellaire de 2018 pour projeter les prédictions de 2019 (4.4.5). Une comparaison des parcellaires de 2018 et de 2019 a donc été effectuée.

Afin d'observer les différences entre les deux parcellaires, la fonction « différenciation symétrique » des polygones a été utilisée sur QGIS. Cette fonction permet la création d'une couche représentant tous les polygones présents dans un des parcellaires, mais pas dans les deux.

La comparaison des deux parcellaires (issus de la classification des cultures) a montré que 316,85 hectares (soit 2,72% de l'aire totale des parcelles), répartis dans 2335 parcelles, sont soit présents dans le parcellaire 2018 mais pas 2019, soit le contraire. Ces terres sont en très grande majorité des morceaux de parcelles ayant peu évolués entre les deux années. Elles n'apportent pas de problèmes majeurs dans l'utilisation du parcellaire de l'année 2018 pour prédire l'année 2019, car elles existent sur les deux parcellaires. Pour se rendre compte des changements plus importants entre années, comme la disparition et l'apparition de parcelles entières (ou des variations plus importantes de tailles de parcelles), les morceaux de parcelles nouvellement créés (pour la comparaison des parcellaires avec l'outil « différenciation symétrique ») plus petits que 1000 m² ont été supprimés (Figure 14). Cette valeur seuil permet

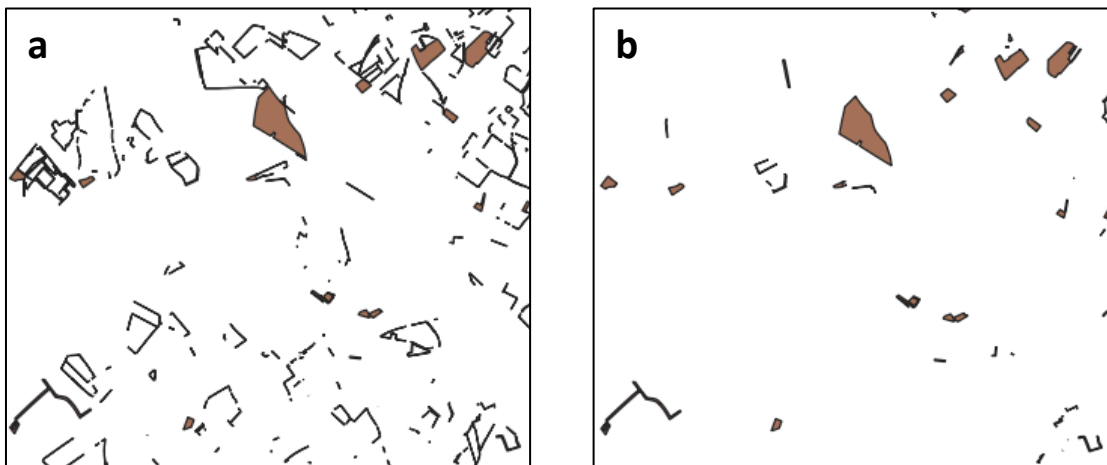


Figure 15 : a) Exemple de la différenciation symétrique entre les parcellaires 2018 et 2019 sur une partie de la zone d'étude. b) Idem sans les parcelles de moins de 1000 m².

d'éliminer les petites variations de parcelles, sans risquer de supprimer les parcelles entières qui auraient pu apparaître ou disparaître entre les deux années. La superficie de ces parcelles à problème ne diminue pas énormément, 303,31 ha, mais le nombre de parcelles concernées descend lui à 309.

La plupart de ces parcelles (38 disparues entre 2018 et 2019 et 175 apparues en 2019) font partie du type prairie. Ce sont sans doute pour beaucoup, des prairies temporaires en 2018 passées à permanentes en 2019, ou des prairies permanentes en 2018 repassées en classe prairie en 2018 en étant identifiées comme « Bande/parcelle aménagée MAE - MC7, MC8,

MB5 » (127 des 175 prairies apparues en 2019 sont classées de cette manière) dans le cadre des mesures agro-environnementales de la Région wallonne.

Pour les autres classes, il reste donc 89 polygones (correspondant à des variations, apparitions, ou disparitions de parcelles). Ce qui représente peu de parcelles par rapport au nombre total de parcelles de 2018 (3037, soit 2.93%) et 2019 (3208, soit 2.77%). L'utilisation des limites de parcelles du parcellaire de 2018 pour prédire celles de 2019 peut donc être considérée comme raisonnable.

5.1.2. Calibration

Cette section a pour but de déterminer dans un premier temps les variables à prendre en compte dans les modèles de prédictions par rotation (4.4.4.3). Et ensuite de définir les modèles donnant les meilleures performances entre les trois modèles (4 classes, 10 classes direct et indirect (4.4.3)) afin de prédire les 4 classes et les 10 sous-classes.

Tout d'abord, les résultats concernant le modèle de 4 classes vont être présentés. Ensuite, ceux des modèles de 10 classes, d'abord ceux de la méthode indirecte, puis ceux de la méthode directe sont exposés et discutés.

5.1.2.1 Modèle 4 classes

Nombre d'années de rotations

Le premier test était destiné à évaluer les exactitudes moyennes en fonction du nombre d'années de rotation en utilisant seulement les cultures antécédentes comme variables du modèle. On observe de meilleures sensibilités pour les durées de rotation de trois et quatre ans. Dans chaque cas, la précision moyenne est plus élevée que la sensibilité moyenne, ce qui signifie que le modèle de prédictions n'a pas tendance à classer de manière superflue des parcelles en « Printemps » pour augmenter l'exactitude du modèle.



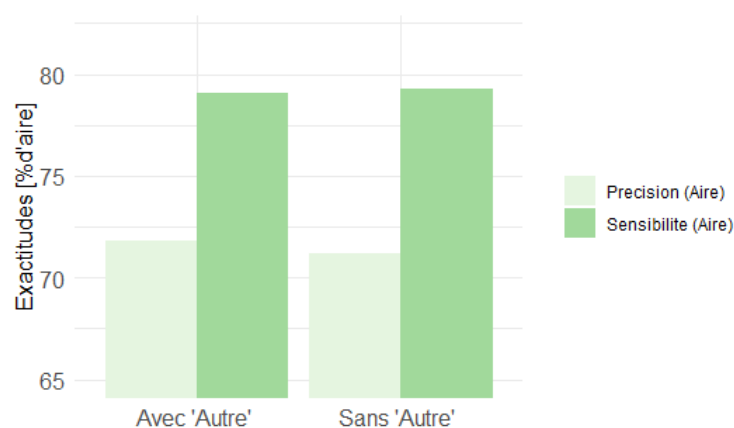
Graph 2 : Evolution de la précision et de la sensibilité de la classe printemps du modèle 4 classes utilisé en prenant compte différentes durées de rotation, et écart-type autour de l'estimation de l'exactitude moyenne.

Résultats

On peut remarquer (Graphe 2) que le modèle utilisant trois années est moins stable que les autres. Néanmoins certains tests pourraient diminuer cette instabilité et il est toujours plus intéressant de garder, pour une même sensibilité, le moins d'années de rotation dans le modèle (cela permet de prédire plus de parcelles). C'est pourquoi la série de tests ci-dessous a été effectuée en parallèle sur les modèles utilisant trois et quatre années.

Rotation sur trois années

Dans un premier temps le test supprimant la classe « Autre » a été effectué pour pouvoir connaître l'influence de ces parcelles sur les prédictions. Comme on peut le voir sur le Graphe 3, les différences de précisions et de sensibilités entre les deux modèles sont minimales. On peut donc estimer que les erreurs amenées par le problème de la classe « autre » ne sont pas assez significatives que pour s'en préoccuper.

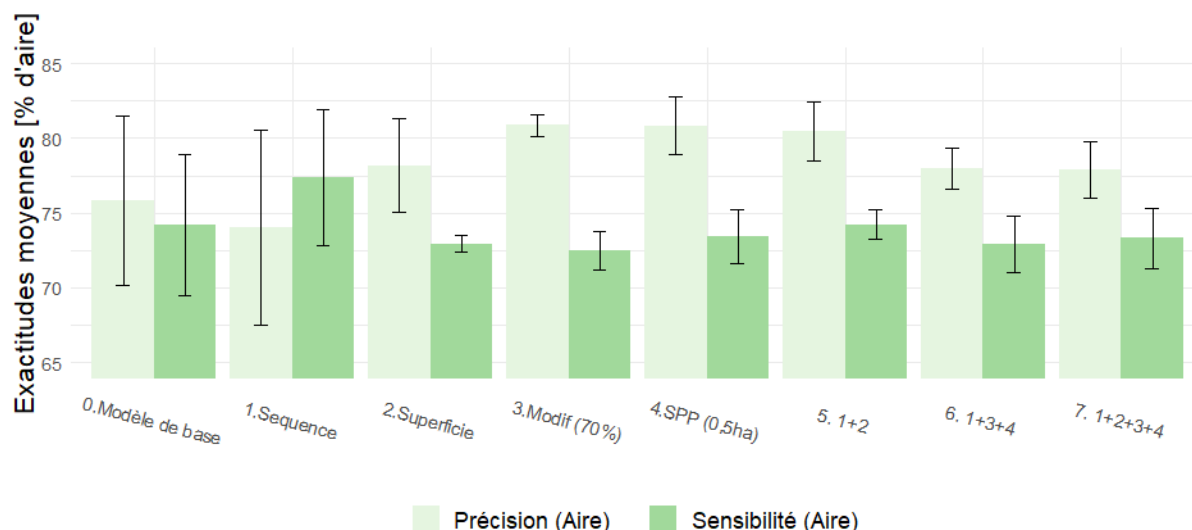


Graphe 3 : Comparaison de la précision de la classe « Printemps » des modèles 4 classes basé sur trois années de rotations, prenant, ou ne prenant pas en compte les parcelles de type « Autre ».

Le Graphe 4 suivant, montre les résultats obtenus avec différentes variantes du modèle. Ceux-ci sont respectivement : 0) le modèle de base, 1) le modèle prenant en compte les séquences, 2) le modèle prenant en compte les superficies des parcelles, 3) le modèle ne prenant pas en compte dans les parcelles d'entraînement, celles ayant été modifiées à plus de 70 %¹, 4) le modèle ne prenant pas en compte dans les parcelles d'entraînement, celles plus petites que 0,5 ha¹, 5) le modèle prenant en compte comme variable la séquence et la superficie, 6) le modèle utilisant la séquence et ne prenant pas en compte les parcelles modifiées ainsi que les plus petites dans les parcelles d'entraînement, et enfin, 7) similaire au modèle 6 en considérant cette fois la surface.

¹ Seuil permettant la meilleure précision (Annexe C, Graphe A-14 et Graphe A-17)

Résultats



Graph 4 : Comparaison des précisions et sensibilités moyennes de la classe « Printemps » par suite des différents tests effectués sur le modèle 4 classes basé sur des rotations de trois ans, et écart-type autour de l'estimation de l'exactitude moyenne. (Modif = test d'évolution des parcelles, SPP= test des parcelles de petites tailles).

On remarque qu'à l'exception du test des séquences, l'ajout de variables, où la suppression des parcelles pouvant contribuer aux erreurs permettent d'obtenir des résultats plus stables. Seule la prise en compte des séquences permet d'augmenter significativement la sensibilité, mais, en plus d'être peu stable, la précision de la classe est alors diminuée.

Afin d'obtenir un modèle permettant d'augmenter la sensibilité sans altérer la précision (défaut du test 1), les tests 5, 6 et 7 ont été essayés en utilisant différentes variables, ou parcelles d'intérêts ensemble. Le modèle 5, utilisant la séquence et la superficie comme paramètres du modèle, permet une augmentation de la précision avec une bonne sensibilité. De plus, il est plus stable que le modèle de base. Ce modèle est donc gardé.

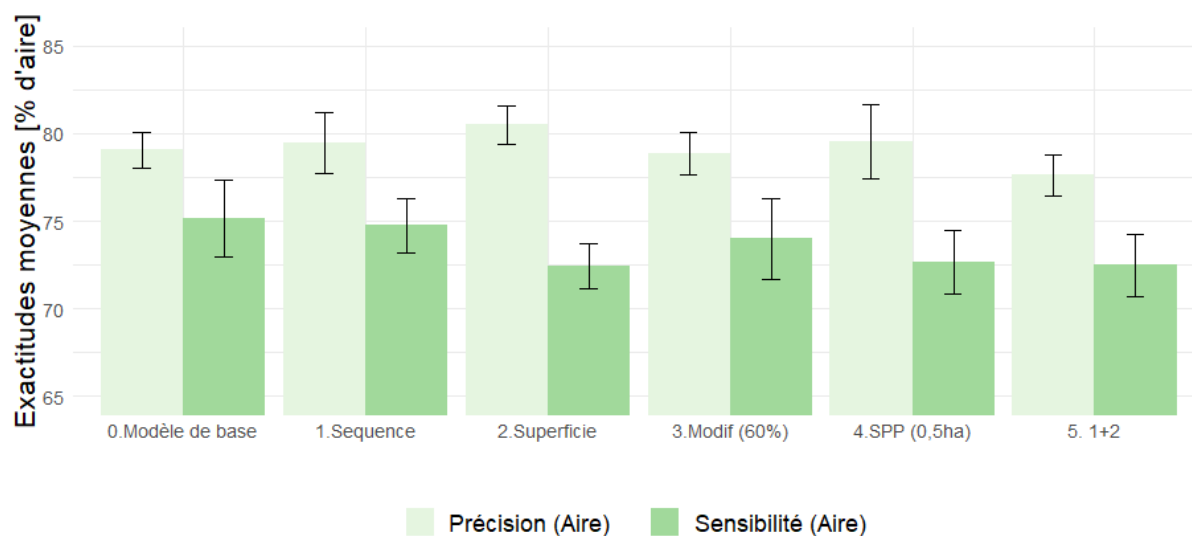
Rotation sur quatre années

Les tests de variables supplémentaires 1, 2, 3, 4, 5 et 6 des rotations sur trois années ont également été effectués sur le modèle prenant en compte quatre années de rotation. Le test 3 ne prend, cette fois-ci, pas en compte les parcelles ayant été modifiées de plus de 60 %¹ lors de la calibration.

Dans ce cas-ci (Graph 5), aucun test ne permet d'améliorer de manière significative la sensibilité. Le modèle retenu sera donc le modèle initial ne tenant compte que des cultures antécédentes comme paramètre.

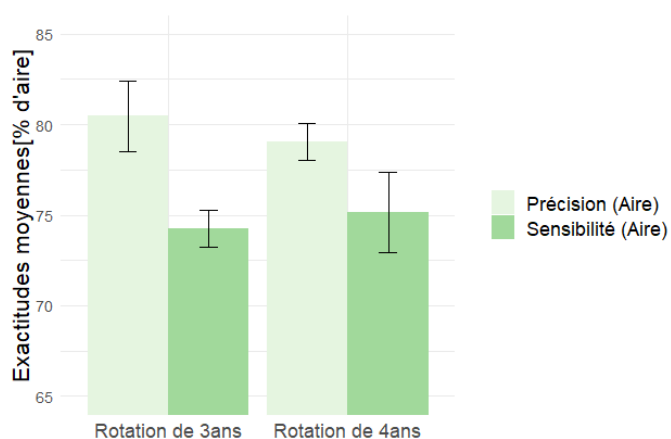
¹ Seuil permettant la meilleure précision (Annexe C, Graph A-15 et Graph A-18)

Résultats



Graphe 5 : Comparaison des précisions et sensibilités de la classe « Printemps » par suite des différents tests effectués sur le modèle 4 classes basé sur des rotations de quatre ans, et écart-type autour de l'estimation de l'exactitude moyenne. (Modif = test d'évolution des parcelles, SPP= test des parcelles de petites tailles).

Comparaison des deux modèles de 4 classes retenus



Il s'observe (Graphe 6) que le modèle sur quatre ans donne une meilleure sensibilité, mais une moins bonne précision. La différence est néanmoins limitée. On va donc préférer utiliser le modèle de rotations sur trois ans car il permet de prédire plus de parcelles.

Graphe 6 : Comparaison des précisions et sensibilités de la classe « Printemps » des deux modèles 4 classes retenus et écart-type autour de l'estimation de l'exactitude moyenne.

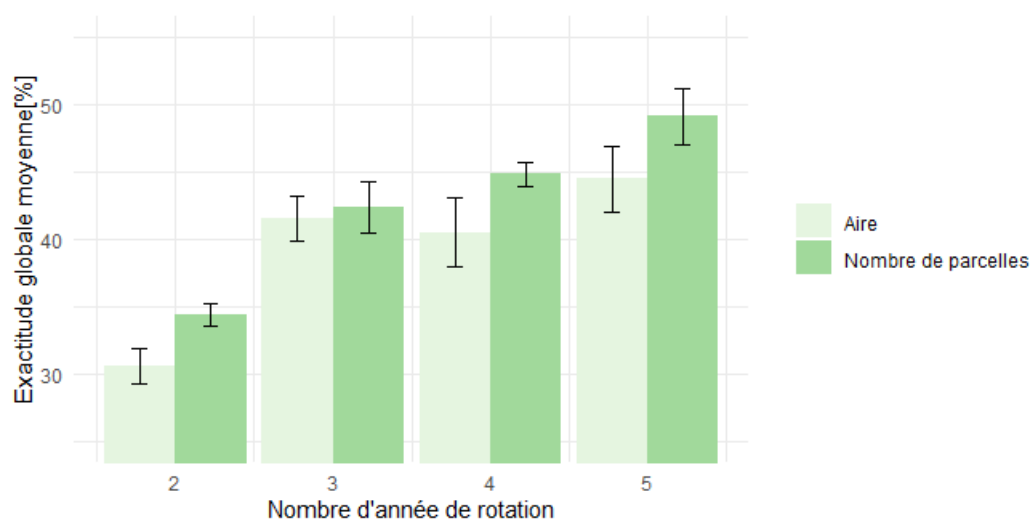
5.1.2.2 Modèle 10 classes

Pour rappel, les modèles 10 classes sont les modèles de prédictions utilisant les tables de calibration créées sur base des classifications en 10 sous-classes. Deux méthodes (directe et indirecte) d'utilisation sont possibles pour prédire les cultures de printemps (4.4.3). La méthode directe, qui se base sur les résultats du modèle 4 classes afin de prédire plus finement les parcelles prédites en printemps, a été utilisée pour effectuer les tests des variables complémentaires (4.4.4.3).

Méthode indirecte

Années de rotations

Sur le Graphe 7, il se voit clairement que dans le cas des modèles de 10 classes indirects, la qualité de la prédiction croît avec le nombre d'années considérées pour définir la rotation, que ce soit en nombre de parcelles ou en aire. Les tests vont donc être appliqués sur le modèle de cinq années de rotations.



Graphe 7 : Evolution de l'exactitude globale moyenne des modèles 10 classes indirect de base, prenant en compte un nombre différent d'années de rotations et écart-type autour de l'estimation de l'exactitude moyenne.

Test ROSE

Pour rappel, l'outil ROSE a permis de suréchantillonner les classes faibles afin d'améliorer leurs prédictions. Les matrices de confusion des Tableaux 12 montrent que le fait de renforcer la proportion des classes faiblement représentées dans les parcelles diminue l'exactitude globale, n'augmente pas significativement la sensibilité des parcelles faiblement représentées et ne permet pas de diminuer suffisamment les erreurs de contaminations comme voulu.

Résultats

Tableaux 12 : Matrices de confusions issues d'une prédiction sur les parcelles tests de calibration avec (gauche) et sans (droite) l'augmentation de la proportion des classes faiblement représentées dans les parcelles d'entraînement.

Avec ROSE										Sans ROSE													
		Référence											Référence										
		Mais	Bet.	PDT	Lin	Pois	Au.P.	Cér.	Total	Conta.			Référence							Total	Conta.		
Prédiction	Mais	74	14	12	0	4	5	1	110	36	Prédiction	Mais	79	15	13	1	6	4	2	120	41		
	Bet.	19	62	13	7	15	16	3	135	73		Bet.	16	69	9	5	16	16	2	133	64		
	PDT	7	13	36	9	7	8	2	82	46		PDT	10	21	48	15	10	14	2	120	72		
	Lin	1	5	4	7	5	3	0	25	18		Lin	0	1	3	5	3	3	0	15	10		
	Pois	3	6	6	2	6	6	0	29	23		Pois	1	1	2	0	2	3	0	9	7		
	Au.P	1	0	4	0	0	2	0	7	5		Au.P	1	2	5	1	1	3	1	14	11		
	Cér	2	9	5	2	1	3	1	23	22		Cér	0	0	0	0	0	0	0	0	0		
	Total	107	109	80	27	38	43	7	411		Total	107	109	80	27	38	43	7	411				
	Omis.	33	47	44	20	32	41	6		188		Omis.	28	40	32	22	36	40	7		206		

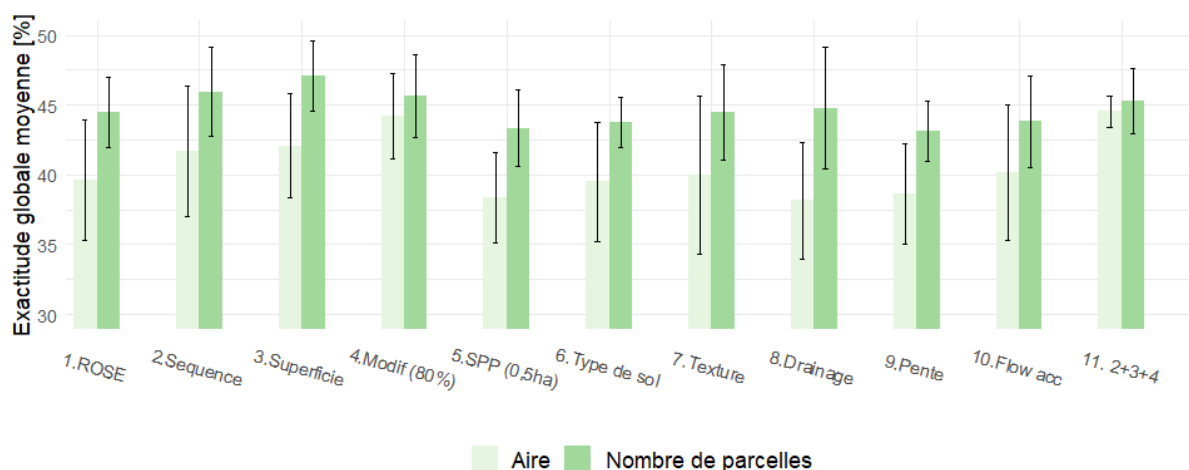
Néanmoins les tests voués à améliorer les prédictions ont tous été répétés sur des données dans lesquelles les classes faibles ont été suréchantillonnées avec ROSE, dans l'espoir que de nouveaux paramètres puissent mieux ajuster la classification de ces parcelles.

Série de tests

Le Graphe 8 suivant présente les exactitudes moyennes obtenues par les prédictions des modèles utilisant pour chacun d'eux ROSE afin d'augmenter artificiellement le nombre de parcelles des classes les plus faibles, ainsi que les différents tests suivants : 1) ROSE seul, 2) Les séquences, 3) la superficie de la parcelle, 4) la suppression des parcelles d'entraînement ayant été modifiées à plus de 80 %¹, 5) la suppression des parcelles d'entraînement plus petites que 0,5 ha¹, 6) l'utilisation du type de sol du CNSW comme paramètre, 7) l'emploi de la texture du CNSW comme paramètre, 8) le recours au drainage du CNSW comme paramètre, 9) l'utilisation de la pente moyenne comme paramètre, 10) l'emploi de l'accumulation de flux moyen comme paramètre, et enfin 11) un test utilisant les séquences et la superficie comme paramètres et en supprimant les parcelles d'entraînement modifiées.

¹ Seuil permettant la meilleure précision (Annexe C, Graphe A-16 et Graphe A-19)

Résultats



Graph 8 : Exactitudes globales moyennes en nombre de parcelles et en aires obtenus suite aux différents tests effectués sur le modèle de 10 classes indirect basé sur cinq années de rotation dont les classes faibles ont été suréchantillonnées, et écart-type autour de l'estimation de l'exactitude moyenne. (Modif = test d'évolution des parcelles, SPP= test des parcelles de petites tailles, flow acc=test de l'accumulation de flux).

Les tests ayant augmentés significativement les exactitudes, en nombres de parcelles et en aires, sont les tests de la séquence, de la superficie et de la modification de parcelles. Les tests reprenant les types de sol du CNSW ou obtenus avec le relief ne sont eux pas concluants, aucun d'eux n'apportant une amélioration significative de l'exactitude.

Comme lors des tests du modèle de type 4, le test de la séquence donne une exactitude globale variant fortement en fonction des parcelles utilisées pour l'entraînement du modèle. En faisant tourner un modèle regroupant tous les tests permettant d'augmenter l'exactitude (le test 11), les résultats obtenus sont meilleurs au niveau des exactitudes et assez stables en fonction des parcelles prises en compte dans la rotation. Ce modèle est dès lors retenu.

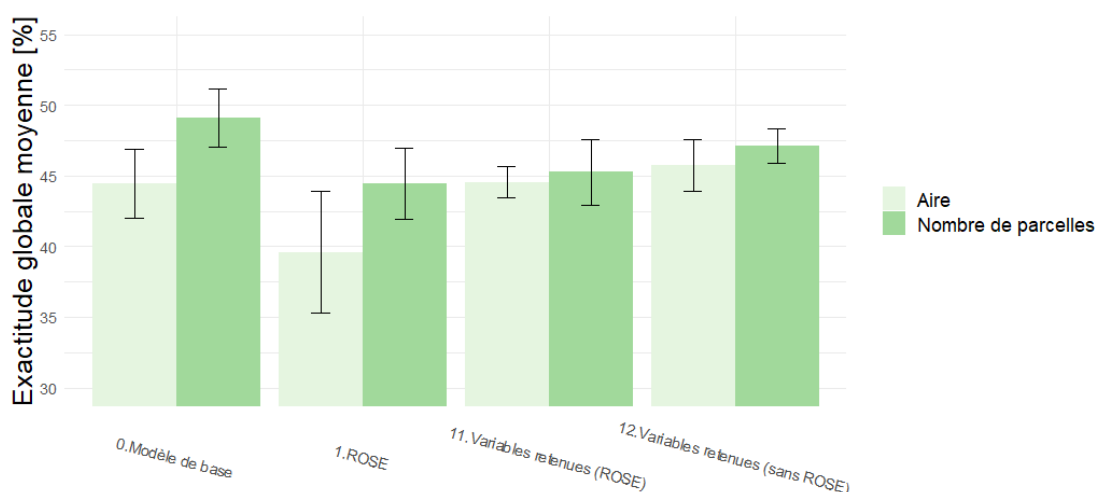
Ensuite, le modèle de 10 classes avec les variables retenues a été testé sur les parcelles d'entraînement sans la sur-représentation des classes faibles. Au regard des matrices de confusion (

Tableaux 13), comme pour les modèles ne prenant pas en compte les variables retenues, l'enrichissement des classes les moins présentes n'a pas permis d'améliorer les prédictions des classes faibles. Le modèle n'utilisant pas ROSE va donc être préféré, d'autant plus qu'il permettrait d'obtenir une exactitude très légèrement supérieure (non significative) au vu du Graph 9.

Résultats

Tableaux 13 : Matrices de confusion issues d'une prédiction sur les parcelles « tests » de calibration en utilisant les paramètres retenus, avec (dessus) et sans (dessous) l'augmentation de la proportion de classes faibles dans les parcelles d'entraînement.

Avec ROSE										Sans ROSE															
		Référence							Total		Conta.				Référence							Total		Conta.	
		Mais	Bet.	PDT	Lin	Pois	Au.P.	Cér.							Mais	Bet.	PDT	Lin	Pois	Au.P.	Cér.				
Prédiction	Mais	67	25	15	5	4	9	2	127	60	Prédiction	Mais	65	22	16	5	4	8	3	123	58				
	Bet.	8	68	17	17	14	13	4	141	73		Bet.	10	71	18	17	13	15	5	149	78				
	PDT	25	19	51	16	16	22	1	150	99		PDT	29	23	53	16	17	25	2	165	112				
	Lin	0	1	1	3	1	2	0	8	5		Lin	0	0	0	3	1	0	0	4	1				
	Pois	0	2	1	1	6	3	0	13	7		Pois	0	2	0	1	6	3	0	12	6				
	Au.P	0	0	0	0	0	0	0	0	0		Au.P	0	0	2	0	0	0	0	2	2				
	Cér	4	3	4	0	0	2	3	16	13		Cér	0	0	0	0	0	0	0	0	0				
	Total	104	118	89	42	41	51	10	455		Total	104	118	89	42	41	51	10	455						
	Omis.	37	50	38	39	35	51	7		198		Omis.	39	47	36	39	35	51	10		198				



Graph 9 : Effet de ROSE sur les exactitudes en aire et en nombre de parcelles des modèles sans et avec les variables retenues du modèle 10 classes indirect basé sur cinq années de rotations, et écart-type autour de l'estimation de l'exactitude moyenne.

Méthode directe

Les variables d'intérêts choisies pour le modèle 10 classes indirect sont ensuite utilisées dans le modèle 10 classes direct, qui de plus, était entraîné sur des données dont la classe « hiver » était sous-échantillonnée (4.4.4.4).

La matrice de confusion (Tableau 14) obtenue montre que seules les cultures des classes betterave, pomme de terre, prairie et hiver vont avoir une sensibilité¹ supérieure à 50% et que les précisions² des sous-classes de printemps vont être très peu élevées car aucune d'entre elles ne dépassent 50%.

¹ Sensibilité = 1- erreur d'omission

² Précision = 1- erreur de contamination (4.4.4.2)

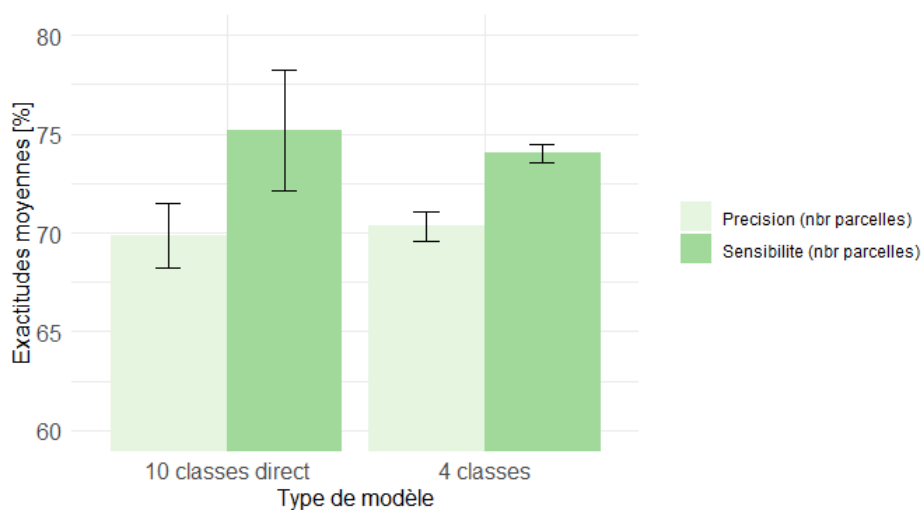
Résultats

Tableau 14 : Matrices de confusion obtenues en utilisant le modèle de 10 classes en méthode direct, en diminuant dans les parcelles d'entraînement la proportion de cultures d'hiver à 25%.

		Référence										Total	Er. cont.
		Mais	Bet.	PDT	Lin	Pois	Au.P.	Cér.	Prair	Hiver	Autre		
Prédiction	Mais	52	12	10	3	2	6	0	8	40	1	134	0,61
	Bet.	7	52	4	17	8	8	2	2	25	0	125	0,58
	PDT	15	20	46	9	8	17	3	2	51	4	175	0,74
	Lin	0	0	0	1	1	0	0	0	0	0	2	0,50
	Pois	0	0	1	0	1	1	0	0	1	0	4	0,75
	Au.P.	0	0	0	0	0	0	0	0	2	0	2	1
	Cér.	0	0	0	0	0	0	0	0	0	0	0	/
	Prair.	1	0	1	0	0	0	1	30	3	0	36	0,17
	Hiver	21	27	24	2	9	11	2	4	247	0	347	0,29
	Autre	0	0	0	0	0	0	0	0	0	0	0	/
Total		96	111	86	32	29	43	8	46	369	5	825	
Er.d'om		0,46	0,53	0,47	0,97	0,97	1	1	0,35	0,33	1		429

En regroupant les cultures de printemps ensemble, la précision et la sensibilité de la classe de printemps semblent similaires à celles obtenues avec le modèle de 4 classes retenu. En observant le Graphe 10, il semblerait même que la sensibilité moyenne de la classe « printemps » obtenue avec le modèle 10 classes soit légèrement supérieure. Une analyse statistique (test de Fisher) a été effectuée afin de savoir si cette amélioration était significative. Il en est sorti que les moyennes des deux indices de qualité (précision et sensibilité) ne sont pas significativement différentes.

Les matrices de confusion (Tableaux 15), montrent que les prédictions des trois autres classes (« Prairie », « Hiver » et « Autre ») sont nettement moins bonnes avec le modèle 10 classes direct. La classe « Autre » n'y est même plus prédites.



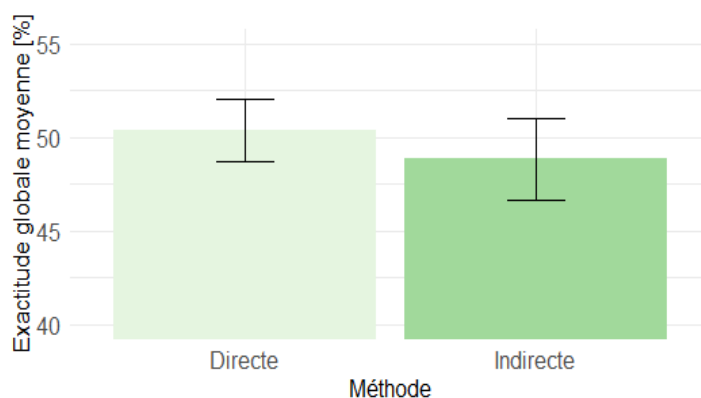
Graphe 10 : Comparaison des prédictions (précision et sensibilité calculées sur base du nombre de parcelles) des sous-classes regroupées ou de la classe de printemps obtenues avec le modèle 4 classes retenu (droite) et le modèle 10 classes utilisé par la méthode directe (gauche) et écart-type autour de l'estimation de l'exactitude moyenne.

Résultats

Tableaux 15 : Comparaison des matrices de confusion obtenues avec le modèle de 4 classes (droite) et le modèle de 10 classes utilisé par la méthode directe (gauche).

10 classes, méthode directe								4 classes							
		Référence								Référence					
		Print	Prair	Hiver	Autre	Total	Er.C.			Print	Prair	Hiver	Autre	Total	Er.C.
Prédiction	Print	306	12	119	5	442	0,31	Prédiction	Print	989	44	333	30	1396	0,29
	Prair	3	30	3	0	36	0,17		Prair	23	185	7	16	231	0,2
	Hiver	96	4	247	0	347	0,29		Hiver	326	17	725	15	1083	0,33
	Autre	0	0	0	0	0	/		Autre	9	10	8	120	147	0,18
	Total	405	46	369	5	825			Total	1347	256	1073	181	2857	
	Er.O	0,24	0,35	0,33	1		583		Er. O	0,27	0,28	0,32	0,34		2019

Afin de comparer les prédictions des sept sous-classes de printemps des deux modèles 10 classes, la partie des matrices de confusions contenant les cultures de printemps de la méthode directe a été isolée. L'exactitude globale moyenne de ces sept sous-classes a ensuite pu être calculée. Il se voit sur le Graphe 11 qu'elle est légèrement supérieure que celle obtenue par la méthode indirecte.



Graphe 11 : Comparaison des exactitudes sur les classes de printemps obtenues avec les modèles de 10 classes direct (gauche) et indirect (droite).

Les matrices de confusion (Tableaux 16) montrent peu de différences entre les prédictions des deux modèles. Le maïs aurait une meilleure sensibilité et une meilleure précision avec la méthode directe contrairement à la betterave. La pomme de terre est, quant à elle, caractérisée d'une meilleure sensibilité mais d'une moins bonne précision.

Résultats

Tableaux 16 : Comparaison des matrices de confusion des cultures de printemps obtenues avec les modèles 10 classes indirect (droite) 10 classes direct (gauche).

10 classes, méthode directe										10 classes, méthode indirecte											
		Référence							Total	Er.C.			Référence							Total	Er.C.
		Mais	Bet.	PDT	Lin	Pois	Au.P.	Cér.					Mais	Bet.	PDT	Lin	Pois	Au.P.	Cér.		
Prédiction	Mais	52	12	10	3	2	6	0	85	0,39	Prédiction	Mais	71	17	16	4	5	5	3	121	0,41
	Bet.	7	52	4	17	8	8	2	98	0,47		Bet.	9	75	12	19	8	16	1	140	0,46
	PDT	15	20	46	9	8	17	3	118	0,61		PDT	16	18	58	8	13	19	4	136	0,57
	Lin	0	0	0	1	1	0	0	2	0,50		Lin	0	0	0	1	1	0	0	2	0,50
	Pois	0	0	1	0	1	1	0	3	0,67		Pois	0	1	0	0	2	3	0	6	1
	Au.P	0	0	0	0	0	0	0	0	/		Au.P	0	0	0	0	0	0	0	0	/
	Cér	0	0	0	0	0	0	0	0	/		Cér	0	0	0	0	0	0	0	0	/
	Total	74	84	61	30	20	32	5	306		Total	104	118	89	42	41	51	10	455		
	Er.O.	0,30	0,38	0,25	0,97	0,95	1	1		152		Er.O.	0,32	0,36	0,35	0,98	0,95	1	1		207

Discussion

Concernant la prédiction des quatre classes, le modèle 10 classes direct (regroupé en 4 classes) a montré une sensibilité faiblement supérieure, toutefois non significative, au modèle de 4 classes. Ils ont donc, tous deux, été retenus pour effectuer la validation par la prédiction sur l'année 2019.

Il est intéressant de noter que le nombre d'années de rotations pris en compte dans le modèle augmente le nombre de cultures non prédites (non prises en compte dans les calculs des indicateurs de performance). Comme le modèle 4 classes n'utilise que trois années de rotations (contre cinq années pour le modèle 10 classes direct), son utilisation permet d'obtenir plus de prédictions. Enfin, il ne faut pas oublier que le modèle de rotations de cinq ans est censé être moins robuste qu'un modèle de trois ans de rotations car il est calibré sur moins de données (4.4.4.1). L'utilisation d'un plus grand nombre de variables dans ce modèle peut également en être la cause. En effet, le RF n'utilise pas toutes les variables dans la construction des arbres décisionnels de la forêt. Un plus grand nombre de variables va donc induire une plus grande diversité d'arbres ce qui peut rendre les modèles moins robustes.

Pour la prédiction des 10 sous-classes, les performances données par les deux méthodes de prédictions (directe et indirecte) ne semblent pas suffisantes pour une utilisation dans le cadre de prédictions d'inondations boueuses. Une prédiction sur l'année 2019 a néanmoins été effectuée.

5.1.3. Validation des modèles

La validation consistait en l'utilisation des modèles calibrés pour prédire les cultures principales de l'année 2019. Dans un premier temps les modèles 4 et 10 classes direct ont été utilisés pour prédire les 4 classes de cultures. Ensuite les 10 sous-classes ont été prédites par les deux méthodes (directe et indirecte)

4 Classes de cultures

Le Graphe 12 montre que l'exactitude globale et la sensibilité de la classe « printemps » obtenue par le regroupement des sept sous-classes sont meilleures pour le modèle de 10 classes direct, à l'inverse de la précision. Les résultats obtenus lors de la calibration, qui n'étaient pas statistiquement significatifs, sont donc confirmés de manière plus nettes (les écarts de performance entre les deux modèles étant beaucoup plus importants dans la validation). Il semble donc que la classification en 4 classes noie l'information de la rotation et qu'il est plus judicieux d'utiliser le modèle 10 classes puis de regrouper les prédictions en 4 classes. Il se voit clairement (sur le Graphe 12) que l'obtention d'une meilleure sensibilité provoque une perte de la précision. Cela signifie que la prédiction de plus de cultures de printemps, induit fatalement une plus grande contamination.



Graph 12 : Comparaison de l'exactitude globale moyenne, de la précision et de la sensibilité de la classe printemps des prédictions sur l'année 2019 par le modèle 4 classes et le modèle 10 classes direct (regroupement des sept sous-classes de printemps).

En regardant la Figure 16, il s'avère que le nombre de cultures non prédites n'augmente pas considérablement entre les deux modèles (183 parcelles non prédites sur 3037 pour le modèle de 10 classes, contre 142 pour le 4 classes). Le modèle de 10 classes direct va donc être préféré pour prédire les cultures de printemps avec une précision et une sensibilité légèrement supérieure à 75%.

Prédiction des 4 classes de cultures en 2019

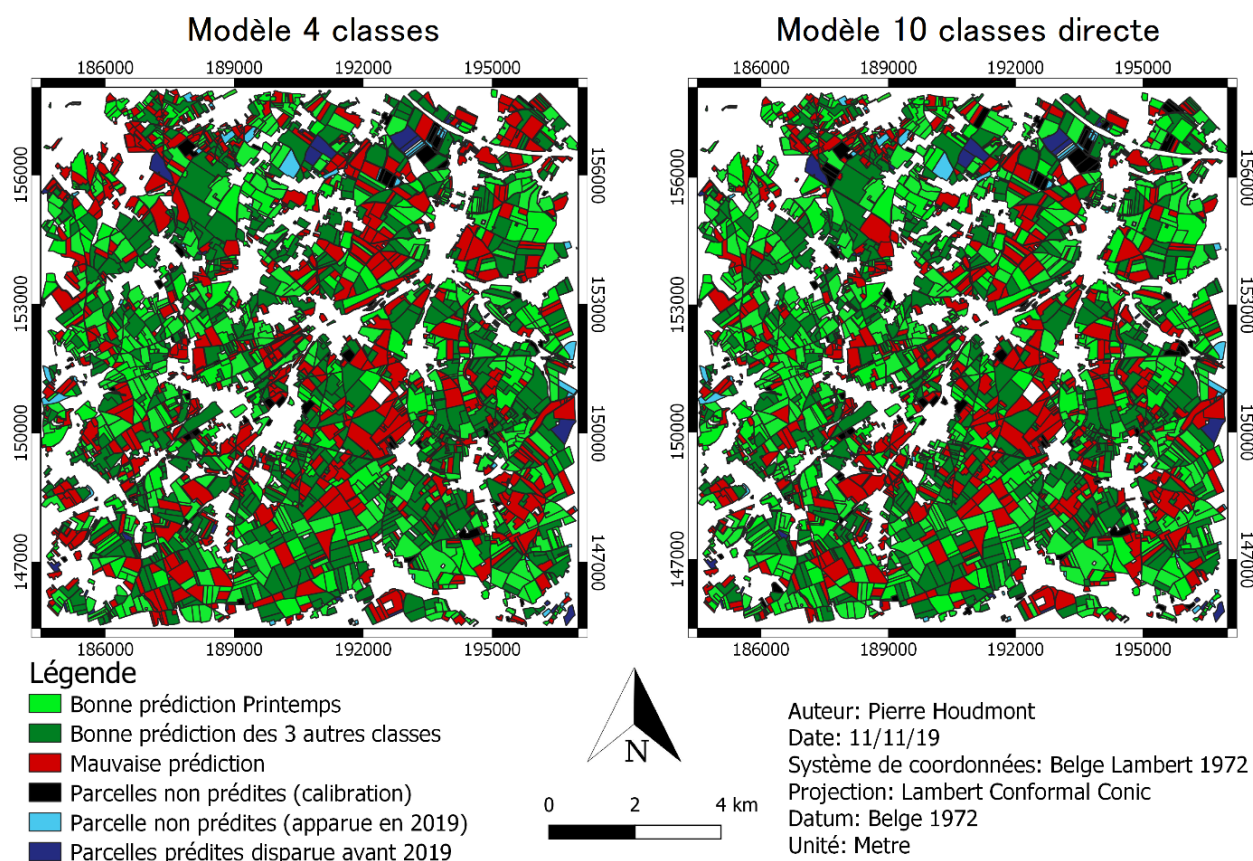


Figure 16 : Comparaison des prédictions des quatre classes pour l'année 2019 obtenues par le modèle 4 classes et le modèle 10 classes direct (regroupement des sept sous-classes de printemps).

10 Classes de cultures

Pour comparer les prédictions des sous-classes de cultures de printemps obtenues par les méthodes directes et indirectes, une mesure de l'exactitude remaniée a été faite (équation 7).

Exactitude des cultures de printemps

$$= \frac{\sum \text{Aires des cultures des 7 sous-classes (Printemps) bien prédites}}{\sum \text{Aires des cultures prédites en printemps (4 classes)}}$$

7

Les exactitudes des cultures de printemps calculées pour les méthodes indirecte et directe sont respectivement de 30,7 % et de 29,88%. Ces résultats signifient que plus de deux-tiers de l'aire des cultures prédites en printemps, ne vont pas être bien prédites dans les modèles de 10 classes. La Figure 17 présente la prédiction des sept sous-classes de printemps par les deux méthodes de prédictions. Elle précise également, les bonnes prédictions, les omissions et les contaminations de la prédiction des cultures de printemps obtenues grâce au modèle de 4 classes (méthode indirecte) et par le regroupement des sept sous-classes (méthode directe).

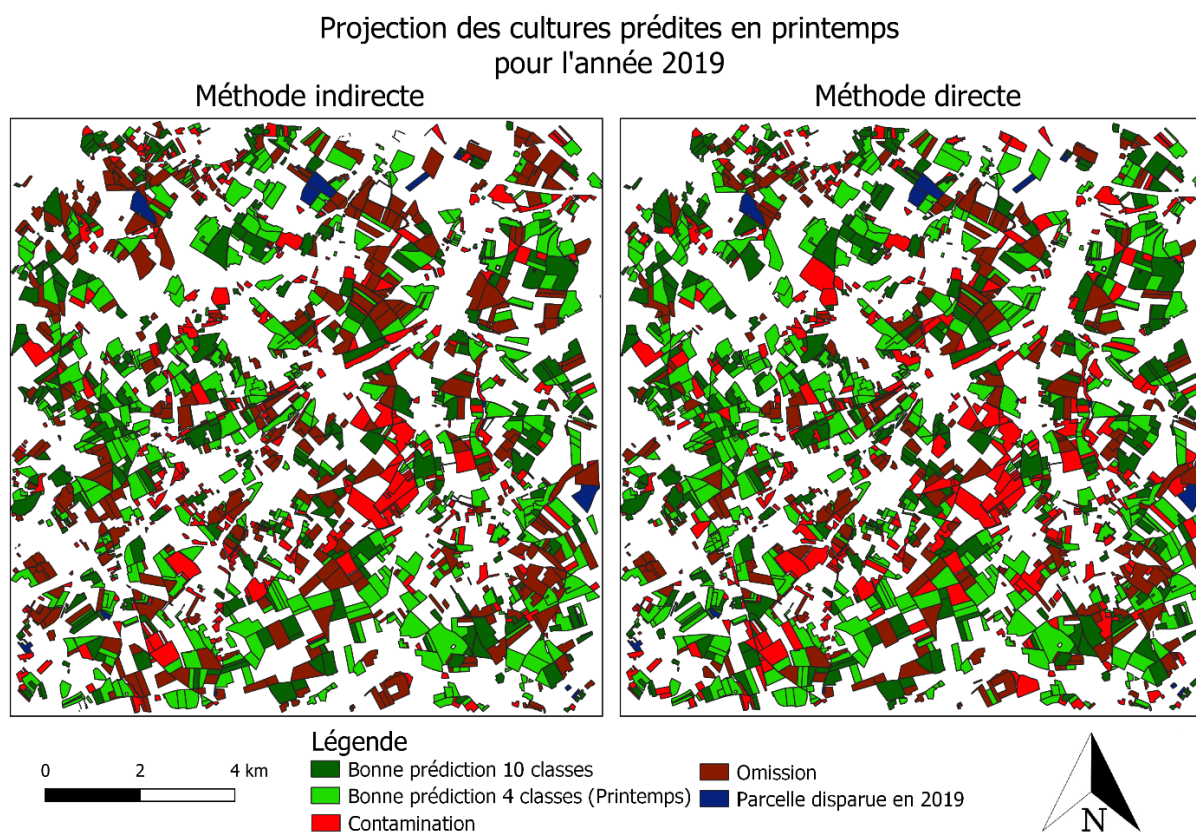


Figure 17 : Comparaison des prédictions sur les cultures de printemps, obtenues par les méthodes indirecte (modèles 4 classes puis 10 classes indirect) et directe (modèle 10 classes direct). Les bonnes prédictions en 4 classes représentent les parcelles mal prédites en dix classes mais bien prédites par le modèle 4 classes ou par le rassemblement des sept sous-classes de printemps (10 classes direct).

Discussion

La Figure 17 montre que le modèle 10 classes direct permet de détecter plus de cultures de printemps que le modèle 4 classes, ce qui lui permet d'avoir une plus grande sensibilité (Graphe 12). Cette plus large détection amène avec elle plus de contaminations ce qui explique la moins bonne précision sur le Graphe 12. En ce qui concerne les sept sous-classes de cultures de printemps (modèle 10 classes), on peut se rendre compte visuellement des mauvaises exactitudes des cultures de printemps, en analysant la Figure 17 et son manque flagrant de cultures bien prédites pour la prédiction des 10 classes.

En partant de ce constat, les rotations ne permettent pas de prédire de manière satisfaisante les sous-classes. La méthode de prédictions de ces 10 classes n'aurait donc pas d'intérêt dans le but de prédire les zones à risques d'inondations boueuses.

En revanche, le modèle des 10 classes direct permet d'obtenir de meilleures prédictions pour la classe « printemps » lorsque les sept sous-classes sont rassemblées. Comme la plupart des cultures de printemps sont des cultures sarclées contribuant aux zones à risques, ce modèle s'avère donc utilisable pour les prédire. Néanmoins, il faut garder à l'esprit qu'environ 25% des cultures de printemps ne seront pas prédites comme telles (précision de 75,25%), et qu'environ 25% des cultures prédites en printemps ne le seront pas en réalité (sensibilité de 75.83%), ce qui reste considérable.

5.2. Détection des couverts hivernaux

La détection des couverts hivernaux a été effectuée grâce au processeur L4C. Dans un premier temps, la calibration a consisté en la détermination de la meilleure période de détection (4.5.2). Ensuite, la validation, effectuée par photo-interprétation, a servi à connaître l'effet de l'utilisation de la période de détection retenue sur la détection des couverts d'une autre année (4.5.3). Enfin, l'amélioration du modèle a permis d'utiliser les résultats et de les associer au modèle de détection des couverts pour augmenter ses performances (4.5.4).

5.2.1. Calibration

Avant tout, au vu des résultats, il est important de noter que l'utilisation des seuils calculés pour les Pays-Bas (Annexe D, Tableau A-30) convient pour une détection des couverts en Wallonie, par le processeur L4C.

Maximisation des performances par le choix de la période de détection

On remarque que la période de détection donnée au processeur influence fortement les détections des couverts hivernaux. Pour rappel, les périodes de détections choisies pour effectuer la calibration (4.5.2) permettent de se rendre compte de l'influence de la date de début, de fin et de la durée de la période sur les détections.

Tableau 17 : Effet du choix de la date de début d'intervalle dans la performance de détection des couverts.

Période	CIPAN présumée	Contamination par des parcelles en culture d'hiver	Culture de printemps 2019 (prédictibles par EO)	
	Sensibilité		Sensibilité	Précision
15/08 – 15/11	33.98%	5.13%	27.71%	91.48%
01/09 – 15/11	51.99%	4.55%	36.49%	92.8%
15/09 – 15/11	60.57%	5.89%	59.03%	92.22%
20/09 – 15/11	71.57%	5.74%	65.53%	92.48%
25/09 – 15/11	85.12%	5.89%	81.08%	92.52%
30/09 – 15/11	85.12%	5.52%	81.15%	92.89%
05/10 – 15/11	85.52%	6.59%	82.52%	91.87%
10/09 – 10/11	63.94%	6.54%	63.13%	91.70%
20/09 – 10/11	75.73%	6.29%	70.17%	92.03%
25/09 – 10/11	90.28%	9.76%	87.59%	88.76%
30/09 – 10/11	90.28%	9.44%	87.67%	89.07%
05/10 – 10/11	91.72%	14.05%	89.48%	84.56%
20/09 – 30/10	77.17%	7.43%	72.72%	90.96%
25/09 – 30/10	92.08%	10.65%	90.63%	87.92%
30/09 – 30/10	92.08%	10.4%	90.7%	88.16%

Résultats

Le Tableau 17 montre, pour des mêmes dates de fin, l'influence du changement du début de la période sur la performance du processeur dans la détection des couverts. Il montre que pour une date de fin située autour du 15 novembre, le début de la période n'apporte que très peu de changements au niveau de la contamination par des cultures d'hiver, ce qui signifie que les intercultures courtes sont bien écartées des détections par les marqueurs 8 et 9. Plus globalement, il s'observe que le fait de retarder la date de début de l'intervalle permet de détecter plus de couverts (les sensibilités augmentant sensiblement). Dans le même temps, il se remarque une augmentation de la contamination par des cultures d'hiver. Cela signifie qu'en retardant la date de début de la période, le processeur va détecter plus de couverts hivernaux, mais cela implique aussi une augmentation des cultures d'hiver détectées incorrectement.

En ce qui concerne la date de fin de la période, une hypothèse de la fixer proche du 15 novembre (date à partir de laquelle les agriculteurs peuvent détruire leurs CIPAN) avait été faite. Dans le Tableau 18, on remarque en effet qu'en dépassant cette date (30/09 – 30/11), une chute se produit dans la détection des couverts. En revanche, avancer la date de fin permet d'augmenter la détection des couverts hivernaux, mais dans le même temps d'augmenter la contamination par des cultures d'hiver.

Tableau 18 : Effet du choix de la date de fin d'intervalle dans la performance de détection des couverts.

Période	CIPAN présumée	Contamination par des parcelles en culture d'hiver	Culture de printemps 2019 (prédictibles par EO)	
	Sensibilité		Sensibilité	Précision
20/09 – 30/10	77.17%	7.43%	72.72%	90.96%
20/09 – 10/11	75.73%	6.29%	70.17%	92.03%
20/09 – 15/11	71.57%	5.74%	65.53%	92.48%
25/09 – 10/11	90.28%	9.76%	87.59%	88.76%
25/09 – 15/11	85.12%	5.89%	81.08%	92.52%
30/09 – 10/11	90.28%	9.44%	87.67%	89.07%
30/09 – 15/11	85.12%	5.52%	81.15%	92.89%
30/09 – 30/11	48.49%	18.17%	54.88%	79.45%
05/10 – 10/11	91.72%	14.05%	89.48%	84.56%
05/10 – 15/11	85.52%	6.59%	82.52%	91.87%

Résultats

Quant à l'effet de la durée de la période, on voit sur le Tableau 19 que plus elle est courte, plus nombreux sont les couverts détectés. Il faut néanmoins admettre que durant la période en question, on retrouve souvent une couverture nuageuse importante en Wallonie. Lors de mauvaises années, le processeur pourrait ne pas obtenir de données de NDVI si la durée de la période choisie est trop courte. La durée minimale pour la période a donc été fixée à 1 mois et 10 jours, soit 40 jours. Cette durée permet le passage du satellite Sentinel-2 sur la zone d'au minimum 10 fois.

Une autre observation révélée par le Tableau 19 est que les performances de détection vont être modifiées considérablement entre deux périodes de durées égales mais décalées de cinq jours dans le temps. Ainsi, une variation d'environ 10 % dans la sensibilité de la détection des CIPAN présumées apparaît pour la durée d'un mois et 20 jours. Elle est aussi d'environ 5% pour les durées d'un mois et 15 jours et d'un mois et 10 jours. Dans le cadre de l'utilisation de la même période sur plusieurs années, cela peut poser problème. En effet, le climat variant d'année en année, les agriculteurs ne sèment pas leurs cultures spécifiquement aux mêmes dates, et la croissance des couverts peut évoluer différemment entre deux années. Si le simple déplacement de la période de détection de 5 jours induit autant de variations dans le modèle de détection, il se peut que le modèle utilisé sur une autre année donne de pareilles variations, sinon pires.

Tableau 19 : Effet de la durée et de la situation de la période dans l'année sur la détection des couverts.

	Période	CIPAN présumée	Contamination par des parcelles en culture d'hiver	Culture de printemps 2019 (prédictibles par EO)	
		Sensibilité		Sensibilité	Précision
2 m	10/09 – 10/11	63.94%	6.54%	63.13%	91.70%
	15/09 – 15/11	60.57%	5.89%	59.03%	92.22%
	30/09 – 30/11	48.49%	18.17%	54.88%	79.45%
1m 20 j	20/09 – 10/11	75.73%	6.29%	70.17%	92.03%
	25/09 – 15/11	85.12%	5.89%	81.08%	92.52%
1m 15 j	25/09 – 10/11	90.28%	9.76%	87.59%	88.76%
	30/09 – 15/11	85.12%	5.52%	81.15%	92.89%
1m 10 j	30/09 – 10/11	90.28%	9.44%	87.67%	89.07%
	05/10 – 15/11	85.52%	6.59%	82.52%	91.87%

Au vu des différents résultats obtenus et de la durée minimum fixée à 1 mois et 10 jours, le choix de garder deux périodes (l'une maximisant la sensibilité et l'autre maximisant la précision des cultures de printemps prédictibles de 2019) a été fait pour représenter géographiquement les prédictions. Les périodes retenues sont donc les périodes allant du **30/09 – 10/11** et du **30/09 – 15/11**. La première d'entre elle correspond aussi à la période présentant la meilleure balance entre la précision et la sensibilité. Dans le cadre de prédictions pour les inondations boueuses, c'est donc la période du **30 septembre au 10 novembre** qui serait utilisée. Les tableaux des matrices permettant de calculer la sensibilité et la précision des cultures de printemps « prédictibles » pour ces deux périodes sont disponibles en annexe (Annexe D, Tableaux A-32).

Comme expliqué dans la méthodologie, toutes les cultures de printemps n'étaient pas prédictibles. Plusieurs raisons sont liées à cela. La première est due aux cultures de printemps qui ne sont tout simplement pas précédées par des couverts car elles sont précédées par des cultures récoltées après le 1^{er} septembre. Les agriculteurs ne sont donc pas tenus d'implanter une CIPAN dans ce cas. La superficie totale des parcelles de ce type représentait **22,46%** des cultures de printemps de 2019. Cela signifie que malgré une détection totale des couverts hivernaux, la sensibilité du modèle de prédictions des cultures de printemps par télédétection ne dépasserait jamais les 77,54%.

Les autres raisons sont, elles, directement liées à la télédétection. Les parcelles plus petites ou ne contenant qu'un seul pixel de Sentinel-1 (400 m²) dans son tampon interne (4.5.1.1) ne sont pas considérées par le processeur. Ensuite, il y a, les parcelles sur lesquelles une récolte n'a pas été détectée, ou détectée plus tard que la date de début de la pratique culturale, pour qui le processeur n'a pas fait tourner les marqueurs de la pratique culturale. Enfin, il y a les parcelles pour lesquelles un nombre insatisfaisant, voir nul, d'images ont pu être utilisées (les images ont été jugées insatisfaisantes lors de l'extraction de données (4.5.1.1)). Ce dernier point varie selon l'intervalle de temps choisi. Pour les deux périodes retenues, la proportion des cultures de printemps de 2019 pour qui le processeur n'a pas effectué de détection est identique, et vaut 1,45%. Le pourcentage total des cultures de printemps de 2019 non prédictibles par la télédétection est donc de **23,91%**.

En prenant en compte ces cultures de printemps non prédictibles dans le calcul de la sensibilité de la classe « Printemps », la détection sur la période du 30/11 au 10/09 amènerait une sensibilité de **66,7%** et de 61,74% pour la période du 30/11 au 15/09.

La Figure 18 montre la représentation cartographique de la qualité des prédictions ainsi obtenues pour les deux périodes retenues. On y observe que les cultures non prédictibles (en gris), sont relativement nombreuses. En bordaux, les omissions sont peu présentes dans les deux cas, surtout concernant la période se clôturant le 10 novembre. Les contaminations (en rouge) y sont en revanche légèrement plus nombreuses.

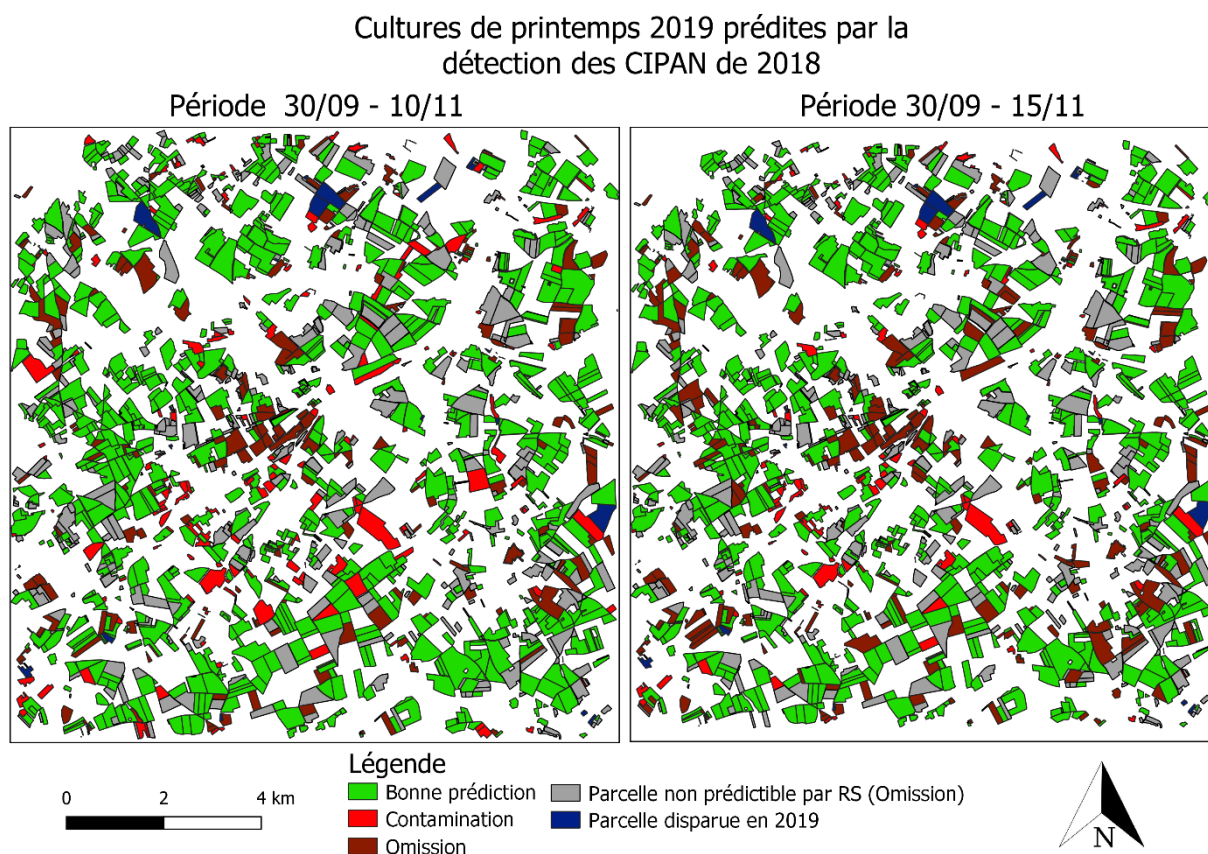


Figure 18 : Comparaison des prédictions des cultures de printemps de 2019 par la détection des couverts hivernaux dans les périodes du 30/09 au 10/11 et du 30/09 au 15/11.

Une analyse approfondie des contaminations a été effectuée. Au niveau des cultures d'hiver, elle a montré que le colza d'hiver est systématiquement détecté comme un couvert hivernal. L'orge d'hiver (dont environ le tiers des parcelles est détecté comme étant un couvert) et le froment (environ 5% des parcelles de froment sont détectées comme couverts) sont les deux autres cultures d'hiver représentées en nombre. Les autres cultures contaminantes sont des cultures classées en prairie, comme les « autres fourrages », « bande/parcelle aménagée MAE - MC7, MC8, MB5 », « autres couverts » ou encore la luzerne et le trèfle. En jetant un œil aux NDVI (Annexe D, Graphes A-20), beaucoup de ces cultures contaminantes seraient difficilement différenciables des CIPAN avec les marqueurs utilisés par le processeur L4C.

En ce qui concerne les omissions, c'est-à-dire les CIPAN présumées non détectées, l'analyse des NDVI montre trois cas récurrents (Annexe D, Prédiction par la détection des CIPAN Graphes A-21). Soit, aucune croissance n'est détectée, la parcelle ne semble pas accueillir de culture, l'agriculteur n'ayant peut-être pas semé de CIPAN malgré l'obligation. Un autre cas observé montre des croissances semblables à celle des couverts typiques. Malheureusement cette croissance est trop tardive pour être détectée par le processeur dans la période allant jusqu'au 10/11. Les autres cas de figures recensés, étaient des cultures semblant être des couverts, ayant été détruits ou ayant subi une décroissance pendant ou avant la période de détection. Ces cas sont étonnants puisque les agriculteurs ne peuvent en principe pas détruire leurs CIPAN avant le 15 novembre.

5.2.2. Validation

Le processeur SEN4CAP L4C a été utilisé pour détecter les couverts présents entre les cultures de 2019 et 2020. Pour ce faire, les périodes de récoltes et la période de détection (30/09-10/11) calibrées sur les détections de couverts en 2018 ont été utilisées. Les résultats ainsi obtenus ont ensuite pu être confrontés aux 50 parcelles photo-interprétées comme des couverts (4.5.3).

Il s'est avéré que seule une parcelle, sur les 50, n'a pas été détectée comme un couvert par le processeur. La Figure 19 montre d'ailleurs qu'elle n'est pas très importante en superficie. La sensibilité (en aire) de cette détection vaudrait donc 99,51%. Pour rappel, la précision, n'a pas pu être calculée. Les périodes utilisées semblent donc utilisables sur plusieurs années. Néanmoins, une validation supplémentaire permettant d'avoir une idée de la contamination devrait encore être effectuée pour s'en assurer.

Validation de la méthode de détection

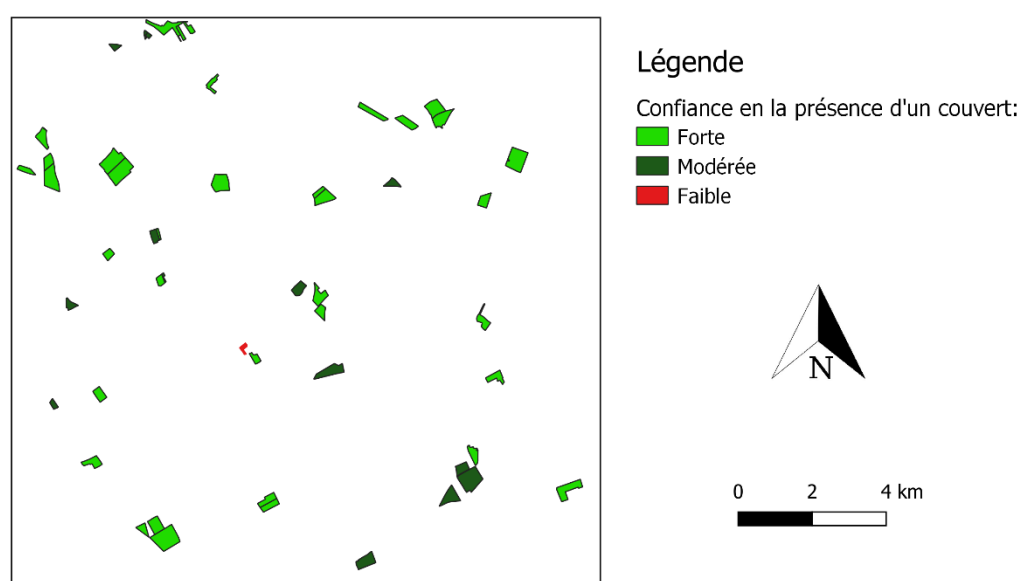


Figure 19 : Confrontation de la détection des couverts dans la période du 30/09 au 10/11 avec des parcelles photo-interprétées comme des couverts.

5.2.3. Amélioration du modèle

Probabilité simple de succession de cultures

Malheureusement, les probabilités calculées sur bases des successions de cultures montrent que les monocultures de maïs ne constituent qu'une petite partie des parcelles de maïs et qu'elles sont difficilement isolables. En effet, en moyenne sur les 5 années (des parcellaires utilisés dans la partie 1), seules 24,8% de l'aire des parcelles de maïs vont être reconduites en maïs l'année suivante. Une manière d'isoler les parcelles de maïs en monoculture était de prendre en compte plusieurs années antécédentes. 54,5 % des aires des parcelles ayant été en maïs pendant deux ans, le seront encore l'année suivante. De même, au regard du Tableau 20, 63% des aires des parcelles ayant été en maïs durant trois ans le seront encore l'année

Résultats

suivante. Néanmoins, l'utilisation des trois années précédentes ne permettraient d'obtenir au mieux que 43,8% des cultures de maïs se trouvant dans un processus de monoculture.

Tableau 20 : Matrice de confusion obtenues par la prédiction des cultures de maïs (précédées par une autre culture de maïs) de 2019 en utilisant les parcelles de monocultures de maïs sur la période 2016-2018.

		Culture 2019		Précision
		Maïs (2018, 2019)	Autre	
Cultures antécédentes	Maïs (2016, 2017, 2018)	48,515 ha	28,454 ha	63%
	Autre	62,296 ha		
Sensibilité		43,8%		

Les cultures de maïs n'étant pas les seules cultures de printemps pouvant suivre une culture récoltée tardivement, le fait d'en détecter seulement 43,8% ne semble pas suffisant dans la recherche de la prédiction des cultures de printemps non précédées d'un couvert.

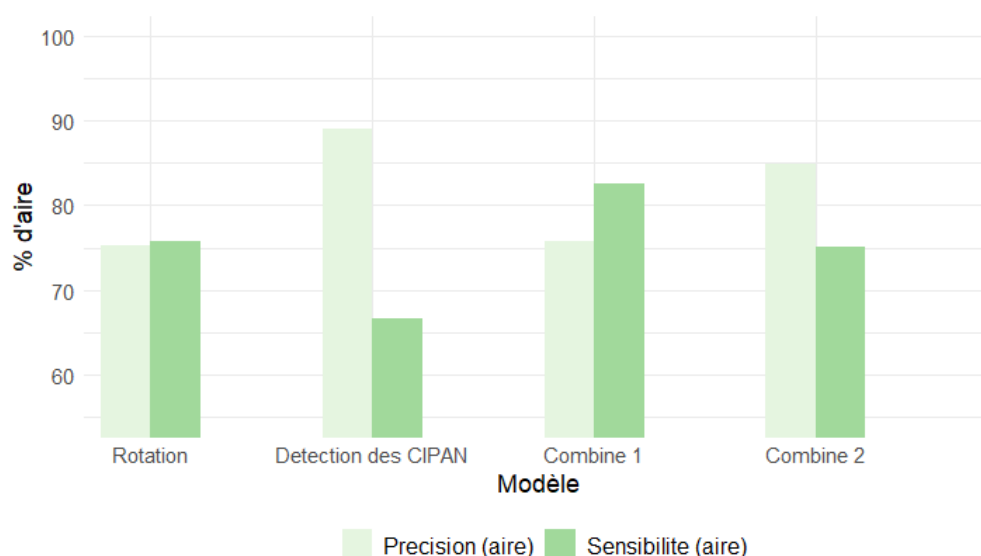
Un fait intéressant remarqué en revanche, est que la probabilité de voir apparaître une culture de printemps après un maïs est d'environ une chance sur deux, et ce pourcentage est en augmentation. Ainsi, si 42,6% des cultures de maïs de 2015 étaient suivies d'une culture de printemps, elles étaient 57,2% en 2017. Si l'expansion des cultures de printemps aux dépens des cultures d'hiver continue de croître, cette probabilité pourrait encore augmenter, ce qui permettrait de réellement pouvoir prédire une bonne partie des cultures de printemps non précédées d'un couvert. En attendant, cette progression semble stagner car **56,2%** des maïs de 2018 ont été suivis d'une culture de printemps en 2019. Utiliser les cultures de maïs comme prédicteur de cultures de printemps peut donc permettre de renforcer le modèle de détection des couverts hivernaux en prédisant une quantité non négligeable de cultures de printemps, mais il apportera avec lui beaucoup d'erreurs.

Combinaison avec le modèle de prédictions par rotation

Le modèle de prédictions par rotation choisi pour être combiné est le modèle ayant donné les meilleures prédictions de la classe « Printemps » dans la rubrique de validation (4.4.5), à savoir le modèle 10 classes direct, dont les prédictions sont regroupées en 4 classes. En ce qui concerne la détection des couverts hivernaux, la période retenue est celle allant du 30 septembre au 10 novembre, cette période permettant de détecter le plus de cultures de printemps.

Résultats

Comme anticipé (4.5.4), une meilleure sensibilité pour le modèle combiné 1 et une meilleure précision pour le modèle combiné 2 sont observées sur le Graphe 13. Il est possible de s'en rendre compte visuellement sur les cartes de la Figure 20, où les parcelles contaminantes (en rouge) sont bien plus nombreuses dans le modèle combiné 1. Et où les parcelles de printemps omises sont plus nombreuses dans le modèle combiné 2.



Graphe 13 : Précisions et sensibilités des prédictions de cultures de printemps de 2019 obtenues par les modèles utilisant la rotation (10 classes direct), la détection des couverts (entre le 30/09 et le 10/11) et les modèles combinant les deux méthodes.

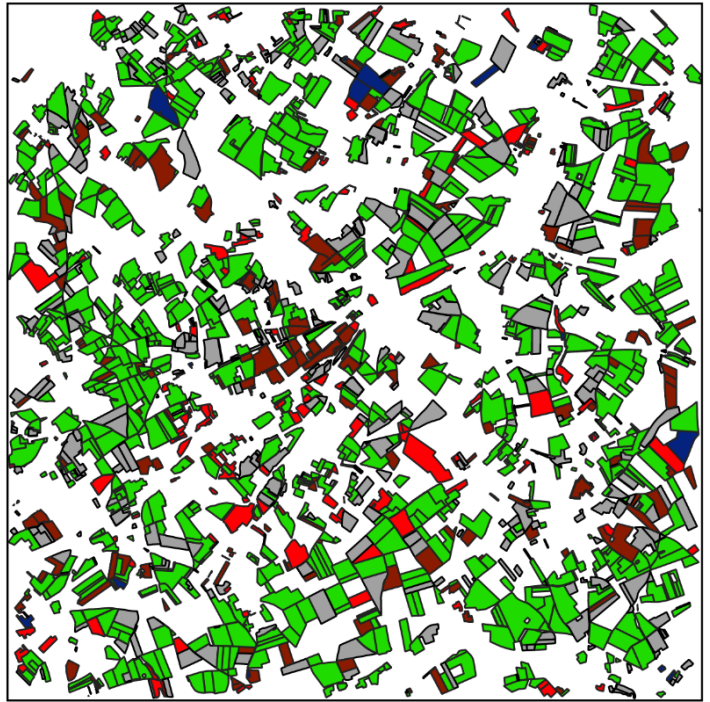
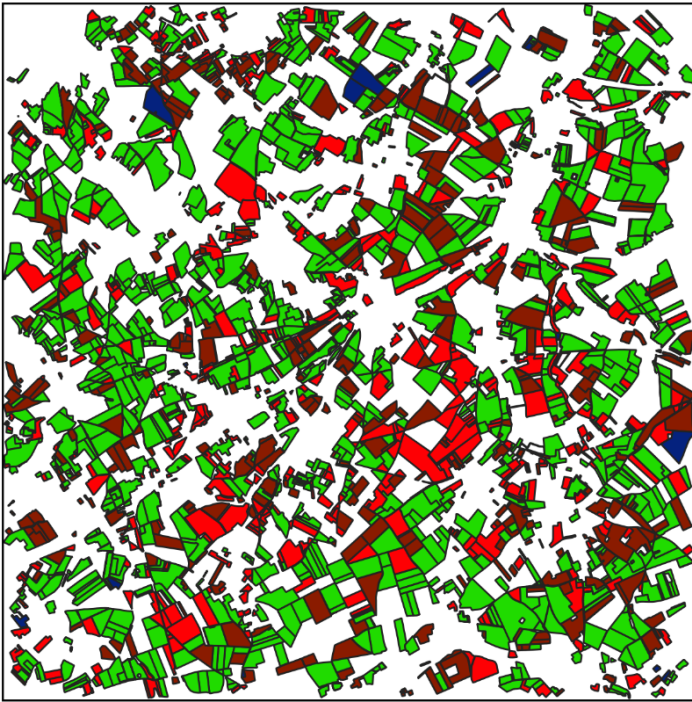
Malheureusement, la Figure 20 montre que la plupart des cultures non prédictibles par détection des couverts hivernaux (gris), ne sont pas prédites en printemps par le modèle de rotation, ce qui empêche d'obtenir des résultats réellement satisfaisants (dans les 90% de précision et de sensibilité) pour les combinaisons des modèles. Il est logique que ce soit le cas. Les parcelles non-prédictibles représentent les parcelles de printemps succédant à une culture de printemps récoltée tardivement. Or, à l'exception des cultures de maïs, la probabilité de voir apparaître une culture de printemps après une autre culture de printemps est assez faible (il est possible de s'en rendre compte en jetant un œil aux occurrences de rotations (Annexe C, Tableau A-28). Le modèle de rotation ne va donc généralement pas prédire une culture de printemps après une culture de printemps. Et *in fine*, les parcelles en question sont donc mal prédites.

Toutes ces parcelles ne vont néanmoins pas être omises, ce qui va permettre de pouvoir augmenter la sensibilité par rapport à la méthode de détection des couverts utilisée seule. Le modèle combiné 1 permet d'obtenir une bonne sensibilité (**82,63%**) pour une précision de **75,82%**. Cette précision, assez basse par rapport à la sensibilité, est due au fait qu'en plus des parcelles omises par la détection des couverts, elle prend en compte également toutes les cultures omises de la prédiction par méthode de rotation. Le modèle combiné 2 ne va pas subir ce problème. Sa précision va monter à **84,95 %**, mais dans le même temps, il va détecter moins de cultures de printemps, sa sensibilité étant de **75,07%**.

Cultures de printemps 2019 prédites par les différents modèles

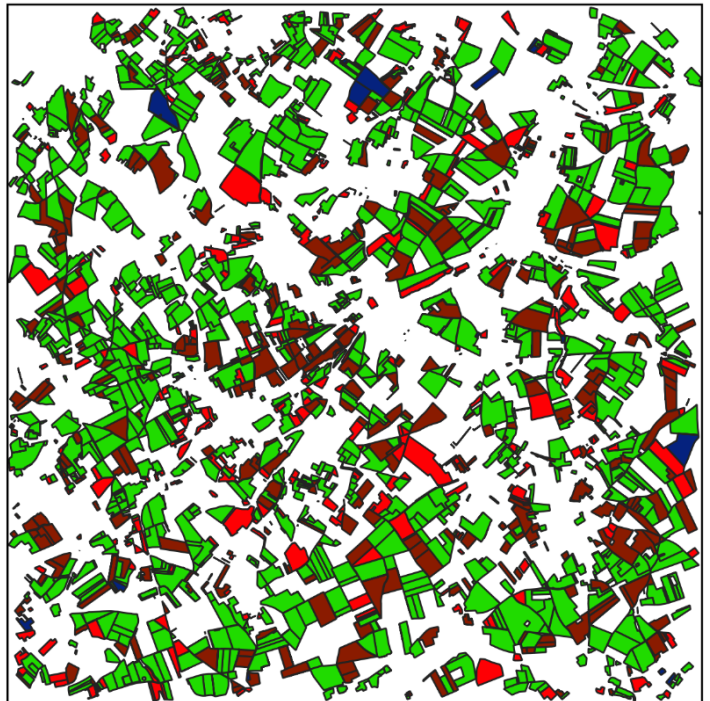
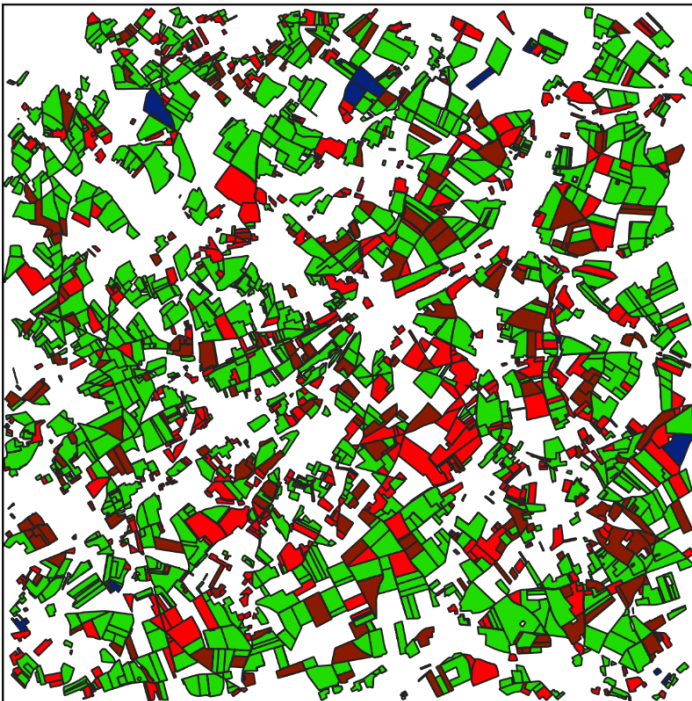
Rotation (10c. directe)

Détection de CIPAN (30/09 - 10/11)



Modèles combinés 1

Modèles combinés 2



0 2 4 km

Légende

■ Bonne prédiction
■ Contamination

■ Omission

■ Parcelle non prédictible par EO

■ Parcelle disparue 2019



Figure 20 : Représentation cartographiques des modèles de prédictions de cultures de printemps obtenus par la méthode de rotation, de détection des couverts et les modèles combinant les deux.

6. Discussion générale et perspectives d'améliorations

Malgré leurs faibles améliorations face aux résultats des modèles obtenus par rotation, les modèles combinés semblent utilisables dans l'ambition de prédire les inondations boueuses. Les deux modèles combinés pourraient être utilisés alternativement. L'utilisation du modèle combiné 1 permettrait de prédire la grande majorité des zones à risques, dont des zones qui ne le seront peut-être pas dans la réalité. Alors que l'utilisation du modèle combiné 2 prédirait moins de zones à risques qu'il y en aurait en réalité. En revanche, celles prédites auront beaucoup de chance de l'être réellement.

Les résultats sont toutefois insatisfaisants par rapport à l'attente qu'il y avait en utilisant les deux méthodes de prédictions. En effet, une sensibilité et une précision de la classe « printemps » situées au-delà des 90% aurait été la bienvenue. De plus, les performances de prédictions obtenues grâce aux modèles combinés pourraient être plus faibles, par le fait que le modèle de détection des couverts utilisé pour l'effectuer était le modèle calibré et non un modèle validé. Dans la réalité, la détection des couverts serait, donc, possiblement plus faible et il s'en suivrait logiquement que les résultats combinés le soient aussi.

Au niveau de la temporalité des prédictions, les modèles combinés pourraient certainement être utilisés assez tôt de manière à anticiper les zones à risques. Les prédictions par le modèle de rotation peuvent être obtenues dès la réception du LPIS de l'année précédant l'année à prédire. En Wallonie il est rempli par les agriculteurs dans le courant de mai-juin, et est disponible au public plus tard dans l'année. Les détections peuvent, elles, être obtenues soit après la période de détection implémentée dans le processeur (mi-novembre), soit après l'obtention du LPIS s'il n'est pas encore disponible. Toutefois, comme l'outil serait utilisé dans l'intérêt des communes et de la Région, si le LPIS n'est pas disponible au public en novembre, il serait sans doute possible de l'obtenir malgré tout. Les prédictions des méthodes combinées pourraient donc être obtenues mi-novembre ce qui laisse largement le temps de prendre des mesures préventives contre les inondations boueuses.

Plusieurs pistes peuvent être suivies afin d'améliorer les performances des modèles pour une utilisation dans un cadre réel de prédictions de zones à risque d'inondations boueuses.

- Les modèles ont été calibrés et validés sur la zone d'étude, il serait intéressant de tester les modèles, après une éventuelle recalibration, sur d'autres zones propices aux inondations boueuses comme dans le Condroz par exemple, où les cultures principales peuvent être différentes. Les performances des modèles dans ces zones pourraient alors être un bon indicateur permettant de se rendre compte de l'utilité que pourraient apporter ces modèles dans le cadre d'une utilisation réelle pour aider à prédire les zones à risque d'inondations boueuses.

- Au niveau du modèle de rotation, les classes n'étaient pas stables dans le temps (4.3.2), avec la croissance de la superficie des cultures de printemps aux dépens des cultures d'hiver. Cela pourrait poser des problèmes si le modèle calibré avec les années 2014 à 2018 était utilisé pour prédire des cultures dans plusieurs années, la répartition des différentes cultures ayant pu fortement évoluer entre-temps. Dans le cadre d'une utilisation réelle du modèle dans un outil de prédictions des inondations boueuses, il vaudrait mieux réentraîner le modèle sur les années qui précèdent immédiatement l'année de prédictions voulue.
- En créant le modèle de prédictions par rotation, le postulat qu'une culture d'hiver n'influence pas la culture suivante avait été posé. Il serait utile de vérifier ce postulat en divisant les cultures d'hiver en plusieurs sous-classes et en observant si cela ne permet pas d'augmenter la performance après regroupements des prédictions en 4 classes.
- Un autre point à discuter concernant le modèle de rotation se situe au niveau des variables utilisées dans le modèle 10 classes direct. Il a été choisi de lui donner les mêmes variables que celles du modèle 10 classes indirect pour obtenir une bonne performance de prédictions des sept sous-classes de printemps. Toutefois, le modèle 10 classes direct est finalement utilisé pour prédire les 4 classes. Les variables composant le modèle ne sont donc possiblement pas celles qui maximiseraient les performances de prédictions. Effectuer à nouveau les tests des variables à prendre en compte (4.4.4.3) pourrait donc aider à obtenir de meilleures performances de prédictions.
- En outre, lors de la phase d'entraînement du modèle, le *Random Forest* traite les données par parcelles dans la création des arbres de décision. Ces derniers sont donc construits de manière à obtenir la meilleure classification selon le nombre de parcelles. Or, comme effectué dans le choix des variables du modèle (4.4.4.3), il serait préférable de maximiser les classifications en termes d'aire de parcelles et non en nombre. Pour ce faire, l'utilisation d'une autre méthode pourrait être envisagée. Lors du mémoire de Urdiain-Arraiza (2018), les chaînes de Markov avaient été utilisées et elles permettaient d'obtenir de meilleures performances que les réseaux de neurones. Une utilisation de celles-ci permettrait de prendre en compte les aires, plutôt que les parcelles, dans le calcul de probabilités, ce qui permettrait d'obtenir un modèle de prédictions par rotation possiblement plus performant que ceux obtenus grâce au *Random Forest*. Ce modèle ne possède néanmoins pas la possibilité de pouvoir prendre en compte d'autres variables que la séquence des cultures successives sur la parcelle.

- Concernant le modèle de prédictions des cultures de printemps par la détection des couverts, il a été observé que la performance de détection du processeur L4C est fortement influencée par la période de détection choisie. Dans le cadre d'une utilisation réelle, les détections pourraient être vérifiées par la confrontation à un jeu de donnée pris sur le terrain (les couverts étant déjà implantés au moment des détections). La période de détection pourrait alors être modifiée si elle ne permet pas d'obtenir une performance de détection satisfaisante.
- Pour améliorer les performances du modèle, l'utilisation d'un modèle de classification comme la méthode du maximum de vraisemblance, en utilisant des dates clés entre septembre et novembre, pourrait permettre d'obtenir de meilleures performances que le processeur L4C. Une utilisation plus tardive après novembre d'un modèle de ce type (à l'instar du modèle de Danzin et al. (2018)) permettrait aussi d'augmenter fortement les performances. Cependant, le modèle ne serait pas un modèle de type logique et donc pas applicable chaque année.
- Une autre manière d'améliorer le modèle de détection, en diminuant l'erreur de contamination, serait d'utiliser un ou plusieurs marqueurs supplémentaires dans le processeur SEN4CAP, dont l'objectif est de détecter une perte de végétation sur la parcelle après le 15 novembre (soit une destruction des couverts). Ce modèle aurait l'avantage, par rapport au point précédent, d'être logique et donc utilisable sur différentes années. Cependant, la période de destruction durant laquelle il faudrait tester ces marqueurs pourrait être longue, les agriculteurs ne détruisant pas systématiquement leurs couverts en laissant parfois le gel faire ce travail. Les résultats obtenus de cette façon le seront donc seulement dans le courant du mois de février, ce qui laisse alors peu de temps pour la prise des mesures avant l'apparition potentielles des inondations boueuses.
- Il a été montré que malgré une détection totale des couverts hivernaux, la sensibilité de la classe printemps ne dépasserait pas les 77.5% pour l'année 2019 (5.2.1). En outre, cette sensibilité semble décroître au fil du temps. En effet, entre 2015 et 2018, la proportion de cultures de printemps précédées d'une culture hivernale n'a cessé de diminuer (Tableau 21). Cette diminution peut s'expliquer par la tendance à augmenter les superficies de cultures de printemps au dépend des cultures d'hiver pour cause de rentabilité économique. Comme moins de cultures d'hiver précèdent les cultures de printemps, moins de couverts hivernaux sont implantés. La sensibilité maximale obtenue par le modèle de prédictions des cultures de printemps basé sur la détection des couverts va donc diminuer si ces modifications agronomiques perdurent. Afin d'augmenter les performances des modèles de prédictions combinés, il serait donc préférable de commencer par améliorer le modèle de prédictions par rotation.

Discussion générale et perspectives d'améliorations

Tableau 21 : Evolution de la proportion de la superficie des cultures de printemps précédées par des cultures d'hiver entre 2015 et 2019 sur la zone d'étude.

2015	2016	2017	2018	2019
77.45%	76.12%	72.55%	66.53%	62.33%

- Dans leurs études d'identification de cultures, Blaes et al. (2005), ont cherché à vérifier la validité du LPIS. Sur 899 parcelles, 56 ne correspondaient pas aux cultures principales mentionnées dans le LPIS, ce qui représente 6,23% d'erreur. Si les LPIS n'ont pas été améliorés depuis, il est fort probable qu'une proportion similaire de parcelles soient encore mal identifiées dans les LPIS utilisés lors de ce mémoire. Par ailleurs, le LPIS de 2019 utilisé est un LPIS provisoire car le PAA de 2019 n'est pas encore disponible au public. Quelques erreurs d'identification de cultures principales ont aussi pu se glisser dans ce parcellaire.

Ces parcelles contenant de l'erreur ont pu se répercuter à plusieurs niveaux. Dans le modèle de rotation, lors de l'entraînement et lors de la prédiction des modèles, ainsi que lors de la confrontation des prédictions au parcellaire de 2019. Dans le modèle de détection, lors des choix de parcelles à détecter, dans les périodes de récoltes attribuées aux parcelles à problèmes, et lors de la génération de données de référence pour l'analyse des détections.

- Il serait intéressant de conceptualiser un modèle permettant de prédire les limites de parcelles à l'avance. En effet, cela permettrait de diminuer l'erreur due à l'utilisation du parcellaire de l'année antécédente (donc aux modifications entre les deux parcellaires) pour projeter les prédictions. C'est le travail d'une mémorante du laboratoire de géomatique de l'UCLouvain cette année. Elle essaiera ainsi de les prédire par imagerie satellite en utilisant des algorithmes de *deep learning*.

Enfin, il faut noter que ces améliorations éventuelles ne pourront, en aucun cas, permettre d'atteindre une prédiction parfaite des cultures de printemps. Une détection totale des couverts ne permet pas de prédire plus de 77.5% des cultures hivernales (de 2019) et le modèle de rotation dépend des choix agronomiques et économiques faits par les agriculteurs. En effet, leurs décisions d'implantation de cultures peuvent différer des schémas habituels retrouvés dans la région. Ils peuvent très bien choisir d'implanter une culture plus rentable qui ne suit plus une logique de rotation typique (observé par la diminution des cultures d'hiver par exemple). Par conséquent, aucun modèle ne permettrait de prédire les choix que décideront tous les agriculteurs, du moins en utilisant les deux approches de ce mémoire.

7. Conclusion

Ce mémoire avait pour objectif de prédire les cultures de printemps, à l'échelle de la parcelle, le plus tôt possible dans l'année. Pour se faire deux approches ont été développées.

La première utilise les successions de cultures sur les années antécédentes pour prédire les cultures de printemps. Trois modèles ont ainsi été construits, permettant d'obtenir quatre prédictions. Deux d'entre elle permettent de prédire les cultures en 4 classes, les deux autres prédisent les 10 sous-classes de cultures. Les performances de prédictions des 10 sous-classes furent pour les deux modèles, bien en deçà des attentes. De plus, la validation a montré des résultats encore moins satisfaisants. En ce qui concerne les prédictions des quatre classes, elles semblaient plus convenables. Comme la grande majorité des cultures composant ces classes sont des cultures sarclées (les plus problématiques dans l'érosion par ruissellement), les prédictions obtenues en 4 classes peuvent être utiles dans le cadre de prédictions de zones à risque d'inondations boueuses. Cependant, environ un quart des parcelles de printemps sont omises dans les prédictions, et un quart des prédictions de cultures de printemps sont contaminées par d'autres cultures. Ces erreurs sont donc à considérer si le modèle est utilisé pour prédire les inondations boueuses.

La deuxième approche utilise la détection des couverts hivernaux pour prédire les cultures de printemps en utilisant le processeur SEN4CAP L4C. Les performances de la détection étaient amplement satisfaisantes. Toutefois, un nombre considérable de cultures de printemps ne sont pas précédées pas des couverts détectables par EO. L'utilisation des couverts hivernaux pour détecter les futures cultures de printemps ne permet donc pas d'obtenir une sensibilité supérieure à celles obtenues par les prédictions par rotations.

Deux méthodes combinant les deux approches ont enfin été effectuées. Elles permettent d'obtenir de meilleures performances de prédictions que lorsque les deux approches sont utilisées séparément. L'une contient plus de parcelles contaminantes, mais donne une meilleure sensibilité que l'autre. Dans le cadre d'une utilisation réelle, ces modèles combinés pourraient être utilisés alternativement afin d'avoir une idée des parcelles représentant potentiellement un risque dans l'apparition d'une inondation boueuse. La possibilité d'obtenir les prédictions mi-novembre répond au besoin de prédire les cultures le plus tôt possible afin de prendre des mesures préventives à ces inondations.

Avant une possible utilisation dans un cadre réel, il faudrait dans un premier temps tester les modèles construits sur une autre zone que celle utilisée pour la création du modèle. De plus, des pistes d'améliorations des deux approches ont été abordées afin d'améliorer les performances de prédictions de chacune d'elles.

8. Bibliographie

- Amore, E., C. Modica, M. A Nearing, et V. C Santoro. 2004. « *Scale Effect in USLE and WEPP Application for Soil Erosion Computation from Three Sicilian Basins* ». *Journal of Hydrology* 293 (1-4): 100-114. <https://doi.org/10.1016/j.jhydrol.2004.01.018>.
- « ASAP - NRT Monitoring Tool ». s. d. Consulté le 30 décembre 2019. <https://mars.jrc.ec.europa.eu/asap/hresolution/?region=774#>.
- Biielders, C., F. Colard, H. Cordonnier, A. Degré, M.-F. Destain, N. Feltz, et A. Maignard. 2011. « *Convention de recherche d'intérêt général Gestion Intégrée Sol Erosion Ruissellement : Rapport Final* ». Convention GISER, Service Pubic de Wallonie, Direction générale opérationnelle agriculture, ressources naturelles et environnement, Département de la ruralité et des cours d'eau, Direction du développement rural.
- Biielders, C., A. Degré, P. Demarcain, A. Dewez, A. Le Roi, JB. Staquet, B. deWalque, et N. Pineux. 2017. « *Projet GISER : Rapport Final-Version provisoire* ». Convention GISER, Service Public de Wallonie, UClouvain, Earth and Life Institute, Uliège-Gembloux Agro-Bio Tech, BIOSE Echange Eau-Sol-Plante.
- Biielders, C., C. Ramelot, et E. Persoons. 2003. « *Farmer Perception of Runoff and Erosion and Extent of Flooding in the Silt-Loam Belt of the Belgian Walloon Region* ». *Environmental Science & Policy* 6 (1): 85-93. [https://doi.org/10.1016/S1462-9011\(02\)00117-X](https://doi.org/10.1016/S1462-9011(02)00117-X).
- Blaes, X, L Vanhalle, et P Defourny. 2005. « *Efficiency of Crop Identification Based on Optical and SAR Image Time Series* ». *Remote Sensing of Environment* 96 (3-4): 352-65. <https://doi.org/10.1016/j.rse.2005.03.010>.
- Boardman, J. 2006. « *Soil Erosion Science: Reflections on the Limitations of Current Approaches* ». *CATENA* 68 (2-3): 73-86. <https://doi.org/10.1016/j.catena.2006.03.007>.
- Boardman, J., L. Ligneau, A. de Roo, et K. Vandaele. 1994. « *Flooding of Property by Runoff from Agricultural Land in Northwestern Europe* ». In *Geomorphology and Natural Hazards*, édité par M. Morisawa, 183-96. Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-444-82012-9.50017-7>.
- Boardman, J., J. Poesen, et R. Evans. 2003. « *Socio-Economic Factors in Soil Erosion and Conservation* ». *Environmental Science & Policy* 6 (1): 1-6. [https://doi.org/10.1016/S1462-9011\(02\)00120-X](https://doi.org/10.1016/S1462-9011(02)00120-X).

Bibliographie

- Boardman, J., G. Verstraeten, et C. Bielders. 2006b. « *Muddy Floods* ». In *Soil Erosion in Europe*, 743-55. John Wiley & Sons, Ltd.
<https://doi.org/10.1002/0470859202.ch53>.
- Borselli, L., P. Cassi, et D. Torri. 2008. « *Prolegomena to Sediment and Flow Connectivity in the Landscape: A GIS and Field Numerical Assessment* ». *CATENA* 75 (3): 268-77.
<https://doi.org/10.1016/j.catena.2008.07.006>.
- Breiman, L. 2001. « Random Forests ». *Machine Learning* 45 (1): 5-32.
<https://doi.org/10.1023/A:1010933404324>.
- Bullock, D. G. 1992. « *Crop Rotation* ». *Critical Reviews in Plant Sciences* 11 (4): 309-26.
<https://doi.org/10.1080/07352689209382349>.
- Burket, J., D. Hemphill, et R. Dick. 1997. « *Winter Cover Crops and Nitrogen Management in Sweet Corn and Broccoli Rotations* ». *HortScience* 32 (4): 664-68. <https://doi.org/10.21273/HORTSCI.32.4.664>.
- Castellazzi, M.S., G.A. Wood, P.J. Burgess, J. Morris, K.F. Conrad, et J.N. Perry. 2008. « *A Systematic Representation of Crop Rotations* ». *Agricultural Systems* 97 (1-2): 26-33.
<https://doi.org/10.1016/j.agry.2007.10.006>.
- Conțiu, Ș., et A. Groza. 2016. « *Improving Remote Sensing Crop Classification by Argumentation-Based Conflict Resolution in Ensemble Learning* ». *Expert Systems with Applications* 64 (décembre): 269-86.
<https://doi.org/10.1016/j.eswa.2016.07.037>.
- Danzin, A. 2018. « *Early warning system against muddy floods in central Belgium* ». Mémoire, Université Catholique de Louvain.
- Defourny, P. 2017. « Géomatique appliquée : Télédétection aéroportée et spatiale ». Cours, Université Catholique de Louvain, Louvain-la-Neuve.
- Delince, J., G. Lemoine, P. Defourny, J. Gallego, A. Davidson, S. Ray, O. Rojas, J. Latham, et F. Achard. 2017. « *Handbook on Remote Sensing for Agricultural Statistics* ». <https://doi.org/10.13140/RG.2.2.13259.69920>.
- Denis, J. 1992. « *Géographie de la Belgique* ». *Géographie de la Belgique, Bulletin trimestriel - Crédit communal de Belgique*. 1992, Num 179, pp 69-78, , no 179: 69-78.
- Destain, J.P., V. Reuter, et J.P. Goffart. 2010. « *Les cultures intermédiaires pièges à nitrate (CIPAN) et engrais verts : protection de l'environnement et intérêt agronomique* ». *Biotechnol. Agron. Soc. Environ.*, 6.

Bibliographie

- Dogliotti, S, W.A.H Rossing, et M.K van Ittersum. 2003. « *Rotat, a Tool for Systematically Generating Crop Rotations* ». *European Journal of Agronomy* 19 (2): 239-50.
[https://doi.org/10.1016/S1161-0301\(02\)00047-3](https://doi.org/10.1016/S1161-0301(02)00047-3).
- « *ESA - Observing the Earth* ». s. d. Consulté le 18 décembre 2019.
https://www.esa.int/Applications/Observing_the_Earth/.
- Evans, R., 2002. « *Rural land use in England and Wales and the delivery to the adjacent seas of nitrogen, phosphorus and atrazine* ». *Soil Use Manage.* 19, 1–7.
- Evrard, O., C. Biielders, K. Vandaele, et B. van Wesemael. 2007. « *Spatial and Temporal Variation of Muddy Floods in Central Belgium, off-Site Impacts and Potential Control Measures* ». *CATENA* 70 (3): 443-54.
<https://doi.org/10.1016/j.catena.2006.11.011>.
- Evrard, O., K. Vandaele, C. Biielders, et B. van Wesemael. 2008. « *Seasonal Evolution of Runoff Generation on Agricultural Land in the Belgian Loess Belt and Implications for Muddy Flood Triggering* ». *Earth Surface Processes and Landforms* 33 (8): 1285-1301.
<https://doi.org/10.1002/esp.1613>.
- « *GISER | Gestion Intégrée Sol – Erosion – Ruissellement* ». s. d. Consulté le 31 décembre 2019. <https://www.giser.be/>.
- Glaude, Q. 2018. « *Téledétection radar* ». Cours, Université de Liège.
https://orbi.uliege.be/bitstream/2268/223142/1/SAR_Remote_Sensing.pdf.
- Hsieh, C.H., R.H. Lu, N.H. Lee, W.T Chiu, M.H. Hsu, et Y.C. Li. 2011. « *Novel Solutions for an Old Disease : Diagnosis of Acute Appendicitis with Random Forest, Support Vector Machines, and Artificial Neural Networks* ». *Surgery* 149 (1): 87-93.
<https://doi.org/10.1016/j.surg.2010.03.023>.
- Huang, J., H. Wang, Q. Dai, et D. Han. 2014. « *Analysis of NDVI Data for Crop identification and Yield Estimation* ». *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (11): 4374-84.
<https://doi.org/10.1109/JSTARS.2014.2334332>.
- Kinnell, P.I.A. 2017. « *A Comparison of the Abilities of the USLE-M, RUSLE2 and WEPP to Model Event Erosion from Bare Fallow Areas* ». *Science of The Total Environment* 596-597 (octobre): 32-42. <https://doi.org/10.1016/j.scitotenv.2017.04.046>.
- Kollas, C., K. Kersebaum, C. Nendel, K. Manevski, C. Müller, T. Palosuo, et C. M. Armas-Herrera. 2015. « *Crop Rotation Modelling—A European Model Intercomparison* ». *European Journal of Agronomy* 70 (octobre): 98-111.
<https://doi.org/10.1016/j.eja.2015.06.007>.

Bibliographie

- Kucera, L., N. Joshi, et J. Slacikova. 2019. « *Sen4CAP - Sentinels for Common Agricultural Policy / Design Justification File ATBD for L4C agricultural monitoring product* ». UCL\ELI-Geomatics 2019.
- Leteinturier, B., B. Tychon, et R. Oger. 2007. « *Diagnostic agronomique et agro-environnemental des successions culturales en Wallonie (Belgique)* ». Biotechnol. Agron. Soc. Environ., 12.
- Mullan, D., T. Matthews, K. Vandaele, I. Barr, G. Swindles, J. Meneely, J. Boardman, et C. Murphy. 2019. « *Climate Impacts on Soil Erosion and Muddy Flooding at 1.5 versus 2°C Warming* ». Land Degradation & Development 30 (1): 94-108. <https://doi.org/10.1002/ldr.3214>.
- Mullan, D., K. Vandaele, J. Boardman, J. Meneely, et L. Crossley. 2016. « *Modelling the Effectiveness of Grass Buffer Strips in Managing Muddy Floods under a Changing Climate* ». Geomorphology 270 (octobre): 102-20. <https://doi.org/10.1016/j.geomorph.2016.07.012>.
- Parikh, R., A. Mathai, S. Parikh, G. Sekhar, et R. Thomas. 2008. « *Understanding and using sensitivity, specificity and predictive values* ». Indian journal of ophthalmology, 56(1), 45–50. doi:10.4103/0301-4738.37595
- « *PROBA-V Satellite Mission / PROBA-V* ». s. d. Consulté le 18 décembre 2019. <http://proba-v.vgt.vito.be/en/about/proba-v-satellite-mission>.
- Protect'eau. 2016. « *CIPAN : Quelles obligations respecter ?* » <https://protecteau.be/resources/shared/publications/articles/160805-CIPAN-SIE-module.pdf>.
- « *Protect'eau-Zone vulnérable* ». s. d. Consulté le 31 décembre 2019. <https://protecteau.be/fr/agriculteurs/legislations/zone-vulnerable>.
- Équipe de développement de QGIS. 2019. « *Système d'information géographique QGIS* ». Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.
- RStudio Team. 2019. « *RStudio: Integrated Development for R* ». RStudio Inc., Boston, MA URL <http://www.rstudio.com/>.
- Renard, K. G., USA, et USA. 1997. *Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE)*. Agriculture Handbook 703. Washington, D. C.
- Schönhart, M., E. Schmid, et U. Schneider. 2011. « *CropRota – A Crop Rotation Model to Support Integrated Land Use Assessments* ». European Journal of Agronomy 34 (4): 263-77. <https://doi.org/10.1016/j.eja.2011.02.004>.

Bibliographie

- « *Sen4Cap* ». s. d. Consulté le 18 décembre 2019. <http://esa-sen4cap.org/>.
- Statbel, et Economie. 2011. « *Chiffres clés de l'agriculture 2011* ». http://www.agripress.be/_STUDIOEMMA_UPLOADS/downloads/keyagr_fr_tcm326-133838.pdf.
- Stone, R., et D. Hilborn. 2012. « *12-051 — Universal Soil Loss Equation (USLE)* », 8.
- Tiwari, A., L. Risse, et M. Nearing. 2000. « *Evaluation of WEPP and its comparison with USLE and RUSLE* ». Transactions of the ASAE 43 (5): 1129-35. <https://doi.org/10.13031/2013.3005>.
- Urdiain-Arraiza, M. 2018. « *Évaluation de deux méthodes de prédictions de l'occupation du sol basées sur les rotations de cultures dans une perspective de prédictions d'inondations boueuses* ». Mémoire, Université Catholique de Louvain.
- « *User Guides - Sentinel Online* ». s. d. Consulté le 18 décembre 2019. <https://sentinel.esa.int/web/sentinel/user-guides/>.
- Van Campenhout, J., E. Hallot, G. Houbrechts, A. Peeters, Y. Levecq, P. Gérard, et F. Petit. 2015. « *Flash Floods and Muddy Floods in Wallonia: Recent Temporal Trends, Spatial Distribution and Reconstruction of the Hydrosedimentological Fluxes Using Flood Marks and Sediment Deposits* ». Belgeo, no 1 (mars). <https://doi.org/10.4000/belgeo.16409>.
- Van den haute, P., J.P. Buylaert, D. Vandenberghe, et F. De Corte. 2002. « the Last Interglacial Palaeosol in the Belgian Loess Belt: TL-age Record. » In Book of Abstracts LED 2002, 185–185.
- Van der Helm, P.P.M. 1987. « *Bodemerosie en wateroverlast; een milieuprobleem in het Zuid-Limburgse heuvelland* ». MSc Thesis, Dept. of Physical Geography, University of Utrecht.
- Vandaele, K., et J. Poesen. 1995. « *Spatial and Temporal Patterns of Soil Erosion Rates in an Agricultural Catchment, Central Belgium* ». CATENA 25 (1-4): 213-26. [https://doi.org/10.1016/0341-8162\(95\)00011-G](https://doi.org/10.1016/0341-8162(95)00011-G).
- Verstraeten, G., J. Poesen, G. Demarée, et C. Salles. 2006. « *Long-Term (105 Years) Variability in Rain Erosivity as Derived from 10-Min Rainfall Depth Data for Ukkel (Brussels, Belgium): Implications for Assessing Soil Erosion Rates* ». Journal of Geophysical Research 111 (D22): D22109. <https://doi.org/10.1029/2006JD007169>.
- « *WalOnMap | Géoportail de la Wallonie* ». s. d. Consulté le 18 décembre 2019. <http://geoportail.wallonie.be/walonmap>.

Bibliographie

- de Walque, B., A. Degré, A. Maignard, et C. Biielders. 2017. « *Artificial Surfaces Characteristics and Sediment Connectivity Explain Muddy Flood Hazard in Wallonia* ». CATENA 158 (novembre): 89-101. <https://doi.org/10.1016/j.catena.2017.06.016>.
- Were, K., D. Tien Bui, Ø. Dick, et B. Ram Singh. 2015. « *A Comparative Assessment of Support Vector Regression, Artificial Neural Networks, and Random Forests for Predicting and Mapping Soil Organic Carbon Stocks across an Afromontane Landscape* ». Ecological Indicators 52 (mai): 394-403. <https://doi.org/10.1016/j.ecolind.2014.12.028>.

Annexe

Table des annexes

Annexe A : Complément de l'état de l'art	II
Annexe B : Classification	III
Culture supprimée.....	III
Prairie	III
Autre.....	IV
Hiver	IV
Printemps	V
Annexe C : Modèle de prédictions par rotation	VI
Occurrence des rotations	VI
Test des modèles ne prenant pas en compte les plus petites parcelles.....	VIII
Test des modèles ne prenant pas en compte les parcelles dont les limites ont été modifiées au cours de la rotation	IX
Test de la variable « type de sol »	XI
Annexe D : Prédiction par la détection des CIPAN	XII
Seuils des marqueurs	XII
Périodes de récoltes des cultures	XIII
Détection des CIPAN pour les deux périodes retenues	XIII
Type de parcelles contaminantes.....	XIV
Type de parcelles omises	XIV

Annexe A. Complément de l'état de l'art

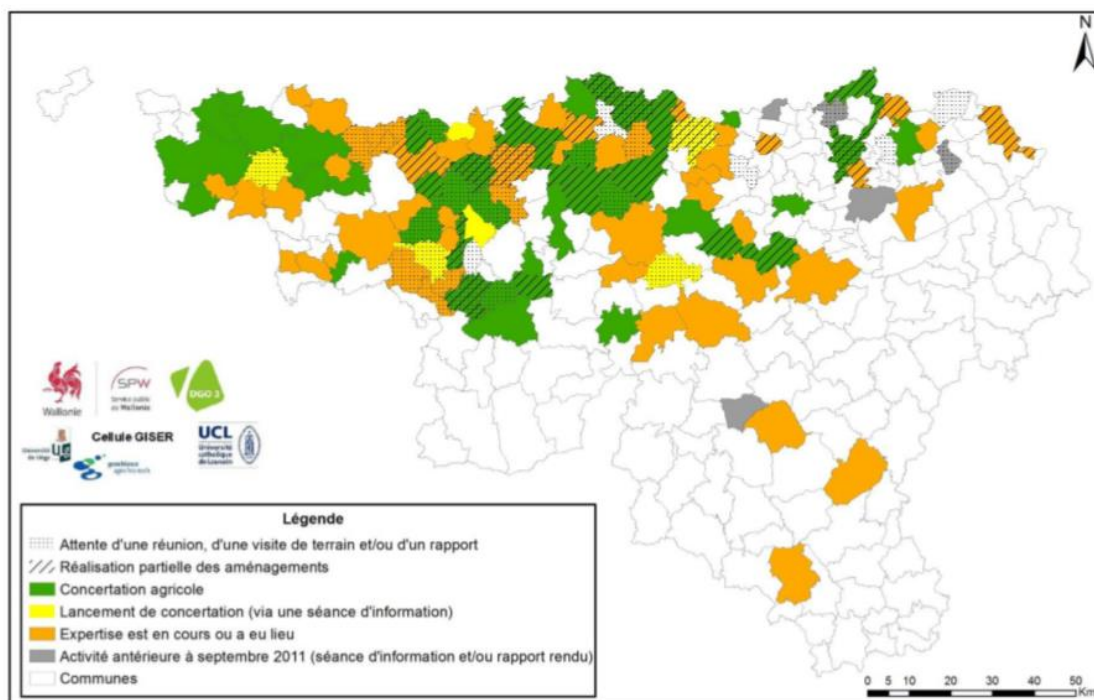


Figure A-21 : Etat des lieux des activités de la cellule GIER (07/04/17) [Biolders et al., 2017].



Figure A-22 : Communes de Wallonie, et zone vulnérable du PGDA [Protect'eau-ZoneVulnérable].

Annexe B. Classification

Culture supprimée

Tableau A-22 : Cultures considérées comme permanente et non prises en compte dans les classifications 4 et 10.

Nom de la culture principale	Code culture du SIGEC
Bande bordures de champs	752
Culture forestière	883
Cultures fruitières pluriannuelles basses-tiges	9741
Cultures fruitières pluriannuelles hautes-tiges	9742
Cultures fruitières annuelles Fraise sous serres	8516
Cultures fruitières annuelles Fraise	9516
Cultures fruitières annuelles Framboise	9517
Cultures maraichères sous verre	952
Pépinières de plants fruitiers ou plantes ornementales	9520
Prairies permanentes 50%<taux de couvertures<90%	678/670
Prairies permanentes >90%	610
Sapins de Noël	962
Tournières enherbées	751
Vignes	9716

Prairie

Tableau A-23 : Cultures classées dans la classe « prairie » des classifications 4 et 10.

Nom de la culture principale	Code culture du SIGEC
Autre couvert semés	85
Autre couvert semencé	83
Autres fourrages	743
Bande parcelles aménagées MAE	754
Couvert à finalité environnementale	874
Couvert favorisant la faune	851
Couvert naturel	81
Jachère herbacée	811
Jachère non herbacée	812
Luzerne	73
Prairie temporaire	62
Trèfles	72

Classification

Autre

Tableau A-24 : Cultures classées dans la classe « autre » des classifications 4 et 10.

Nom de la culture principale	Code culture du SIGEC
Autre	99
Céréales et légumineuses	39
Miscanthus	884
Orge de brasserie	323

Hiver

Tableau A-25 : Cultures classées dans la classe « hiver » des classifications 4 et 10.

Nom de la culture principale	Code culture du SIGEC
Autre	99
Avoine d'hiver	341
Colza/ navette d'hiver	4111
Epeautre	36
Froment d'hiver	311
Mélange protéagineux d'hiver + céréales ou autre	541
Orge d'hiver	321
Pois protéagineux d'hiver	511
Triticales d'hiver	351

Classification

Printemps

Tableau A-26 : Cultures classées dans la classe « printemps » des classifications en 4 classes et composition des sept sous-classes de printemps de la classification en 10 classes.

10 classes	Nom de la culture principale	Code culture du SIGEC
Maïs	Maïs ensilage	201
	Maïs grain	202
Betterave	Betterave fourragère	71
	Betterave sucrière	91
Pomme de terre	Pomme de terre non hâtive	901
	Pomme de terre plants	902
	Pomme de terre féculière	903
	Pomme de terre hâtive	904
Pois	Pois (autre que récoltés à sec)	831
	Pois récoltés à l'état frais	931
Lin	Lin textile	921
Céréale de printemps	Avoine de printemps	342
	Froment de printemps	312
	Orge de printemps	322
Autre printemps	Autres légumes de plein air	951
	Carotte	7433
	Carotte hâtive	9564/8564
	Carotte non hâtive	9535
	Chanvre non textile	872
	Chanvre textile	922
	Chicorée à inuline	9811
	Choux - légume	9548
	Endives (chicons)	9515
	Endives de Bruxelles	8561
	FR-Asperges-vers	9511
	Fèves et féveroles de printemps	522
	Haricot de conserve	9410
	Haricot vert	8410
	Légume légumineuse	966
	Mélange légumineuse fourragère + céréale ou autre	541
	Mélange protéagineux de printemps + céréale ou autre	542
	Oignon non hâtif	9514
	Oignon hâtif	9514
	Epinard	9519
	Poireau	9538/8538
	Pois protéagineux de printemps	512
	Tournesol	42

Annexe C. Modèle de prédictions par rotation

Occurrence des rotations

4 Classes

Tableau A-27 : Occurrence des rotations observées dans la zone d'études entre 2014 et 2018, classification en quatre classes. (Hiv = hiver, Pri = Printemps, Aut = Autre et Pra = Prairie).

2 années		3 années		4 années		5 années	
Rotation	Occurrence [%]	Rotation	Occurrence [%]	Rotation	Occurrence [%]	Rotation	Occurrence [%]
Hiv-Pri	28,52	Pri-Hiv-Pri	22,11	Hiv-Pri-Hiv-Pri	18,62	Pri-Hiv-Pri-Hiv-Pri	15,02
Pri-Hiv	26,70	Hiv-Pri-Hiv	20,86	Pri-Hiv-Pri-Hiv	15,38	Hiv-Pri-Hiv-Pri-Hiv	12,74
Pri-Pri	13,67	Hiv-Pri-Pri	7,86	Pri-Hiv-Pri-Pri	7,09	Hiv-Pri-Hiv-Pri-Pri	6,03
Hiv-Hiv	10,50	Hiv-Hiv-Pri	7,06	Hiv-Hiv-Pri-Hiv	6,47	Hiv-Hiv-Pri-Hiv-Pri	4,21
Aut-Aut	6,72	Pri-Hiv-Hiv	6,16	Pri-Pri-Hiv-Pri	4,70	Pri-Hiv-Hiv-Pri-Hiv	4,00
Pra-Pra	6,53	Pri-Pri-Hiv	6,05	Hiv-Pri-Hiv-Hiv	4,63	Pri-Hiv-Pri-Pri-Hiv	3,71
Hiv-Pra	1,49	Pri-Pri-Pri	6,05	Pri-Hiv-Hiv-Pri	4,46	Hiv-Pri-Pri-Hiv-Pri	3,39
Pri-Pra	1,27	Pra-Pra-Pra	4,33	Hiv-Pri-Pri-Hiv	4,36	Hiv-Hiv-Pri-Hiv-Hiv	3,21
Pra-Pri	0,85	Aut-Aut-Aut	4,14	Pri-Pri-Pri-Pri	3,70	Pri-Pri-Hiv-Pri-Hiv	2,86
Pri-Aut	0,79	Hiv-Hiv-Hiv	3,82	Hiv-Pri-Pri-Pri	2,69	Pri-Pri-Pri-Pri-Pri	2,71
Hiv-Aut	0,79	Pri-Pra-Pra	0,98	Hiv-Hiv-Hiv-Pri	2,21	Hiv-Pri-Hiv-Hiv-Pri	2,61
Aut-Pri	0,50	Hiv-Pra-Pra	0,91	Hiv-Hiv-Hiv-Hiv	2,08	Pri-Pri-Hiv-Pri-Pri	2,39
Pra-Aut	0,50	Pri-Hiv-Pra	0,77	Pra-Pra-Pra-Pra	2,01	Pri-Hiv-Pri-Pri-Pri	2,28
Aut-Pra	0,46	Pra-Pri-Pri	0,53	Hiv-Hiv-Pri-Pri	1,99	Hiv-Hiv-Hiv-Pri-Hiv	2,18
Pra-Hiv	0,40	Pri-Hiv-Aut	0,42	Pri-Pri-Pri-Hiv	1,93	Pri-Hiv-Pri-Hiv-Hiv	2,07
Aut-Hiv	0,33	Hiv-Pri-Aut	0,41	Pri-Hiv-Hiv-Hiv	1,65	Pri-Hiv-Hiv-Pri-Pri	1,71
		...	...	Pri-Pri-Hiv-Hiv	1,28	Hiv-Pri-Pri-Pri-Hiv	1,36
				Pri-Pra-Pra-Pra	1,18	Hiv-Hiv-Hiv-Hiv-Hiv	1,32
				Hiv-Pra-Pra-Pra	0,95	Hiv-Pri-Hiv-Hiv-Hiv	1,32
				...	...	Pri-Pri-Pri-Hiv-Pri	1,32
						Hiv-Hiv-Pri-Pri-Hiv	1,11
						Hiv-Pri-Pri-Pri-Pri	1,11
						...	...

Modèle de prédictions par rotation

10 Classes

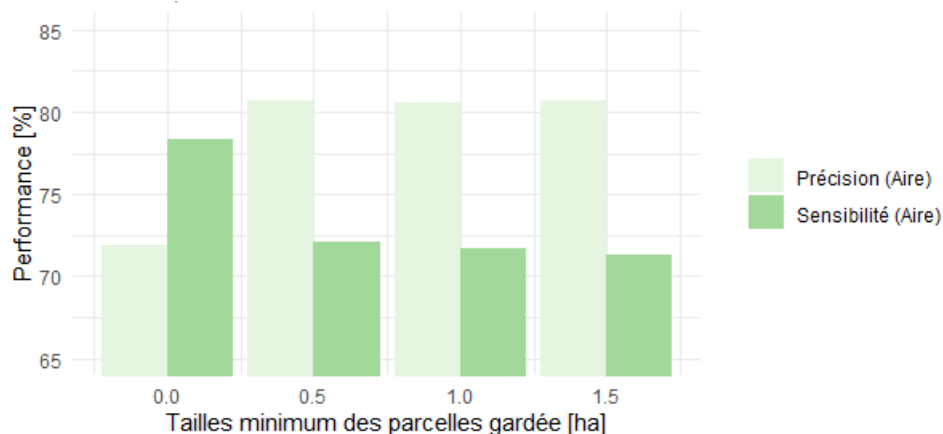
Tableau A-28 : Occurrence des rotations observées dans la zone d'études entre 2014 et 2018, classification en dix sous-classes. (Hiv = hiver, Bet = Betterave, Maï = Maïs, PDT = Pomme de terre, Lin = Lin textile, Poi = Pois, AuP = Autre Printemps, CéP = Céréale de printemps, Aut = Autre et Pra = Prairie).

2 années		3 années		4 années		5 années	
Rotation	Occurrence [%]	Rotation	Occurrence [%]	Rotation	Occurrence [%]	Rotation	Occurrence [%]
Hiv-Hiv	10,50	Hiv-Bet-Hiv	5,63	Hiv-Hiv-Bet-Hiv	2,28	Maï-Maï-Maï-Maï-Maï	1,93
PDT-Hiv	7,45	Hiv-PDT-Hiv	5,63	Bet-Hiv-PDT-Hiv	2,27	Hiv-Bet-Hiv-PDT-Hiv	1,64
Hiv-Bet	7,35	Pra-Pra-Pra	4,33	Maï-Maï-Maï-Maï	2,21	Hiv-Hiv-Hiv-Hiv-Hiv	1,32
Bet-Hiv	7,21	Aut-Aut-Aut	4,14	Hiv-Hiv-Hiv-Hiv	2,08	Bet-Hiv-PDT-Hiv-Bet	1,28
Aut-Aut	6,72	Hiv-Hiv-Hiv	3,82	Pra-Pra-Pra-Pra	2,01	Hiv-Hiv-Bet-Hiv-Hiv	1,18
Pra-Pra	6,53	Hiv-Maï-Hiv	3,31	Hiv-PDT-Hiv-Bet	2,01	Hiv-PDT-Hiv-Bet-Hiv	1,11
Hiv-PDT	6,01	Maï-Maï-Maï	2,61	Hiv-Bet-Hiv-PDT	1,69	Bet-Hiv-Hiv-Bet-Hiv	0,96
Hiv-Maï	5,16	Hiv-AuP-Hiv	2,46	Hiv-Hiv-PDT-Hiv	1,58	Pra-Pra-Pra-Pra-Pra	0,93
Maï-Hiv	4,76	Hiv-Hiv-Bet	2,46	PDT-Hiv-Bet-Hiv	1,47	Hiv-AuP-Hiv-PDT-Hiv	0,93
Maï-Maï	4,04	Bet-Hiv-PDT	2,21	Hiv-Bet-Hiv-Hiv	1,43	Hiv-Hiv-Maï-Hiv-Hiv	0,86
Hiv-AuP	3,32	PDT-Hiv-Bet	2,14	Hiv-Hiv-Maï-Hiv	1,38	Hiv-Hiv-PDT-Hiv-Hiv	0,79
Hiv-Lin	3,19	Bet-Hiv-Hiv	1,94	Hiv-PDT-Hiv-Hiv	1,19	Hiv-Maï-Hiv-Maï-Hiv	0,75
AuP-Hiv	3,10	Hiv-Lin-Hiv	1,94	Maï-Hiv-Maï-Hiv	1,06	Maï-Hiv-Maï-Hiv-Maï	0,71
Hiv-Poi	2,74	Hiv-Hiv-Maï	1,76	Hiv-Maï-Hiv-Maï	1,06	Hiv-AuP-Hiv-Bet-Hiv	0,71
Lin-Hiv	2,05	Maï-Hiv-Maï	1,64	Hiv-Maï-Hiv-Hiv	1,02	Hiv-Hiv-Hiv-Bet-Hiv	0,71
Poi-Hiv	1,69	PDT-Hiv-Hiv	1,57	Hiv-AuP-Hiv-Bet	0,99	PDT-Hiv-Bet-Hiv-PDT	0,68
Hiv-Pra	1,49	Hiv-Poi-Hiv	1,53	Hiv-Pra-Pra-Pra	0,95	PDT-Hiv-Lin-Hiv-PDT	0,68
...	...	PDT-Hiv-Lin	1,49	Hiv-PDT-Hiv-Lin	0,94	Hiv-Poi-Hiv-PDT-Hiv	0,68
		Maï-Hiv-Hiv	1,43	Hiv-AuP-Hiv-PDT	0,89	Hiv-Lin-Hiv-PDT-Hiv	0,64
		Hiv-Hiv-PDT	1,38	AuP-Hiv-PDT-Hiv	0,87	Bet-Hiv-AuP-Hiv-Bet	0,57
		PDT-Hiv-Poi	1,06	Bet-Hiv-Hiv-Bet	0,83	Hiv-Maï-Hiv-PDT-Hiv	0,54
		AuP-Hiv-Bet	1,02	Hiv-Lin-Hiv-PDT	0,80	Hiv-Hiv-Hiv-PDT-Hiv	0,54
		...	...	Bet-Hiv-AuP-Hiv	0,75	Maï-Hiv-Hiv-Maï-Hiv	0,50
				AuP-Hiv-Bet-Hiv	0,75	Hiv-Maï-Hiv-Bet-Hiv	0,50
				Hiv-PDT-Hiv-Poi	0,75	Hiv-Bet-Hiv-Hiv-Bet	0,50
				PDT-Hiv-Lin-Hiv	0,73	Bet-Hiv-PDT-Hiv-Lin	0,46
				Maï-Hiv-Hiv-Maï	0,72	Bet-Hiv-Hiv-PDT-Hiv	0,46
				Hiv-Maï-Hiv-Bet	0,70	Lin-Hiv-PDT-Hiv-Bet	0,46
				...	...	...	...

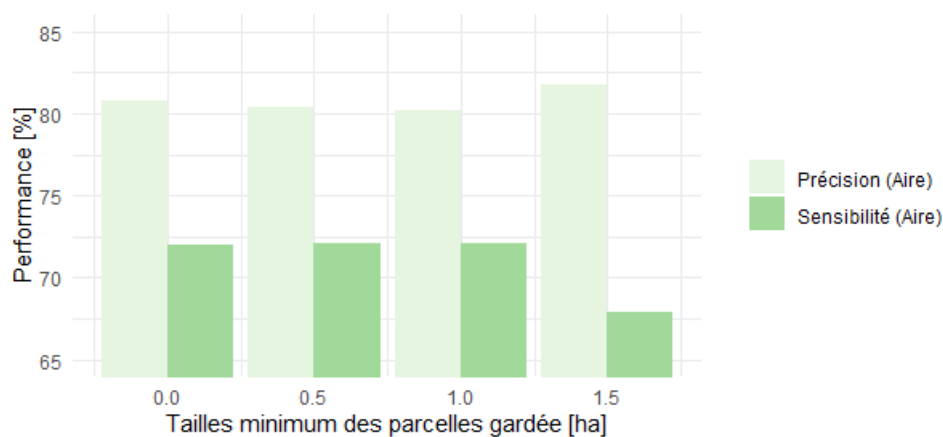
Modèle de prédictions par rotation

Test des modèles ne prenant pas en compte les plus petites parcelles

Modèle 4 classes



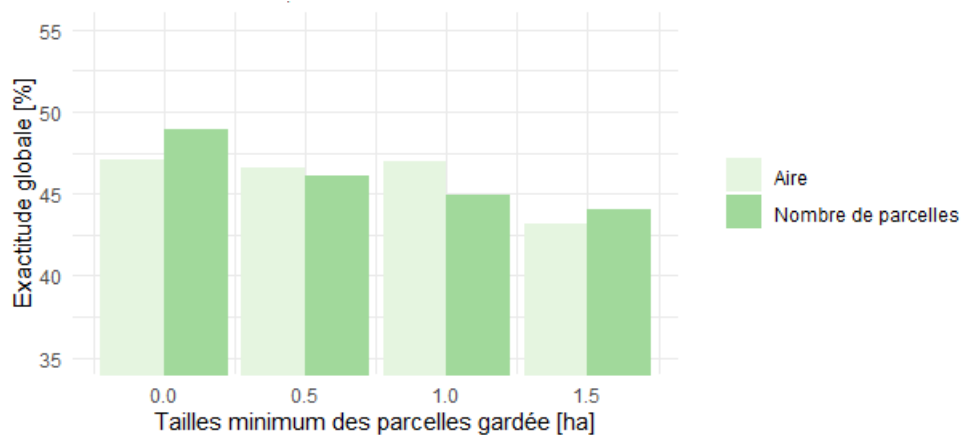
Graph A-14 : Effet de la suppression des petites parcelles sur les performances de prédictions du modèle 4 classes basé sur trois années de rotations.



Graph A-15 : Effet de la suppression des petites parcelles sur les performances de prédictions du modèle 4 classes basé sur quatre années de rotations.

Modèle de prédictions par rotation

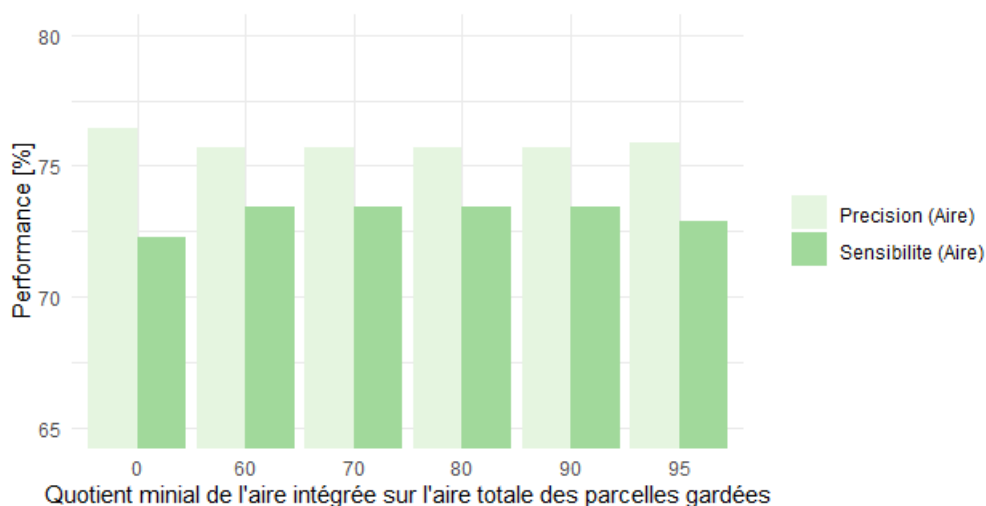
Modèle 10 classes



Graph A-16 : Effet de la suppression des petites parcelles sur les performances de prédictions du modèle 10 classes indirect basé sur cinq années de rotations.

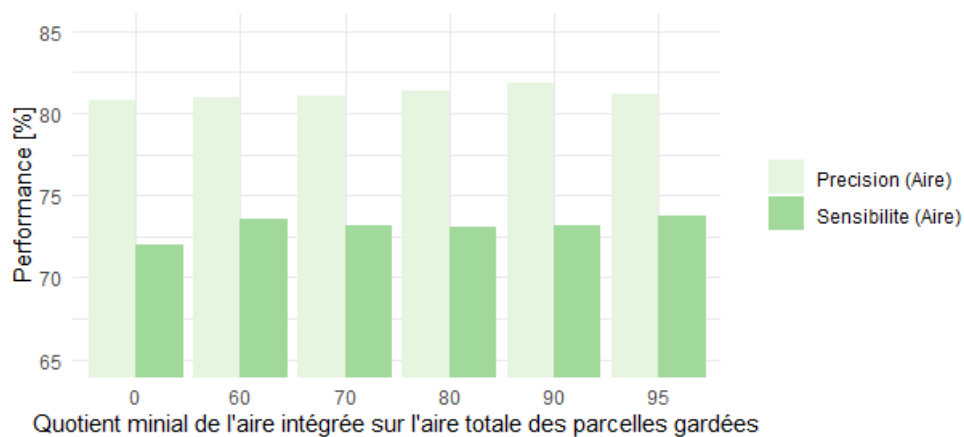
Test des modèles ne prenant pas en compte les parcelles dont les limites ont été modifiées au cours de la rotation

Modèle 4 classes



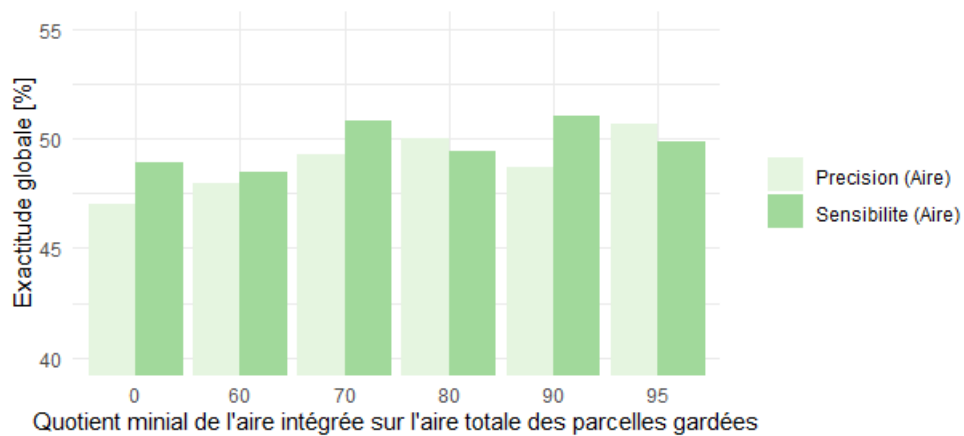
Graph A-17 : Effet de la suppression des parcelles dont les limites ont été modifiées selon des degrés différents dans le modèle 4 classes basé sur trois années de rotations.

Modèle de prédictions par rotation



Graph A-18 : Effet de la suppression des parcelles dont les limites ont été modifiées selon des degrés différents dans le modèle 4 classes basé sur quatre années de rotations.

Modèle 10 classes



Graph A-19 : Effet de la suppression des parcelles dont les limites ont été modifiées selon des degrés différents dans le modèle 10 classes basé sur cinq années de rotations.

Modèle de prédictions par rotation

Test de la variable « type de sol »

Tableau A-29 : 15 classes retenues dans le test basé sur le type de sol issus du CNSW.

aba
AbB
abp
ac
ada
adp
aep
Afp
Ahp
Aip
EDx
GAx
L
S
<i>Non Classées</i>

Annexe D. Prédiction par la détection des CIPAN

Seuils des marqueurs

Tableau A-30 : Seuils entrés dans le processeur L4C utilisé par les différents marqueurs. Ces seuils ont été repris de ceux optimisés pour les Pays-Bas.

Marqueur(s)	Type de seuil	Encodage
1	NDVI value defining the presence of vegetation on the parcel	OPTTHRVEGCYCLE=350
2	Lower limit of the NDVI loss threshold	NDVIDW=300
2	Upper limit of the NDVI loss threshold	NDVIUP=350
2/7/8	Value to which the computed NDVI loss threshold is round up	NDVISTEP=5
2	Minimum NDVI threshold	OPTTHRMIN=100
5	Basic increase in coherence threshold	COHTRBASE=0.05
5	High increase in coherence threshold	COHTRHIGH=0.15
5	Absolute coherence threshold	COHTHRABS=0.75
3	Minimum backscatter (CR) loss threshold	AMPTHRMIN=0.1
6	NDVI value defining the presence of EFA practice vegetation on the parcel	EFANDVITHR=325
7/8	Upper limit of the NDVI loss threshold (can be specified for EFA practice)	CC_NDVIUP=500
7/8	Lower limit of the NDVI loss threshold (can be specified for EFA practice)	EFANDVIDW=300
7	minimal value of NDVI (default 300)	EFANDVIMIN="-10000"
9	slope threshold (default 0.01)	CC_EFAAMPTHR=0.03
10	High increase in coherence threshold	CC_COHTRBASE=0.1
10	Absolute coherence threshold	CC_COHTHRABS=0.7

Périodes de récoltes des cultures

Tableau A-31 : Période de récoltes calibrées sur les récoltes des cultures de l'année 2018.

Cultures	Début de la période de récolte	Fin de la période de récolte
Froment d'hiver	01-juil	25-août
Orge d'hiver	15-juin	20-juil
Avoine de printemps	15-juil	30-août
Triticales d'hiver	25-juin	25-août
Autres légumes	01-juin	30-août
Colza d'hiver	01-juil	05-août
Lin textile	20-juin	15-août
Pois	30-juin	30-août
Haricot	01-juin	30-août
Epeautre	15-juil	25-août
Luzerne	01-juin	20-juil
Autres fourrages	01-juin	30-juin
Mélange protéagineux	25-juin	25-août
Autres couverts semés	01-juin	30-juin
Fèves et féverolles	30-juin	30-août

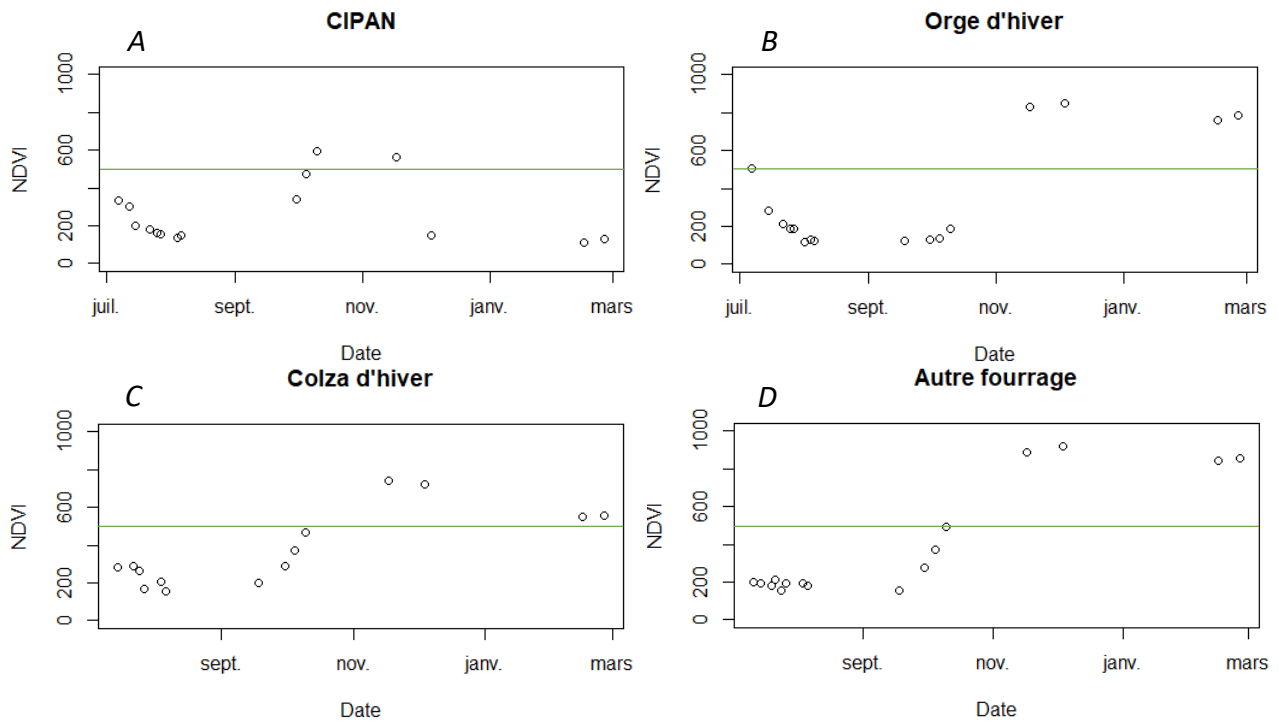
Détection des CIPAN pour les deux périodes retenues

Tableaux A-32 : Matrices de confusions obtenues par le croisement des détections de CIPAN et les cultures de 2019 pour les deux périodes retenues [ha].

30/09 – 10/11					30/09 – 15/11				
	Print	Hiver	Prair	Autre		Print	Hiver	Prair	Autre
Forte	2491,8	85,8	30,4	1,9	Forte	2038,4	66	12,1	1,9
Modérée	1330,9	319,2	31,6	0	Modérée	1499,9	144,3	46,6	0
Faible	491,2	918,6	15,1	26,7	Faible	775,6	1113,3	18,4	26,7
Pauvre	46,6	57,6	6,7	0	Pauvre	46,6	57,6	6,7	0

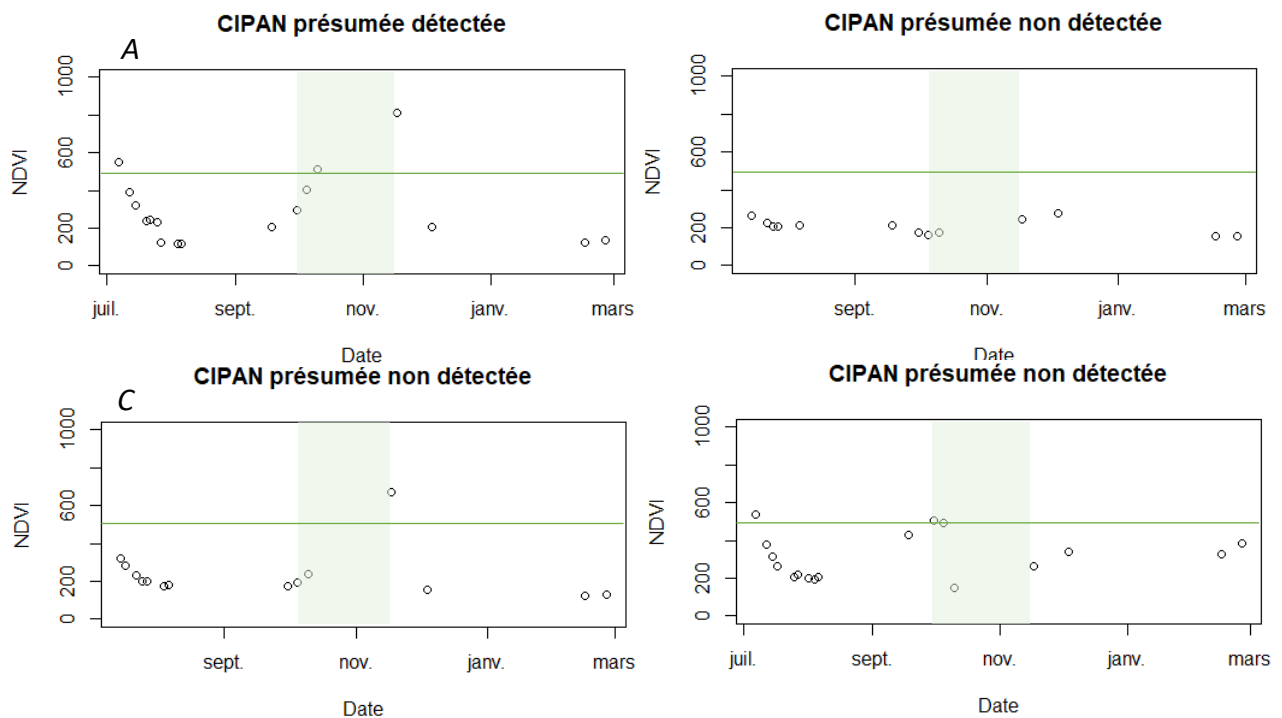
Prédiction par la détection des CIPAN

Type de parcelles contaminantes



Graphes A-20 : Exemple de l'évolution du NDVI de parcelles détectées comme CIPAN dans la période du 30/09 au 10/11. Avec ; A : une parcelle contenant bien un couvert, B : une culture d'orge d'hiver détectée comme couvert, C : une culture de colza d'hiver et D : une culture « autre fourrage » détectés comme des couverts. Le segment horizontal vert montre la limite à laquelle le marqueur 6 du processeur détecte de la végétation sur base du NDVI (500).

Type de parcelles omises



Graphes A-21 : Exemple de l'évolution du NDVI de parcelles de couverts détectées (A) et non détectées (B, C et D) par le processeur dans la période du 30/09 au 10/11. Le segment horizontal vert montre la limite à laquelle le marqueur 6 du processeur détecte de la végétation sur base du NDVI (500). L'intervalle en vert clair représente la période de détection.

Prédiction des cultures de printemps dans le contexte de la création d'un système d'alerte précoce pour les inondations boueuses

Présenté par Pierre Houdmont

Les inondations boueuses sont des problèmes récurrents dans le centre de la Belgique, notamment en régions limoneuses, sablo-limoneuses et du Condroz. Elles impliquent d'importants coûts de réparation dans les zones urbanisées, au niveau des habitations et des voiries, en plus d'être néfastes pour les rendements agricoles et l'environnement. La conception d'un système d'alerte précoce pourrait s'avérer très utile aux communes pour les aider à installer judicieusement et de manière préventive, des aménagements de lutte. Un facteur les influençant grandement est la couverture du sol. Dans les zones cultivées, cette dernière est variable durant et entre les années à la suite des successions de cultures dans les parcelles. Pour différentes raisons, il s'avère que les parcelles accueillant les cultures de printemps sont bien plus susceptibles de provoquer ces inondations.

L'objectif de ce mémoire s'est inséré dans ce contexte, en essayant de prédire, à l'échelle de la parcelle, ces cultures de printemps le plus tôt possible dans l'année afin de permettre la conception du système d'alerte par la prédiction des zones à risque. Afin d'y arriver, deux approches ont été utilisées.

La première consiste en la création d'un modèle de prédictions utilisant les successions de cultures sur les parcelles. Celles-ci ont été obtenues par la superposition de parcellaire de plusieurs années. Ces superpositions ont ensuite pu être utilisées pour entraîner un algorithme de *machine learning*, un *Random Forest*. Le modèle obtenu par cette méthode a permis d'obtenir une prédiction des classes de printemps constituée d'une précision et d'une sensibilité d'environ 75% lors de la validation.

La seconde approche consiste à prédire les cultures de printemps via la détection des couverts hivernaux grâce aux images satellites, ceux-ci précédant généralement des cultures de printemps. Pour y arriver, un processeur fonctionnant sous forme d'un modèle logique, développé dans le cadre du projet SEN4CAP a été utilisé. Une calibration de la période de détection des couverts a été effectuée et a permis d'obtenir des résultats satisfaisants. Cependant, toutes les parcelles de printemps ne sont pas précédées d'un couvert. Ainsi, la prédiction obtenue en utilisant les détections comme prédicteur de cultures de printemps, n'a pas permis d'obtenir une très grande sensibilité (66,7%). La précision de cette prédiction était, quant à elle, de 89,07%.

Deux modèles combinant les deux approches ont finalement été réalisés. Ceux-ci ont permis d'obtenir des prédictions de sensibilités et de précisions, respectivement de 82,63% et 75,82% pour le premier et de 75,07% et 84,95% pour le deuxième. Dans le cadre d'une utilisation réelle pour prédire les zones à risque d'inondations boueuses, il serait intéressant d'utiliser les deux modèles combinés alternativement, l'un induisant plus de contamination et l'autre plus d'omission. Les deux approches permettraient d'obtenir les prédictions mi-novembre, ce qui répond au besoin de les obtenir suffisamment tôt dans l'année.